UNIVERSIDAD NACIONAL DE MAR DEL PLATA

TESIS DOCTORAL

Diseño de Algoritmos basados en la Teoría de Juegos Cuántica para el Modelado de Redes de Comunicación y Aprendizaje por Refuerzo Multi-Agente

---

Autor:
Agustin SILVA

Director:
Dr. Constancio Miguel ARIZMENDI

Co-Director:
Dr. Omar Gustavo ZABALETA

Una tesis presentada en cumplimiento de los requisitos para obtener el grado de Doctor en Ingeniería, orientación Simulación y Modelado Computacional

desarrollada en

Laboratorio de Sistemas Complejos y Computación Cuántica
Departamento de Física

26 de febrero de 2024

UNIVERSIDAD NACIONAL DE MAR DEL PLATA

TESIS DOCTORAL

Diseño de Algoritmos basados en la Teoría de Juegos Cuántica para el Modelado de Redes de Comunicación y Aprendizaje por Refuerzo Multi-Agente

---

Autor:
Agustin SILVA

Director:
Dr. Constancio Miguel ARIZMENDI

Co-Director:
Dr. Omar Gustavo ZABALETA

Una tesis presentada en cumplimiento de los requisitos para obtener el grado de Doctor en Ingeniería, orientación Simulación y Modelado Computacional

desarrollada en

Laboratorio de Sistemas Complejos y Computación Cuántica
Departamento de Física

26 de febrero de 2024

Declaración de Autoría

Yo, Agustin SILVA, declaro que esta tesis titulada, «Diseño de Algoritmos basados en la Teoría de Juegos Cuántica para el Modelado de Redes de Comunicación y Aprendizaje por Refuerzo Multi-Agente» y el trabajo presentado en ella son míos. Lo confirmo:

- Este trabajo fue realizado en su totalidad durante mi candidatura a un posgrado de investigación en esta Universidad.
- Cuando alguna parte de esta tesis se haya presentado previamente para obtener un título o cualquier otra calificación en esta Universidad o cualquier otra institución, esto se ha indicado claramente.
- Cuando he consultado el trabajo publicado de otros, esto siempre se atribuye claramente.
- Donde he citado el trabajo de otros, siempre se da la fuente. Con la excepción de tales citas, esta tesis es enteramente mi propio trabajo.

Fecha: 26 de febrero de 2024
Firma:

Aclaración: Agustin Silva

«Those swirls in the cream mixing into the coffee? That's us. Ephemeral patterns of complexity, riding a wave of increasing entropy from simple beginnings to a simple end. We should enjoy the ride.»

Sean Carroll

UNIVERSIDAD NACIONAL DE MAR DEL PLATA

Resumen

Facultad de Ingeniería
Departamento de Física

Doctor en Ingeniería, orientación Simulación y Modelado Computacional

Diseño de Algoritmos basados en la Teoría de Juegos Cuántica para el Modelado de Redes de Comunicación y Aprendizaje por Refuerzo Multi-Agente

por Agustin SILVA

Con el objetivo de abordar problemas fundamentales en el transporte de datos en redes de comunicación y el diseño de algoritmos de aprendizaje por refuerzo multi-agente se han utilizado métodos de computación cuántica en conjunto con la teoría de juegos. A través de un enfoque novedoso, esta investigación demuestra cómo la aplicación de la teoría de juegos cuántica, haciendo uso del entrelazamiento y la superposición, supera las limitaciones de los enfoques clásicos. Provee soluciones avanzadas para mitigar los problemas de congestión en redes y encuentra estrategias óptimas en sistemas multi-agente. La investigación se basa en la combinación de la computación cuántica con la teoría de juegos, con el objetivo de permitir el desarrollo de protocolos más eficientes en el ruteo de redes de comunicación y algoritmos para la coordinación entre agentes autónomos. Al implementar estrategias cuánticas en el modelo de congestión propuesto, se logra una reducción en la latencia de las redes, evidenciando una mejora en el rendimiento en comparación con las estrategias clásicas. Además, la integración de algoritmos de aprendizaje por refuerzo adaptativos permite a los sistemas autónomos en ambientes cuánticos optimizar la toma de decisiones estratégicas en tiempo real, abriendo el camino hacia aplicaciones prácticas en otros campos como: la logística, la gestión de tráfico urbano y la simulación de sistemas económicos. Estos algoritmos demuestran ser particularmente efectivos en la exploración de estrategias cuánticas, puras o mixtas, y en la adaptación a entornos con información imperfecta incluso al evaluarlos tanto en ambientes ideales como ruidosos, revelando la potencialidad de la interacción entre agentes autónomos en contextos cuánticos reales. La metodología empleada, que combina modelado matemático y simulaciones computacionales, proporciona una validación sólida de las propuestas teóricas, permitiendo comprobar la efectividad de los algoritmos desarrollados.

Palabras claves: Computación Cuántica, Teoría de Juegos, Redes de Comunicación, Aprendizaje por Refuerzo.

NATIONAL UNIVERSITY OF MAR DEL PLATA

Abstract

Faculty of Engineering
Departamento de Física

Doctor in Engineering, Computational Simulation and Modeling orientation

Design of Algorithms based on Quantum Game Theory for Modeling Communication Networks and Multi-Agent Reinforcement Learning

by Agustin SILVA

Aiming to address fundamental problems in data transport in communication networks and the design of multi-agent reinforcement learning algorithms, quantum computing methods have been used in conjunction with game theory. Through a novel approach, this research demonstrates how the application of quantum game theory, making use of entanglement and superposition, surpasses the limitations of classical approaches. It provides advanced solutions for mitigating congestion problems in networks and finds optimal strategies in multi-agent systems. The research is based on the combination of quantum computing with game theory, aiming to enable the development of more efficient protocols in the routing of communication networks and algorithms for the coordination among autonomous agents. By implementing quantum strategies in the proposed congestion model, a reduction in network latency is achieved, showing an improvement in performance compared to classical strategies. Moreover, the integration of adaptive reinforcement learning algorithms allows autonomous systems in quantum environments to optimize strategic decision-making in real time, paving the way for practical applications in other fields such as logistics, urban traffic management, and the simulation of economic systems. These algorithms prove to be particularly effective in exploring quantum strategies, pure or mixed, and in adapting to environments with imperfect information, even when evaluated in both ideal and noisy environments, revealing the potential of the interaction between autonomous agents in real quantum contexts. The methodology employed, which combines mathematical modeling and computational simulations, provides a solid validation of the theoretical proposals, allowing the effectiveness of the developed algorithms to be verified.

Keywords: Quantum Computing, Game Theory, Communication Networks, Reinforcement Learning.

Acknowledgements

Haber tomado la decisión de intentar hacer un doctorado, pese a los momentos difíciles, inevitables, ha convertido los últimos cinco años en los mejores de mi vida. He tenido la suerte de trabajar en cuatro laboratorios distintos, durante mis estudios, y he conocido a las personas más increíbles que hicieron de esta experiencia un viaje intelectual y emocional que nunca me hubiese imaginado.

Cronológicamente, mi primer año lo pasé en el laboratorio de Cómputos. Fui tratado amablemente desde el principio a pesar de mi inexperiencia. José, Diego y Leo, mi agradecimiento más grande por la buena onda siempre y la experiencia ganada, trabajando con Linux, redes, caos (el caos que había en el laboratorio con todos las computadoras y monitores por todos lados) y cambiando el balde del aire acondicionado pinchado. Los quiero y les deseo lo mejor, no quiero dejar de mencionar a Felipe, una de las personas más apasionadas que conocí, con quién todos los días intercambiábamos ideas y quedaron múltiples proyectos pendientes que me dejarán pensando el resto de mi vida.

El siguiente año y medio, lo viví en el laboratorio de Componentes. Allí me reconcilié con la electrónica que había dejado de lado, siempre me dieron la libertad para trabajar en mi tesis aunque intentando fusionar los temas cuando fuese posible [1] y [2]. Claudio, Leo y Lucas, muchísimas gracias por la compañía, siempre fuimos un grupo silencioso, se recuerdan con alegría los días con horas sin que vuele una mosca por el laboratorio hasta que llegue Celeste con su energía para hacernos reír e interactuar.

Tuve la enorme fortuna de tener la posibilidad de viajar cuatro años consecutivos al International Centre for Theoretical Physics en Trieste, Italia, estando un total de 13 meses como estudiante, tutor y becario. Fue una experiencia inesperada que intenté aprovechar en cada segundo, como ya he dicho anteriormente, ningunas palabras serán suficientes para agradecer todo lo que me dieron, principalmente a María Liz y Andrés, pero destacando el trabajo del DreamTeam: Werner, Bruno, Romina, Iván, Luis, Maynor, Long y Cristian. Los siguientes pasos en mi carrera profesional no serían los que son sin todo lo que me enseñaron. ¡Aguante las FPGAs, las tortas de Romi, el ron y el ceviche guatemalteco!

El último año y medio tuve la suerte de estar en el lugar donde más cómodo me siento, el Departamento de Física de la Facultad de Ingeniería. Allí pude, por primera vez, dedicarme exclusivamente a los temas de mi doctorado aprovechando toda la experiencia ganada anteriormente. Luciana, Raúl, Juan Pablo y Maxi, ojalá en el futuro podamos seguir colaborando, muchas gracias por siempre estar y hacer lo posible para generar un clima de trabajo agradable.

¡A mis directores, los mejores que me podrían haber tocado! Aunque a pesar de que nunca compartimos oficina, ni laboratorio, nunca dejamos de encontrarnos. Extrañaré nuestras reuniones semanales por zoom durante los cinco años del doctorado, donde hablábamos de deporte, política, economía, el clima, chusmerío y un poco de física al final. Miguel, muchas gracias por la libertad que me diste para trabajar. Siempre me dejaste abrir mi camino, aunque también me guiaste, con tu criterio y experiencia que siempre respetaré, cuando creías que debía avanzar en una dirección y cuando no. Fue un orgullo enorme trabajar con vos y espero que podamos seguir haciéndolo, sos una gran persona y un gran científico. Gustavo, en 2016 pude convencerte de que quieras trabajar conmigo y desde ahí nunca dejamos de interactuar, me has ayudado tanto que nunca podré agradecerte suficiente, hemos tenido infinitas charlas, como alumno/docente, becario/director, colegas, paciente/psicólogo,

pero principalmente amigos. Sos una de las mejores personas que conocí en mi vida y mi deseo es que sigamos en contacto tanto como sea posible.

No quiero agradecer particularmente a familiares ni amigos porque la lista sería interminable, pero sí quiero agradecer a mi mamá. Durante estos años logré independizarme y sé que nada de esto hubiese sido posible sin ella, ¡nada! Gracias de nuevo, nunca me voy a cansar de decírtelo, te quiero mucho.

También a la educación pública. Tengo la fortuna de ser egresado de Escuela Técnica pública, Facultad de Ingeniería pública y ahora doctorando en la Universidad Nacional becado por CONICET. Durante estos cinco años fui docente en cuatro materias de grado de la Universidad Nacional de Mar del Plata y estaré eternamente agradecido. ¡Aguante la educación pública!

Por último, lo mejor para el final, mi esposa, Meli, Vidamí, gracias por acompañarme en todo momento. Todo es tan gratificante cuando estamos juntos, caminando, comiendo, viendo películas y discutiendo horas sobre las diferencias entre una 'suprema' y una 'milanesa de pollo'. Me apoyaste cuando tuvimos que tomar decisiones difíciles, la distancia no es fácil pero siempre hacemos lo mejor posible sabiendo que es lo mejor para los dos. Estoy muy orgulloso de que estés a mi lado, sos una hermosa persona.

Estos cinco años fueron maravillosos, ¡mañana es mejor!

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

En la vanguardia de la transformación tecnológica, la computación cuántica y la teoría de juegos emergen como un faro de innovación, prometiendo transformar el diseño de algoritmos para redes de comunicación congestionadas y el aprendizaje por refuerzo multi-agente. Esta convergencia, que da como resultado un nuevo área denominado teoría de juegos cuántica, no solo abre nuevas avenidas en la investigación, sino que también redefine los límites del desarrollo tecnológico. Al adentrarnos en esta exploración, nos enfrentamos al desafío de descifrar complejidades nunca antes vistas, con el potencial de desbloquear soluciones a problemas que, hasta ahora, parecían insuperables. La promesa de esta sinergia radica tanto en su capacidad para optimizar el ruteo en redes de comunicación sobrecargadas como también en reconfigurar el paisaje del aprendizaje automático, marcando el inicio de una era donde física cuántica, la teoría de la toma de decisiones, el internet cuántico y los sistemas autoadaptativos se entrelazan para crear algoritmos que permitan importantes avances.

En un mundo donde la digitalización y las redes de comunicación son esenciales, enfrentamos una creciente demanda de soluciones innovadoras para optimizar la congestión de redes de comunicación. La investigación en teoría de juegos cuántica y aprendizaje por refuerzo multi-agente aborda estos desafíos al explorar estrategias que las metodologías clásicas no pueden alcanzar. Este enfoque promete mejorar la eficiencia y la capacidad de las redes de comunicaciones y abrir nuevas posibilidades para sistemas autónomos más inteligentes. En este contexto, nuestra investigación se vuelve crítica, ofreciendo la promesa de superar algunas de las limitaciones contemporáneas.

La teoría de juegos cuántica, al fusionar los principios de la mecánica cuántica con la teoría de juegos tradicional, inaugura un campo de posibilidades inexploradas para enfrentar desafíos en redes de comunicación y sistemas multi-agente. Al introducir estados de superposición y entrelazamiento, esta teoría permite la concepción de estrategias que trascienden las limitaciones clásicas, ofreciendo soluciones más eficientes a la congestión de redes y promoviendo un aprendizaje más robusto y adaptativo en entornos multi-agente. Este paradigma no solo redefine las estrategias óptimas en situaciones de conflicto y cooperación, sino que también abre el camino hacia algoritmos de comunicación y aprendizaje más poderosos y eficaces en sistemas distribuidos.

El modelado de redes de comunicación mediante teoría de juegos, antes clásico y ahora cuántico, representa un avance crucial frente a los desafíos de congestión que limitan la eficiencia de las infraestructuras actuales. Al aplicar principios cuánticos a la teoría de juegos, este enfoque permite explorar soluciones innovadoras que superan las restricciones de los modelos clásicos. La investigación en este dominio promete mitigar los problemas de congestión mediante estrategias óptimas de enrutamiento y asignación de recursos y, al mismo tiempo, mejorar la seguridad y la

capacidad de adaptación de las redes frente a demandas dinámicas y fluctuantes. La relevancia de este modelo radica en su potencial para transformar las comunicaciones globales, elevando la robustez, eficiencia y flexibilidad de las redes del futuro.

La sección 3.2.1 introduce un modelo innovador que aplica la teoría de juegos al problema de congestión en redes de datos, provocado por el aumento exponencial de usuarios móviles y el volumen de datos transmitidos [3]. Este modelo, en su versión cuántica, representa un salto cualitativo frente a su contraparte clásica y permite a los paquetes de datos, que actuan de manera egoísta, seleccionar las rutas más eficientes minimizando el tiempo de transmisión por paquete. Mediante la implementación de estrategias cuánticas, basadas en un modelo de compuertas cuánticas de un qubit con tres parámetros, se logra una notable reducción en la latencia, principalmente en condiciones de alta demanda de datos. Este enfoque ofrece tanto una solución creativa al desafío persistente de la congestión de red, como evidencias de las mejoras significativas en el rendimiento general de la red. La aplicación de la teoría de juegos cuántica en este contexto destaca el potencial transformador de la mecánica cuántica en la optimización de redes de comunicación, marcando un avance importante hacia el aprovechamiento de tecnologías cuánticas en la mejora de los sistemas de comunicación globales.

La sección 3.2.2 profundiza en la aplicación avanzada de la teoría de juegos cuántica para abordar el problema persistente de la congestión en redes de datos [4]. Este análisis detallado revela cómo la implementación de estrategias cuánticas, tanto puras como mixtas, puede ofrecer soluciones efectivas a la congestión, explorando conceptos claves como el equilibrio de Nash y la optimalidad de Pareto. La investigación se adentra en la complejidad añadida por el entrelazamiento cuántico parcial y el ruido cuántico, evaluando su impacto en la eficacia de los protocolos de enrutamiento. A pesar de los desafíos presentados por la naturaleza NISQ [5] de los dispositivos cuánticos actuales, que introduce variables como la decoherencia, se demuestra que los protocolos cuánticos conservan sus ventajas, manteniendo un rendimiento sólido incluso en condiciones adversas. Esta sección no solo valida la aplicabilidad de la teoría de juegos cuántica en el entorno desafiante de las redes congestionadas sino que también abre puertas a futuras investigaciones y aplicaciones prácticas en el ámbito de la comunicación cuántica, subrayando la robustez frente a sistemas no ideales y el potencial transformador de esta aproximación frente a la congestión de redes.

El capítulo 3.3 aborda el desarrollo de un protocolo basado en aprendizaje por refuerzo para el enrutamiento en redes cuánticas, destacando cómo este enfoque facilita la adaptación de estrategias en respuesta a las dinámicas de la red, como la variabilidad en la congestión y el volumen de tráfico [6]. En primer lugar, a través de una comparativa detallada con los métodos clásicos, se evidencian las condiciones específicas bajo las cuales el protocolo cuántico es superior al clásico, enfocándose en el número mínimo de paquetes en la red y la distancia máxima entre nodos. Por otro lado, se incorpora un algoritmo de aprendizaje por refuerzo al protocolo cuántico que se distingue por su habilidad para superar los desafíos inherentes a los sistemas estáticos, optimizando las de estrategias cuánticas de ruteo de manera dinámica basándose en las fluctuaciones en la carga de la red y mejorando la latencia global en la transmisión de datos. Este analisis se enfoca en la relevancia de la adaptabilidad en los algoritmos de aprendizaje, que pueden ser diseñados para explorar (invertir mucho tiempo en buscar la estrategia óptima) o explotar (invertir poco tiempo en buscar la estrategia óptima) dependiendo de la estabilidad o volatilidad del entorno de red. Este trabajo sienta un precedente importante para el avance en el campo de la comunicación cuántica y el enrutamiento adaptativo, marcando un paso adelante

en la búsqueda de soluciones eficientes para la gestión de redes de datos cuánticas.

El aprendizaje por refuerzo multi-agente, cuando se complementa con algoritmos cuánticos, abre un nuevo horizonte de eficiencia y posibilidades en el ámbito del aprendizaje automático. Esta fusión permite a los sistemas aprender y adaptarse a ambientes complejos con dinámicas sin precedentes, explorando un espacio de soluciones mucho más amplio que el accesible por métodos clásicos. Más allá de las redes de comunicación, esta aproximación tiene el potencial de aplicarse a campos como la optimización de sistemas logísticos, la gestión de tráfico urbano y la simulación de sistemas económicos, demostrando su versatilidad y el valor transformador de los principios cuánticos aplicados al aprendizaje automático.

En la sección 4.2.1, se aborda el desafío de aprender estrategias mixtas en juegos cuánticos caracterizados por la presencia de información imperfecta [7]. Este segmento resalta la manera en que las propiedades cuánticas, como las superposiciones y el entrelazamiento, amplían el espectro de estrategias disponibles, introduciendo una complejidad sin precedentes en la toma de decisiones estratégicas. A través de la implementación de un algoritmo descentralizado de aprendizaje por refuerzo multi-agente, se muestra cómo los agentes pueden desarrollar y adaptar estrategias puras o mixtas sin conocimiento previo sobre las acciones o recompensas de otros participantes, ni siquiera del juego específico en el que están involucrados. Este método innovador no solo navega eficazmente por el complejo espacio estratégico cuántico, sino que también demuestra ser resiliente frente a desafíos como el entrelazamiento parcial y el ruido en los canales cuánticos. La habilidad de adaptación de los agentes en entornos con información incompleta y condiciones ruidosas es vital, reflejando la realidad operativa de los sistemas cuánticos y marcando un avance significativo hacia la aplicación práctica de la teoría de juegos cuántica en escenarios de información imperfecta.

La sección 4.2.2, presenta un avance significativo en la intersección de la teoría de juegos cuántica y el aprendizaje por refuerzo, introduciendo el algoritmo QESRL (Exploración Egoísta de Aprendizaje por Refuerzo Cuántico). Este innovador enfoque se centra en juegos cuánticos repetidos de suma no cero, donde se busca maximizar las recompensas individuales de los agentes y asegurar una distribución equitativa de estas recompensas. A través de este algoritmo, los agentes aprenden estrategias cuánticas que les permiten adaptarse y optimizar su rendimiento en el contexto de interacciones estratégicas complejas. Lo que distingue a QESRL es su énfasis en la equidad, logrando una distribución de recompensas más justa entre los agentes en comparación con las técnicas convencionales de aprendizaje por refuerzo en entornos cuánticos. Este enfoque mantiene la adaptabilidad y las dinámicas de convergencia de las estrategias en juegos cuánticos y, al mismo tiempo, ofrece soluciones al desafío de equilibrar beneficios individuales y colectivos en escenarios multiagente. La implementación de QESRL representa un paso adelante en la comprensión y gestión de la competencia y cooperación en juegos cuánticos, abriendo nuevas vías para la exploración de comportamientos estratégicos en sistemas cuánticos complejos.

Finalmente, la sección 4.3, presenta un nuevo algoritmo de aprendizaje por refuerzo multi-agente diseñado para optimizar el rendimiento en juegos cuánticos, utilizando estrategias de aprendizaje basadas en gradientes [8]. Este enfoque avanzado permite a los agentes maximizar sus recompensas locales a través de una adaptación precisa y efectiva a los dinámicos entornos de juego cuántico, donde la complejidad de la competencia por las recompensas incrementa junto con el número de participantes. Caracterizado por su capacidad para utilizar el ruido cuántico inherente al sistema como una ventaja estratégica, este método abre camino hacia una

mejor comprensión de cómo la interacción entre los principios de la mecánica cuántica y los algoritmos de aprendizaje automático puede ser explotada para mejorar la toma de decisiones y la adaptabilidad de los agentes. Al establecer un marco teórico y práctico sólido, este capítulo ilustra el potencial de los algoritmos basados en la estimación del gradiente para facilitar una optimización efectiva en escenarios de juegos cuánticos multi-agentes, subrayando el impacto significativo de estas técnicas en el avance de la integración entre dispositivos cuánticos reales y el aprendizaje por refuerzo multiagente.

Nuestra investigación ha marcado metas significativas, destacando la creación de algoritmos cuánticos para el modelado y la optimización de redes de comunicación y el aprendizaje por refuerzo. Estos avances permiten una modelización de redes más eficaz, abordando la congestión de manera innovadora, y facilitan un aprendizaje por refuerzo multi-agente que adapta y optimiza las estrategias en tiempo real. Estos logros no solo evidencian el poder transformador de la computación cuántica aplicada a problemas complejos sino que también establecen un precedente para futuras investigaciones y aplicaciones en tecnologías emergentes y sistemas inteligentes.

La estructura de esta tesis se articula en capítulos diseñados para explorar exhaustivamente la intersección de la teoría de juegos cuántica con el modelado de redes de comunicación y el aprendizaje por refuerzo multi-agente. El Capítulo 2 establece el marco teórico, introduciendo los fundamentos necesarios de la computación cuántica, la teoría de juegos, el ruteo en redes de comunicación y el aprendizaje por refuerzo multi-agente. El Capítulo 3 se centra en la aplicación de la teoría de juegos cuántica para abordar técnicas de ruteo en redes de comunicación congestionadas, mientras que el Capítulo 4 avanza sobre el uso de algoritmos descentralizados para aprendizaje por refuerzo multi-agente en juegos cuánticos. Cada capítulo contribuye a los objetivos generales de la investigación, construyendo un entendimiento multifacético del potencial de la computación cuántica en la solución de problemas contemporáneos.

La metodología de esta investigación se fundamenta en un enfoque bipartito que combina modelado matemático y simulaciones computacionales. Este enfoque multidimensional permite una exploración profunda y rigurosa de las dinámicas complejas en la intersección de la teoría de juegos y la computación cuántica. El análisis teórico proporciona la base conceptual, mientras que las simulaciones computacionales ofrecen evidencia empírica de las teorías propuestas. El modelado matemático, por su parte, permite una abstracción y generalización de los resultados, asegurando su validez y aplicabilidad en diversos contextos.

Este trabajo aporta a la computación cuántica y la teoría de juegos, ofreciendo nuevos algoritmos y modelos para las telecomunicaciones y el aprendizaje automático. Destaca por su potencial en aplicaciones prácticas, como optimización de tráfico en redes y desarrollo de sistemas inteligentes más eficaces. Estas contribuciones no solo avanzan el conocimiento teórico sino que también prometen impactos significativos en tecnologías emergentes, facilitando soluciones innovadoras a desafíos actuales y futuros en la era de la información.

Esta investigación pretende trascender el ámbito académico, proyectándose como un catalizador para futuros avances en la teoría de juegos cuántica aplicada. Al desentrañar complejidades de redes de comunicación y sistemas autoadaptativos, establece un precedente para la exploración de soluciones cuánticas a problemas relevantes. Pensamos que es posible que sus hallazgos inspiren nuevos desarrollos tecnológicos y permitan avanzar en el corpus de conocimiento, motivando la investigación en de la toma de decisiones y las tecnologías cuánticas.

1.1. Objetivos Generales

- Modelar sistemas de redes de comunicación utilizando la teoría de juegos cuántica.
- Diseñar algoritmos cuánticos de ruteo que minimicen la latencia en redes de comunicación congestionadas.
- Modelar sistemas de multiples agentes aprendiendo independientemente en juegos cuánticos.
- Diseñar algoritmos de aprendizaje por refuerzo que maximicen las recompensas de agentes en juegos cuánticos de suma no cero.

1.2. Objetivos Particulares

- Desarrollar un modelo idealizado de redes de comunicaciones congestionadas para comparar su rendimiento bajo la implementación de diversos protocolos, tanto clásicos como cuánticos.
- Investigar exhaustivamente las implicaciones de aplicar protocolos cuánticos de enrutamiento en redes de comunicación, enfocándose en a) condiciones ideales, b) diversos grados de entrelazamiento cuántico, y c) distintos niveles de ruido cuántico.
- Implementar un protocolo de enrutamiento autoadaptativo en redes de comunicación que se optimice a partir de la experiencia, reduciendo la latencia a través de algoritmos de aprendizaje por refuerzo.
- Diseñar y validar algoritmos distribuidos de aprendizaje por refuerzo multi-agente para juegos cuánticos, donde los agentes operen sin información previa del entorno y sean capaces de aprender tanto estrategias puras como mixtas.
- Explorar la eficacia y equidad en la distribución de recompensas de distintos algoritmos de aprendizaje por refuerzo multi-agente en juegos cuánticos, donde los agentes busquen maximizar exclusivamente su propia recompensa local.
- Desarrollar y aplicar en diversas situaciones algoritmos de aprendizaje por refuerzo multi-agente basados en la estimación del gradiente de la función de recompensa de juegos cuánticos, seguido de un ascenso en la dirección del gradiente, por parte de cada agente de manera independiente.

1.3. Publicaciones

1.3.1. Como primer autor

Revistas internacionales

1. “Mitigation of Routing Congestion on Data Networks: A Quantum Game Theory Approach”, *Agustin Silva*, Omar Gustavo Zabaleta y Constancio Miguel Arizmendi. **Quantum Reports** MDPI Journal [4] (2021).

2. “Learning Mixed Strategies in Quantum Games with Imperfect Information”, *Agustin Silva*, Omar Gustavo Zabaleta y Constancio Miguel Arizmendi. **Quantum Reports** MDPI Journal [7] (2022).

3. "Maximizing Local Rewards on Multi-Agent Quantum Games through Gradient-Based Learning Strategies", *Agustin Silva*, Omar Gustavo Zabaleta y Constancio Miguel Arizmendi. **Entropy** MDPI Journal [8] (2023).

Congresos internacionales

1. “Quantum Game Theory approach for data network routing: a solution for the congestion problem”, *Agustin Silva*, Omar Gustavo Zabaleta y Constancio Miguel Arizmendi. CCP 2021 (Londres, UK, online) and **Journal of Physics** Proceedings [3].

2. “Learning-based Protocol for Routing in Quantum Networks”, *Agustin Silva*, Omar Gustavo Zabaleta y Constancio Miguel Arizmendi. COSY 2022 (Bolonia, Italia, presencial) and **IFAC** Proceedings [6].

3. "QESRL: Exploring Selfish Reinforcement Learning for Repeated Quantum Games", *Agustin Silva*, Omar Gustavo Zabaleta y Constancio Miguel Arizmendi. IC-MSQUARE 2023 (Belgrado, Serbia, presencial) and **Journal of Physics** Proceedings [falta cita, artículo aceptado pero todavía no publicado].

Congresos nacionales

1. “SoC FPGA implementation of Hopfield Neural Network for solving the Shortest Path Problem”, *Agustin Silva* y Claudio Gonzalez. **CASE** 2021 Proceedings [1].

2. “Acceleration of Parameterized Quantum Circuits Simulation in SoC FPGA”, *Agustin Silva*, Omar Gustavo Zabaleta y Claudio Gonzalez. **CASE** 2022 Proceedings [2].

1.3.2. Otros

1. “Looking for suitable rules for true random number generation with asynchronous cellular automata”, séptimo autor. **Nonlinear Dynamics** Springer Journal [9].

2. "HyperFPGA: An Experimental Testbed for Heterogeneous Supercomputing", cuarto autor. **The Journal of Supercomputing** Springer [10].

1.4. Materias cursadas durante el doctorado

1.4.1. En Universidad Nacional de Mar del Plata

1. Procesamiento Cuántico de Datos, 2019.

2. Inteligencia Computacional, 2020.

3. Diseño Digital con Técnicas de Alto Nivel y Principios de Diseño VLSI, 2021.

4. Teoría de la Información y Codificación, 2022.

5. Filosofía de la Ciencia, 2022.

1.4.2. En otras Universidades Nacionales

1. Redes Neuronales, Universidad Nacional de Córdoba (FAMAF-UNC).
2. Teoría de la Información Cuántica, Universidad Nacional de La Plata (Exactas-UNLP).
3. Mecánica cuántica y Elementos de Computación Cuántica, Universidad de Buenos Aires (FI-UBA).
4. Inteligencia Artificial y Aprendizaje Profundo en Física, Universidad de Buenos Aires (Exactas-UBA).

1.4.3. En ICTP, Trieste, Italia

1. Quantum Photonics and Information, 2020 | (smr 3424).
2. Machine Learning for Condensed Matter, 2021 | (smr 3589).
3. Quantum Dynamics: From Electrons to Qbits, 2022 | (smr 3733).
4. Collaborative Scientific Software Development and Management of Open Source Scientific Packages, 2023 | (smr 3894).

Capítulo 2

Marco Teórico

2.1. Computación Cuántica

2.1.1. Introducción

¿Qué caracteriza a una computadora cuántica y cómo se diferencia de las computadoras clásicas? Esta cuestión no solo nos invita a explorar la mecánica cuántica, la teoría de la información y las ciencias de la computación, sino que también nos lleva a una comprensión de cómo estas áreas convergen en la tecnología de la computación cuántica. Aunque las computadoras clásicas, aquellas que operan sin emplear directamente principios cuánticos, también se rigen por las leyes de la mecánica cuántica que fundamentan nuestro universo físico, no explotan las propiedades únicas que esta ofrece para el procesamiento de la información. En cambio, una computadora cuántica se distingue por su habilidad de aprovechar estas propiedades específicas de la mecánica cuántica, abriendo así nuevos horizontes en el campo del cálculo computacional.

Según los principios de la mecánica cuántica, los sistemas se establecen en un estado definido solo una vez que se miden. Antes de una medición, los sistemas están en un estado indeterminado; después de medirlos, están en un estado definido. Si tenemos un sistema, por ejemplo, que puede adoptar uno de dos estados discretos al medirse, podemos representar los dos estados en notación de Dirac como $|0\rangle$ y $|1\rangle$. La notación de Dirac, utiliza "kets"$|\psi\rangle$ para representar estados y "bras"$\langle\phi|$ para sus conjugados hermitianos. Este tipo de notación facilita cálculos como el producto escalar $\langle\phi|\psi\rangle$, que mide la amplitud de probabilidad entre estados cuánticos. Luego podemos representar una superposición de estados como una combinación lineal de estos estados, como: $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$. Todos estos conceptos mencionados se desarrollarán de manera más rigurosa en el resto del capítulo.

El Principio de Superposición postula que la combinación lineal de dos o más vectores pertenecientes a un espacio de Hilbert, una generalización del espacio euclideano, es otro vector de en el mismo espacio de Hilbert.

Como ejemplo, todavía en el mundo clásico, consideremos una propiedad intrínseca de la luz que ilustra una superposición de estados: la polarización. En casi toda la luz que vemos en la vida cotidiana, como la del sol, no hay una dirección preferida para la polarización. Los estados de polarización se pueden seleccionar mediante un filtro polarizador, una película delgada con un eje que solo permite pasar la luz con polarización paralela a ese eje.

Con un solo filtro polarizador, podemos seleccionar una polarización de la luz, por ejemplo, la polarización vertical, que podemos denotar como $|\uparrow\rangle$. La polarización horizontal, que podemos denotar como $|\rightarrow\rangle$, es un estado ortogonal a la

polarización vertical (podríamos haber usado $|0\rangle$ y $|1\rangle$ para denotar los dos estados de polarización; las etiquetas utilizadas en los kets son arbitrarias). Juntos, estos estados forman una base para cualquier polarización de la luz. Es decir, cualquier estado de polarización $|\psi\rangle$ se puede escribir como combinación lineal de estos estados. Utilizamos la letra griega ψ para denotar un estado genérico del sistema: $|\psi\rangle = \alpha|\uparrow\rangle + \beta|\rightarrow\rangle$.

Los coeficientes α y β son números complejos conocidos como amplitudes. En este ejemplo, el coeficiente α está asociado con la polarización vertical y el coeficiente β con la horizontal. La amplitud tiene una interpretación importante en la mecánica cuántica que veremos en breve. Después de seleccionar la polarización vertical con un filtro polarizador, podemos introducir un segundo filtro polarizador después del primero. Imaginemos que orientamos el eje del segundo filtro perpendicular al eje del primero. En este caso, el estado horizontal $|\rightarrow\rangle$ es ortogonal al primero, por lo que no hay cantidad de polarización horizontal después del primer filtro vertical.

Supongamos ahora que orientamos el eje del segundo filtro polarizador a 45° (a lo largo de la diagonal $|\nearrow\rangle$ entre $|\uparrow\rangle$ y $|\rightarrow\rangle$) en lugar de horizontalmente. En este caso, tal vez para la sorpresa de algunos veríamos pasar algo de luz por el segundo filtro. ¿Cómo podría ser esto si toda la luz después del primer filtro tiene polarización vertical? La razón es que podemos expresar la polarización vertical como una superposición de componentes diagonales. Dejando que $|\nearrow\rangle$ denote la polarización de 45° y $|\nwarrow\rangle$ denote -45°, podemos escribir $|\uparrow\rangle = \frac{|\nearrow\rangle + |\nwarrow\rangle}{\sqrt{2}}$.

Como se puede esperar desde la intuición geométrica, el estado vertical consta de partes iguales de $|\nearrow\rangle$ y $|\nwarrow\rangle$. Es por esta razón que vemos pasar algo de luz por el segundo filtro. Es decir, la polarización vertical se puede escribir como una superposición de estados, uno de los cuales es precisamente el estado diagonal de 45° $|\nearrow\rangle$ que estamos permitiendo pasar a través del segundo filtro. Dado que el estado $|\nearrow\rangle$ es solo un término en la superposición, no toda la luz pasa por el filtro, pero algo sí lo hace. La intensidad de la luz transmitida es precisamente la mitad de la luz incidente. Este valor se determina a partir de las amplitudes del estado de superposición mediante una ley conocida como la regla de Born.

2.1.2. Regla de Born

La regla de Born establece que, en una superposición de estados cuánticos, el módulo al cuadrado de la amplitud de un estado es la probabilidad de que la medición dé como resultado ese estado. Además, la suma de los cuadrados de las amplitudes de todos los estados posibles en la superposición es igual a 1. Así, para el estado $|\psi\rangle = \alpha|\uparrow\rangle + \beta|\rightarrow\rangle$, tenemos $|\alpha^2| + |\beta^2| = 1$:

Utilizando los valores del ejemplo de la luz, dado que la amplitud es $\frac{1}{\sqrt{2}}$, la probabilidad de obtener ese estado es $|\frac{1}{\sqrt{2}}|^2 = 0{,}5$, por lo que la probabilidad de medir la luz en el estado de polarización vertical u horizontal sería del 50 %. Mientras que en este ejemplo tenemos una división de probabilidad 50/50 para cada uno de los dos estados, si examináramos algún otro sistema físico podría tener cualquier otra distribución arbitraria de probabilidad. Una diferencia crítica entre la mecánica clásica y cuántica es que las amplitudes (no las probabilidades) pueden ser números complejos.

2.1.3. Ecuación de Schrödinger

Como vimos anteriormente, podemos representar el estado de un sistema con un vector de estado usando la notación de Dirac. Por ejemplo, continuando con el análisis de la polarización, podemos representar el vector de estado de un fotón con la letra griega $|\Psi\rangle$ en el ket como: $|\Psi\rangle = \frac{|\uparrow\rangle+|\rightarrow\rangle}{\sqrt{2}}$. Si midiésemos este fotón en cuanto a polarización, tendríamos una probabilidad del 50 % de encontrar el fotón en un estado de polarización vertical y un 50 % de probabilidad de encontrarlo en un estado de polarización horizontal. La escuela de Copenhague de la Mecánica Cuántica dice que la función de onda se ha colapsado en uno de los dos estados.

Ahora que tenemos un método para representar el estado de un sistema con una función de onda, podemos representar la evolución de este sistema a lo largo del tiempo. Consideremos la función de onda de una partícula que se mueve en un espacio unidimensional, la cual podemos representar como $\psi(x,t)$, donde x y t representan la posición y el tiempo respectivamente. La manera en que esta función de onda evoluciona con el tiempo puede describirse mediante la ecuación de Schrödinger (ES) dependiente del tiempo, la cual se puede escribir de la siguiente manera para una partícula a lo largo de una dimensión:

$$i\hbar\frac{\partial}{\partial t}\Psi(x,t) = \hat{H}\Psi(x,t) \tag{2.1}$$

En el lado izquierdo de esta ecuación, vemos las constantes i y $\hbar$. i es el símbolo que representa la unidad imaginaria en matemáticas. $\hbar$ es la forma reducida de la constante de Planck h, llamada así porque se obtiene dividiendo h entre 2π. Los siguientes símbolos representan la derivada parcial de la función de onda con respecto al tiempo t. El lado derecho de la ecuación representa la aplicación del operador Hamiltoniano a la función de onda $\Psi(x,t)$. El Hamiltoniano representa la energía total de un sistema; esta energía total incluye toda la energía cinética y potencial de las partículas en el sistema. En resumen, la ecuación de Schrödinger nos dice que si queremos saber cómo cambiará la función de onda $\Psi(x,t)$ con el tiempo, necesitamos conocer la energía total del sistema. Por lo tanto, como veremos a continuación, la ES fundamenta la dinámica de los qubits (bits cuánticos) y los operadores (compuertas cuánticas). En este marco, los qubits evolucionan bajo la influencia de operadores que, a su vez, son regidos por las soluciones de la ecuación de Schrödinger, integrando así los principios de la mecánica cuántica en la arquitectura de la computación cuántica.

2.1.4. Bits cuánticos

Un qubit es similar a un bit clásico en que puede tomar los estados 0 o 1, pero se diferencia de un bit en que también puede tomar un rango continuo de valores que representan una superposición de estados. Aunque generalmente utilizamos sistemas de qubits de dos niveles para construir computadoras cuánticas (CCs), también podemos elegir otras arquitecturas de computación. Por ejemplo, podríamos construir una compuradora cuántica con qutrits, que son sistemas de tres niveles. Podemos pensar en estos como teniendo estados de 0, 1 o 2 o una superposición de estos estados.

Un sistema de qubits de, digamos, 100 qubits puede manejar 2^{100} estados (1,26e30), mientras que un sistema de qutrits puede manejar 3^{100} estados (5,15e47), un número que es 17 órdenes de magnitud mayor. Dicho de otra manera, para representar el mismo espacio numérico que un sistema de 100 qubits, solo necesitamos 63 qutrits

($\log_3(2^{100})$). Dado que es más difícil construir sistemas de qutrits, las computadoras cuánticas convencionales se basan actualmente en qubits. Ya sea que elijamos qubits, qutrits u otro número qudit (d niveles), cada uno de estos sistemas puede ejecutar cualquier algoritmo que los demás puedan, es decir, pueden simularse entre sí.

En el marco de la Mecánica Cuántica, los estados se modelan como vectores y los operadores como matrices. Empleamos la notación de Dirac en lugar de los símbolos tradicionales del álgebra lineal para describir vectores y otras abstracciones. Comencemos con la definición de un qubit: Un qubit físico es un sistema cuántico de dos niveles de energía y que podemos representar en un espacio de Hilbert complejo bidimensional, $\mathbb{C}^2$. El estado de un qubit en cualquier momento puede ser representado por un vector en este espacio de Hilbert complejo.

El espacio de Hilbert está equipado, por definición, con un producto interno que nos permite determinar la posición relativa de dos vectores que representan estados de qubits. Denotamos el producto interno de los vectores $|u\rangle$, $|v\rangle$ como $\langle u|v\rangle$; esto será igual a 0 si $|u\rangle$ y $|v\rangle$ son ortogonales y 1 si $|u\rangle$ = $|v\rangle$. Para representar dos o más qubits podemos tensorizar espacios de Hilbert juntos para representar los estados combinados de los qubits utilizando el producto de Kronecker. Como veremos, tenemos métodos para representar estados separables, donde los qubits son independientes entre sí, y estados entrelazados, donde no podemos separar los estados de los dos qubits.

Los estados $|0\rangle$ y $|1\rangle$ pueden ser representados mediante vectores, como se ilustra a continuación. Estos dos estados son conocidos como la base computacional de un sistema de dos niveles. Luego, podemos aplicar operadores en forma de matrices a los vectores en el espacio de estados para modificar su valor.

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, |1\rangle = \begin{pmatrix}0\\1\end{pmatrix} \tag{2.2}$$

Hemos establecido que un qubit puede estar en uno de los estados de la base computacional de $|0\rangle$ o $|1\rangle$ o en una superposición de estos dos estados. Ahora, representamos la superposición de múltiples estados como una combinación lineal de las bases computacionales del espacio de estados, cada término en la superposición tiene un coeficiente complejo o amplitud. Utilizando los vectores de la base computacional podemos representar estados superpuestos de la siguiente manera:

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}, |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \tag{2.3}$$

Estos dos estados difieren por un signo menos en el estado $|1\rangle$. De manera más formal, llamamos a esta diferencia una fase relativa. El término fase tiene numerosos significados en física; en este contexto, se refiere a un ángulo. El signo menos está relacionado con el ángulo π a través de la identidad de Euler: $e^{i\pi} = -1$. Las fases relativas son de fundamental importancia para los algoritmos cuánticos ya que permiten la interferencia constructiva y destructiva. Por ejemplo, si evaluamos la suma de los estados anteriores, podemos eliminar el eliminar el estado $|1\rangle$ utilizando interferencia destructiva:

$$\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) = \frac{|0\rangle + |1\rangle}{2} + \frac{|0\rangle - |1\rangle}{2} = |0\rangle \tag{2.4}$$

La habilidad de las computadoras cuánticas para realizar interferencia constructiva y destructiva durante el cómputo es fundamental para las ventajas potenciales

de los algoritmos cuánticos sobre los clásicos, permitiendo un procesamiento y exploración más eficientes de soluciones complejas.

2.1.5. Entrelazamiento Cuántico

El entrelazamiento cuántico es un fenómeno fundamental en la mecánica cuántica, donde pares o grupos de partículas interactúan de tal manera que el estado cuántico de cada partícula no puede describirse de forma independiente respecto a las demás, incluso cuando están separadas por grandes distancias. Esta propiedad contrasta marcadamente con las expectativas clásicas y es clave para numerosos avances en la computación y la comunicación cuántica.

Para formalizar el concepto, consideremos un sistema compuesto de dos qubits. El espacio de estados de este sistema es el producto tensorial de los espacios de estados de cada qubit individual, denotado como $\mathbb{C}^2 \otimes \mathbb{C}^2$. Un estado no entrelazado o separable en este espacio se puede escribir como el producto tensorial de dos estados de qubits individuales, es decir:

$$|\psi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \tag{2.5}$$

donde $|\phi_1\rangle$ y $|\phi_2\rangle$ son estados de los qubits individuales. Sin embargo, un estado entrelazado no puede descomponerse de esta manera. Un ejemplo típico de un estado entrelazado es unos de los estados de Bell:

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{2.6}$$

Este estado describe una superposición en la que los dos qubits están perfectamente correlacionados: si uno se mide en el estado $|0\rangle$, el otro también se encontrará en el estado $|0\rangle$, y lo mismo ocurre para el estado $|1\rangle$. El entrelazamiento es una propiedad no local, lo que significa que las mediciones en un qubit afectan instantáneamente al estado del otro qubit entrelazado, independientemente de la distancia que los separe.

El grado de entrelazamiento de un sistema bipartito puede cuantificarse, por ejemplo, mediante medidas como la Entropía de Von Neumann:

$$S(\rho_A) = -\mathrm{Tr}(\rho_A \log \rho_A)$$

, siendo $\rho_A = \mathrm{Tr}_B(\rho)$ el estado reducido de uno de los subsistemas la matriz densidad del sistema completo. Estas medidas ayudan a entender la "fuerza"del entrelazamiento y su utilidad potencial en diferentes aplicaciones cuánticas. El entrelazamiento cuántico desafía nuestras nociones intuitivas de separabilidad y localidad, y es un recurso esencial en la tecnología cuántica emergente, abriendo puertas a nuevas formas de procesamiento y transmisión de información que son imposibles en el marco clásico.

2.1.6. Operadores cuánticos

Ahora analicemos el conjunto de operadores cuánticos comúnmente utilizados. Denotamos un operador de un solo qubit con una caja que contiene la letra que representa ese operador y un cable que la atraviesa. Denotamos una compuerta de dos qubits con una caja que abarca dos cables y una compuerta de tres qubits con una caja con tres cables, etc.

Operadores de un qubit

Comenzaremos cubriendo un conjunto de operadores cuánticos de un solo qubit. Los primeros tres operadores que examinaremos son los operadores de Pauli. Estas tres matrices, junto con la matriz identidad y todos sus múltiplos de +/-1, constituyen lo que se conoce como el grupo de Pauli. Las tres puertas cuánticas son X, Y y Z (también conocidas como σ_X, σ_Y y σ_Z), y se definen de la siguiente manera:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.7}$$

El resultado de aplicar, por ejemplo, la compuerta X al estado $|0\rangle$ es:

$$X|0\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0+0 \\ 1+0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |1\rangle \tag{2.8}$$

El resultado de aplicar, por ejemplo, la compuerta Y al estado $|1\rangle$ es:

$$Y|1\rangle = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0-i \\ 0+0 \end{pmatrix} = \begin{pmatrix} -i \\ 0 \end{pmatrix} = -i|0\rangle \tag{2.9}$$

El resultado de aplicar, por ejemplo, la compuerta Z al estado $|0\rangle$ es:

$$Z|0\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1+0 \\ 0+0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle \tag{2.10}$$

La representación gráfica de las compuertas de Pauli puede verse en la Fig. 2.1.

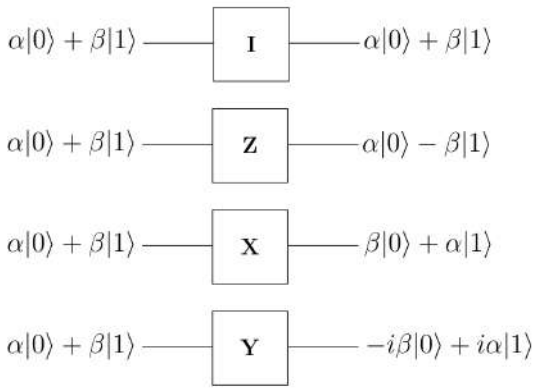


FIGURA 2.1: Representación visual de las compuerta cuánticas de un qubit de Pauli.

Luego, las matrices de rotación son operadores que modifican el estado del qubit utilizando un parámetro. Las rotaciones en X, Y y Z son versiones continuas de los operadores de Pauli denominadas $R_X(\theta)$, $R_Y(\theta)$ y $R_Z(\theta)$ y definidas como:

$$R_X(\theta) = \begin{pmatrix} cos(\frac{\theta}{2}) & -isin(\frac{\theta}{2}) \\ -isin(\frac{\theta}{2}) & cos(\frac{\theta}{2}) \end{pmatrix}; R_Y(\theta) = \begin{pmatrix} cos(\frac{\theta}{2}) & -sin(\frac{\theta}{2}) \\ sin(\frac{\theta}{2}) & cos(\frac{\theta}{2}) \end{pmatrix}; R_Z(\theta) = \begin{pmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{pmatrix}$$

Finalmente, el operador de Hadamard es crucial en la computación cuántica, ya que nos permite llevar un qubit de un estado de base computacional a una superposición de estados. La matriz de Hadamard es: $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

Si aplicamos la Hadamard al estado $|0\rangle$ obtenemos:

$$H|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1+0\\1+0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \tag{2.11}$$

Y si la aplicamos al estado $|1\rangle$, obtenemos:

$$H|1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}0+1\\0-1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \tag{2.12}$$

Entonces podemos ver que el operador H toma un estado de base computacional y lo proyecta en una superposición de estados $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ o $\frac{|0\rangle-|1\rangle}{\sqrt{2}}$, dependiendo del estado inicial.

Operadores de dos qubits

La base computacional en sistemas de dos qubits que se suele utilizar por convención es:

$$|00\rangle = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}; |01\rangle = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}; |10\rangle = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}; |11\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix} \tag{2.13}$$

En esta sección definiremos 3 compuertas de 2 qubits para presentar su funcionamiento.

En primer lugar la compuerta SWAP definida como: $SWAP = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$. La compuerta SWAP intercambia las amplitudes entre los estados de dos qubits. A continuación se puede observar el resultado de aplicar la compuerta SWAP al estado $|01\rangle$:

$$SWAP|01\rangle = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}0\\1\\0\\0\end{pmatrix} = \begin{pmatrix}0+0+0+0\\0+0+0+0\\0+1+0+0\\0+0+0+0\end{pmatrix} = \begin{pmatrix}0\\0\\1\\0\end{pmatrix} = |10\rangle \tag{2.14}$$

La compuerta CNOT está definida como: $CNOT = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}$. En este operador, identificamos al primer qubit como el qubit de control y al segundo como el qubit objetivo. Si el qubit de control está en el estado $|0\rangle$, entonces no hacemos nada al qubit objetivo. Sin embargo, si el qubit de control está en el estado $|1\rangle$, entonces aplicamos el operador X al qubit objetivo. Generalmente, utilizamos la puerta CNOT en computación cuántica para entrelazar dos qubits. A continuación se puede observar el resultado de aplicar la compuerta CNOT al estado $|10\rangle$:

$$CNOT|10\rangle = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0+0+0+0 \\ 0+0+0+0 \\ 0+0+0+0 \\ 0+0+1+0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = |11\rangle \tag{2.15}$$

Finalmente presentaremos el operador J, el cual utilizaremos con frecuencia en el resto de la tesis. El operador J se puede definir para N qubits y tiene la capacidad de convertir un estado con entrelazamiento mínimo en otro de entrelazamiento máximo. El operador J se define como $J = \frac{1}{\sqrt{2}}(\mathbb{I}^{\otimes N} + iX^{\otimes N})$ y para $N = 2$ tiene la siguiente forma: $J = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & i \\ 0 & 1 & i & 0 \\ 0 & i & 1 & 0 \\ i & 0 & 0 & 1 \end{pmatrix}$. A continuación se puede observar el resultado de aplicar la compuerta J al estado $|00\rangle$:

$$J|00\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & i \\ 0 & 1 & i & 0 \\ 0 & i & 1 & 0 \\ i & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ i \end{pmatrix} = \frac{|00\rangle + i|11\rangle}{\sqrt{2}} \tag{2.16}$$

2.1.7. Teoría de la Complejidad Cuántica

La Teoría de la Complejidad Cuántica explora cómo la computación cuántica redefine los límites de lo computable, se adentra en el análisis de la eficiencia de los algoritmos cuánticos y su capacidad para resolver problemas desafiantes dentro del marco de la computación clásica.

Un aspecto fundamental en la Teoría de la Complejidad Cuántica es la clasificación de los problemas según su grado de dificultad computacional en el contexto cuántico. Las clases de complejidad más importantes para nuestro análisis son: P (aquellos problemas que puede ser resueltos en tiempo polinómico por un ordenador clásico), NP (aquellos cuya solución puede ser verificada en tiempo polinómico por un ordenador clásico) y BQP (Bounded-Error Quantum Polynomial time) que representan conjuntos de problemas que pueden ser eficientemente solucionados por una computadora cuántica. Un problema está estrictamente en la clase BQP si existe un algoritmo cuántico que lo resuelve con una probabilidad de error menor a $1/3$ en un tiempo polinómico [11].

La superioridad cuántica se manifiesta, por ejemplo, en problemas como la factorización de números grandes, donde no se ha encontrado hasta el momento una solución eficiente en computadoras clásicas pero sí en computadoras cuánticas (BQP) mediante el algoritmo de Shor. Este algoritmo tiene una complejidad temporal polinómica de $O((\log N)^2(\log\log N)(\log\log\log N))$, contrastando marcadamente con las mejores soluciones clásicas conocidas que tienen una complejidad subexponencial de $O(e^{1,9(logN)^{\frac{1}{3}}(loglogN)^{\frac{2}{3}}})$ [12]. La capacidad de los algoritmos cuánticos para explotar el entrelazamiento y la superposición de estados permite un paralelismo y una eficiencia inalcanzables en la computación clásica.

Estos avances sugieren que el estudio de la complejidad cuántica no solo redefine

nuestra comprensión de lo computable, sino que también tiene implicaciones profundas en campos como la criptografía, optimización y simulación de sistemas físicos. Uno de los grandes objetivos de la computación cuánticas es encontrar problemas útiles donde la complejidad de las soluciones cuánticas sean significativamente menores que las mejores soluciones clásicas. Por lo tanto, la Teoría de la Complejidad Cuántica es una pieza crucial en la comprensión y el desarrollo de la computación cuántica y sus aplicaciones futuras.

Para más información sobre Computación Cuántica, referirse al libro [13].

2.2. Teoría de Juegos

2.2.1. Introducción

La teoría de juegos puede considerarse como una rama tanto de las matemáticas como de las ciencias aplicadas. Se ha utilizado en ciencias sociales, especialmente en economía, pero también ha penetrado en una variedad de otras disciplinas como ciencia política, biología, filosofía y, recientemente, ciencia de la computación y redes de comunicación.

La teoría de juegos no cooperativos es una de las ramas más importantes de la teoría de juegos, centrada en el estudio y análisis de la toma de decisiones competitivas que involucran a varios agentes. Proporciona un marco analítico adecuado para caracterizar las interacciones y el proceso de toma de decisiones. Este involucra a varios agentes con intereses parcial o totalmente en conflicto sobre el resultado de un proceso de decisión que se ve afectado por sus acciones. Ejemplos de juegos no cooperativos son ubicuos. En economía, las empresas que operan en el mismo mercado compiten por estrategias de precios, control del mercado, comercio de bienes y similares. En redes inalámbricas y de comunicación, los nodos inalámbricos están involucrados en numerosos escenarios no cooperativos como la asignación de recursos, elección de rangos de frecuencias o potencia de transmisión, reenvío de paquetes y gestión de interferencias.

Un juego no cooperativo involucra a varios agentes que tienen intereses totalmente o parcialmente en conflicto en el resultado de un proceso de decisión. Por ejemplo, consideremos varios nodos inalámbricos que intentan controlar su potencia de transmisión, dada la interferencia generada por otros nodos. En tal situación, aunque todos los nodos tienen incentivos para transmitir, la presencia de interferencia presenta un conflicto, acoplando las decisiones de los nodos: cada nodo quiere transmitir a su máximo nivel de potencia para mejorar su rendimiento; sin embargo, hacerlo aumenta la interferencia general en el sistema, lo que, a su vez, impacta negativamente en el rendimiento de todos los nodos inalámbricos involucrados.

Un juego no cooperativo es un juego que refleja una situación competitiva donde cada agente necesita tomar su decisión independientemente de los demás agentes, dadas las posibles elecciones de los otros agentes y su efecto en los objetivos o recompensas del agente. Cabe destacar que el término no cooperativo no siempre implica que los agentes no cooperen, sino que significa que cualquier cooperación que pueda surgir debe ser autoimpuesta, sin comunicación o coordinación de elecciones estratégicas entre los agentes.

2.2.2. Definiciones básicas

Al describir un juego no cooperativo estático o dinámico, la noción de forma estratégica (o normal) resulta ser una de las representaciones más populares. En este

sentido, un juego no cooperativo en forma estratégica (o normal) tiene tres componentes: el conjunto de agentes, sus estrategias y las recompensas. Más formalmente, un juego estratégico se define de la siguiente manera: $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$, donde:

- N es un conjunto finito de agentes, es decir, $N = \{1, ..., N\}$.
- S_i es el conjunto de estrategias disponibles para el agente i.
- $u_i : S \to \mathbb{R}$ es la función de recompensas para el agente i, donde $S = S_1 \times ... \times S_i \times ... \times S_N$ (producto cartesiano de los conjuntos de estrategias).

Dado el concepto de un juego estratégico, para cualquier agente i, cada elemento $s_i \in S_i$ es la estrategia de i. $s_{-i} = [s_j] j \in N; i \neq j$ denota el vector de estrategias de todos los agentes excepto i, y $s = (si, s_{-i}) \in S$ se conoce como un perfil de estrategia. Siempre que los conjuntos de estrategias S_i sean finitos para todos $i \in N$, el juego se denomina finito. En un juego en forma estratégica, cada agente debe seleccionar una estrategia para optimizar su función de recompensa. Siempre que cada agente $i \in N$ seleccione una estrategia $s_i \in S_i$ de manera determinista, es decir, con probabilidad 1, entonces esta estrategia se conoce como una estrategia pura.

Un juego se dice que tiene información completa si todos los elementos del juego son de conocimiento común entre todos los agentes. De lo contrario, se dice que el juego es de información incompleta. Así, en un juego con información completa, cada agente conoce las identidades de todos los demás agentes, sus estrategias y las recompensas que resultarían de cualquier combinación de estrategias. Para un juego con información incompleta, los agentes pueden no conocer las identidades de todos los demás agentes, sus recompensas o sus estrategias.

Uno de los tipos más comunes de juegos no cooperativos es el juego de suma cero de dos agentes, que involucra a dos agentes donde las ganancias de un agente son las pérdidas del otro agente. Por otro lado, los juegos de suma no cero describen situaciones en las que todos los agentes pueden considerarse maximizadores de su propia recompensa, sin tener ninguna restricción en la suma total de las recompensas. En cierto sentido, los juegos de suma no cero describen escenarios donde los participantes podrían ganar o perder juntos.

2.2.3. Juegos matriciales

Los juegos no cooperativos en forma estratégica son elementos fundamentales para comprender la toma de decisiones basada en la teoría de juegos. En estos juegos, el objetivo es determinar si existe un resultado o solución razonable para el juego. Una solución implica un conjunto de estrategias que los agentes, al actuar racionalmente, es decir, para optimizar su propia recompensa, seleccionarían.

Para analizar un juego no cooperativo en forma estratégica, primero se deben especificar claramente los agentes, sus estrategias y sus posibles recompensas. En este contexto, cualquier juego finito no cooperativo de dos agentes se puede representar en formato matricial, donde las estrategias de los agentes constituyen las filas y columnas de la matriz, y cada elemento es un par de números que representan las recompensas para los dos agentes cuando se usa una cierta combinación de estrategias. Un juego representado por una matriz se conoce como un juego matricial. La representación matricial de un juego se compone principalmente de lo siguiente:

- Cada fila representa una estrategia para el primer agente en el juego, a veces referido como el agente de la fila. Por lo tanto, el número de filas es igual al número de estrategias para el agente de la fila.

	Confesar (C)	No confesar (NC)
Confesar (C)	(-4, -4)	(0, -5)
No confesar (NC)	(-5, 0)	(-2, -2)

CUADRO 2.1: Dilema del prisionero

- Cada columna representa las estrategias del segundo agente en el juego, a veces referido como el agente de la columna. Por lo tanto, el número de columnas es igual al número de estrategias para el agente de la columna.
- Cada entrada de la matriz es un par (x, y), donde x representa la recompensa para el primer agente, es decir, el agente de la fila, y y representa la recompensa para el segundo agente, es decir, el agente de la columna.

El Dilema del Prisionero es posiblemente el ejemplo más famoso de juegos matriciales no cooperativos: Dos sospechosos son arrestados por un crimen y colocados en dos habitaciones aisladas. Cada uno de los sospechosos debe decidir si confesar o no e implicar al otro. Las reglas que rigen son las siguientes. Si ninguno de los sospechosos confiesa, cada uno cumplirá 2 años de cárcel. Si ambos confiesan y se implican mutuamente, ambos irán a prisión durante 4 años. Sin embargo, si un prisionero confiesa e implica al otro mientras el otro no confiesa, aquel que ha cooperado con la policía, es decir, ha confesado, será puesto en libertad, mientras que el otro pasará 5 años en prisión. En esta situación, se puede formular un juego no cooperativo en forma estratégica con los agentes siendo los dos prisioneros, y cada prisionero teniendo dos estrategias: confesar (estrategia C) o no confesar (estrategia NC). La recompensa para cada prisionero es simplemente el número de años que pasará en prisión. Una representación matricial de este juego se da en la Tabla 2.1. Las recompensas mostradas en la Tabla 2.1 son números negativos ya que tratamos con juegos donde los agentes buscan maximizar una recompensa. Finalmente, la Tabla 2.1 muestra claramente que este no es un juego de suma cero.

En el Dilema del Prisionero, observamos que los agentes tienen información completa, es decir, están al tanto de todos los elementos de la representación matricial. Una vez que un juego se expresa en forma estratégica o matricial, el siguiente paso es resolverlo. Resolver un juego implica predecir las estrategias que podrían ser adoptadas por cada agente y los posibles resultados del juego. En el resto de esta sección, discutimos cómo resolver juegos no cooperativos utilizando una variedad de conceptos.

2.2.4. Estrategias dominantes

Una noción útil para resolver juegos no cooperativos en forma estratégica es el concepto de estrategias dominantes. El uso de estrategias dominantes simplifica la solución de un juego eliminando algunas estrategias, es decir, filas o columnas en un juego matricial, que se sabe desde el principio que no tendrán efecto en el resultado del juego. Primero, examinamos el concepto de una estrategia dominante: Una estrategia $s_i \in S_i$ se dice que es dominante para el agente i si $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}), \forall s_i' \in S_i y \forall s_{-i} \in S_{-i}$, donde $S_{-i} = \prod_{j \neq i} S_j$ es el conjunto de todos los perfiles de estrategia para todos los agentes excepto i. Por lo tanto, una estrategia dominante es la mejor estrategia de un agente, es decir, la estrategia que produce la mayor recompensa para el agente independientemente de las estrategias que elijan los otros agentes.

Siempre que un agente tiene una estrategia dominante, un agente racional no tiene incentivo para elegir ninguna otra estrategia. En consecuencia, si cada agente posee una estrategia dominante, entonces todos los agentes elegirán sus estrategias dominantes. Esta elección intuitiva da lugar al siguiente concepto de solución para un juego no cooperativo: Un perfil de estrategia $s^* \in S$ es el equilibrio de estrategia dominante si cada elemento s_i^* de s^* es una estrategia dominante del agente i.

El concepto de equilibrio de estrategia dominante es un resultado natural para un juego dado. Por ejemplo, en El Dilema del Prisionero en la Tabla 2.1, cada agente obtiene un mejor resultado confesando, es decir, eligiendo C, independientemente de la elección estratégica del otro agente. Por lo tanto, (C, C) es un equilibrio de estrategia dominante que produce un vector de recompensas de (-4, -4). Cabe destacar que, aunque este punto es una solución para el juego, los recompensas recibidos no son los mejores para ambos agentes, ya que (-2, -2) proporcionaría una mayor recompensa a ambos agentes que (-4, -4). Revisitaremos el tema de la eficiencia en el resultado de un juego más adelante. Aunque el equilibrio de estrategia dominante es una solución intuitiva para un juego dado, no se garantiza la existencia de este punto de equilibrio. De hecho, hay muchos juegos en los que ningún agente tiene una estrategia dominante.

Más allá de la idea de una estrategia dominante, es útil definir el concepto opuesto, una estrategia estrictamente dominada, de la siguiente manera: Una estrategia $s_i' \in S_i$ de un agente i se dice que está estrictamente dominada por una estrategia $s_i \in S_i$ si $u_i(s_i, s_{-i}) > u_i(s_i', s_{-i}), ; \forall s_{-i} \in S_i$. Así, una estrategia está estrictamente dominada para un agente si este agente tiene otra estrategia que se desempeña mejor, independientemente de lo que elijan los demás agentes. Por lo tanto, dado que se tiene información completa, es natural que un agente racional elimine todas las estrategias estrictamente dominadas antes de tomar una decisión. Esto lleva al concepto de dominación estricta iterada, que puede ser utilizado para ayudar a resolver un juego matricial. La dominación estricta iterada implica eliminar todas las estrategias estrictamente dominadas en un juego dado. Al hacerlo, reducimos el número de posibilidades y, en algunos casos, podemos llegar a un resultado razonable para el juego. En El Dilema del Prisionero en la Tabla 2.1, claramente NC está estrictamente dominada por C para ambos agentes; así, al eliminar NC, se encuentra que (C, C) es un resultado razonable del juego.

2.2.5. Equilibrio de Nash

La mayoría de los juegos no cooperativos no son solubles por dominancia iterada, por lo que se deben investigar conceptos alternativos de solución. En este sentido, el concepto de solución más aceptado para un juego no cooperativo es el de un equilibrio de Nash, introducido por John F. Nash en su trabajo seminal [14]. En términos generales, un equilibrio de Nash es un estado de un juego no cooperativo donde ningún agente puede mejorar su recompensa cambiando su estrategia, si los otros agentes mantienen sus estrategias actuales. Formalmente, cuando se trata de estrategias puras, es decir, elecciones determinísticas por parte de los agentes, el equilibrio de Nash se define de la siguiente manera: Un equilibrio de Nash en estrategia pura de un juego no cooperativo $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ es un perfil de estrategia $s^* \in S$ tal que $\forall i \in N$ tenemos lo siguiente: $u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*), \forall s_i \in S_i$.

En otras palabras, un perfil de estrategia es un equilibrio de Nash en estrategia pura si ningún agente tiene incentivo para desviarse unilateralmente a otra estrategia, dado que las estrategias de los otros agentes permanecen fijas. En el caso de que

tengamos $u_i(s_i^*, s_{-i}^*) > u_i(s_i, s_{-i}^*), \forall s_i \in S_i, s_i \neq s^*, \forall i \in N$, el equilibrio de Nash se dice que es estricto.

Con esta definición, podemos verificar si se puede encontrar una solución de equilibrio de Nash para el Dilema del Prisionero estudiando posibles desviaciones de los agentes para cada combinación de estrategias. Al inspeccionar la Tabla 2.1 podemos encontrar fácilmente que (C, C) es el único equilibrio de Nash de este juego. Si repetimos este análisis para diferentes juegos, podríamos deducir las siguientes declaraciones con respecto al concepto de un equilibrio de Nash en estrategia pura:

- Existencia y multiplicidad: Un juego no cooperativo puede admitir cero, uno o múltiples equilibrios de Nash.
- Eficiencia: Un equilibrio de Nash no es necesariamente el mejor resultado, desde la perspectiva de la recompensa.

Por lo tanto, al estudiar los equilibrios de Nash de un juego, los puntos clave de interés son la existencia, multiplicidad y eficiencia. Por ejemplo, juegos como el Dilema del Prisionero admiten un único equilibrio de Nash en estrategia pura, juegos como el Juego de la Gallina (que se presentará más adelante) admiten múltiples equilibrios de Nash, mientras que un juego como el Juego de Descoordinación (que también se presentará más adelante) no tiene equilibrio de Nash en estrategia pura.

Con respecto a la eficiencia, en el Dilema del Prisionero, como se mencionó en la subsección anterior, el equilibrio de Nash en estrategia pura del juego (-4, -4) es ineficiente. Por ejemplo, los dos prisioneros podrían hacerlo mejor, es decir, lograr (-2, -2), si ambos eligen no confesar (NC); sin embargo, este resultado no es estable en un entorno no cooperativo, es decir, no es un punto de equilibrio, ya que existen posibles desviaciones unilaterales. Esto demuestra que, aunque cooperar no confesando daría a cada agente una mejor recompensa de -2, la avaricia de cada prisionero conduce a un resultado ineficiente. Esto demuestra que la solución de equilibrio de Nash en estrategia pura de un juego no cooperativo puede ser ineficiente.

2.2.6. Estrategias mixtas

Hasta ahora, el principal enfoque en el estudio de juegos estratégicos ha sido en estrategias puras y equilibrios de Nash puros. Como se mencionó anteriormente, una estrategia pura es una selección determinística de una estrategia por un agente dado. Sin embargo, en general, un agente puede ser capaz de seleccionar cada estrategia pura con cierta probabilidad, que es la base del concepto de una estrategia mixta. Para un agente dado, una estrategia mixta consiste en un número de posibles movimientos que son elegidos con una distribución de probabilidad.

Dado un juego estratégico $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$, para cada agente i definimos $\sum_i$ como el conjunto de distribuciones de probabilidad sobre su conjunto de estrategias S_i. Una estrategia mixta $\sigma_i(s_i) \in \sum_i$ de un agente i es una distribución de probabilidad sobre las estrategias puras $s_i \in S_i$. Por ejemplo, cuando el conjunto S_i es finito, entonces σ_i es una función de densidad de probabilidad de las estrategias puras. Dado el perfil de estrategias mixtas $\sigma \in \sum = \prod_{i=1}^{N} \sum_i$ y asumiendo que los conjuntos de estrategias puras S_i son finitos, dejamos que $sup(\sigma_i) = \{s_i \in S_i | \sigma_i(s_i) > 0\}$ denote el soporte del conjunto de estrategias que se les asignan probabilidades positivas. Por consiguiente, la recompensa para una estrategia mixta corresponde al valor esperado de los perfiles de estrategia pura en su soporte, es decir, $u_i(\sigma) = \sum_{s \in S} (\prod_{j=1}^{N} \sigma_j(s_j)) u_i(s_i, s_{-i})$, donde $u_i(s_i, s_{-i})$ es la recompensa de estrategia pura para una tupla N de estrategias (s_i, s_{-i}).

Ahora podemos definir el concepto de equilibrio de Nash de estrategias mixta MSNE: Un perfil de estrategia mixta $\sigma^* \in \sum$ es un equilibrio de Nash de estrategia mixta si, para cada agente $i \in N$, tenemos: $u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(\sigma_i, \sigma_{-i}^*), \forall \sigma_i \in \sum$.

Note que un equilibrio de Nash de estrategia pura también puede considerarse como un equilibrio de Nash de estrategia mixta con un perfil de estrategia mixta en el que cada agente selecciona una estrategia con probabilidad 1 (una estrategia pura) mientras asigna cero probabilidad a todas las demás estrategias.

Es crucial entender el concepto de estrategias mixtas, ya que nos permite presentar el resultado más importante y crucial de la teoría de juegos, que es el siguiente teorema de Nash: "Todo juego no cooperativo finito en forma estratégica tiene al menos un equilibrio de Nash de estrategia mixta"[14].

Para concluir, el equilibrio de Nash en estrategias mixtas o puras proporciona un poderoso concepto de solución para juegos estratégicos no cooperativos que ha revolucionado la teoría de juegos desde el trabajo de Nash. Como se verá en el resto de este trabajo, muchas aplicaciones, conceptos y clases de juegos tratan con modelos y soluciones que, de una forma u otra, se basan en conceptos inspirados en el equilibrio de Nash.

2.2.7. Eficiencia de equilibrios

En las subsecciones anteriores, estudiamos principalmente la existencia y caracterización de los equilibrios de Nash, tanto en estrategias puras como mixtas. Por ejemplo, el teorema de Nash afirma que, en una amplia clase de juegos, siempre existe al menos un equilibrio de Nash de estrategia mixta. Sin embargo, una vez que hemos verificado la existencia (para estrategias puras) y número de equilibrios, es importante seleccionar un equilibrio que sea deseado en el juego, por ejemplo, óptimo o eficiente. Una medida importante de eficiencia se puede encontrar en el concepto de optimalidad de Pareto, definido de la siguiente manera: Un perfil de estrategia $s \in S$ es Pareto-superior a otro perfil de estrategia $s' \in S$ si, para cada agente $i \in N$, tenemos: $u_i(s_i, s_{-i}) \geq u_i(s_i', s'_{-i})$, con desigualdad estricta para al menos un agente. En consecuencia, un perfil de estrategia $s^o \in S$ es Pareto-óptimo si no existe otro perfil de estrategia que sea Pareto-superior a s^o.

El resultado de un juego es Pareto-óptimo si no hay otro resultado que haga que cada agente esté al menos igual de bien y al menos un agente estrictamente mejor. Por lo tanto, un resultado Pareto-óptimo no se puede mejorar sin perjudicar al menos a un agente. Como resultado, en juegos donde existe un gran número de equilibrios de Nash, es deseable seleccionar un equilibrio Pareto-óptimo, si es posible. Sin embargo, se debe notar que a menudo los equilibrios de Nash no son Pareto-óptimo y los Pareto-óptimo no son equilibrios de Nash. Por ejemplo, en el Dilema del Prisionero, un punto Pareto-óptimo es el punto (NC, NC), es decir, $(-2, -2)$, porque no se puede mejorar la recompensa para un prisionero sin disminuir la recompensa para el otro, al pasar de $(-2, -2)$ a $(-5, 0)$, el prisionero 2 mejora mientras que el prisionero 1 tiene un peor desempeño. Note que los puntos $(-5, 0)$ y $(0, -5)$ también son puntos Pareto-óptimos. Además, notamos que el equilibrio de Nash $(-4, -4)$ del Dilema del Prisionero no es Pareto-óptimo ya que podemos mejorar las recompensas para ambos prisioneros moviéndonos al punto $(-2, -2)$, que es Pareto-óptimo. El Dilema del Prisionero ilustra una situación donde 3 de los 4 resultados posibles son Pareto-óptimos, sin embargo, el equilibrio de Nash resulta ser el que no es eficiente.

Para más información sobre Teoría de Juegos, referirse al libro [15].

2.3. Modelado de Redes de Comunicación con Teoría de Juegos

2.3.1. Introducción

Los avances recientes en tecnología y la creciente necesidad de computación y comunicación ubicuas han llevado a una necesidad incesante de nuevos marcos analíticos que sean adecuados para abordar los numerosos desafíos presentes en la telecomunicaciones. Como resultado, en los últimos años, la teoría de juegos se ha convertido en una herramienta central para el diseño de futuras redes inalámbricas y de comunicación. Esto se debe principalmente a la necesidad de incorporar reglas y técnicas de toma de decisiones en los nodos de comunicación e inalámbricos de próxima generación, para permitirles operar de manera eficiente y satisfacer las necesidades de los usuarios en términos de servicios de comunicación (por ejemplo, transmisión de video a través de redes móviles, acceso a Internet ubicuo, uso simultáneo de múltiples tecnologías, intercambio de archivos peer-to-peer, etc.).

Uno de los ejemplos más populares de la teoría de juegos en redes inalámbricas se relaciona con el modelado del problema de control de potencia en redes celulares mediante juegos no cooperativos. Por ejemplo, en la subida de un sistema celular, los investigadores e ingenieros se han preocupado por el problema de diseñar un mecanismo que permita a los usuarios regular su potencia de transmisión, dado la interferencia que causan (o que es causada por otros usuarios) en la red. Al hacerlo, los investigadores pudieron trazar una sorprendente similitud entre los problemas de control de potencia y la teoría de juegos no cooperativos. En un juego no cooperativo, varios agentes están involucrados en una situación competitiva en la cual, cada vez que un agente elige una estrategia, esta jugada tiene un impacto en la recompensa (por ejemplo, una medida de beneficio o ganancia) de los otros agentes. De manera similar, en un juego de control de potencia, tenemos una situación competitiva en la cual el nivel de potencia de transmisión (estrategia) de un usuario inalámbrico puede impactar positiva o negativamente (debido a la interferencia) en la tasa de transmisión y la calidad del servicio de los otros usuarios. Como resultado, se ha demostrado que resolver un juego de control de potencia es equivalente a resolver un juego no cooperativo, por ejemplo, encontrando un equilibrio de Nash. El control de potencia es solo un ejemplo en el cual la teoría de juegos puede ser utilizada para diseñar redes inalámbricas y de comunicación de próxima generación. De hecho, tras los primeros trabajos sobre juegos no cooperativos en control de potencia, han surgido una plétora de nuevas áreas de aplicación para la teoría de juegos en las comunidades de inalámbricas, comunicaciones y procesamiento de señales.

El desafío clave al aplicar la teoría de juegos en un contexto de comunicaciones radica en el hecho de que la teoría de juegos fue desarrollada esencialmente como una herramienta para ser utilizada en economía y ciencias sociales. Por lo tanto, el aprovechamiento de la teoría de juegos para su uso en aplicaciones de ingeniería viene acompañado de muchas dificultades. Por ejemplo, los investigadores interesados en aplicar modelos de teoría de juegos a problemas en redes inalámbricas y de comunicación enfrentan muchos obstáculos para encontrar modelos y soluciones precisos. Esto se debe al hecho de que los modelos de teoría de juegos existentes no están adaptados para hacer frente a problemas específicos de ingeniería, como modelar canales inalámbricos variables en el tiempo, desarrollar funciones de rendimiento (es decir, recompensas) que dependen de métricas de comunicación restrictivas (por ejemplo, tasa de transmisión, retraso de colas, relación señal a ruido), y cumplir con ciertos estándares. Esto ha hecho necesaria una fuente de referencia

oportuna y completa que pueda guiar a los investigadores e ingenieros de comunicaciones en su búsqueda para encontrar modelos analíticos efectivos de la teoría de juegos que puedan aplicarse al diseño de futuras redes inalámbricas y de comunicación.

Los avances recientes en comunicación inalámbrica han hecho posible la implementación a gran escala de redes inalámbricas, que consisten en nodos pequeños y de bajo costo con capacidades simples de procesamiento y redes. Para alcanzar el destino deseado, como el sumidero de datos, son necesarias transmisiones que dependen de múltiples saltos. Como resultado, la optimización del enrutamiento es un problema crítico que involucra muchos aspectos como la calidad del enlace, la eficiencia energética y la seguridad. Además, los nodos pueden no estar dispuestos a cooperar completamente. Desde la perspectiva del nodo, reenviar los paquetes que llegan consume su limitada energía de batería, por lo que puede no estar en el interés del nodo reenviar todos los paquetes que llegan. Además, hacerlo afectará negativamente la conectividad de la red. Por lo tanto, es crucial diseñar un mecanismo de control distribuido que fomente la cooperación entre los nodos multi-salto participantes.

2.3.2. Conceptos básicos de los juegos de enrutamiento

Una red se define por un grafo dirigido $G = (V, E)$, con un conjunto de vértices V y un conjunto de canales E. Un conjunto $\{(s_1, d_1), ..., (s_k, d_k)\}$ consiste en pares de vértices fuente-destino. Cada agente se identifica con un paquete. Diferentes agentes pueden originarse de diferentes vértices fuente y pasar información a diferentes vértices destino. Usamos P_i para denotar las rutas $s_i - d_i$ de una red multi-salto y definimos $\mathbb{P} = \cup_{i=1}^{K} P_i$. Permitimos que el grafo G contenga canales paralelos, y un vértice puede participar en múltiples pares fuente-destino.

Cada canal e de una red tiene una función de costo w_e y son siempre positivas. Todas estas suposiciones son razonables en aplicaciones inalámbricas. Note que el costo de cada canal está determinado tanto por la naturaleza (como la calidad del enlace) como por las acciones de los agentes. Por ejemplo, el costo representa una cantidad que aumenta con la congestión de la red (cuando los agentes utilizan demasiado el canal). La recompensa para un agente U_i generalmente se formula como la suma del costo sobre las rutas seleccionadas, y depende de los flujos.

En el juego de enrutamiento de una red multi-salto, la estrategia s_i para cada agente es elegir el flujo óptimo (por ejemplo, con el costo total mínimo). El flujo es un vector no negativo indexado por el conjunto $\mathbb{P}$ de rutas fuente-destino. En tal juego de enrutamiento, varios tomadores de decisiones independientes (agentes) interactúan para formar un grafo de red. Dependiendo de los objetivos de cada agente, los flujos finales de la red resultan de las decisiones individuales de los agentes. Denotamos por $G_{s_i, s_{-i}}$ al grafo G formado cuando el agente i juega una estrategia s_i mientras que todos los demás nodos mantienen sus estrategias $s_{-i} = [s_1, ..., s_{i-1}, s_{i+1}, ..., s_K]$. Para analizar el resultado de tal juego, primero podemos recordar la definición de un equilibrio de Nash como el punto estable en el que ningún agente puede mejorar unilateralmente su rendimiento cambiando solo su propia estrategia.

A continuación, presentamos un ejemplo de la paradoja de Braess [16] y sus aplicaciones en redes inalámbricas, para mostrar cómo formular el juego en una red multi-salto. La paradoja de Braess afirma que agregar capacidad adicional a una red, cuando las entidades en movimiento eligen egoístamente su ruta, puede, en algunos casos, reducir el rendimiento general. Esto se debe a que el equilibrio de Nash de tal

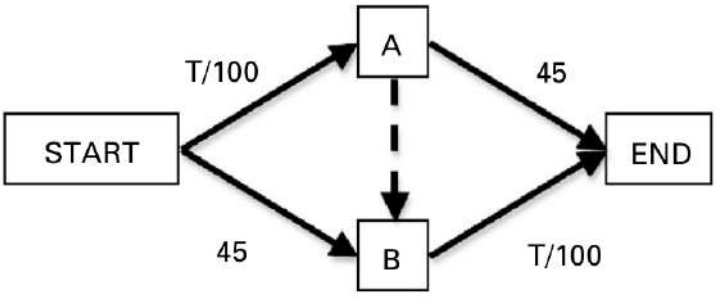


FIGURA 2.2: Ejemplo de la paradoja de Braess.

sistema no es necesariamente óptimo a nivel global. Formalmente, la paradoja se establece de la siguiente manera [16]: Para cada punto de una red de carreteras, se da el número de coches que parten de él y el destino de los coches. Bajo estas condiciones, se desea estimar la distribución de congestión de tráfico. Si una calle es preferible a otra depende no solo de la calidad de la carretera, sino también de la congestión. Si cada conductor toma el camino que le parece más favorable, los tiempos de recorrido resultantes no tienen por qué ser mínimos. Además, se indica con un ejemplo que agregar una carretera nueva a la red puede causar una redistribución del tráfico que resulte en tiempos de recorrido individuales más largos.

Consideremos la red mostrada en la Fig. 2.2 como un ejemplo, en la que 4000 conductores desean viajar de Inicio a Fin. El tiempo de viaje en minutos en la carretera Inicio-A es un número proporcional al número de viajeros, T minutos dividido por 100, y en Inicio-B es un constante de 45 minutos. Si la carretera punteada no existe (así que la red de tráfico tiene un total de cuatro carreteras), el tiempo necesario para conducir la ruta Inicio-A-Fin con A conductores sería $\frac{A}{100} + 45$, y el tiempo necesario para conducir la ruta Inicio-B-Fin con B conductores sería $\frac{B}{100} + 45$. Si una ruta es más corta que la otra, esto no sería un equilibrio de Nash porque cualquier conductor racional cambiaría de la ruta más larga a la más corta. Como hay 4000 conductores, el hecho de que $A + B = 4000$ puede ser utilizado para resolver que $A = B = 2000$ cuando el sistema está en equilibrio, y por lo tanto cada ruta toma $\frac{2000}{100} + 45 = 65$ minutos en ambas rutas.

A continuación, suponemos que la línea punteada es una carretera con un tiempo de viaje muy pequeño, de aproximadamente 0 minutos. En esta situación, todos los conductores elegirán el camino Inicio-A-B, porque Inicio-A-B tomará 40 minutos en el peor de los casos, mientras que Inicio-B siempre es 45 minutos. Al llegar a A, cada conductor racional decidirá tomar la carretera "gratuita" a B y continuar hacia Fin, ya que A-Fin siempre es 45 minutos, mientras que A-B-Fin es en el peor de los casos 40 minutos. El tiempo de viaje de cada conductor es $\frac{4000}{100} + \frac{4000}{100} = 80$ minutos, un aumento de los 65 minutos requeridos cuando la rápida carretera A-B no existía. Ningún conductor tiene incentivo para cambiar, ya que las dos rutas originales (Inicio-A-Fin y Inicio-B-Fin) son ahora ambas 85 minutos. Si todos los conductores acordaran no usar la ruta A-B, todos se beneficiarían, reduciendo sus tiempos de viaje en 15 minutos. Sin embargo, debido a que cualquier conductor individual siempre se beneficiará tomando la ruta A-B, la distribución socialmente óptima no es estable, y por eso ocurre la paradoja de Braess.

Ahora consideramos un escenario de red inalámbrica en el que ocurre la paradoja de Braess. Consideramos una red de una sola celda con dos puntos de acceso (APs) a los cuales están conectados varios usuarios móviles, como se muestra en la Fig. 2.3. Supongamos que el primer punto de acceso, denotado por AP_1, ofrece una tasa fija r_F

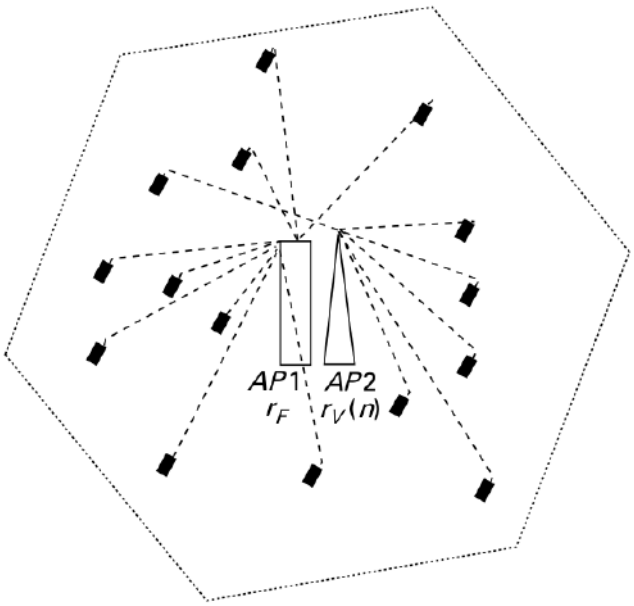


FIGURA 2.3: Celda única con dos puntos de acceso diferentes, uno que ofrece una tarifa fija r_F y el otro que ofrece una tarifa $r_V(n)$ que depende del número n de usuarios conectados a ella.

y que la tasa ofrecida por el segundo punto de acceso, denotado por AP_2, ofrece una tasa $r_V(n)$ que depende del número de usuarios conectados a él. Un ejemplo de tal sistema en la práctica sería una red que tiene dos tipos de puntos de acceso. El primer tipo es un punto de acceso que utiliza un esquema de acceso múltiple ortogonal (por ejemplo, TDMA) en el que cada usuario obtiene una tasa fija. El segundo tipo es un punto de acceso que ofrece una tasa variable que depende del número de usuarios conectados a él. Un ejemplo de tal punto de acceso podría basarse en un sistema CDMA, en el cual cuanto más usuarios estén conectados al sistema, mayor es la interferencia. Esto conduce a una relación señal-a-interferencia-y-ruido (SINR) más baja y, por lo tanto, a una tasa menor. Podemos demostrar que si intentamos mejorar el sistema permitiendo una conexión intersistema, es decir, una conexión entre los dos tipos de APs, eventualmente puede llevar a un rendimiento que es peor que en el sistema original. Esto sucede debido al egoísmo de los usuarios, y es similar a la paradoja de Braess original que se estudió en redes de transporte, en la que agregar una nueva carretera rápida no necesariamente mejora el rendimiento de los vehículos.

Finalmente, existen diferentes formas de reforzar la cooperación en la red inalámbrica multi-salto. El hecho de que los nodos distribuidos no tienen información exacta sobre los demás, actúan egoístamente para optimizar sus propios rendimientos, nos lleva a aplicar un enfoque de teoría de juegos al problema de enrutamiento [15, 17]. En [18], la teoría de juegos repetidos se aplica a problemas de enrutamiento. En [19], los autores proponen un marco de juego repetido para el acceso múltiple utilizando el mantenimiento de carteles. Una solución Tit-for-Tat se propone en [20] para redes inalámbricas multi-salto. En [21], la asignación de recursos de acceso múltiple se estudia utilizando un enfoque de teoría de juegos. En [22], los autores consideran una política de castigo menos agresiva, en la que el nodo usa la probabilidad de reenvío mínima entre sus vecindarios como su probabilidad de reenvío después de detectar el mal comportamiento. Felegyhazi et al. [23] consideran un modelo para mostrar la cooperación entre los nodos participantes y proporcionan

condiciones suficientes sobre la topología de la red bajo las cuales cada nodo que emplea la estrategia de castigo resulta en un equilibrio de Nash. Srinivasan et al. [24] proporcionan un marco matemático para la cooperación en redes ad-hoc, que se centra en los aspectos energéticamente eficientes de la cooperación. En [25], los autores se centran en las propiedades de los mecanismos de imposición de cooperación utilizados para detectar y prevenir el comportamiento egoísta de los nodos en una red ad hoc. En [26] y [27], los autores definen protocolos basados en un sistema de reputación. En [28, 29], se estudian la dinámica evolutiva y los juegos potenciales para el enrutamiento no cooperativo. En [30, 31], se propone un modelo de teoría de juegos para protocolos colaborativos en redes inalámbricas multi-salto egoístas y libres de tarifas. Un enfoque de juego bayesiano dinámico se estudia en [32] para el enrutamiento en redes ad hoc inalámbricas.

Para más información sobre Modelado de Redes de Comunicación con Teoría de Juegos, referirse a los libros [33, 34].

2.4. Algoritmos de Aprendizaje por Refuerzo Multi-Agente

2.4.1. Introducción

Las secciones anteriores introdujeron modelos de juego como conceptos de solución para definir lo que significa que los agentes actúen de manera óptima en un juego. En esta sección, presentaremos métodos para calcular soluciones para juegos. El método principal por el cual buscamos calcular soluciones es mediante el aprendizaje por refuerzo (RL), en el que los agentes prueban repetidamente acciones, hacen observaciones y reciben recompensas. Análogo a la terminología estándar de RL, usamos el término episodio para referirnos a cada ejecución independiente de un juego. Los agentes aprenden sus políticas basadas en datos (es decir, observaciones, acciones y recompensas) obtenidos de múltiples episodios en un juego.

Para establecer el contexto de los algoritmos presentados en este trabajo, esta sección comenzará delineando un marco general de aprendizaje por refuerzo multi-agente (MARL). Luego introduciremos dos enfoques básicos de aplicación de RL en juegos, llamados aprendizaje centralizado e independiente, ambos reducen el problema multi-agente a un problema de agente único. El aprendizaje centralizado aplica RL de agente único directamente al espacio de acciones conjuntas para aprender una política central que elige acciones para cada agente, mientras que el aprendizaje independiente aplica RL de agente único a cada agente de forma independiente para aprender políticas de agente, ignorando esencialmente la presencia de otros agentes.

El aprendizaje centralizado e independiente sirven como un punto de partida útil para discutir varios desafíos importantes enfrentados por los algoritmos MARL. Un desafío característico de MARL es la no estacionariedad del entorno debido a múltiples agentes aprendiendo, lo que puede llevar a un aprendizaje inestable. La selección de equilibrio es el problema de qué solución de equilibrio deben acordar los agentes y cómo pueden lograr un acuerdo. Otro desafío es la asignación de crédito multi-agente, en la que los agentes deben inferir qué acciones contribuyeron a una recompensa recibida. Finalmente, los algoritmos MARL suelen enfrentar un crecimiento exponencial del espacio de acción conjunta a medida que aumenta el número de agentes, lo que lleva a problemas de escalabilidad.

Comenzamos definiendo el aprendizaje en juegos y el resultado de aprendizaje deseado. En aprendizaje automático, el aprendizaje es un proceso que optimiza un modelo o función basado en datos. En nuestro contexto, el modelo es una política conjunta que generalmente consiste en políticas para cada agente, y los datos (o

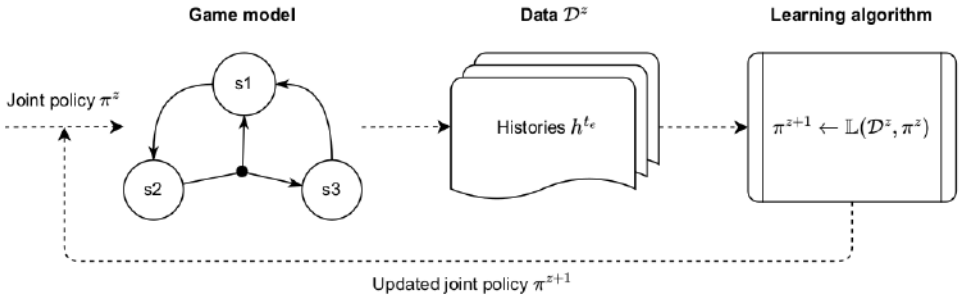


FIGURA 2.4: Elementos de un proceso de aprendizaje general en MARL.

"experiencias") consisten en el feedback recibido durante el juego. El objetivo del aprendizaje es una solución del juego, definida por un concepto de solución elegido. Así, este proceso de aprendizaje implica varios elementos, mostrados en la Fig. 2.4 y detallados a continuación.

El modelo de juego define el entorno multi-agente y cómo los agentes pueden interactuar. Los modelos de juego, en nuestro contexto, incluyen juegos en forma normal no repetidos y juegos en forma normal repetidos. Los datos utilizados para el aprendizaje consisten en un conjunto de z historias, $D^z = \{h^{t_e} | e = 1, ..., z\}, z \geq 0$. Cada historia h^{t_e} fue producida por una política conjunta π^e utilizada durante el episodio e. A menudo, D^z contiene la historia hasta ahora del episodio en curso z y las historias de episodios anteriores $e < z$. El modelo de juego elegido determina la condición de la política conjunta aprendida. En un juego en forma normal no repetido (donde los episodios terminan después de un paso de tiempo), las políticas π_i no están condicionadas en historias, es decir, son distribuciones de probabilidad simples sobre acciones. En un juego en forma normal repetido, las políticas están condicionadas en historias de acciones $h^t = \{a^0, ..., a^{t-i}\}$.

Un algoritmo de aprendizaje $\mathbb{L}$ toma los datos recolectados D^z y la política conjunta actual π^z, y produce una nueva política conjunta, $\pi^{z+1} = \mathbb{L}(D^z, \pi^z)$. La política conjunta inicial π^0 es típicamente aleatoria. El objetivo del algoritmo de aprendizaje es una política conjunta π^* que satisface las propiedades de un concepto de solución elegido. Anteriormente, por ejemplo, introdujimos un posible concepto de solución, el equilibrio de Nash.

El algoritmo $\mathbb{L}$ puede consistir en sí mismo de múltiples algoritmos de aprendizaje que aprenden políticas de agentes individuales, como un algoritmo $\mathbb{L}_i$ para cada agente i. Cada uno de estos algoritmos puede usar diferentes partes de los datos en D^z, o usar sus propios datos D_i^z como en el aprendizaje independiente. Además, una característica importante del RL es que el algoritmo de aprendizaje está activamente involucrado en la generación de los datos explorando acciones, en lugar de simplemente consumir pasivamente los datos. Por lo tanto, las políticas producidas por el algoritmo de aprendizaje pueden aleatorizar activamente sobre las acciones para generar datos útiles para el aprendizaje.

2.4.2. Reducción a Agente Único

El enfoque más típico para usar RL para aprender políticas de agentes en sistemas multi-agente es esencialmente reducir el problema de aprendizaje multi-agente a un problema de aprendizaje de agente único. En esta sección, presentaremos dos

de estos enfoques: el aprendizaje centralizado aplica RL de agente único directamente al espacio de acciones conjuntas para aprender una política central que elige acciones para todos los agentes; y el aprendizaje independiente aplica RL de agente único de manera independiente a cada agente para aprender políticas independientes, ignorando esencialmente la presencia de otros agentes.

Aprendizaje Centralizado

El aprendizaje centralizado entrena una única política central π_c que recibe las observaciones locales de todos los agentes y selecciona una acción para cada uno de ellos, eligiendo acciones conjuntas de $A = A_1 \times ... \times A_n$. Esto esencialmente reduce el problema multi-agente a un problema de agente único, y podemos aplicar algoritmos existentes de RL de agente único para entrenar π_c. Un ejemplo de aprendizaje centralizado se llama Q-learning centralizado (CQL, inspirando en el famoso algoritmo Q-learning de aprendizaje por refuerzo de agente único [35]) y se observa en el Algoritmo 1. Este algoritmo mantiene valores de acción conjunta $Q(a)$ para acciones conjuntas $a \in A$. El aprendizaje centralizado puede ser útil porque evita los aspectos multi-agente de la no estacionariedad (la adaptación colectiva continua de las políticas de aprendizaje de los agentes crea dinámicas cíclicas y cambiantes) y problemas de asignación de crédito (la identificación de las contribuciones individuales de los agentes a las recompensas colectivas entre múltiples agentes interactuantes). Sin embargo, en la práctica, este enfoque tiene también una serie de limitaciones.

Algorithm 1 Q-learning centralizado (CQL) para juegos repetitivos

Require: Inicializar: $Q(a) = 0$ y $a \in A = A = A_1 \times ... \times A_n$
Repetir para cada episodio:
for $t = 0$ **to** t_{max} **do**
Si un número aleatorio generado $>\epsilon$: Elegir una acción conjunta $a^t \in A$ aleatoria.
Sino: Elegir una acción conjunta $a^t \in argmax_a Q(a)$
Aplicar a^t en el juego y observar las recompensas $r_1^t, ..., r_n^t$
Transformar las $r_1^t, ..., r_n^t$ individuales en un valor escalar de recompensas r^t
$Q(a^{t+1}) = Q(a^t) + \alpha[r^t + \gamma(\max_{a'}(Q(a'))) - Q(a^t)]$
end for

La primera limitación a tener en cuenta es que, para aplicar RL de agente único, el aprendizaje centralizado requiere transformar la recompensa conjunta $(r_1, ..., r_n)$ de cada uno de los jugadores en una única recompensa escalar r. En el caso de juegos de recompensa común, en los que todos los agentes reciben recompensas idénticas, podemos usar $r = r_i$ para cualquier i. En este caso, si usamos un algoritmo de RL de agente único que garantiza aprender una política óptima, entonces está garantizado que aprenda una política central π_c para el juego de recompensa común de tal manera que π_c sea un equilibrio correlacionado Pareto-óptimo. La optimalidad del algoritmo de RL de agente único significa que π_c logra rendimientos esperados máximos. Por lo tanto, ya que la recompensa se define como $r = r_i$ para todos los i, sabemos que π_c es Pareto-óptimo porque no puede haber otra política que logre un mayor rendimiento esperado para cualquier agente. Esto también significa que ningún agente puede desviarse unilateralmente de su acción dada por π_c para mejorar sus rendimientos, lo que hace de π_c un equilibrio correlacionado.

Lamentablemente, para juegos de suma cero y suma general, no está claro cómo debería realizarse la escalarización de la recompensa. Si uno está interesado en maximizar el bienestar social en juegos de suma general, una opción es usar $r = \sum_i r_i$. Sin embargo, si la solución deseada es un tipo de solución de equilibrio, entonces puede que no exista ninguna transformación escalar que conduzca a políticas de equilibrio.

La segunda limitación es que al entrenar una política sobre el espacio de acción conjunta, ahora tenemos que resolver un problema de decisión con un espacio de acción que crece exponencialmente en el número de agentes. Por ejemplo, en un juego donde hay tres agentes que eligen entre seis acciones, esto lleva a un algoritmo de aprendizaje central que tiene que aprender en un espacio de acción conjunta con $6^3 = 216$ acciones. Incluso para este simple ejemplo, la mayoría de los algoritmos estándar de RL de agente único no escalan fácilmente a espacios de acción de este tamaño.

Finalmente, una limitación fundamental del aprendizaje centralizado es debido a la estructura inherente de los sistemas multi-agente. Los agentes son a menudo entidades localizadas que están distribuidas física o virtualmente. En tales configuraciones, la comunicación de una política central π_c a los agentes y viceversa puede no ser posible o deseable, por varias razones. Por lo tanto, tales sistemas multi-agente requieren políticas de agentes locales π_i para cada agente i, que actúan sobre las observaciones locales del agente i e independientemente de otros agentes. La próxima sección introducirá el enfoque de aprendizaje independiente que elimina estas limitaciones, a expensas de introducir otros desafíos.

Aprendizaje Descentralizado

En el aprendizaje independiente (a menudo abreviado como IL), cada agente i aprende su propia política π_i utilizando solo su historia local de observaciones, acciones y recompensas propias, mientras ignora la existencia de otros agentes [36, 37]. Los agentes no observan ni usan información sobre otros agentes, y los efectos de las acciones de otros agentes son simplemente parte de la dinámica del entorno desde la perspectiva de cada agente de aprendizaje. Así, similar al aprendizaje centralizado, el aprendizaje independiente reduce el problema multi-agente a un problema de agente único desde la perspectiva de cada agente, y se pueden usar algoritmos existentes de RL de agente único para aprender las políticas de los agentes. Un ejemplo de aprendizaje independiente basado, otra vez, en Q-learning, llamado Q-learning independiente (IQL), se muestra en el Algoritmo 2. Aquí, cada agente utiliza su propia copia del mismo algoritmo.

Algorithm 2 Q-learning independiente (IQL) para juegos repetitivos

(Algoritmo para cada agente i)
Require: Inicializar: $Q_i(a_i) = 0$ para todas las $a_i \in A_i$
Repetir para cada episodio:
for $t = 0$ **to** t_{max} **do**
Si un número aleatorio generado $>\epsilon$: Elegir una acción $a_i^t \in A_i$
Sino: Elegir una acción $a_i^t \in argmax_{a_i} Q_i(a_i)$
(mientras tanto, otros agentes $j \neq i$ eligen sus acciones a_j^t)
Se juega el juego y se observa la recompensa propia r_i^t
$Q_i(a_i^{t+1}) = Q(a_i^t) + \alpha[r_i^t + \gamma(\max_{a_i'}(Q_i(a_i'))) - Q_i(a_i^t)]$
end for

Subclass	1a	1b	2a	2b	3a	3b
# deterministic NE	0	0	2	2	1	1
# probabilistic NE	1	1	1	1	0	0
Dominant action?	No	No	No	No	Yes	Yes
Det. joint act. > NE?	No	Yes	No	Yes	No	Yes
IQL converges?	Yes	No	Yes	Y/N	Yes	Y/N

FIGURA 2.5: Convergencia de Q-learning (IQL) independiente "infinitesimal" en juegos de forma normal de suma general con dos agentes y dos acciones.

El aprendizaje independiente evita naturalmente el crecimiento exponencial en los espacios de acción que afecta al aprendizaje centralizado, y puede ser utilizado cuando la estructura del sistema multi-agente requiere políticas de agentes locales. También no requiere una transformación escalar de la recompensa conjunta, como es el caso en el aprendizaje centralizado, ya que cada agente desea maximizar únicamente su propia recompensa. La desventaja del aprendizaje independiente es que puede verse significativamente afectado por la no estacionariedad causada por el aprendizaje concurrente de todos los agentes. En un algoritmo de aprendizaje independiente como IQL, desde la perspectiva de cada agente i, las políticas π_j de otros agentes $j \neq i$ se convierten en parte de la dinámica del entorno.

A medida que cada agente j continúa aprendiendo y actualizando su política π_j, las probabilidades de acción de π_j en cada iteración pueden cambiar. Así, desde la perspectiva del agente i, el entorno parece no estacionario, cuando realmente las únicas partes que cambian con el tiempo son las políticas π_j de los otros agentes. Como resultado, los enfoques de aprendizaje independiente pueden producir un aprendizaje inestable y pueden no converger a ninguna solución del juego.

Las dinámicas del aprendizaje independiente han sido estudiadas en varios modelos idealizados [38, 39, 40, 41, 42, 43]. Por ejemplo, en [39] estudian un modelo idealizado de IQL con exploración epsilon-greedy que utiliza pasos de aprendizaje infinitamente pequeños $\alpha \to 0$, lo que hace posible aplicar métodos de la teoría de sistemas dinámicos lineales para analizar la dinámica del modelo. Basado en este modelo idealizado, se pueden hacer predicciones sobre los resultados de aprendizaje de IQL para diferentes clases de juegos en forma normal de suma general con dos agentes y dos acciones, que se resumen en la Figura 2.5. Estas clases se caracterizan principalmente por el número de equilibrios de Nash deterministas y probabilísticos que los juegos poseen. Como se puede ver, esta versión idealizada de IQL predice que convergerá a un equilibrio de Nash en algunas de las clases de juegos, mientras que en otras puede no converger en absoluto o solo bajo ciertas condiciones. Un hallazgo interesante de este análisis es que en juegos como el Dilema del Prisionero, que es miembro de la clase 3b, IQL puede tener un comportamiento caótico no convergente que resulta en recompensas que promedian por encima de la recompensa esperada bajo el único equilibrio de Nash del juego.

A pesar de su relativa simplicidad, los algoritmos de aprendizaje independiente todavía sirven como importantes referencias en la investigación de MARL. De hecho, a menudo pueden producir resultados que son competitivos con los algoritmos de MARL del estado del arte, como se muestra en el estudio [44].

Para más información sobre Aprendizaje por Refuerzo Multi-Agente, referirse al libro [45].

2.5. Estado del Arte

2.5.1. Teoría de Juegos Cuántica

Introducción

El surgimiento de la teoría de juegos cuántica tuvo lugar en la convergencia en la intersección de dos de las disciplinas más destacadas de la ciencia moderna: la mecánica cuántica y la teoría de juegos. Esta confluencia de ideas, que se originó a finales del siglo XX, marcó el nacimiento de un campo de estudio que promete desvelar misterios tanto en el ámbito de las partículas subatómicas como en la complejidad de las interacciones de agentes macroscópicos.

La mecánica cuántica, con sus orígenes a principios del siglo XX, revolucionó la comprensión física del universo al nivel más fundamental. Por otro lado, la teoría de juegos, formalizada por John von Neumann y Oskar Morgenstern en su obra de 1944 [46], introdujo un marco formal para el análisis de estrategias en situaciones de conflicto y cooperación. La fusión de estos dos campos, sin embargo, no fue inmediata ni obvia.

El verdadero impulso para la teoría de juegos cuántica surgió con el creciente interés en la computación cuántica y la información cuántica en las últimas décadas del siglo XX. Los investigadores comenzaron a preguntarse si las estrategias y los modelos de la teoría de juegos clásica podían ser mejorados o transformados por los principios de la mecánica cuántica. La idea era audaz: ¿podrían las propiedades cuánticas como la superposición y el entrelazamiento ofrecer nuevas dimensiones estratégicas en los juegos?

El hito inicial en este viaje fue la formulación del primer juego cuántico. En un trabajo pionero a finales de los años 90, David A. Meyer [47] demostró cómo un juego clásico podía obtener un nuevo aspecto al incorporar elementos cuánticos. Este experimento mental, más que un ejercicio teórico, sugería posibilidades fascinantes: los juegos cuánticos no solo podían replicar todos los juegos clásicos sino también ofrecer estrategias y equilibrios completamente nuevos.

A medida que el siglo XXI avanza, la teoría de juegos cuántica se expandió rápidamente, alimentada por el desarrollo tecnológico en computación cuántica y los avances teóricos en la comprensión de sistemas cuánticos complejos. La investigación en este campo ha desafiado y enriquecido nuestra comprensión de la teoría de juegos, llevando a interrogantes y descubrimientos que von Neumann y Morgenstern difícilmente podrían haber imaginado.

Hoy, la teoría de juegos cuántica no solo es un área de investigación activa en física y matemáticas, sino que también está encontrando aplicaciones en campos tan diversos como economía [48], seguridad [49], e ingeniería [50]. Con cada nuevo avance, nos acercamos más a entender cómo los fundamentos cuánticos del universo pueden influir en la toma de decisiones, estrategias y, en última instancia, en nuestra comprensión del comportamiento de múltiples agentes en condiciones de incertidumbre y competencia.

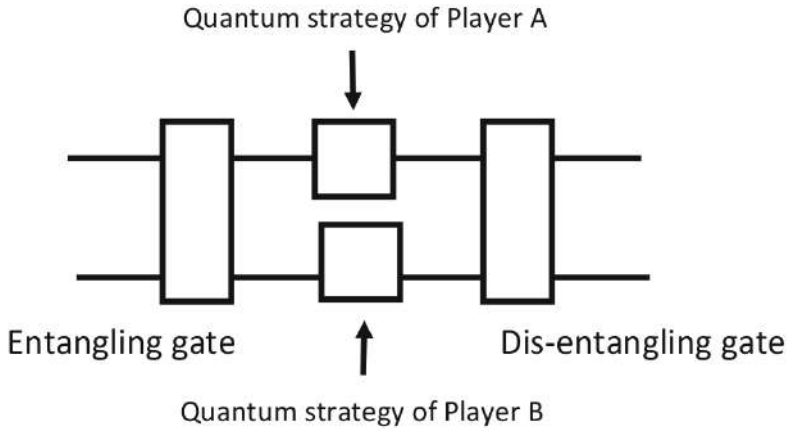


FIGURA 2.6: Circuito cuántico para el esquema de cuantificación EWL.

Juegos Cuánticos No-cooperativos

La teoría de juegos no cooperativos es la base matemática para tomar decisiones óptimas en situaciones competitivas basadas en la información disponible. El término "juego cuántico"parece haber sido utilizado por primera vez por Eisert, Wilkens y Lewenstein en su artículo [51], publicado poco después del trabajo de Meyer. Estos autores estaban interesados en la cuantización de juegos de suma no cero. Eisert et al. muestran que en su implementación computacional cuántica del juego del Dilema del Prisionero, seguido de una medición cuántica, los agentes pueden alcanzar un equilibrio de Nash que también es óptimo de Pareto.

Más precisamente, Eisert et al. proporcionan una implementación computacional cuántica del Dilema del Prisionero. Esta implementación se reproduce en la Fig. 2.6. Definimos un juego cuántico de estrategia pura (en forma normal) como cualquier función unitaria $\mathbb{Q} : \otimes_{i=1}^{n} \mathbb{C}P^{d_i} \rightarrow \otimes_{i=1}^{n} \mathbb{C}P^{d_i}$ donde $\mathbb{C}P^{d_i}$ es el espacio de Hilbert proyectivo complejo de dimensión d_i de estados cuánticos puros que constituye las estrategias cuánticas puras del agente i. Un juego cuántico mixto sería entonces cualquier función $\mathbb{R} : \Delta(\otimes_{i=1}^{n} \mathbb{C}P^{d_i}) \rightarrow \Delta(\otimes_{i=1}^{n} \mathbb{C}P^{d_i})$ donde Δ representa el conjunto de distribuciones de probabilidad sobre el argumento.

En [52], Benjamin et al. argumentan que el equilibrio de Nash en el protocolo EWL, aunque es teóricamente correcto en términos de juegos, tiene un interés físico cuántico limitado. Estos autores proceden a demostrar que cuando se considera una implementación más general y significativa desde el punto de vista físico cuántico del EWL (al extender de 2 a 3 el número de parámetros que los agentes pueden modificar en sus estrategias cuánticas), ¡el Dilema del Prisionero cuántico no tiene equilibrio de Nash con estrategias puras! Sin embargo, una vez que se introduce la aleatorización a través de distribuciones de probabilidad (estrategias mixtas) en el Dilema del Prisionero cuántico, un equilibrio de Nash que es casi óptimo, pero aún así más óptimo que el disponible en el juego clásico, vuelve a materializarse. Esta crítica fue de hecho abordada por el mismo Eisert en una publicación posterior [53].

En las casi dos décadas desde el inicio de la teoría de los juegos cuánticos, el protocolo de cuantización EWL ha asumido el papel de una definición de trabajo para juegos cuánticos no cooperativos para físicos. Varios resultados notables sobre la implementación física cuántica de juegos siguieron al trabajo de Eiset et al., como [54, 55, 56, 57, 58, 59, 60, 61, 62, 63]. Esto puede parecer extraño ya que uno pensaría

que la comunidad de físicos estaría más interesada en el comportamiento de equilibrio u óptimo de los sistemas cuánticos que en la implementación física cuántica de juegos. Por otro lado, este escenario tiene sentido desde un punto de vista práctico, ya que con la llegada de realizaciones tecnológicas de computadoras cuánticas y sistemas de comunicación cuántica, la jugabilidad de los juegos computacionalmente cuánticos sería de fundamental importancia para la toma de decisiones financieras y económicas. La literatura sobre juegos cuánticos es considerable. Los artículos de Guo et al. [64] y Khan et al. [65] son un excelentes revisiones del tema.

El entrelazamiento cuántico en un sistema físico cuántico implica correlaciones no clásicas entre las observables del sistema. Mientras que Eisert et al. demostraron que su implementación computacional cuántica del Dilema del Prisionero producía correlaciones no clásicas y resolvía el dilema, en [66], Shimumura et al. establecen un resultado más fuerte: que las correlaciones habilitadas por el entrelazamiento cuántico siempre resuelven dilemas en juegos de suma no cero, y que las correlaciones clásicas no necesariamente hacen lo mismo. El entrelazamiento cuántico es claramente el recurso más importante para los juegos cuánticos.

Los trabajos publicados y presentados durante esta tesis de doctorado utilizan un modelo similar al EWL para los juegos cuánticos no-cooperativos. La descripción específica de los algoritmos utilizados para modelar cada juego cuántico estarán presentados en detalle en cada una de las secciones correspondientes a dichos trabajos.

Realizaciones Experimentales

La implementación de juegos cuánticos en hardware puede verse como un pequeño cálculo cuántico, en ese sentido, los requisitos para una buena plataforma en la que realizar un juego cuántico son los mismos que para una computadora cuántica. La computadora cuántica puede no necesitar ser una computadora universal, pero requiere puertas de un solo qubit y de dos qubits. Las computadoras cuánticas con las capacidades requeridas para juegos cuánticos apenas están comenzando a estar disponibles, y las redes cuánticas completas están en su infancia, por lo tanto, muchas de las demostraciones experimentales hasta la fecha se han realizado en hardware que no es ideal desde el punto de vista del criterio mencionado anteriormente. Aunque notablemente, a diferencia de muchos algoritmos de computación cuántica interesantes, los juegos cuánticos típicamente se realizan con muy pocos qubits, lo que los hace una aplicación atractiva para demostraciones tempranas en hardware cuántico emergente.

Las implementaciones de juegos cuánticos, más allá de la computación cuántica, requieren características específicas debido a su naturaleza. Esto incluye varios agentes independientes, posiblemente ubicados remotamente, lo que implica la necesidad de una red capaz de transmitir recursos cuánticos, como pares entrelazados distribuidos remotamente, generalmente con fotones. Hasta ahora, no se han realizado experimentos con agentes independientes reales en hardware cuántico en tiempo real, por lo que las implementaciones actuales son parciales, utilizando hardware cuántico para ejecutar el circuito del juego con estrategias determinadas teóricamente. Finalmente, los resultados obtenidos se comparan con los resultados teóricos para validar el experimento. Los juegos que se han implementado siempre se configuran para tener una recompensa mayor en el equilibrio cuántico que en el caso clásico, presumiblemente porque estos son los juegos interesantes para los físicos cuánticos. Debido a esto, el efecto del ruido o la decoherencia es casi siempre reducir la recompensa de los agentes. Generalmente se observa que las recompensas en el equilibrio

de los juegos cuánticos todavía superan al juego clásico con cierta cantidad de decoherencia.

Los investigadores han estado implementando juegos cuánticos con circuitos ópticos, ya que tienen varias características atractivas. No sufren de la incertidumbre en el entrelazamiento y pueden tener fidelidades muy altas. Las puertas se implementan con elementos ópticos estándar como divisores de haz y placas de onda. Además, dado que se realizan con fotones, pueden adaptarse naturalmente para trabajar con agentes remotos.

Una posible implementación es usar un solo fotón y utilizar múltiples grados de libertad. Típicamente, el estado de polarización del fotón se entrelaza con su modo espacial. En [67], se utilizó un láser He-Ne altamente atenuado como fuente de fotón único. Los qubits se entrelazan dividiendo el fotón en dos caminos dependiendo de su polarización. Las puertas se realizan mediante rotaciones de polarización de fotones individuales, es decir, agregando placas de onda al camino de los fotones. Informan un error en la recompensa determinada experimentalmente al teórico del 1-2 %. En otras implementaciones [68, 69], haces de luz inciden en máscaras holográficas para producir modos transversales de orden superior de una manera dependiente de la polarización. Otra posible implementación con óptica lineal realizan compuertas mediante mediciones de fotones para jugar al Dilema del Prisionero con dos agentes, y reportan una fidelidad del 62 % [70].

También se pueden tomar los pares de fotones entrelazados que salen de un cristal no lineal y realizar puertas de la misma manera que se realizan en el caso de fotón único [71, 72]. Estos enfoques han reportado fidelidades de alrededor del 70-80 %. Un juego cuántico de cuatro agentes ha sido implementado con un proceso de conversión paramétrica descendente espontánea que produce cuatro fotones, en dos pares entrelazados [71]. La información se codifica en la polarización y el modo espacial de los fotones. Nuevamente, las fidelidades se reportan cerca del 75 %, lo que resulta en errores en la recompensa en el equilibrio de alrededor del 10 %.

Las implementaciones ópticas lineales son prometedoras debido a su capacidad para realizar juegos con agentes ubicados de forma remota. Sin embargo, también tienen desventajas. Para ejecutar un circuito diferente, uno debe reorganizar físicamente los elementos ópticos lineales, como las placas de onda y los divisores de haz, lo que podría hacerse con pantallas de cristal líquido u otros dispositivos fotónicos, aunque esto podría ser difícil de escalar a implementaciones de juegos más complicados. Además, la producción de mayores cantidades de pares de fotones entrelazados es experimentalmente desafiante. Estos hacen que sea difícil escalar implementaciones con circuitos ópticos lineales a juegos más complicados.

Potenciales Aplicaciones

Las aplicaciones de la teoría de juegos convencional han jugado un papel importante en muchos procesos modernos de toma de decisiones estratégicas, incluyendo la economía, la seguridad y las comunicaciones. Estas aplicaciones típicamente reducen un problema específico del dominio a un escenario de teoría de juegos, de tal manera que se puede desarrollar una solución específica del dominio estudiando la solución teórica de juegos. Un ejemplo bien conocido de aplicación es el juego de la “Gallina” aplicado a estudios de política internacional durante la Guerra Fría [73]. Schelling ha citado el estudio de este juego como influyente en la comprensión de la crisis de los misiles cubanos. Más ampliamente, estudiar estos juegos permite entender cómo los agentes racionales e irracionales seleccionan estrategias.

El método de reconocer soluciones teóricas formales de juegos dentro de aplicaciones específicas del dominio también puede extenderse a conceptos de teoría de juegos cuántica. Esto requiere un modelo para el juego que tenga en cuenta la inclusión de recursos cuánticos únicos, incluyendo estados entrelazados compartidos. Por ejemplo, Zabaleta y Arizmendi [74, 75] han investigado un modelo de juego cuántico para el problema de compartir el espectro en entornos de comunicación inalámbrica en el que los transmisores compiten por el acceso. Su aplicación se plantea como una versión del juego de la minoría propuesto por Challet y Zhang [76] y estudiado por primera vez en forma cuantizada por Benjamin y Hayden [55] y también por Flitney y Hollenberg [77]. Para Zabaleta y Arizmendi, una estación base distribuye un estado cuántico entrelazado n-partito entre n transmisores individuales, es decir, agentes, quienes luego aplican estrategias locales a cada parte del estado cuántico antes de medir. Basado en los resultados correlacionados observados, los agentes seleccionan si transmitir (1) o esperar (0). Zabaleta y Arizmendi mostraron que usar el recurso cuántico en este juego reduce la probabilidad de colisión de transmisión por un factor de n mientras se mantiene la equidad en la gestión de acceso.

En una aplicación relacionada, Solmeyer et al. investigaron un juego de enrutamiento cuántico para enviar transmisiones a través de una red de comunicación [78]. El juego de enrutamiento convencional ha sido ampliamente estudiado como una representación de estrategias de flujo en redes del mundo real, por ejemplo, la paradoja de Braess que agrega más rutas no siempre mejora el flujo [79]. Solmeyer et al. desarrollaron una versión cuantizada del juego de enrutamiento modificada para incluir un estado cuántico distribuido entre agentes que representan los nodos dentro de la red. Cada agente tiene permitido aplicar una estrategia cuántica local a su parte del estado en forma de una rotación unitaria antes de medir. Solmeyer et al. simularon el costo total del flujo de la red en términos de latencia total y encontraron que el costo mínimo se realiza cuando se usa un estado parcialmente entrelazado cuánticamente entre nodos. Notablemente, sus resultados han demostrado que la paradoja de Braess sigue existiendo pero solo para el caso de entrelazamiento cuántico máximo y nulo. Cuando las redes cuánticas se conviertan en una realidad, con múltiples agentes cuánticos independientes operando aplicaciones distribuidas, la teoría de juegos cuántica puede no solo proporcionar aplicaciones posibles, sino que también puede ser necesaria para su análisis.

En otras aplicaciones más recientes, [80] explora la aplicación de la teoría de juegos cuántica en redes cuánticas. Destaca cómo los juegos cuánticos superan a los clásicos en aspectos como recompensas y probabilidades de ganar, aplicando estas ventajas a retos de las redes cuánticas como la distribución de entrelazamiento y la topología de enrutamiento. Los resultados muestran mejoras en la fidelidad de los enlaces y reducción de latencia en comunicaciones, proponiendo un marco para futuras investigaciones en optimización de redes cuánticas mediante teoría de juegos. [81] se centra en el análisis del juego del Coronel Blotto cuántico (QCBG) en el contexto de redes de comunicación cuántica. Examina estrategias clásicas y cuánticas en el QCBG, explorando los equilibrios de Nash mediante un sistema de aprendizaje por refuerzo adversarial multiagente. Además, discute optimizaciones y problemas abiertos en la modelación y análisis del atascamiento en redes cuánticas, marcando un avance importante en la transición de la teoría cuántica a aplicaciones prácticas en comunicaciones de próxima generación.

[82] explora cómo la teoría de juegos cuántica puede optimizar el trading de alta frecuencia (HFT) en computadoras cuánticas. Se enfoca en la implementación del Dilema del Prisionero Cuántico en HFT, demostrando mejoras potenciales en velocidad y ganancias mediante la comunicación cuántica. Propone que la combinación

de teoría de juegos cooperativos y tecnología cuántica puede conducir a decisiones más eficientes y equilibrios Nash Pareto-óptimos en los mercados financieros. [83] explora dinámicas de aprendizaje en juegos cuánticos, introduciendo el modelo "Follow the Quantum Regularized Leader"(FTQL). Se analiza cómo las dinámicas de estado cuántico mezclado se descomponen en componentes clásicos y cuánticos, destacando que sólo los equilibrios cuánticos puros pueden ser estables y atractivos en FTQL. Además, se demuestra que las dinámicas FTQL tienen recurrencia de Poincaré en juegos cuánticos de suma cero, lo que sugiere un nuevo enfoque para comprender el aprendizaje y la toma de decisiones en entornos cuánticos.

Además del estudio de procesos clásicos, la teoría de juegos cuántica también pueden ser usada para el estudio de procesos estrictamente cuánticos. En particular, varios procesos no cooperativos subyacentes a los enfoques existentes para el desarrollo de tecnología cuántica, incluyendo control cuántico [84], corrección de errores cuánticos [65] y criptografía cuántica [85]. Cada una de estas áreas de aplicación requiere una solución a la competencia entre el usuario y el entorno, que puede considerarse un 'agente' en el escenario teórico de juegos. Las soluciones a estas aplicaciones específicas requieren un modelo de los procesos cuánticos mecánicos para dinámicas e interacciones que son más adecuados para la teoría de juegos cuántica.

La disponibilidad actual de prototipos de procesadores cuánticos de propósito general proporciona oportunidades para el estudio continuo de la teoría de juegos cuántica. Esto incluirá estudios experimentales de cómo los usuarios interactúan con juegos cuánticos, así como la traducción de estrategias cuánticas a escenarios del mundo real. Sin embargo, es probable que se necesiten redes cuánticas para las pruebas de campo de aplicaciones de juegos cuánticos, ya que la mayoría requiere la distribución de un recurso cuántico entre varios agentes. Junto con operaciones de entrelazamiento de alta fidelidad, estas redes cuánticas también deben proporcionar a los agentes marcos de control clásicos sincronizados e infraestructura. Estas redes de juegos cuánticos prototipo pueden entonces evolucionar hacia métodos de enrutamiento más robustos [65].

2.5.2. Implementaciones de Computadoras Cuánticas

En la búsqueda de avanzar hacia la implementación práctica de algoritmos basados en la Teoría de Juegos Cuántica, es crucial analizar las implementaciones actuales de computadoras cuánticas. Una de las aproximaciones más prometedoras y avanzadas en este campo es la utilización de qubits superconductores [86]. Estas implementaciones no solo lideran en términos de escalabilidad, con hitos significativos en número de qubits, sino que también presentan avances tecnológicos en términos de fidelidades y optimización energética. Centrarse en las computadoras cuánticas basadas en qubits superconductores no solo refleja el estado actual del arte en la computación cuántica, sino que también resulta esencial para comprender las perspectivas futuras de la computación cuántica aplicada a la teoría de juegos y su integración en sistemas de comunicación avanzados y algoritmos de aprendizaje por refuerzo multi-agente. Esta sección alinea la investigación con las tendencias y desarrollos más prometedores del campo, asegurando que los resultados y análisis presentados en esta tesis sean relevantes, actuales y aplicables a los desafíos futuros en física, computación e ingeniería.

Fundamentos y Desarrollo Histórico

Los qubits superconductores representan una confluencia revolucionaria entre la física cuántica y la ingeniería, emergiendo como la tecnología líder en el espacio comercial de la computación cuántica. Su historia y evolución, que arranca en los años ochenta, se entrelaza con avances fundamentales en la física de baja temperatura y la teoría cuántica. Empresas líderes como Google, así como una serie de startups innovadoras, están explotando esta tecnología que promete redefinir los límites del procesamiento de información. Al día de la escritura de este trabajo IBM tiene el record con una computadora de 433 qubits, aunque, hasta el momento, la calidad de estos qubits sigue siendo insuficiente para que sean útiles a nivel práctico [87].

El viaje hacia la comprensión y aplicación de los qubits superconductores comienza con la teoría BCS en 1957 [88, 89], que explica cómo los pares de Cooper se comportan a bajas temperaturas, generando el efecto superconductor. Este descubrimiento fue seguido por el efecto Josephson en 1962 y su verificación experimental en Bell Labs en 1963 [90], proporcionando un fundamento teórico crucial para la manipulación de fenómenos cuánticos en circuitos superconductores.

En 1985, se realizó un salto significativo cuando John Clarke, Michel Devoret y John Martinis demostraron el Tunelaje Cuántico Macroscópico de una unión Josephson sesgada por corriente, allanando el camino para la creación del primer "átomo eléctrico artificial"[91]. Este hito no solo demostró la existencia de niveles cuánticos discretos en un sistema macroscópico, sino que también estableció la base para futuras investigaciones en qubits superconductores.

Los qubits transmon, que dominan el campo actualmente, son osciladores anarmónicos no lineales cuya peculiaridad deriva de la unión Josephson. Esta unión permite una mejor separación de dos estados de energía en el superconductor que un simple resonador lineal. Esta anharmonicidad es crucial ya que permite la cuantización del flujo de corriente con niveles de energía discretos, controlables mediante pulsos de microondas en el régimen de 5 GHz, enmarcados en la electrodinámica cuántica de circuitos (cQED). Esta elección de frecuencias de operación está dictada por un equilibrio entre el costo de la electrónica para frecuencias superiores a 8 GHz y la interferencia del ruido térmico ambiental para frecuencias inferiores a 4 GHz.

En estos qubits, la superposición lineal de los dos primeros niveles de energía, que presentan diferentes funciones de onda relacionadas con la fase y las probabilidades de corriente, es clave para su operación. Estos estados superpuestos corresponden a un flujo de corriente oscilante en el rango de 10 MHz. La separación de estos niveles de energía se facilita por la diferente energía requerida para las transiciones entre los estados base $|0\rangle$ y $|1\rangle$ y otros niveles superiores. En los qubits transmon, la inductancia no lineal de la unión Josephson proporciona un oscilador anharmónico en el bucle Josephson, asegurando que la energía necesaria para acceder al estado $|1\rangle$ sea mayor que la requerida para acceder a niveles superiores, como se puede observar en la figura 2.7. Esto se alinea con las temperaturas de enfriamiento del procesador y el ruido ambiental, creando un entorno operativo ideal para los qubits.

Se requiere cierta cantidad de energía, conocida como el gap de energía, para romper los pares de Cooper que circulan en un qubit superconductor. Ésta es la razon por la cual los qubits que cuentan con aluminio como el material típico para la creación de la unión Josephson y sus alrededores, operan a alrededor de 15 mK. En el aluminio, el gap de energía corresponde a 90 GHz a 20 mK. Esto es un orden de magnitud mayor que la diferencia de energía entre los niveles $|0\rangle$ y $|1\rangle$. Significa que el qubit puede ser operado con energías más bajas (en el rango de 4–8 GHz) sin

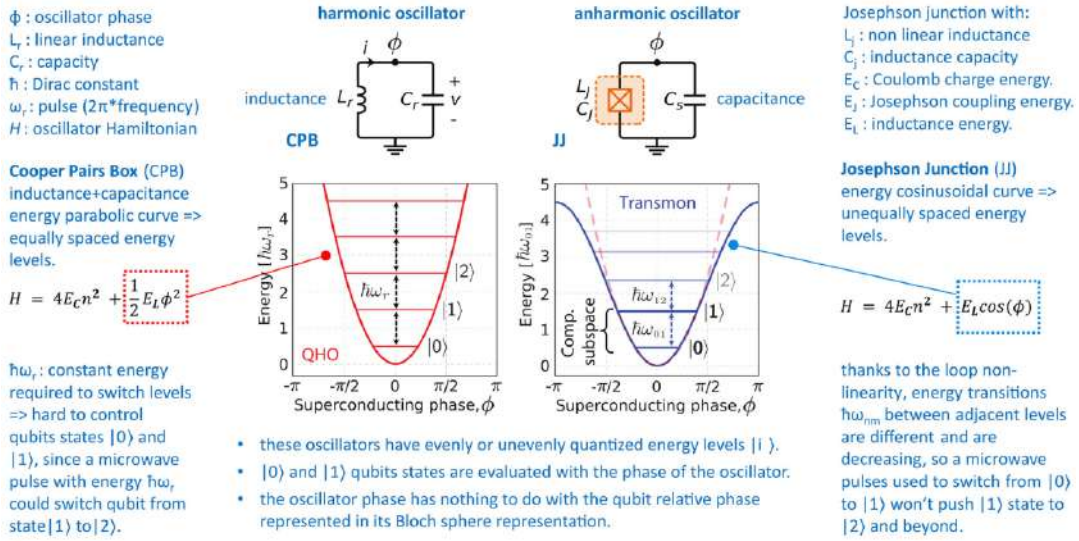


FIGURA 2.7: Los qubits superconductores utilizan un oscilador anarmónico para diferenciar dos niveles de energía correspondientes al estado fundamental y excitado del qubit [92, 86].

romper los pares de Cooper de la corriente superconductora y alterar la coherencia cuántica del qubit [93].

Los qubits transmon han visto varias innovaciones y variaciones, incluyendo qubits con frecuencias fijas o ajustables, acopladores sintonizables para entrelazar múltiples qubits, y puertas de fase controladas con amplitud y frecuencia variables. Estas variaciones han permitido reducir significativamente la profundidad de los circuitos cuánticos, especialmente para la implementación de la transformada de Fourier cuántica necesaria en muchos algoritmos [94]. Las técnicas para una lectura más rápida de qubits y el uso de qutrits en lugar de qubits también están siendo exploradas.

El qubit en sí está acoplado a una cavidad que contiene un resonador, generalmente implementado como un resonador de guía de onda coplanar. La energía del sistema se modeliza mediante un Hamiltoniano de Jaynes-Cummings [95], que implica conceptos como el espectro de Jaynes-Cummings, estados vestidos y un régimen dispersivo para la lectura de qubits [96].

Los qubits transmon y sus variantes representan un área de investigación vibrante y en rápida evolución. Con cada avance en el diseño y la implementación de estos qubits, nos acercamos un paso más a la realización de computadoras cuánticas prácticas y eficientes. Los desafíos en el camino incluyen mejorar la fidelidad de los qubits, aumentar su tiempo de coherencia y desarrollar métodos más eficientes para la lectura y manipulación de qubits. Además, la búsqueda de nuevos materiales y arquitecturas de qubits sigue siendo un área de investigación activa, prometiendo descubrimientos emocionantes en el futuro.

Tecnología y Operaciones de Qubits Superconductores

Los qubits transmon, como se mencionó anteriormente, son la variante más comúnmente utilizada y explorada por gigantes tecnológicos como IBM, Google y IQM. Estos dispositivos operan como osciladores anharmónicos no lineales, permitiendo controlar la frecuencia resonante de cada qubit individualmente. Sin embargo, la sintonización de la frecuencia conlleva una reducción en la vida útil (T1) del

qubit. Por tanto, resulta más efectivo mantener frecuencias fijas y distintas para cada qubit, minimizando la interferencia entre ellos.

Las operaciones de un solo qubit se realizan mediante pulsos de microondas transmitidos a través de cables coaxiales. Estos pulsos se ajustan al nivel $\hbar w_{01}$ (energía necesaria para pasar del estado $|0\rangle$ al $|1\rangle$), y su amplitud y fase se controlan cuidadosamente para realizar operaciones específicas. Estas operaciones incluyen varias compuertas cuánticas con ángulos de rotación ajustables que se generan a partir de señales de microondas en fase y en cuadratura (I y Q), lo que permite una manipulación precisa del estado cuántico del qubit [97].

Las puertas de dos qubits se logran mediante un circuito de acoplamiento situado entre los qubits, que puede ser un simple capacitor o un sistema controlable dinámicamente. IBM utiliza puertas de resonancia cruzada, mientras que Google y sus equivalentes chinos emplean un qubit intermedio en su procesador para gestionar este acoplamiento.

La lectura de los qubits varía según su tipo. En los transmon, un resonador acoplado al qubit se utiliza para transmitir un pulso de microondas y medir su efecto en la frecuencia y fase del resonador. Esta técnica, conocida como "lectura dispersiva", protege al qubit de todas las radiaciones excepto del pulso de lectura de microondas, amplificando la señal saliente con el menor ruido añadido posible.

Las configuraciones actuales de computación cuántica superconductora implican desafíos significativos en términos de escalabilidad. Los criostatos, necesarios para reducir la temperatura de los qubits hasta temperaturas tan bajas, se encuentran equipados con una compleja infraestructura de cables y atenuadores de microondas, con aproximadamente 4-5 cables por qubit físico, Fig. 2.8. La implementación de la corrección de errores cuánticos requerirá miles de qubits físicos por cada qubit lógico, un número que depende de factores como las fidelidades de los qubits físicos, su conectividad y las fidelidades objetivo de los qubits lógicos. Con códigos de corrección de errores como los códigos de superficie, se crearán desafíos importantes para ampliar la arquitectura, al menos, con las limitaciones de los criostatos actuales, fundamentales en la operación de los qubits superconductores, en términos de la cantidad de cableado y la generación externa de microondas necesarios para operar y leer los qubits.

Esta complejidad plantea desafíos significativos para la escalabilidad de los ordenadores cuánticos. Además, el procesamiento de los datos generados por los qubits requiere convertidores digital-analógico (DAC) y analógico-digital (ADC) que manejan grandes volúmenes de datos a velocidades de 8-14 Gbits/s, un diagrama de la arquitectura utilizada por Google para controlar los qubits se puede observar en la Fig. 2.9. Aunque los sistemas actuales se benefician del uso de equipos de laboratorio estándar para la generación de microondas, la escalabilidad sigue siendo un desafío que empresas como IBM están abordando mediante el desarrollo de su propia electrónica de control de qubits.

La fabricación de qubits superconductores implica técnicas que guardan similitud con la producción de circuitos analógicos clásicos y, en cierta medida, con la electrónica digital CMOS. Los materiales utilizados para fabricar qubits superconductores incluyen generalmente aluminio (para la unión Josephson, al menos para el dieléctrico), niobio (para condensadores y resonadores y, a veces, la unión Josephson) e indio (para los conectores del chipset), nitruro de titanio (para condensadores, con mejor factor de calidad) y ocasionalmente selenio (asociado al niobio y al orificio en los condensadores), silicio o zafiro (para el sustrato de la oblea) y tantalio (ver Fig. 2.10) [98].

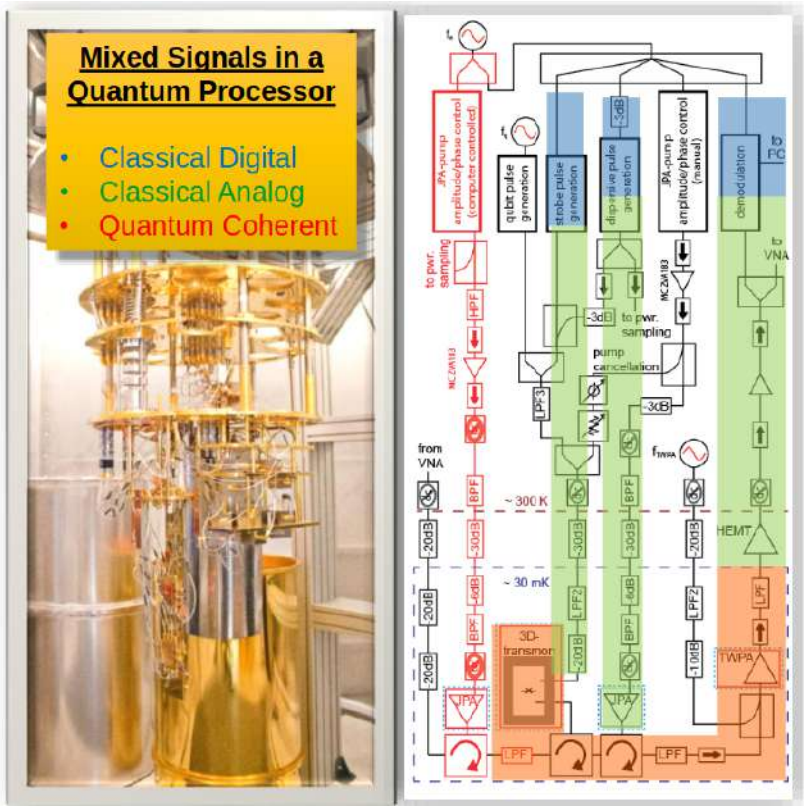


FIGURA 2.8: La "tiranía"de los cables en los qubits superconductores. Cuando las QPU alcancen miles de qubits, se necesitarán soluciones de multiplexación innovadoras ya que no es escalable la forma en la que los qubits están conectados actualmente con el mundo exterior [86].

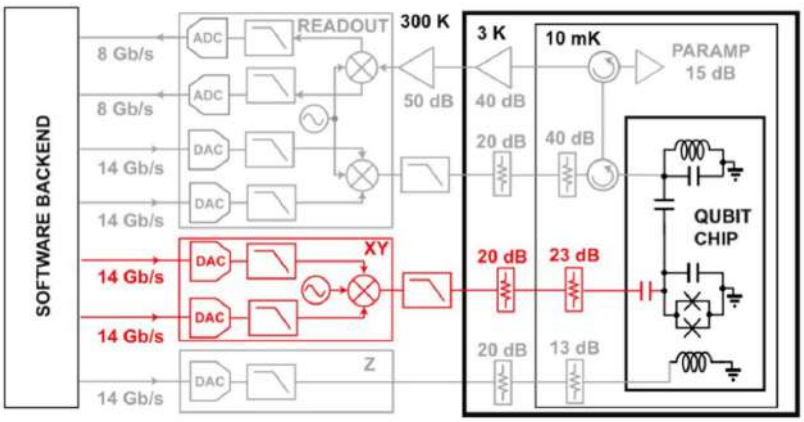


FIGURA 2.9: Arquitectura de lectura y control de qubit de Sycamore que muestra los 4 cables que impulsan un qubit. Fuente: Google.

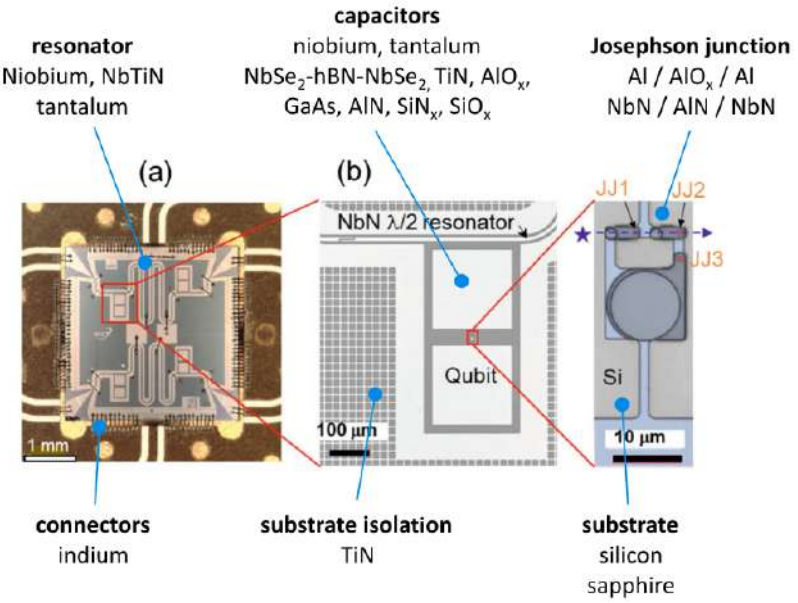


FIGURA 2.10: Los diversos componentes y materiales utilizados en un qubit superconductor.

La miniaturización de los qubits superconductores representa un área activa de investigación, ya que su tamaño es considerable, principalmente por sus resonadores que suelen medir alrededor de $1mm^2$, pero con potencial de reducción hasta $0{,}04mm^2$ [99, 100]. Se están explorando técnicas de fabricación avanzadas para disminuir el tamaño de los resonadores, lo que a su vez reduciría la energía requerida para enfriar los qubits a temperaturas extremadamente bajas. Además, se está trabajando en mejorar la conectividad entre los qubits superconductores mediante circuitos 3D, así como en la separación de los qubits de los controles de microondas utilizando vias a través de silicio (TSV), como las desarrolladas por IBM.

Las fidelidades de los qubits superconductores, inferiores a otros tipos de tecnologías, como por ejemplo los iones atrapados, se ven afectadas por diversos orígenes de ruido y aún no alcanzan para implementar códigos de corrección de errores, aunque se proponen métodos para mejorar la fidelidad de lectura [101, 102]. Por otro lado, aunque el tamaño en el rango de micrones de estos qubits complica la fabricación de chips de millones de qubits, y su miniaturización reduce su calidad [103], proveedores como IBM planean desarrollar chipsets de hasta 133 qubits, interconectándolos con guías de microondas y/o enlaces fotónicos entrelazados capaces de convertir estados cuánticos de qubits en estados cuánticos fotónicos.

Investigación Actual y Preparación del Mercado

En la vanguardia de la computación cuántica, numerosos laboratorios en todo el mundo, incluyendo instituciones en EE.UU., Europa, Asia y la industria, están investigando activamente los qubits superconductores. Con un enfoque en variantes como transmon y fluxonium, los esfuerzos de investigación se centran en prolongar el tiempo de coherencia de los qubits superconductores, un parámetro crítico que actualmente limita la cantidad de operaciones cuánticas que se pueden realizar de manera confiable.

En lugares como la Universidad de Princeton, se han logrado avances significativos, extendiendo el tiempo de coherencia T1 hasta 1.6 ms, una mejora notable en

comparación con los tiempos de coherencia anteriores de unos pocos cientos de microsegundos. Sin embargo, estos logros a menudo se limitan a un pequeño número de qubits funcionales. La investigación también explora el uso de materiales innovadores como el nitruro de titanio y el tantalio en sustratos de zafiro, así como el diseño de cavidades resonantes superconductoras de radiofrecuencia (SRF) que han demostrado tiempos de vida del qubit de hasta 2 segundos [104].

Un área de investigación particularmente intrigante es la conversión de fotones de microondas superconductores en fotones en el espectro visible/infrarrojo, lo que facilitaría su transporte a larga distancia y compartir el entrelazamiento, fundamental para la computación cuántica distribuida. Además, se están investigando métodos para simplificar la lectura de los qubits, como el uso de contadores de fotones de microondas y multiplicadores de fotones Josephson (JPM) integrados directamente en el chipset del qubit [105].

En el panorama actual de la computación cuántica, los ordenadores cuánticos superconductores se clasifican generalmente en las categorías pre-NISQ y NISQ (computadores cuánticos de escala intermedia y ruidosos). Los sistemas NISQ, con más de 50 qubits físicos, pueden ofrecer, en algunas casos muy específicos, algunas ventajas en velocidad de cálculo, calidad de resultados y eficiencia energética en comparación con los mejores computadores clásicos. Sin embargo, los sistemas pre-NISQ, con menos de 50 qubits, generalmente no alcanzan el umbral de ventaja cuántica.

Los sistemas NISQ utilizan algoritmos cuánticos que deben ser resistentes al ruido y tener un bajo número de ciclos de puertas cuánticas. Según una fórmula general, las fidelidades requeridas para un algoritmo dado dependen del error de puerta de dos qubits del sistema (ϵ), el número de qubits utilizados (n) y la profundidad del algoritmo cuántico (d), $\epsilon = \frac{1}{n*d}$. Este requisito de alta fidelidad presenta un desafío significativo para los algoritmos que requieren un gran número de ciclos de compuertas cuánticas.

Los algoritmos típicos para plataformas NISQ incluyen el solucionador de autovalores cuánticos variacionales (VQE) [106], algoritmos de optimización aproximada cuántica (QAOA) [107] y aprendizaje automático cuántico (QML) [108]. Estos algoritmos son esenciales en campos como la simulación química cuántica, optimización combinatoria y tareas de clasificación, agrupación automática y predicción en aprendizaje automático.

En términos de preparación para el mercado, la mayoría de los ordenadores cuánticos superconductores actuales aún no alcanzan la zona utilizable de NISQ, como lo muestra el gráfico de dispersión de fidelidades de puertas de dos qubits y número de qubits de la Fig. 2.11. Empresas como IBM planean lanzar nuevas unidades de procesamiento cuántico (QPUs) con fidelidades en el rango del 99.9 %, suficientes para ejecutar algoritmos cuánticos NISQ con alguna ventaja cuántica. Sin embargo, existe una competencia entre dos enfoques: 1) qubits de muy alta fidelidad y mitigación de errores cuánticos versus 2) una gran cantidad de qubits de alta fidelidad y corrección de errores cuánticos.

Paralelamente al desarrollo de sistemas NISQ, se anticipa el advenimiento a largo plazo de los ordenadores cuánticos tolerantes a fallas (FTQC). Estos QPUs se basarán en grupos de qubits lógicos con tasas de error reducidas gracias al uso de códigos de corrección de errores cuánticos y redundancia. Los qubits lógicos tendrán una tasa de error más baja dependiendo de los códigos de corrección de errores, las fidelidades de los qubits físicos, su conectividad y su número. Se espera que los qubits lógicos basados en superconductores requieran entre mil y millones de qubits físicos dependiendo de la tasa de error objetivo de la aplicación. Esta tasa de error objetivo varía según la profundidad y amplitud del algoritmo cuántico del usuario

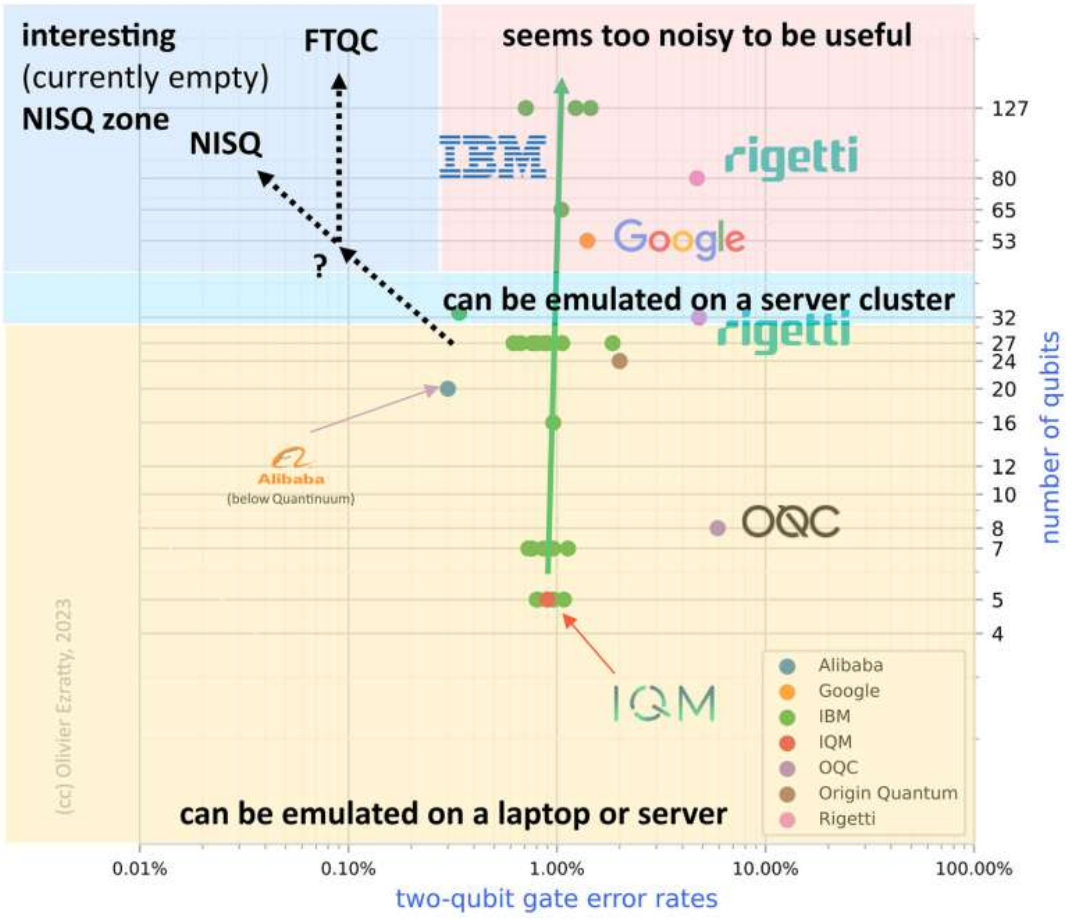


FIGURA 2.11: Fidelidades actuales de puertas de dos qubits de computadoras qubit superconductoras de proveedores comerciales. La zona azul corresponde al área donde las QPU podrían aportar alguna ventaja computacional ya sea en el régimen NISQ o FTQC. El régimen FTQC requiere al menos un 99,9 % de fidelidad y una escala a millones de qubits, mientras que el régimen NISQ se basa en unos pocos cientos o miles de qubits. Fuente: datos de proveedores y compilación en 2023 en [109].

y no es necesariamente estática en un QPU dado. Los algoritmos FTQC que proporcionan un aumento exponencial de velocidad utilizan muchas compuertas para implementar la transformada de Fourier cuántica (QFT), una primitiva utilizada en la estimación de fase cuántica y las estimaciones de amplitud cuántica, algoritmos de álgebra lineal y factorización de enteros.

La transición de sistemas NISQ a computadoras cuánticas tolerantes a fallas representará un cambio significativo en la computación cuántica, con un enfoque en qubits lógicos y corrección de errores cuánticos. Los avances en este campo serán esenciales para superar las limitaciones actuales y desbloquear el verdadero potencial de la computación cuántica. A medida que la tecnología continúa desarrollándose, se espera que los qubits superconductores desempeñen un papel fundamental en la realización de esta prometedora y revolucionaria tecnología.

Desafíos y Perspectivas Futuras

Como todos los tipos de qubits, los qubits superconductores tienen sus desafíos para posibilitar la creación de computadoras cuánticas útiles, tanto en el ámbito de NISQ como de FTQC. En teoría, la tecnología podría escalar a miles o incluso millones de qubits. Las mejores fidelidades de clase fueron obtenidas por IBM con su procesador Egret de 33 qubits en noviembre de 2022, mostrando una fidelidad de puerta de dos qubits del 99.7 %. Crear una computadora cuántica tolerante a fallas requeriría al menos unos 100,000 qubits físicos con una fidelidad del 99.9 %. Esto permitiría 100 qubits lógicos.

Estos requisitos crean desafíos inmensos: ¿se puede contener la interferencia entre qubits a esta escala? ¿Es posible entrelazar de manera máxima y controlada sistemas cuánticos de muchos cuerpos tan grandes? ¿Cómo diseñar códigos de corrección de errores cuánticos con un mínimo de sobrecarga de qubits físicos y requisitos de fidelidad? ¿Es posible crear electrónica de control de baja potencia, cableado, multiplexación y criogenia adecuados para alcanzar dicha escala [110]? ¿Se podrá contener el consumo de energía relacionado, una pregunta interdisciplinaria que la Iniciativa de Energía Cuántica propone abordar de manera sistémica [111]? ¿Será posible interconectar varios procesadores cuánticos con recursos de microondas o fotones entrelazados? Todos estos desafíos científicos y tecnológicos son gigantescos.

Otro desafío es mejorar las herramientas de software utilizadas para diseñar estos chipsets de qubits. Las herramientas de Automatización de Diseño Electrónico (EDA) funcionan en computadoras clásicas. Hay algunas herramientas para diseñar y simular digitalmente chipsets de qubits, pero aún no están suficientemente integradas. Entre estas se encuentra Qiskit Metal de IBM. Anunciado en 2021, actualmente se encuentra en versión alfa. En febrero de 2023, Amazon AWS presentó Palace (PArallel, LArge-scale Computational Electromagnetics), un código de fuente abierta de elementos finitos para simulaciones electromagnéticas de onda completa capaz de simular un único qubit transmon. Otros marcos de Python se utilizan para simular en varios niveles de abstracción un chipset de qubit, desde el funcionamiento interno del qubit hasta el chipset integrado completo.

Mientras tanto, proveedores como IBM y Google intentan crear sistemas NISQ con cientos de qubits que podrían ofrecer alguna ventaja en la computación cuántica gracias a la técnica de mitigación de errores cuánticos que funciona con algoritmos de poca profundidad, particularmente, los variacionales que trabajan en modo híbrido junto con supercomputadoras [86].

Capítulo 3

Modelado de Redes de Comunicación utilizando Teoría de Juegos Cuántica

3.1. Introducción

La era moderna de las telecomunicaciones ha presenciado un crecimiento sin precedentes en la transmisión de paquetes de datos, impulsado en gran medida por la globalización y la proliferación de usuarios móviles. Este aumento constante ha planteado serios desafíos en términos de congestión, afectando negativamente el rendimiento de las redes modernas. La congestión, que se manifiesta en diversos entornos, desde colas en supermercados hasta tráfico urbano y redes 5G, surge cuando múltiples usuarios compiten por recursos limitados, lo que lleva a un aumento en la latencia para todos los contendientes.

Abordar el problema de la congestión no es una tarea trivial; se puede describir como un equilibrio entre la latencia y el costo de transmisión. Tradicionalmente, la teoría de juegos ha ofrecido un marco para modelar tales problemas, especialmente cuando las decisiones egoístas de los agentes, tales como los paquetes en una red, pueden afectar adversamente el rendimiento del sistema en su conjunto. No obstante, la teoría de juegos clásica presenta limitaciones, y es en este contexto donde la teoría de juegos cuántica se revela como una solución promisoria.

La teoría de juegos cuántica, que combina los principios fundamentales de la mecánica cuántica con la teoría de juegos, amplía las posibilidades estratégicas disponibles para los agentes. Al aprovechar propiedades cuánticas únicas, como la superposición y el entrelazamiento, los agentes pueden potencialmente obtener resultados más óptimos que en juegos clásicos. Esta extensión cuántica de la teoría de juegos ha demostrado ser particularmente efectiva en el modelado y solución de problemas de congestión en redes de comunicación.

En el contexto de las redes de comunicación, se ha propuesto un marco basado en la teoría de juegos cuántica donde los paquetes de la red compiten de manera egoísta por la ruta más corta. Las simulaciones y estudios han demostrado que los tiempos de enrutamiento y viaje finales logrados con la versión cuántica superan significativamente a los obtenidos con modelos clásicos. Estos avances abren la puerta al desarrollo de protocolos de comunicación que pueden aprovechar al máximo las propiedades cuánticas para optimizar los sistemas de comunicación.

Más allá de las redes clásicas, se vislumbra la llegada de internet cuántico en el horizonte. Con proyectos en desarrollo en todo el mundo, desde Estados Unidos [112] hasta Europa [113] y China [114], es esencial investigar cómo se puede lograr la comunicación cuántica de la manera más eficiente. Los protocolos basados en la

teoría de juegos cuántica ofrecen soluciones para las redes actuales, contribuyendo así al avance global de las comunicaciones.

Una característica distintiva del protocolo cuánticos propuesto es su capacidad de adaptación. Al combinar la teoría de juegos con el aprendizaje por refuerzo, se ha diseñado un sistema que, al acumular recompensas, aprende a minimizar la latencia en una red congestionada en función de las propiedades del sistema en tiempo real. Esta capacidad de autoadaptación es esencial para garantizar que las redes cuánticas puedan responder dinámicamente a las condiciones cambiantes y garantizar un rendimiento óptimo.

Sin embargo, tal como ocurre con cualquier tecnología emergente, se presentan desafíos. Uno de los principales obstáculos es la influencia del ruido en el comportamiento de los sistemas cuánticos. A pesar de la ausencia de computadoras cuánticas ideales, los estudios han demostrado que los beneficios de los protocolos cuánticos aún prevalecen incluso bajo la influencia de niveles de ruido bajos tanto en dispositivos cuánticos simulados como reales.

La teoría de juegos cuántica ofrece un nuevo paradigma para abordar los persistentes problemas de congestión en las redes de comunicación. Al fusionar la mecánica cuántica, la teoría de juegos y el aprendizaje automático, se abre un mundo de oportunidades para diseñar sistemas de comunicación más eficientes y adaptativos, sentando las bases para la próxima generación de Internet Cuántico [115, 116, 117].

3.2. Mitigación de la congestión de redes mediante la teoría de juegos cuánticos

3.2.1. Teoría cuántica de juegos para el enrutamiento de redes de datos: una solución al problema de la congestión

Introducción

En esta sección, abordamos las bases del desafío de la congestión en las redes de datos, un problema que se intensifica con el crecimiento sostenido de usuarios móviles y la transmisión de una gran cantidad de paquetes de datos, tal y como se lo presentó en el trabajo [3]. La gestión de la congestión se presenta como un delicado equilibrio entre la latencia y el costo de transmisión, un dilema que se ha vuelto cada vez más prominente en el rendimiento de las redes modernas. En este escenario, proponemos un marco novedoso para resolver el problema de la congestión en redes de comunicación utilizando un modelo de teoría de juegos cuánticos, donde los paquetes de red compiten egoístamente por la ruta más corta. Este enfoque no solo aborda la eficiencia en la selección de rutas sino que también introduce una nueva dimensión en la toma de decisiones estratégicas a través de la mecánica cuántica.

El modelo que presentamos se basa en la aplicación de estrategias cuánticas, representadas por un modelo de compuertas de un qubit con tres parámetros. Al expandir las posibilidades de los agentes de ser clásicos a cuánticos, demostramos cómo se puede evitar la congestión en la red y reducir el tiempo de viaje total de todos los paquetes en un sistema con un elevado volumen de paquetes que deciden autónomamente cuál es la mejor decisión para el beneficio propio. A continuación se realiza una presentación detallada del problema y el sistema, seguida de una explicación de las estrategias posibles para modelar la red y el desarrollo en profundidad del modelo cuántico. Los resultados de cada protocolo se comparan y su rendimiento se analiza gráficamente, culminando en una discusión sobre las implicaciones de nuestro enfoque.

Modelado del problema de la congestión

El objetivo principal es minimizar el tiempo total de transmisión, el cual está formado por la suma del tiempo de enrutamiento y el tiempo de viaje. El tiempo de enrutamiento se define como el intervalo necesario para que un paquete determine una ruta desde su punto de origen hasta su destino. Más precisamente, en el modelo presentado, el tiempo de enrutamiento es proporcional al número de juegos que un paquete debe jugar antes de encontrar su camino final. Cuanto mayor sea el número de posibles rutas que un paquete considere, mayor será el tiempo de enrutamiento. El tiempo de viaje, por su parte, mide cuánto tarda un paquete en viajar desde el origen hasta el destino una vez elegida la ruta. Esto significa, la suma de los costos o pesos, $w(e, e+1)$, de todos los canales que forman parte de un camino final.

Por lo tanto, $TiempoTotal_i = K_G G_i + \sum_{e=1}^{E_i} w(e, e+1)$, donde el primer término representa el tiempo de enrutamiento y el segundo término el tiempo de viaje. K_G es el tiempo que se tarda en jugar un juego, G_i es el número de juegos jugados por el agente i, E_i es el número de canales de la ruta final del agente i y $w(e, e+1)$ es el peso del canal que conecta los nodos e y $e+1$.

El modelo de red de comunicación tiene $n1$ nodos y $n2$ paquetes que se enviarán entre nodos. La red se genera utilizando el modelo de Erdős–Rényi–Gilbert [118] con $G(nodes = n1, probability = 0{,}5)$ (donde *nodes* es el número de nodos y *probability* la probabilidad de que dos nodos estén conectados entre sí) con la condición de que cada nodo tenga al menos una conexión con otro nodo. Se opta, para las simulaciones, que la capacidad máxima de cada canal es de un paquete, con un rendimiento que disminuye linealmente a partir de ahí. Este declive se reflejará en nuestro modelo aumentando el peso correspondiente, $w(e, e+1)$, de cada canal proporcionalmente al número de paquetes que lo atraviesan. Un ejemplo de una red de $n1 = 10$ nodos se muestra en la Figura 3.1, donde los paquetes que viajan a través de ella están representados con diferentes colores.

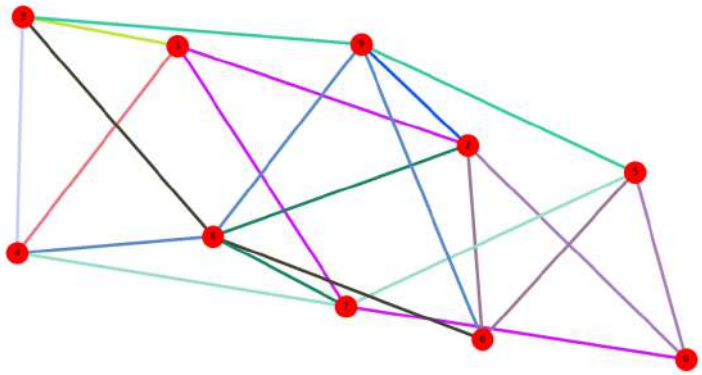


FIGURA 3.1: Ejemplo de modelo de red para $n1 = 10$.

En un canal congestionado, el tiempo de viaje para un paquete dado será mucho más largo que en uno desocupado. En otras palabras, a medida que aumenta el número de paquetes que viajan a través del mismo canal, el tiempo de viaje para cada paquete aumenta. Esta situación es análoga al transporte en la ciudad, donde el tiempo que un vehículo tarda en recorrer una ruta aumenta a medida que aumenta el número de vehículos en esa ruta. Si esta situación no se gestiona adecuadamente, puede resultar en una grave congestión de tráfico [119].

Cada paquete se considera como un agente que desea encontrar la de ruta más

corta desde su origen hasta su destino. La congestión (cuando la cantidad de paquetes supera la capacidad de los canales) hace que el tiempo de viaje a través de los canales congestionados sea mucho más largo. Finalmente, cada vez que un agente desea tomar un canal que está congestionado, se juega un juego en el nodo, donde todos los agentes que deseaban tomar el canal tiene dos opciones: buscar otro canal alternativo o viajar a través del canal a pesar de estar congestionado.

Si cada agente decide elegir su ruta más corta, algunos canales de la red estarán sobrecargados con todos los paquetes que están disputándolos, ralentizando la red. Por lo tanto, si los agentes se comportan de manera egoísta, siempre eligiendo su ruta más corta, toda la red se verá perjudicada. En el otro extremo, si cada agente decide buscar canales desocupados, el tiempo de enrutamiento aumenta y la red también se ralentizaría. Este modelo sencillo resume el considerable desafío que las tecnologías de comunicación modernas enfrentan, planteando una seria amenaza a las estrategias de enrutamiento existentes [120]. Este fenómeno es conocido como el dilema de la congestión, como exploraremos en las secciones siguientes.

Estrategias clásicas y cuánticas

Es bien sabido que un juego se define por tres elementos: agentes, estrategias y recompensas. En este caso, el juego planteado tiene una naturaleza no cooperativa, es decir, es un juego con competencia entre agentes individuales, siendo los agentes los paquetes que viajan a través de la red. Cuando la cantidad de paquetes alcanza niveles que podrían conducir a la congestión del canal, las estrategias son tomar o no dicho canal. Por último, la recompensa es el tiempo total que se especifica sumando el tiempo de enrutamiento y el tiempo de viaje (con un signo negativo ya que cuanto menor sea el tiempo, mayor será la recompensa).

Si más de un paquete está interesado en un canal, porque es parte de su ruta actualmente más corta, cada paquete tiene dos estrategias posibles: elegir este canal preferido (arriesgándose a que otros paquetes también lo seleccionen y luego congestionen el canal) *o* buscar su siguiente ruta más corta (más larga que la anterior pero posiblemente menos congestionada). Estas dos estrategias se llamarán opción 1 y 0, respectivamente. Por ejemplo, en un escenario con dos agentes, cada agente tiene dos opciones: 1) Tomar la ruta más corta arriesgándose a que la congestión aumente significativamente su tiempo de viaje. 0) Probar con su segundo camino más corto, donde podría no haber congestión. En la Tabla 3.1, se presenta una situación en la que dos paquetes están interesados en el mismo canal.

CUADRO 3.1: Ejemplo de dos paquetes interesados en el mismo canal.

Agentes		Agente 1	
Agente 0	**Acciones**	0	1
	0	Ninguno de los dos toma el canal y ambos van a buscar otro.	El agente 1 toma el canal y el agente 0 va a buscar otro.
	1	El agente 0 toma el canal y el agente 1 va a buscar otro.	Ambos toman el canal creando un camino congestionado.

Los juegos pueden ser de diferente tipo, clásicos o cuánticos. Consideremos primero los juegos clásicos. Las estrategias clásicas mixtas del agente son probabilísticas, es decir, la opción 0 se elige con una probabilidad p y la opción 1 con una probabilidad $(1-p)$. Entonces, una probabilidad p cercana a cero corresponde a agentes codiciosos, ya que tenderán a tomar siempre la ruta más corta incluso cuando muchos agentes compiten por ese canal. Un valor de p más cercano a uno creará agentes que buscan con más paciencia otra ruta menos congestionada.

Para estudiar la elección de canal utilizando juegos cuánticos se sigue el protocolo *EWL* para 2 agentes [51] y luego la extensión para N agentes propuesta en [55]. El *primer* paso es asignar un estado cuántico a cada una de las estrategias posibles. El protocolo cuántico es exactamente el mismo que el clásico, la única diferencia es que las estrategias 0 y 1 anteriormente se representaban en un bit y ahora se representan en un qubit. La estrategia 0 (dejar la ruta preferida) se mapea al estado cuántico $|0\rangle$ y la estrategia 1 (tomar la ruta preferida) al estado cuántico $|1\rangle$. El *segundo* paso es crear un circuito cuántico donde a cada agente se le asignará un qubit que comenzará en el estado $|0\rangle$. El *tercer* paso es crear un estado máximamente entrelazado entre todos los agentes. Esto se hace aplicando el operador de entrelazamiento $J = \frac{1}{\sqrt{2}}(\mathbb{I}^{\otimes N} + i\sigma_x^{\otimes N})$, como se ve en la Figura 3.2, donde el número de agentes es $N = 2$.

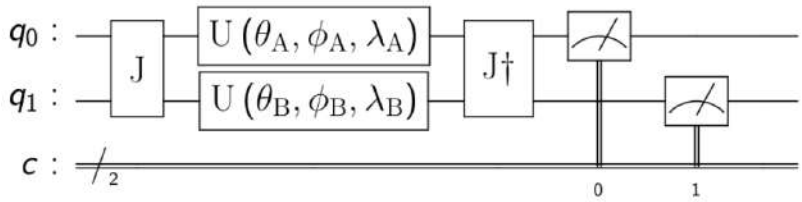


FIGURA 3.2: Modelo de juego EWL para 2 agentes. Donde q_0 y q_1 son los estados cuánticos iniciales de los agentes y c es un registro clásico donde se almacenan las mediciones de los qubits.

En el *cuarto* paso, cada agente elige su estrategia más adecuada de manera individual e independiente. Esto se hace modificando el estado de su propio qubit localmente. Para hacer esto, cada agente aplica una o más puertas de un solo qubit, modificando el estado de su qubit. Una puerta general de un solo qubit es una matriz unitaria que puede representarse como [121]:

$$U(\theta,\phi,\lambda) = \begin{pmatrix} \cos(\frac{\theta}{2}) & -e^{i\lambda}\sin(\frac{\theta}{2}) \\ e^{i\phi}\sin(\frac{\theta}{2}) & e^{i(\phi+\lambda)}\cos(\frac{\theta}{2}) \end{pmatrix} \tag{3.1}$$

Ya podemos destacar el hecho de que mientras los agentes clásicos tienen solo un parámetro para elegir su estrategia mixta: p, los agentes cuánticos tienen tres parámetros para elegir con el fin de seleccionar su propia estrategia pura: θ, ϕ y λ. El *quinto* paso, según el protocolo EWL, es aplicar el operador $J^\dagger$ (transpuesto conjugado de J) después de las estrategias de los agentes. Finalmente, el *sexto* paso consiste en medir el estado de los qubits para conocer el resultado del circuito y, por lo tanto, la acción final de cada agente.

Resultados

En esta sección, comparamos los resultados de los protocolos clásicos y cuánticos. Las simulaciones se realizaron promediando el rendimiento de 50 simulaciones para diferentes configuraciones de red aleatorias generadas automáticamente.

Se realizan simulaciones numéricas del caso clásico con estrategias mixtas, barriendo en p, mientras que en el caso cuántico con estrategias puras, como se explica a continuación, barriendo en φ_X, φ_Y y φ_Z.

Las simulaciones del juego clásico, donde todos los agentes tienen la probabilidad p, se muestran en la Figura 3.3. La Figura 3.3a muestra cómo el tiempo promedio de viaje por paquete aumenta con el incremento del número de paquetes para diferentes probabilidades p. La variación del tiempo promedio de enrutamiento con el número de paquetes se muestra a su lado en la Figura 3.3b.

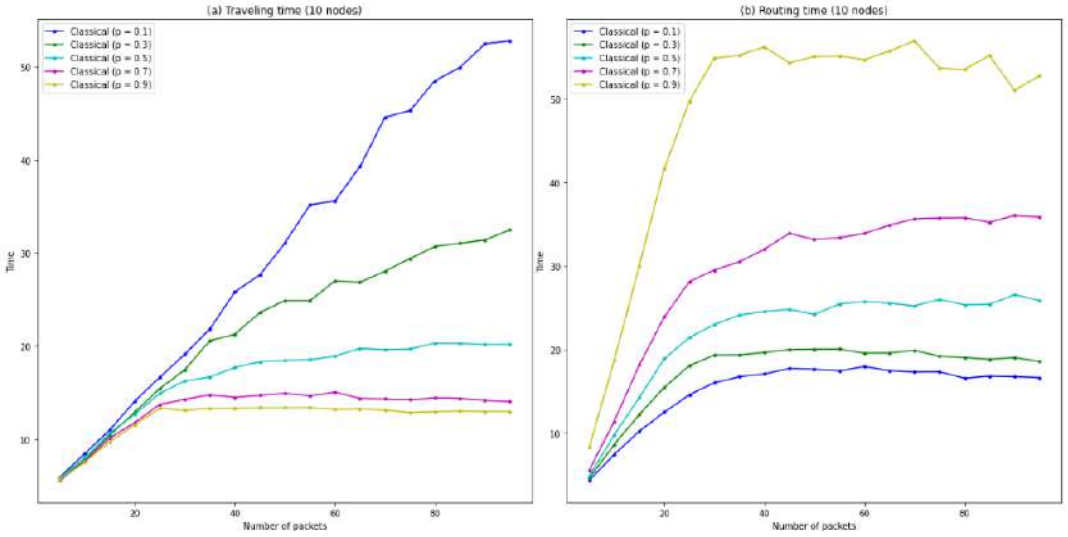


FIGURA 3.3: Gráficas para diferentes probabilidades p de: (a) Tiempo de viaje en función del número de paquetes. (b) Tiempo de enrutamiento en función del número de paquetes.

Las simulaciones se realizaron en redes de $n1 = 10$ nodos y un valor de $n2$ paquetes desde $n2 = 5$ hasta $n2 = 100$ paquetes, promediando el rendimiento de 50 casos diferentes. El comportamiento cualitativo de la red, es decir, la dinámica reflejada en los gráficos, es independiente del número de nodos. La única diferencia es el número de paquetes necesarios para que la red se congestione; cuanto mayor sea la red, mayor será el número de paquetes requeridos para saturarla. Analizando los gráficos es posible concluir que existe una especie de compensación entre el tiempo de viaje y el tiempo de enrutamiento una vez que la red está congestionada. Dado un número fijo de paquetes y un p bajo, el tiempo de viaje por paquete aumenta pero el tiempo de enrutamiento disminuye. Por otro lado, a medida que p aumenta, el tiempo de viaje por paquete disminuye pero el tiempo de enrutamiento aumenta.

Este efecto se puede ver en la Figura 3.4 donde se muestran los tiempos de viaje y enrutamiento para diferentes probabilidades p. Estos resultados se obtienen cuando el sistema alcanza un estado estacionario con un alto número de paquetes ($n2 = 100$).

Como se mencionó anteriormente, la estrategia del agente cuántico consiste en una secuencia de puertas cuánticas de 1 qubit. Para demostrar el potencial que tienen las estrategias cuánticas, vamos a empezar estudiando un caso particular: el sistema con rotaciones en los ejes X, Y y Z. Las matrices de puertas cuánticas de un qubit para rotaciones son:

$$R_X(\varphi) = \begin{pmatrix} \cos(\frac{\varphi}{2}) & -i\sin(\frac{\varphi}{2}) \\ -i\sin(\frac{\varphi}{2}) & \cos(\frac{\varphi}{2}) \end{pmatrix} \tag{3.2}$$

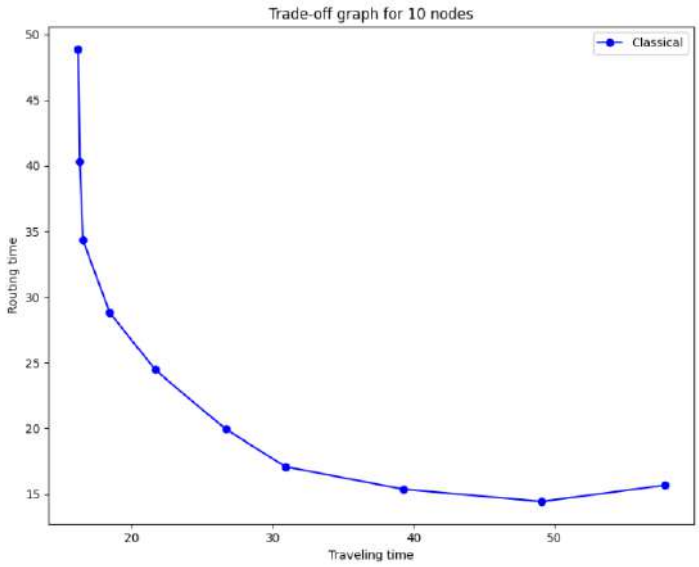


FIGURA 3.4: Trade-off entre el tiempo de viaje y de enrutamiento para diferentes valores de p entre 0 y 0,9. Los valores de p más cercanos a 0 dan un tiempo de viaje alto y un tiempo de enrutamiento bajo. Los valores de p más cercanos a 1 dan un tiempo de viaje bajo y un tiempo de enrutamiento alto.

$$R_Y(\varphi) = \begin{pmatrix} \cos(\frac{\varphi}{2}) & -\sin(\frac{\varphi}{2}) \\ \sin(\frac{\varphi}{2}) & \cos(\frac{\varphi}{2}) \end{pmatrix} \tag{3.3}$$

$$R_Z(\varphi) = \begin{pmatrix} e^{-i\frac{\varphi}{2}} & 0 \\ 0 & e^{i\frac{\varphi}{2}} \end{pmatrix} \tag{3.4}$$

Por lo tanto, cada agente debe elegir 3 ángulos que llamaremos φ_X, φ_Y y φ_Z. En nuestro diseño, proponemos la estrategia $\varphi_X = \frac{\pi}{2}$, $\varphi_Y = \frac{\pi}{4}$ y $\varphi_Z = 0$ para cada agente, como se muestra en la Figura 3.5.

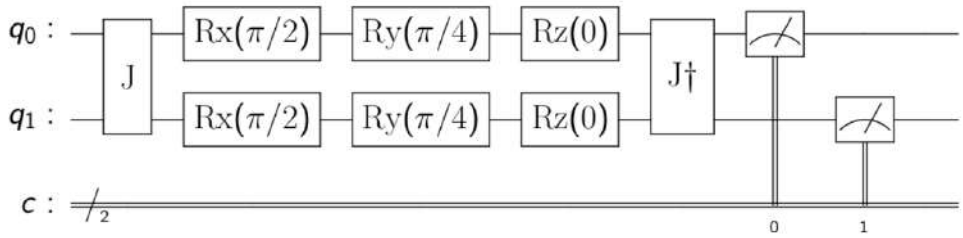


FIGURA 3.5: Circuito correspondiente al protocolo de enrutamiento para 2 agentes.

En las Fig 3.6a) y Fig 3.6b) se muestra el rendimiento del protocolo cuántico cuando todos los agentes están usando $S_1 = (\frac{\pi}{2}, \frac{\pi}{4}, 0)$ en comparación con el rendimiento clásico. El juego cuántico muestra el menor tiempo de viaje y un tiempo de enrutamiento medio en comparación con el clásico. Con esto en mente, podemos recalcular la Figura 3.4 y agregar el caso cuántico. Al hacer esto, obtenemos la Figura 3.7 donde podemos observar cómo la barrera de compensación de tiempo de viaje-enrutamiento clásica es superada por este protocolo cuántico. Así, obtendremos un

rendimiento que mejora cualquier rendimiento clásico alcanzando simultáneamente menos tiempo de enrutamiento y menos tiempo de viaje.

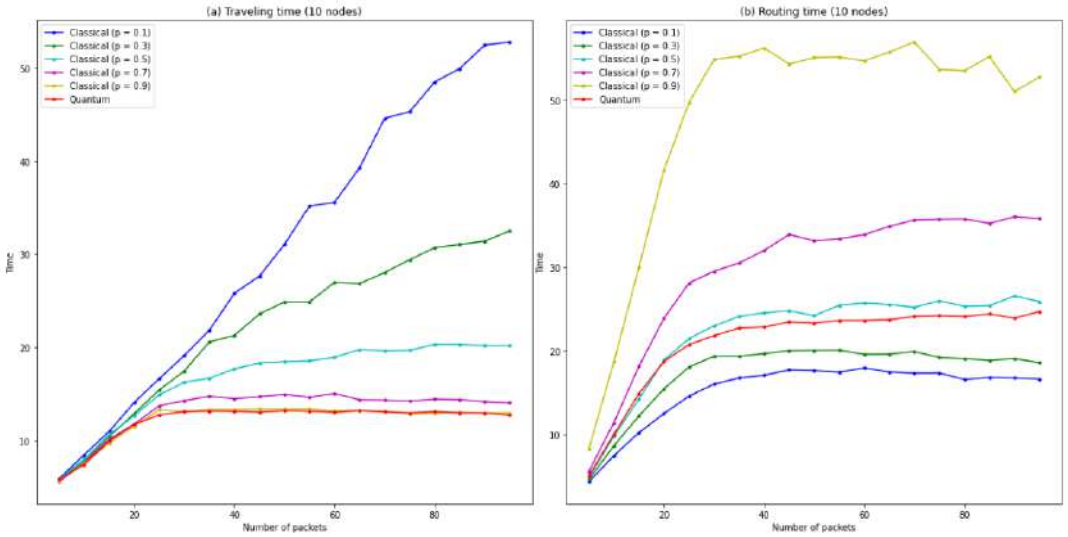


FIGURA 3.6: Gráficas para diferentes probabilidades p y el caso cuántico: (a) Tiempo de viaje en función del número de paquetes. (b) Tiempo de enrutamiento en función del número de paquetes.

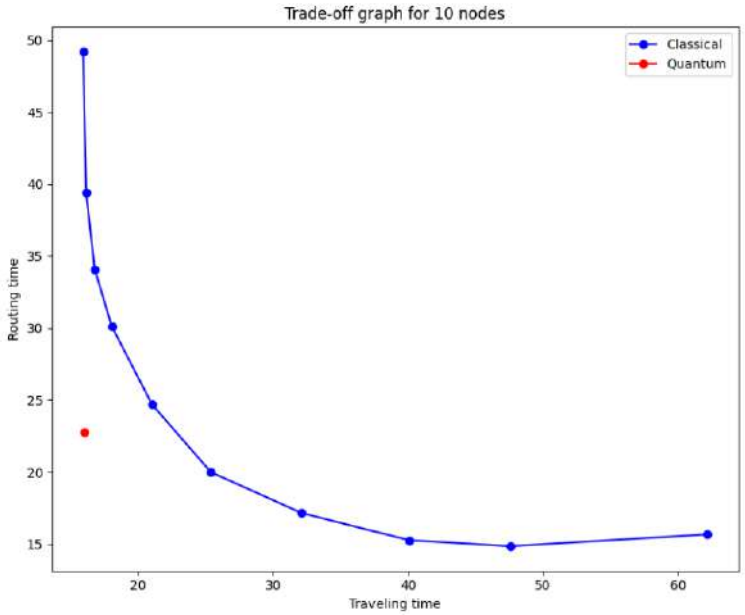


FIGURA 3.7: Barrera de trade-off rota por el protocolo cuántico. En rojo la estrategia cuántica, en azul diferentes estrategias clásicas mixtas con valores de p entre 0 y 0,9.

Otra forma de visualizar la ventaja del protocolo cuántico es midiendo el total_time = routing_time + traveling_time. En la Figura 3.8, es claro que cuando se mide el tiempo total, el rendimiento del protocolo cuántico supera el rendimiento del protocolo clásico a medida que aumenta el número de paquetes y la red se congestiona cada vez más.

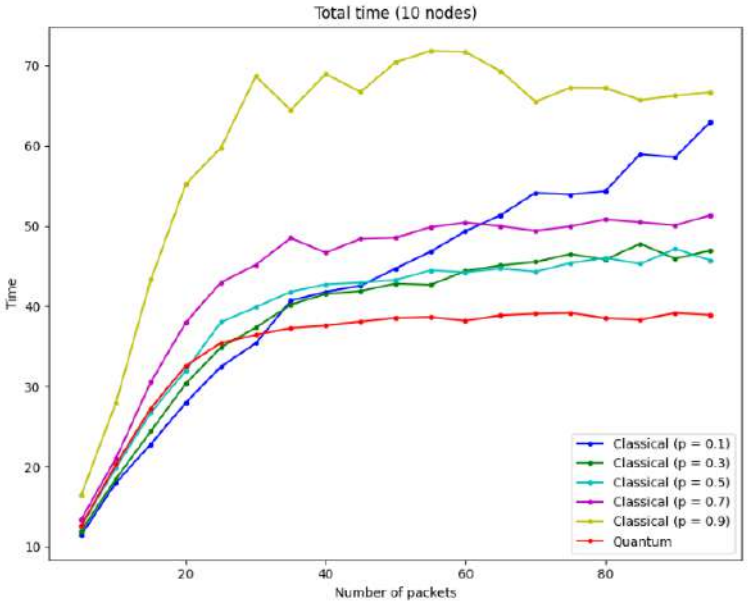


FIGURA 3.8: Tiempo total = tiempo de enrutamiento + tiempo de viaje, es evidente que el tiempo total mínimo corresponde al caso cuántico cuando aumenta el número de paquetes.

Esta ventaja puede entenderse considerando el estado cuántico del circuito para dos agentes (Figura 3.5) justo antes de medir, que es: $|\psi_{out}\rangle = J^{\dagger}(R_z(0) \otimes R_z(0))(R_y(\frac{\pi}{4}) \otimes R_y(\frac{\pi}{4}))(R_x(\frac{\pi}{2}) \otimes R_x(\frac{\pi}{2}))J|00\rangle = -j\frac{|01\rangle+|10\rangle}{\sqrt{2}}$.

Es decir, el estado $|01\rangle$ se medirá con una probabilidad del 50 % y el estado $|10\rangle$ con una probabilidad del 50 %. Esta estrategia resulta ser óptima en Pareto ya que ningún agente puede mejorar su rendimiento sin empeorar el de alguien más. Este estado $|\psi_{out}\rangle = -j\frac{|01\rangle+|10\rangle}{\sqrt{2}}$ significa que uno de los dos agentes siempre tomará el canal y el otro no. Al evitar el estado $|11\rangle$ estamos evitando el caso en que los dos agentes toman el canal, por lo tanto, evitando la congestión.

Aprovechando el entrelazamiento generado al principio del circuito en el juego cuántico, se obtiene un comportamiento inalcanzable en el caso clásico: tener un tiempo de enrutamiento medio con un tiempo de viaje mínimo. El mejor caso clásico, como se ve en la figura 3.8, sucede para cuando $p = 0{,}5$ en ambos jugadores, es decir, las combinaciones 00, 01, 10 y 11 suceden con una misma probabilidad de 25 %. En este caso, el 50 % de las veces un solo agente tomaría el canal (01 o 10), el 25 % ninguno tomaría el canal (00) y el otro 25 % ambos tomarían el canal (11) generando un canal congestionado.

Finalmente, $S_1 = (\varphi_X, \varphi_Y, \varphi_Z) = (\frac{\pi}{2}, \frac{\pi}{4}, 0)$ es solo una de las posibles estrategias puras. En la Figura 3.9, se muestra el rendimiento de diferentes estrategias cuánticas. El protocolo cuántico puede mejorar (puntos bajo la curva de trade-off clásica) o empeorar (puntos por encima de la curva de trade-off clásica) el rendimiento de la red dependiendo de las diferentes estrategias cuánticas seleccionadas por los agentes si cambian los valores de φ_X, φ_Y y φ_Z.

Conclusión

La sección presentada ofrece una visión innovadora y detallada sobre la aplicación de la teoría de juegos cuántica en la gestión de redes de datos para abordar el

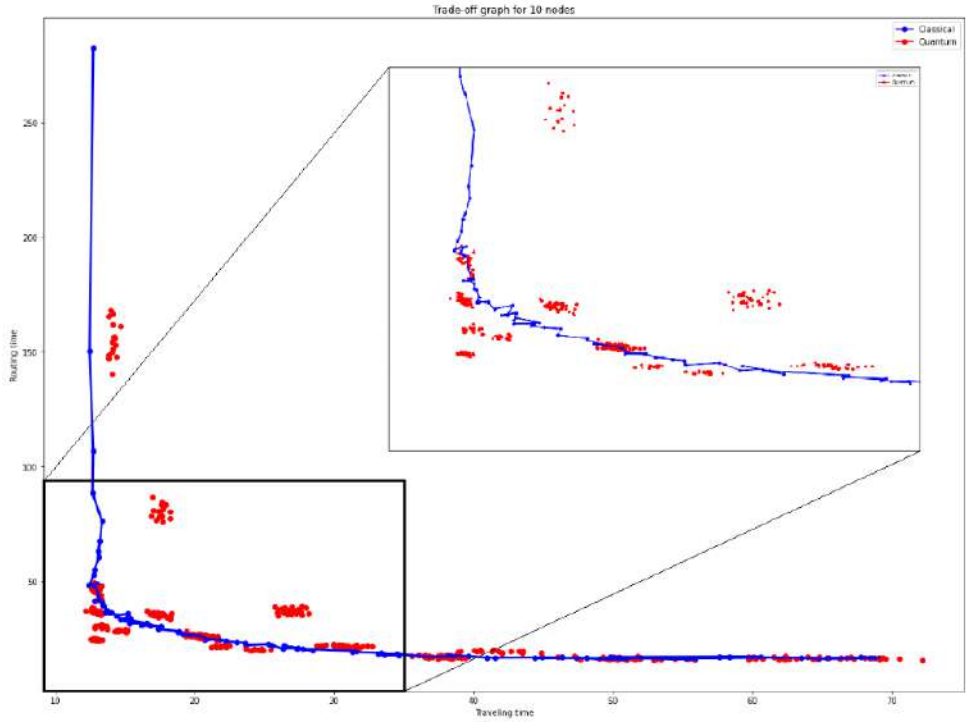


FIGURA 3.9: Barrera de compensación rota por el protocolo cuántico. En rojo diferentes estrategias cuánticas puras, en azul diferentes estrategias clásicas mixtas con valores de p entre 0 y 0.99.

persistente problema de la congestión. A través de un enfoque cuántico, este estudio propone una solución creativa y prometedora, superando los enfoques clásicos y demostrando mejoras significativas en el rendimiento de la red. La utilización de estrategias cuánticas, representadas por un modelo de compuertas cuánticas, permite una reducción efectiva en el tiempo total de transmisión, tiempo de viaje + tiempo de enrutamiento, evidenciando una notable optimización en la gestión de congestiones de red. Los resultados obtenidos subrayan un valor óptimo de Pareto, proponiendo un enfoque que es simultáneamente beneficioso para el individuo y la red en su conjunto.

En resumen, los resultados obtenidos en este estudio subrayan la capacidad de la teoría cuántica de juegos para superar las limitaciones de los enfoques clásicos en la gestión de redes de datos. La eficiente resolución del dilema de la congestión, evidenciada en las simulaciones y el análisis gráfico, sugiere un avance significativo hacia la optimización de redes en un mundo cada vez más digitalizado y conectado. Este enfoque cuántico no solo ofrece una solución al problema de la congestión en las redes de datos, sino que también abre caminos hacia nuevas investigaciones y aplicaciones de la teoría cuántica en la solución de problemas complejos en la informática y las telecomunicaciones.

3.2.2. Mitigación de la congestión de enrutamiento en redes de datos: un enfoque de teoría de juegos cuántica

Introducción

En esta sección abordaremos la expansión del modelo de Teoría de Juegos Cuántica aplicada a la mitigación de la congestión en redes de datos como se publicó en

[4]. Este trabajo completa el trabajo anterior en distintos aspectos. En primer lugar, el modelo actual permite que los agentes utilicen estrategias cuánticas mixtas y da un cierre al análisis de los equilibrios de Nash y óptimos de Pareto. En segundo lugar, introduce y examina la influencia del entrelazamiento cuántico y el ruido cuántico en el rendimiento de los protocolos de enrutamiento.

El análisis se profundiza al considerar la realidad de los dispositivos cuánticos actuales, caracterizados por su naturaleza NISQ (Quantum Noisy Intermediate-Scale Quantum), donde la coherencia cuántica puede verse comprometida por el entorno, llevando a una pérdida del estado cuántico debido a la decoherencia. A pesar de estas limitaciones, se demuestra que el protocolo cuántico todavía mantiene beneficios en presencia de una cantidad no muy alta tanto de ruido simulado y en dispositivos cuánticos reales como los ofrecidos por IBM. Este enfoque práctico y realista subraya la viabilidad de la teoría de juegos cuántica como una solución robusta frente a los desafíos de la congestión en las redes de datos, marcando un camino prometedor para futuras investigaciones y aplicaciones en sistemas de comunicación complejos.

Estrategias cuánticas mixtas

Es interesante saber si la estrategia óptima de Pareto $S_1 = (\frac{\pi}{2}, \frac{\pi}{4}, 0)$ que se propuso en el trabajo anterior es también un equilibrio de Nash. Para que una estrategia $S_x = (\varphi_X, \varphi_Y, \varphi_Z)$ sea un equilibrio de Nash, ningún agente tendrá incentivos para modificar individualmente su estrategia. Resulta que existe una estrategia $S_2 = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{2})$ que es beneficiosa para un agente individual asumiendo que el resto aplica S_1, por lo tanto S_1 no es un equilibrio de Nash. También resulta que S_2 es óptimo de Pareto y no es un equilibrio de Nash, ya que existe una $S_3 = (\frac{\pi}{2}, \frac{\pi}{4}, \pi)$ que es beneficiosa para un agente individual asumiendo que el resto aplica S_2. Nuevamente, S_3 es óptimo de Pareto y no un equilibrio de Nash, ya que existe otra $S_4 = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{3\pi}{2})$ que es beneficiosa para un agente individual asumiendo que el resto aplica S_3. Finalmente, S_4 es óptimo de Pareto y no un equilibrio de Nash ya que un agente individual puede aplicar S_1 para beneficiarse a sí mismo.

En resumen, S_1, S_2, S_3 y S_4 son todas estrategias puras. Todas ellas son óptimas de Pareto en nuestro problema, ya que ningún agente puede mejorar su rendimiento sin empeorar el de alguien más y si todos los agentes aplican la misma estrategia se puede obtener el máximo rendimiento y evitar la congestión de la red. Sin embargo, ninguna de ellas es un equilibrio de Nash ya que siempre hay una estrategia dominante que aumentaría el beneficio de un agente individual, incentivando así a los agentes a modificar individualmente su estrategia.

Finalmente, podemos construir una estrategia mixta S_5 que aplica las cuatro estrategias S_1, S_2, S_3 y S_4 con igual probabilidad $w = 0{,}25$. S_5 resulta ser óptima de Pareto, ya que nadie puede mejorar su propio rendimiento sin empeorar el de alguien más y también tiene potencial de ser un equilibrio de Nash, ya que no pudimos encontrar ninguna estrategia S_X que un agente pueda aplicar individualmente para aumentar su propio rendimiento asumiendo que los otros agentes están aplicando S_5.

$$S_5 = \begin{cases} S_1 = (\frac{\pi}{2}, \frac{\pi}{4}, 0) & \text{with } w = 0{,}25 \\ S_2 = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{2}) & \text{with } w = 0{,}25 \\ S_3 = (\frac{\pi}{2}, \frac{\pi}{4}, \pi) & \text{with } w = 0{,}25 \\ S_4 = (\frac{\pi}{2}, \frac{\pi}{4}, \frac{3\pi}{2}) & \text{with } w = 0{,}25 \end{cases} \tag{3.5}$$

La prueba rigurosa de que S_5 es un equilibrio de Nash se dejará para trabajos futuros y su dificultad reside en el hecho de que el espacio de estrategias es infinito. φ_X, φ_Y y φ_Z ya son variables continuas, y para probar que S_5 es un equilibrio de Nash, es necesario demostrar que no existe ninguna estrategia mixta en todo el espacio (tres funciones de densidad de probabilidad, sobre φ_X, φ_Y y φ_Z) que tenga el incentivo de desviarse.

Observando el resultado $|\psi_{out}\rangle$ del sistema cuando todos los agentes aplican la estrategia mixta S_5, vemos en la Ecuación (3.6) cómo evitamos la congestión. Los estados $|00\rangle$ y $|11\rangle$ están ausentes, por lo tanto, se evita la congestión.

$$|\psi_{out}\rangle = \begin{cases} \frac{|01\rangle+|10\rangle}{\sqrt{2}} & \text{with } 50\,\% \\ |10\rangle & \text{with } 25\,\% \\ |01\rangle & \text{with } 25\,\% \end{cases} \tag{3.6}$$

Influencia del ruido

Hasta el momento, hemos estado estudiando el sistema sin considerar el ruido y la decoherencia de los estados cuánticos. En las siguientes subsecciones, se analiza el protocolo cuántico bajo condiciones no ideales, modelando un canal ruidoso y también utilizando los dispositivos cuánticos disponibles de IBM.

Simulación de la decoherencia

Para añadir ruido cuántico al modelo, utilizamos el canal de depolarización cuántica [121]. Este canal depende de un parámetro C que mapea desde el estado ideal $\rho_{out} = |\psi_{out}\rangle\langle\psi_{out}|$ al estado mixto con máxima aleatoriedad ($\frac{I}{d}$) siguiendo la ecuación: $\Delta_C(\rho_{out}) = C\rho_{out} + (1-C)\frac{I}{d}$, con I la matriz identidad y d la dimensión del estado cuántico.

La Figura 3.10 se obtiene trazando el rendimiento del sistema para un caso óptimo de Pareto y diferentes valores de C (desde $C = -\frac{1}{3}$ hasta $C = 1$, respetando la condición de positividad completa). El rendimiento se asemeja cada vez más al caso clásico a medida que el estado cuántico original tiende al estado mixto máximo. Finalmente, como se esperaba, el estado mixto máximo ($C = 0$) y la estrategia mixta clásica con $p = 0{,}5$ tienen el mismo rendimiento.

Ruido de dispositivos reales

Para probar nuestra propuesta, se utilizaron las computadoras cuánticas de IBM [122] para simular los juegos cuánticos. Las computadoras cuánticas de IBM se denominan dispositivos NISQ (Noisy Intermediate-Scale Quantum) [5], lo que significa que los procesadores cuánticos son muy sensibles al entorno y pueden perder su estado cuántico debido a la decoherencia cuántica. En la era NISQ, los procesadores cuánticos no son lo suficientemente sofisticados para implementar continuamente la corrección de errores cuántica, por eso es importante probar nuestro algoritmo en este tipo de dispositivos cuánticos. Los resultados pueden observarse en la Figura 3.11.

Es importante destacar que el rendimiento del protocolo es todavía relativamente alto. Esto se debe a que los juegos se implementaron jugando en lotes de 2 agentes, es decir, circuitos cuánticos de 2 qubits. Aunque los computadores cuánticos disponibles todavía son muy ruidosos para sistemas cuánticos grandes, se desempeñan

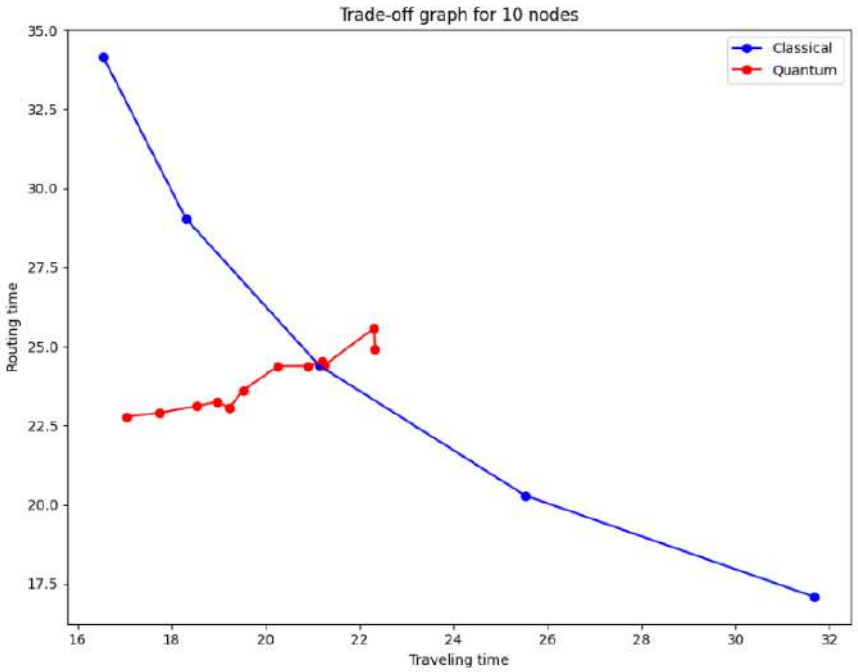


FIGURA 3.10: Efecto de la decoherencia en la barrera de compensación por protocolo cuántico. A medida que el valor de C se aleja de C = 1 (caso ideal), el caso cuántico se parece cada vez más al caso clásico. Valores de p entre 0,3 y 0,7.

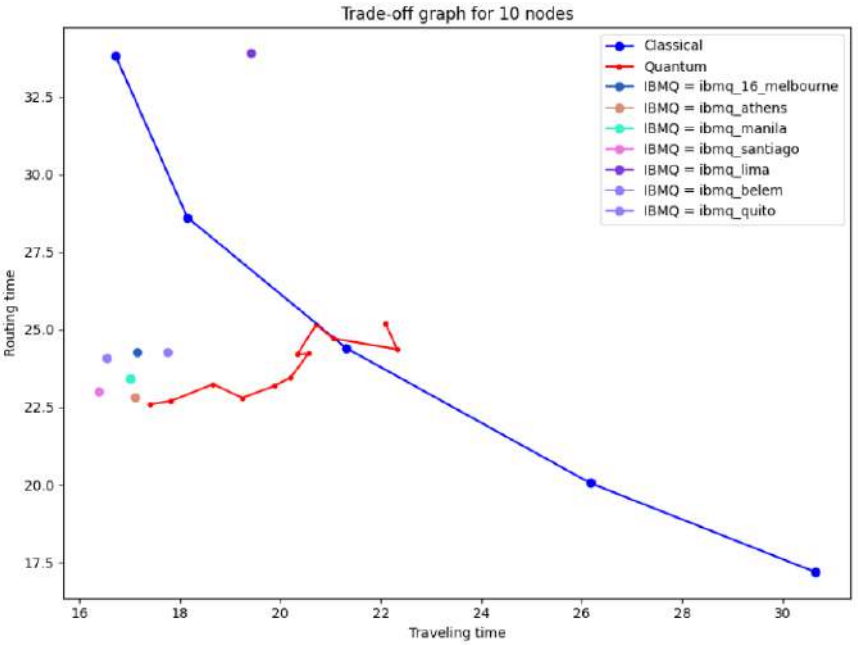


FIGURA 3.11: Efecto de dispositivos reales en la barrera de compensación por protocolo cuántico. La congestión se puede mitigar mediante el uso de computadoras cuánticas IBM NISQ. Valores de p entre 0,3 y 0,7.

bien cuando se trata de pocos qubits y circuitos de puertas cuánticas poco profundas.

Conclusión

En esta sección, hemos explorado la aplicación de la Teoría de Juegos Cuántica en la mitigación de la congestión de redes de datos, un paso significativo en el desarrollo de estrategias de enrutamiento más eficientes. Hemos visto cómo el modelo cuántico, al permitir estrategias mixtas y puras tanto para agentes clásicos como cuánticos, supera las limitaciones inherentes a los modelos clásicos, especialmente en la compensación entre el tiempo de enrutamiento y el tiempo de viaje. A pesar de la ausencia de un equilibrio de Nash claro en las estrategias puras, la estrategia mixta S_5 emerge como una solución potencialmente óptima de Pareto y candidata a equilibrio de Nash, destacando la complejidad y la riqueza del paisaje estratégico en la teoría cuántica de juegos. Además, el análisis bajo condiciones no ideales, incluyendo el impacto del ruido y la decoherencia cuántica, proporciona una perspectiva realista de la aplicación de estos modelos en el contexto de los dispositivos cuánticos actuales.

La implementación de estos modelos en computadoras cuánticas IBM NISQ ha demostrado la viabilidad práctica de nuestras estrategias cuánticas, incluso frente a desafíos como la decoherencia y el ruido ambiental. Los resultados obtenidos, que muestran un rendimiento significativamente alto en escenarios de pocos qubits y circuitos de puertas cuánticas poco profundas, son prometedores para el futuro de las comunicaciones y el enrutamiento de redes. En resumen, este trabajo representa un paso significativo hacia la comprensión y la aplicación práctica de la teoría de juegos cuántica en la mitigación de la congestión en redes de datos.

3.3. Protocolo basado en aprendizaje para enrutamiento en redes cuánticas

Introducción

En esta sección, continuamos sobre la construcción de un protocolo para el enrutamiendo de redes de comunicaciones cuánticas incorporando técnicas de aprendizaje por refuerzo para adaptar las estrategias que los agentes eligen en función del estado de la red (cantidad de paquetes, congestión de la red, etc), como se realizó en [6]. Este avance, crucial en un mundo globalizado y con crecientes demandas de seguridad en telecomunicaciones, anticipa una futura interconexión global a través de canales cuánticos. La propuesta, que extiende investigaciones previas en enrutamiento al incorporar técnicas de inteligencia artificial, se centra en mitigar la congestión en telecomunicaciones modernas. Mediante un protocolo que combina teoría de juegos y aprendizaje por refuerzo, se permite que los paquetes adapten estrategias cuánticas en respuesta a la dinámica de la red. Este enfoque busca no solo mejorar el rendimiento frente a protocolos clásicos, sino también adaptarse a variaciones en la red, empleando juegos no cooperativos y aprendizaje para un sistema autoajustable y eficiente en distintos escenarios.

En el presente trabajo, se expande el protocolo basado en la Teoría de Juegos Cuántica introducido en los trabajos anteriores aportando dos contribuciones principales en comparación con nuestro estudio anterior. Primero, se definen las condiciones exactas bajo las cuales el protocolo cuántico supera a su equivalente clásico, analizando: a) el número mínimo de paquetes en la red en función de su distancia máxima y b) la mínima distancia máxima entre dos nodos en función de la cantidad de nodos en la red. Segundo, dada la creciente relevancia de la intersección entre la teoría del aprendizaje, la teoría de juegos y el diseño de mecanismos, se incorpora

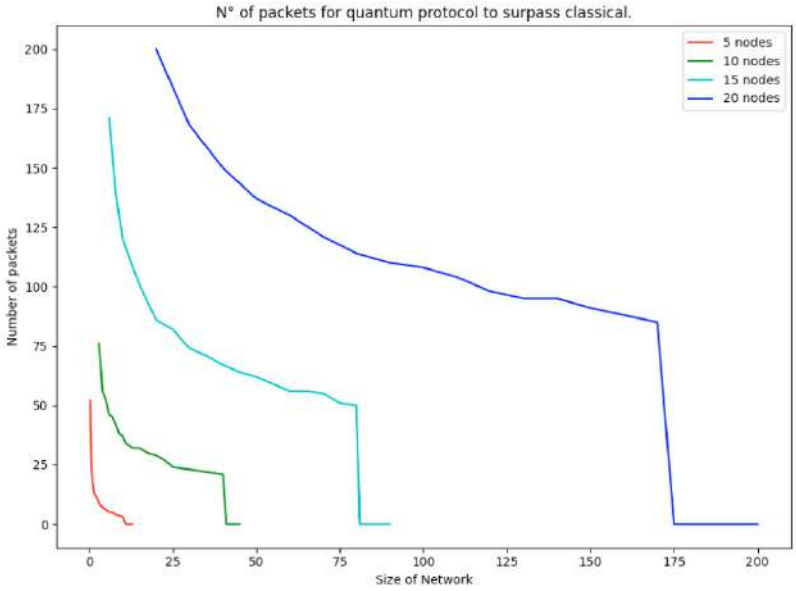


FIGURA 3.12: Número mínimo de paquetes para que el rendimiento del protocolo cuántico supere al clásico en función del tamaño de la red.

la capacidad de autoadaptación en nuestro protocolo. El rendimiento del sistema se estudia comparando diferentes algoritmos de aprendizaje en diversos escenarios.

Modelo clásico versus cuántico

En sección 3.2.1 se demostró que el protocolo cuántico para una estrategia cuántica particular supera a cualquier estrategia clásica si se alcanza un número mínimo de paquetes que garantice la congestión. Este fenómeno podía observarse en la Fig. 3.8 donde se representan el caso cuántico (en rojo) y los casos clásicos (con diferentes colores para distintas probabilidades p de buscar otro camino que no sea el más corto). En esa figura, es evidente que el caso cuántico tiene un tiempo total menor que el resto de los casos clásicos, pero solo después de un cierto punto. Ese punto corresponde con el número mínimo de paquetes que una red debe tener para congestionar la red y explotar las propiedades del protocolo cuántico.

Tamaño de la red: El número mínimo de paquetes que garantiza la ventaja del protocolo cuántico sobre el protocolo clásico depende de la distancia máxima entre dos nodos conectados (tamaño del canal más largo en la red) y del número de nodos en la red.

En la Fig. 3.12, ilustramos el resultado de calcular el mencionado número de paquetes en función de la distancia máxima entre nodos correspondiente a cuatro tipos de redes dependiendo de su número de nodos (5, 10, 15 y 20). En todos los casos, es posible visualizar que el número de paquetes necesarios disminuye con la distancia máxima entre dos nodos conectados.

Esto se debe a que a medida que aumenta el tamaño de la red, la influencia de la congestión se hace mayor, siempre y cuando la cantidad de nodos en la red sea fija (5, 10, 15 o 20): un canal congestionado largo afecta mucho más al rendimiento del sistema que uno corto. Esta es la razón por la cual el número mínimo de paquetes necesarios para congestionar una red disminuye a medida que aumenta la distancia del canal más largo en una red.

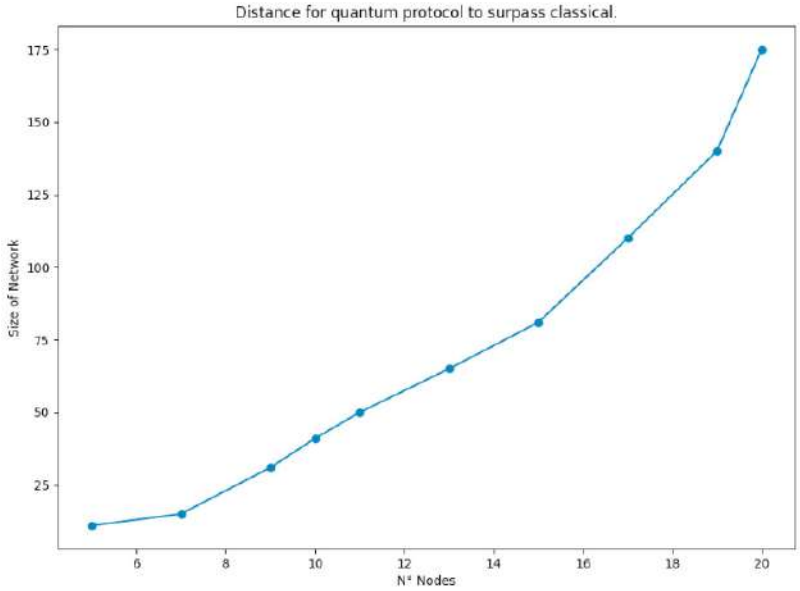


FIGURA 3.13: Distancia mínima para que el rendimiento del protocolo cuántico supere al clásico en función del número de nodos.

Cantidad de nodos: En la Fig. 3.12 también es posible observar otro fenómeno. Existe un punto más allá del cual las distancias máximas en la red son tan largas que la congestión allí siempre influye significativamente en la performance del sistema. Por lo tanto, el número mínimo de paquetes necesario para que el protocolo cuántico supere al clásico es 0 (parte plana en la Fig. 3.12).

Ese punto depende exclusivamente del número de nodos en la red y es posible obtener una idea de esa relación en la Fig. 3.13.

Precisamente, estos puntos representan la mínima distancia máxima que conecte dos nodos presente en una red, en función del número de nodos, que asegura que el rendimiento del protocolo cuántico supera a su equivalente clásico para cualquier número de paquetes.

Estrategias basadas en aprendizaje por refuerzo

Los algoritmos basados en aprendizaje automático aprovechan los datos para mejorar el rendimiento. En nuestro caso, los datos son recompensas obtenidas de la experiencia, por eso aprovechamos las herramientas del aprendizaje por refuerzo para encontrar las mejores estrategias. El aprendizaje por refuerzo se ocupa específicamente de cómo los agentes deben seleccionar estrategias en un entorno para maximizar una señal de recompensa numérica [35].

En las secciones previas, se propusieron estrategias basadas en el conocimiento del problema. Sin embargo, en este caso, el sistema deducirá las estrategias basándose únicamente en la experiencia (feedback). Es importante aclarar que el sistema se tratará como centralizado, por lo tanto, la señal de recompensa a minimizar será la suma de los tiempos totales de toda la red y las estrategias serán las mismas para todos los agentes.

Se emplea aprendizaje por refuerzo para abordar este problema, aunque cabe destacar que también podrían aplicarse otras técnicas. Este trabajo representa un paso inicial hacia la consideración del problema como un sistema multiagente, donde

cada jugador busque minimizar su propio tiempo total aprendiendo su propia estrategia.

El primer obstáculo al intentar aprender la estrategia óptima es el hecho de que el espacio que necesita ser explorado es infinito. Todas las estrategias posibles son todas las posibles matrices unitarias de 2x2 $U(\theta, \phi, \lambda)$ [ec. 3.1], es decir, explorar sobre 3 variables de valor real independientes: θ, ϕ y λ. Una de las formas de abordar este problema es utilizar Tile Coding (Codificación por Mosaicos). La idea principal de Tile Coding es agrupar el espacio en particiones (*mosaicos*) y representar cada estrategia del espacio continuo a partir de una combinacion finita de mosaicos, un ejemplo de cómo pasar de un espacio continuo de 2 dimensiones a un espacio discreto de 4 mosaicos se puede ver en la Fig. 3.14. En nuestro caso, se utilizaron 512 mosaicos para representar el espacio continuo de estrategias en 3 dimensiones (θ, ϕ, λ).

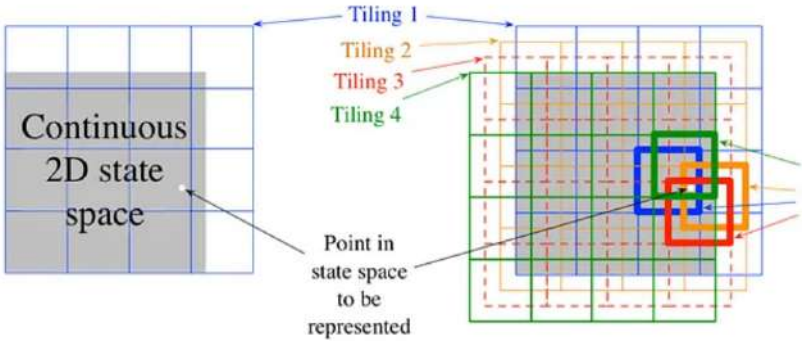


FIGURA 3.14: Tile coding aplicado a un espacio continuo de 2 dimensiones.

El valor $Q(A)$ es una estimación de lo beneficioso que es tomar cualquier acción/estrategia y se actualizará basándose en la experiencia. Esto se hará utilizando diferentes reglas de actualización y estudiando cuál es la más apropiada.

1. $Q_{t+1}(A) = Q_t(A) + \alpha[R_t - Q_t(A)]$
2. $\begin{cases} Q_{t+1}(A) = Q_t(A) + \alpha(R_t - \bar{R}_t)(1 - \pi_t(A)) & \text{y} \\ Q_{t+1}(a) = Q_t(a) - \alpha(R_t - \bar{R}_t)\pi_t(a) & \text{para todo } a \neq A \end{cases}$

   where:

   - α denota la tasa de aprendizaje.
   - R_t denotes la recompensa inmediata en el tiempo t.
   - $\bar{R}_t$ denota el valor medio de R_t hasta el tiempo t.
   - $\pi_t(a) = \frac{e^{\frac{Q_t(a)}{\tau}}}{\sum_{b=1}^{k} e^{\frac{Q_t(a)}{\tau}}}$ denota la distribución de probabilidad de Boltzmann para seleccionar la acción a y τ un parámetro de temperatura.

Por un lado, el caso *1)*, se basa en el aprendizaje por diferencia temporal, y el caso *2)*, por otro lado, se basa en la idea del ascenso de gradiente estocástico [35]. Se consideraron dos tipos diferentes de tasas de aprendizaje: $\alpha = const$ y $\alpha = \frac{1}{N(A)}$, donde $N(A)$ es el número de veces que la estrategia A ha sido seleccionada hasta ahora.

Una segunda consideración a tener en cuenta es la regla de selección de estrategia a utilizar. En cada momento, se debe seleccionar una estrategia basada en la función

de valor actual estimado $Q(A)$. Si decidimos siempre usar la estrategia con el valor estimado más alto, nunca exploraremos suficiente como para asegurarnos de que la estrategia con el valor estimado más alto corresponda a la mejor estrategia real. Esto se llama selección de estrategia codiciosa y puede ocurrir en el caso *1* (en el caso *2* las estrategias se seleccionan muestreando la función de Boltzmann $\pi_t(a)$). Para resolver este problema, una alternativa simple es comportarse de manera codiciosa la mayoría del tiempo, pero de vez en cuando, con una probabilidad ϵ, el agente selecciona una estrategia aleatoriamente.

La cuestión de equilibrar el egoísmo con la aleatorización es un problema abierto y se conoce como el dilema de exploración-explotación. Para nuestro caso, se consideran dos enfoques diferentes: $\epsilon = const$ y $\epsilon = \epsilon_0 * \gamma^t$, γ siendo una constante de diseño y t el número de iteración.

En la Fig. 3.15, es posible observar el tiempo total inmediato (donde $R_t = -t_{tot}$, cuanto más corto sea el tiempo, mayor será la recompensa) y su valor medio ($-\bar{R}_t$) para 6 variaciones de algoritmos de aprendizaje (cada paso siendo diferentes iteraciones de selecciones de estrategias aleatorizadas del juego para una red fija y promediando los valores en 10 ejecuciones diferentes) donde sus hiperparámetros fueron ajustados adecuadamente:

- $\epsilon = const$ y $\alpha = const$
- $\epsilon = const$ y $\alpha = \frac{1}{N(A)}$
- $\epsilon = \epsilon_0 * \gamma^t$ y $\alpha = const$
- $\epsilon = \epsilon_0 * \gamma^t$ y $\alpha = \frac{1}{N(A)}$
- Ascenso por gradiente y $\alpha = const$
- Ascenso por gradiente y $\alpha = \frac{1}{N(A)}$

Previo al análisis, es importante aclarar un aspecto: el entorno es estacionario. La información de los paquetes (origen y destino) y la configuración de la red varían en cada iteración, sin embargo, la cantidad de paquetes de paquetes y nodos en la red se mantiene constante. Asimismo, la distancia máxima entre dos nodos conectados permanece invariable.

Después de aclarar esto, se pueden extraer algunas conclusiones de estas simulaciones. Primero, todos los algoritmos de aprendizaje convergen hacia una solución eficiente y, aunque algunos tardaron más que otros, la mayoría de ellos convergen hacia la misma solución. En segundo lugar, dado que el sistema es estacionario, los algoritmos que convergieron más rápido a la solución son los que tienen $\alpha = \frac{1}{N(A)}$. Esto tiene sentido debido a que un $\alpha = \frac{1}{N(A)}$ da igual peso a todas las recompensas desde el principio, lo cual es bueno siempre que las características principales del sistema se mantengan relativamente constantes (estacionario).

Finalmente, y lo más importante, ¡la estrategia aprendida por los algoritmos que superan a todos los demás es la misma que la propuesta en nuestro trabajo anterior! La estrategia es $S_{best} = (\varphi_X, \varphi_Y, \varphi_Z) = (\frac{\pi}{2}, \frac{\pi}{4}, 0)$ (o alguna variación que mantenga esa relación entre ángulos, ya que lo que importa es la relación entre los ángulos y no su valor absoluto) que corresponde con el estado $|\psi_{out}\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}$ al final del circuito. Este estado de salida garantiza minimizar la congestión (como se explicó en las secciones 3.2.1 y 3.2.2), pero esta vez el resultado provino de la evolución del protocolo en sí y no fue impuesto externamente como en las secciones anteriores.

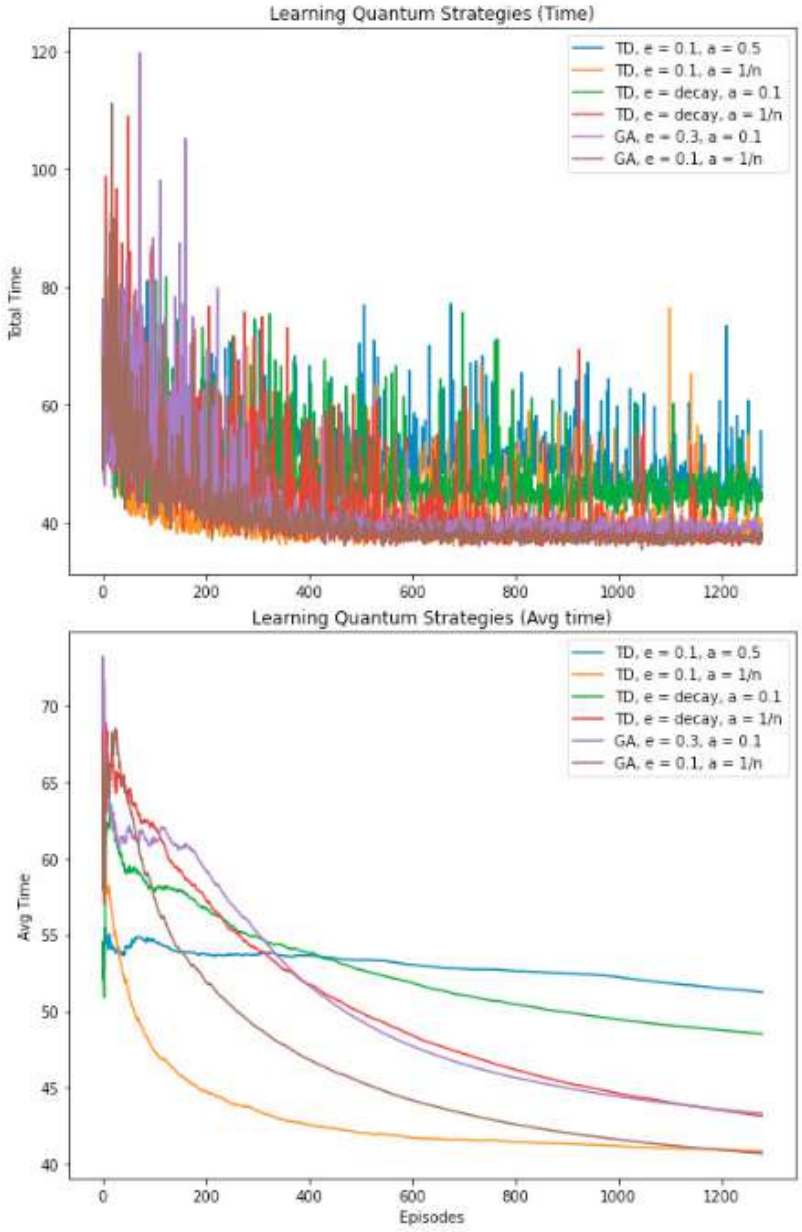


FIGURA 3.15: Tiempos totales de red mientras aprendes acumulando experiencia. Tiempo de estrategia aprendido arriba y su valor medio abajo. TD = diferencia temporal. GA = ascenso de gradiente (e = tau1).

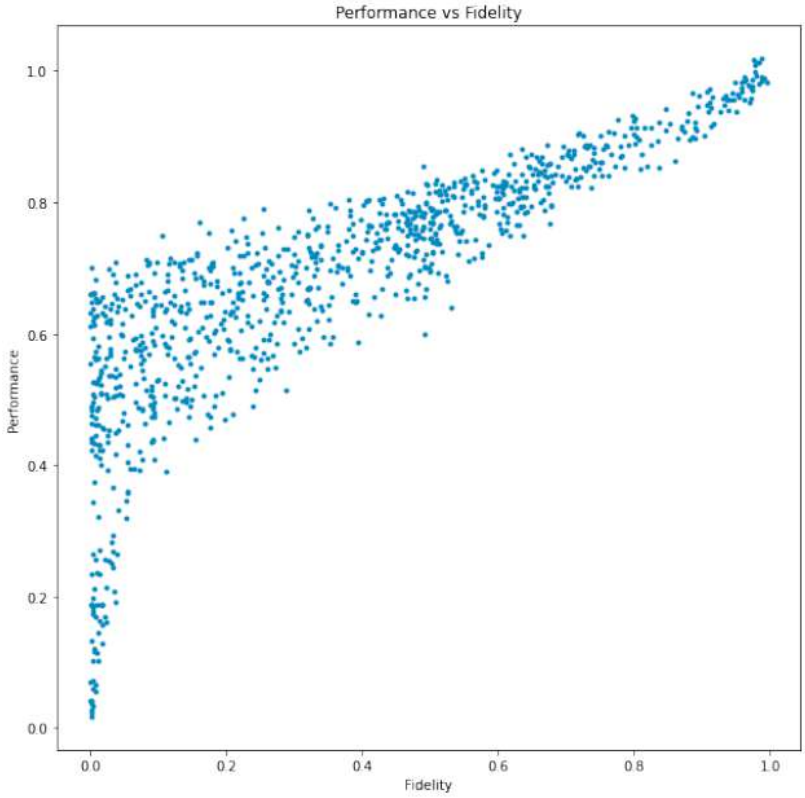


FIGURA 3.16: Correlación entre fidelidad con el estado $|\psi_A\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$ y el rendimiento del protocolo.

Eficiencia

Este resultado indica que podría haber una correlación entre el estado $|\psi_A\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$ y el rendimiento del sistema cuando la red está congestionada. Para visualizar esta idea, el rendimiento se midió para mil estrategias diferentes, siendo el rendimiento el tiempo ideal precalculado mínimo que el protocolo podría alcanzar dividido por el tiempo medido para cada estrategia, Rendimiento $= \frac{\text{Tiempo total ideal}}{\text{Tiempo total medido}}$. Al mismo tiempo, se utilizó el estado de salida $|\psi_B\rangle$ generado por cada una de estas estrategias para calcular la similitud o fidelidad entre los estados $|\psi_A\rangle$ y $|\psi_B\rangle$, $F(\psi_A, \psi_B) = |\langle\psi_A|\psi_B\rangle|^2$. El resultado de trazar Rendimiento versus Fidelidad para cada estrategia se muestra en la Fig. 3.16.

Efectivamente, en la Fig. 3.16, es claro cómo la fidelidad y el rendimiento están positivamente correlacionados. Sin embargo, para cuantificar esta relación, se calcularon el coeficiente de covarianza, el coeficiente de Pearson y el coeficiente de Spearman.

- $Cov(F,P) = \frac{\sum_1^n (F_i-\bar{F})*(P_i-\bar{P})}{n} = 0{,}04423$
- $Pea(F,P) = \frac{Cov(F,P)}{\sigma(F)*\sigma(P)} = 0{,}85274$
- $Spe(F,P) = \frac{Cov(rank(F),rank(P))}{\sigma(rank(F))*\sigma(rank(P))} = 0{,}92393$

donde σ significa desviación estándar y $rank()$ es una función de transformación que ordena y reemplaza un valor por su rango entero.

En el primer caso, el coeficiente de covarianza es positivo, 0.04423, sugiriendo que las variables cambian en la misma dirección como esperamos. Un problema con la covarianza como herramienta estadística por sí sola es que es difícil de interpretar. Esto nos lleva al coeficiente de correlación de Pearson a continuación. El coeficiente de Pearson devuelve un valor entre -1 y 1 que representa los límites de correlación desde una correlación negativa completa hasta una correlación positiva completa; un valor de 0.85274 sugiere un alto nivel de correlación. Además, el coeficiente de correlación de Spearman se utiliza cuando las dos variables pueden estar relacionadas por una relación no lineal. Al igual que con el coeficiente de correlación de Pearson, los puntajes están entre -1 y 1 para variables perfectamente correlacionadas negativamente y perfectamente correlacionadas positivamente respectivamente [123]. Finalmente, el enfoque basado en rangos muestra una fuerte correlación entre las variables de 0.92393.

Adaptabilidad

Los escenarios de la vida real suelen ser no estacionarios. Para emular estos casos, creamos cambios repentinos en la red y observamos cómo responden ahora todos los algoritmos. Hay muchos cambios que podrían provocarse: cambios repentinos en el número de paquetes, el número de nodos o la distancia máxima entre dos nodos consecutivos. Lo importante es que la red cambie, por un momento, de las condiciones donde el protocolo cuántico es mejor que el clásico a otra donde sucede lo contrario y luego vuelva a las condiciones normales.

En la Fig. 3.17, se provocó un cambio en la distancia máxima entre dos nodos consecutivos (simulando la caída momentánea del canal más grande de la red).

Después de realizar un análisis similar al de la subsección anterior, es posible ver cómo los algoritmos que se desempeñaron mejor cuando el entorno era estacionario son, aproximadamente, los que peor se adaptan cuando el entorno es no estacionario. Los algoritmos de aprendizaje que priorizan la explotación (rojo y verde) son los que no se adaptan en absoluto después de cambios repentinos. Por otro lado, aquellos que priorizan la exploración (naranja y azul) son los que mejor pueden aprovechar la retroalimentación recibida cuando la red cambia.

Además, cuando la red vuelve a la normalidad, aquellos que se adaptaron al cambio anterior en la red tienen que readaptarse y olvidar todo lo aprendido durante ese período. Por otro lado, aquellos que no lograron adaptarse, no tienen que readaptarse ya que no cambiaron durante la ruptura en la red.

La selección del algoritmo a utilizar en una red no estacionaria dependerá de la frecuencia con la que ocurran los cambios y cuánto duren. En una red con interrupciones cortas, la adaptación no es crucial ya que centrarse en aprender las características de la red en condiciones normales será mejor. Si se presta demasiada atención a ajustarse a pequeñas variaciones en la red, dicha sensibilidad podría impedir la obtención de la estrategia óptima para un funcionamiento eficiente de la red en condiciones normales. Con base en estas premisas, se dará preferencia a los algoritmos que prioricen la explotación (TD con $\epsilon = \epsilon_0 * \gamma^t$ o GA con bajo τ).

De lo contrario, si la red es inestable y puede cambiar recurrentemente o permanecer lo suficientemente largo en esas pausas como para afectar el rendimiento general de la red, los algoritmos que mejor se adaptan y reducen los problemas de fluctuaciones son los que nunca dejan de explorar (TD con $\epsilon = const$ o GA con alto τ).

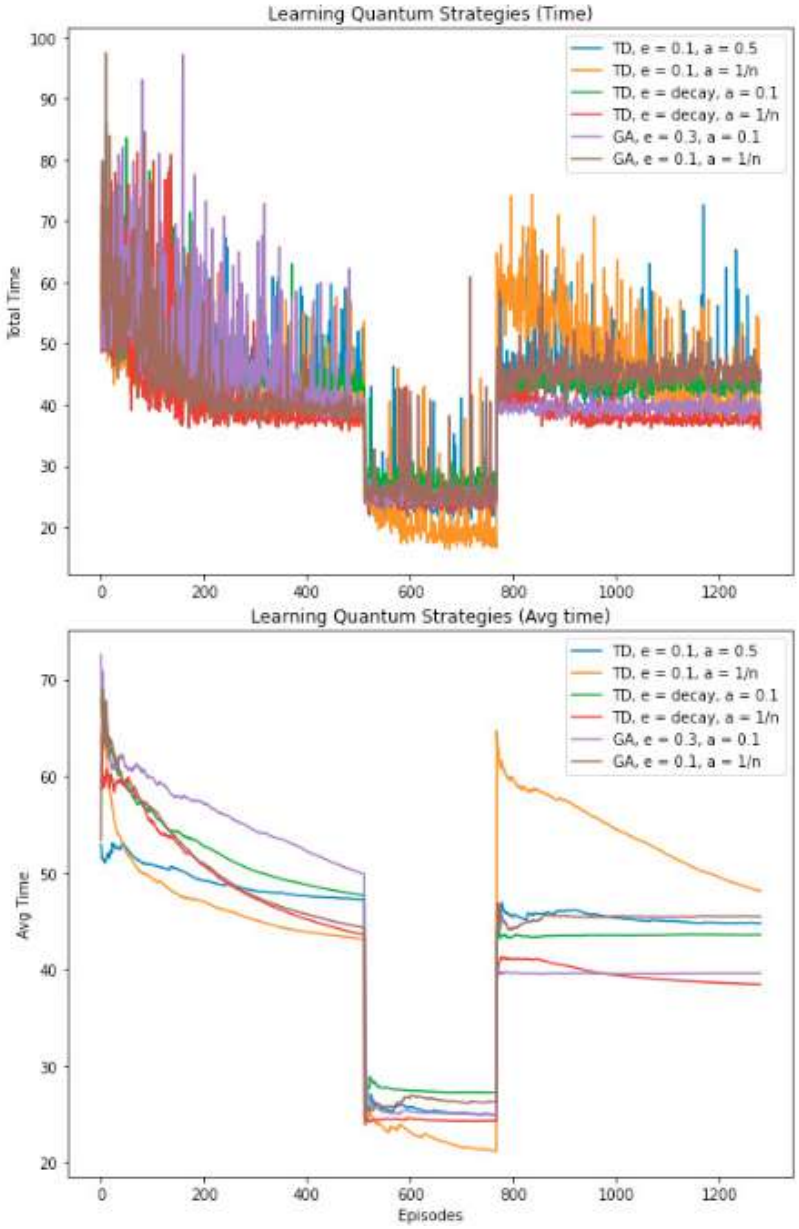


FIGURA 3.17: Adaptabilidad de los diferentes algoritmos de aprendizaje a cambios bruscos en la red.

Conclusión

En esta sección, hemos explorado la intersección de la teoría de juegos cuántica y las técnicas de aprendizaje por refuerzo aplicadas al enrutamiento en redes de comunicaciones cuánticas. Este enfoque innovador no solo promete mejorar el rendimiento frente a los protocolos de enrutamiento clásicos, sino también adaptarse a los cambios dinámicos en la red. A través de un modelo de aprendizaje centralizado, para optimizar el tiempo total en la red, hemos investigado cómo los paquetes pueden adaptar estrategias cuánticas en respuesta a la congestión y otras variaciones en la red, abordando así los desafíos actuales de eficiencia en telecomunicaciones.

Nuestro estudio profundizó sobre las condiciones específicas bajo las cuales el protocolo cuántico supera al clásico, considerando el número mínimo de paquetes y la distancia del canal más largo de la red. Adicionalmente, se integró la capacidad de autoadaptación mediante algoritmos de aprendizaje por refuerzo, permitiendo al sistema responder a variaciones estacionarias y no estacionarias. La adaptabilidad de los algoritmos resultó ser clave, ya que aquellos que priorizan la explotación mostraron limitaciones en entornos no estacionarios, mientras que los que favorecen la exploración se adaptaron mejor a los cambios. La elección del algoritmo de aprendizaje depende así de la naturaleza de la red, ya sea una red con interrupciones esporádicas o una propensa a cambios prolongados. Este trabajo sienta las bases para futuras investigaciones en el campo de la comunicación cuántica y abre el camino hacia sistemas de enrutamiento más avanzados y adaptables.

Capítulo 4

Algoritmos Cuánticos de Aprendizaje por Refuerzo Multi-Agente

4.1. Introducción

La confluencia de la teoría de juegos, la mecánica cuántica y el aprendizaje por refuerzo ha abierto un fascinante dominio de investigación, donde las dinámicas de toma de decisiones entre agentes racionales se manifiestan tanto en configuraciones clásicas como cuánticas. Desde la primera formulación de los juegos cuánticos a principios de este milenio, se ha desarrollado un marco robusto que describe cómo se comportan los agentes cuando los juegos son cuantizados. Estos juegos cuánticos, que expanden el espacio estratégico de los agentes, han demostrado tener aplicaciones en diversos campos, desde la economía hasta el diseño de redes de comunicación.

Una de las características más intrigantes de la teoría de juegos cuántica es la emergencia de equilibrios adicionales, que a menudo superan a sus contrapartes clásicas en términos de recompensas. Sin embargo, encontrar estos equilibrios en un espacio de estrategia ampliado, especialmente cuando se permite a los agentes utilizar estrategias mixtas, presenta desafíos significativos. En este contexto, se propuso un método innovador para aprender y visualizar estrategias cuánticas mixtas, permitiendo a los agentes adaptar sus estrategias sin información completa sobre la matriz de pagos del juego, las estrategias seleccionadas por sus oponentes, ni sus recompensas en cada paso.

A medida que el mundo avanza hacia la coexistencia de múltiples IA [124] que aprenden de su interacción, que puede ser colaborativa, estratégica o adversarial, el aprendizaje en juegos se vuelve cada vez más relevante [125]. Existen estrategias de aprendizaje establecidas, como el juego ficticio [126] y el aprendizaje sin arrepentimientos [127]. Sin embargo, el enfoque propuesto en estos trabajos se centra en un algoritmo descentralizado que permite a los agentes ajustar sus estrategias basándose únicamente en una señal de recompensa de retroalimentación individual.

Más allá de los juegos cuánticos en sí, el aprendizaje por refuerzo en un entorno cuántico presenta oportunidades y desafíos únicos. Se introdujo, adicionalmente, un algoritmo unificado de aprendizaje que se aplica sin problemas tanto a juegos clásicos como cuánticos que incorpora el concepto de justicia en la distribución de las recompensas entre los agentes a la hora de la toma decisiones. Este algoritmo, basado en Exploring Selfish Reinforcement Learning (ESRL) [128], se extendió para manejar juegos cuánticos de suma no nula con información imperfecta, destacando

la capacidad del algoritmo para promover la equidad dentro de las interacciones estratégicas.

La mecánica cuántica, con su inherente incertidumbre y entrelazamiento, introduce un reino que se desvía de los paradigmas deterministas de la mecánica clásica. El entrelazamiento, donde estados correlacionados emergen independientemente de su separación espaciales, presenta tanto un desafío como una oportunidad en el ámbito de la teoría de juegos cuántica. Esta capacidad de los juegos cuánticos para explorar cómo estos fenómenos impactan en las interacciones estratégicas ha sido fundamental para su desarrollo.

A medida que la teoría de juegos cuántica ha avanzado, también ha surgido la necesidad de algoritmos de aprendizaje más sofisticados que puedan navegar por el complejo paisaje de los espacios de estrategia cuántica. En este sentido, se propuso, finalmente, un algoritmo basado en la estimación del gradiente de la función de recompensa para juegos cuánticos en entornos multiagente. Este enfoque basado en políticas y descenso de gradientes proporciona flexibilidad al permitir que los agentes aprendan directamente acciones a partir de recompensas, asegurando mejoras continuas de la política y resultando en estrategias optimizadas localmente.

Sin embargo, uno de los hallazgos más sorprendentes de estos estudios fue el descubrimiento de una relación no trivial entre el ruido del circuito cuántico y el rendimiento del algoritmo. En contradicción con la concepción generalizada de que el ruido cuántico siempre es un obstáculo para el rendimiento, se constató que en condiciones específicas una pequeña cantidad de ruido cuántico puede aumentar las recompensas individuales, especialmente en juegos con un gran número de agentes.

La exploración de juegos cuánticos con aprendizaje multiagente ha revelado una riqueza de dinámicas complejas y fenómenos fascinantes. Estos descubrimientos no solo tienen implicaciones teóricas, sino que también pueden tener aplicaciones prácticas, especialmente dadas las limitaciones inherentes de los ordenadores cuánticos intermedios ruidosos (NISQ) contemporáneos. A medida que la comunidad científica avanza en la exploración de la interacción entre la teoría de juegos, la mecánica cuántica y el aprendizaje por refuerzo, los algoritmos y métodos propuestos en estos trabajos emergen como piedras angulares fundamentales para desentrañar las complejidades de las interacciones estratégicas en diversos dominios.

4.2. Algoritmos sin cálculo de gradiente para aprendizaje automático en juegos cuánticos repetidos

4.2.1. Aprendizaje de estrategias mixtas en juegos cuánticos con información imperfecta

Introducción

En esta sección se aborda la notable intersección de la teoría de juegos y la mecánica cuántica, explorando cómo las estrategias mixtas se aprenden en juegos cuánticos con información imperfecta. Este campo emergente desafía nuestra comprensión de la toma de decisiones estratégicas en un entorno cuántico, donde las superposiciones y el entrelazamiento expanden exponencialmente el espacio de estrategias posibles. [7] introduce un método innovador para navegar por este vasto paisaje, buscando equilibrios que, aunque difíciles de encontrar debido a la complejidad cuántica, prometen una nueva comprensión de la competencia y la colaboración en escenarios de múltiples agentes.

El enfoque propuesto se basa en el aprendizaje por refuerzo, una técnica bien establecida en la teoría de juegos clásica, pero aquí se extiende al dominio cuántico. A pesar de la falta de información completa, como la matriz de pagos y las acciones de otros agentes, el algoritmo desarrollado demuestra una flexibilidad y eficiencia notables, adaptándose a las peculiaridades de los juegos cuánticos. Este algoritmo no solo adquiere la capacidad de explorar el espacio de estrategias mixtas sino que también se ajusta a los efectos del entrelazamiento y el ruido del canal cuántico. La capacidad de los agentes para aprender y adaptarse en tales entornos imperfectos y ruidosos es crucial, ya que refleja las condiciones reales en las que se esperaría que operen los sistemas cuánticos.

Juego clásicos y cuánticos

El objetivo de la teoría de juegos es analizar sistemas de toma de decisiones que involucran a dos o más agentes cooperando o no entre sí. Una característica importante en los juegos es que la recompensa que recibe un agente depende no solo de la acción que elija, sino también de las acciones adoptadas por los otros agentes. Es bien sabido que un juego se define por tres elementos: agentes, estrategias y recompensas. Este trabajo se basa en juegos de dos agentes con dos estrategias puras cada uno. No obstante, se otorga la posibilidad a los agentes de usar estrategias mixtas, donde pueden asignar probabilidades a cada estrategia pura y luego seleccionar aleatoriamente entre ellas. Dado que las probabilidades son continuas, hay infinitamente muchas estrategias mixtas disponibles para un agente. Luego, las recompensas se definen por una matriz de pagos de 2×2. En la Tabla 4.1, es posible observar una representación general de una matriz de pagos de 2×2 (donde [a,c,e,g] y [b,d,f,h] son las recompensas de los agentes 0 y 1, respectivamente). Las siguientes tablas representan todos los juegos que se estudiarán en el resto de la sección (Tabla del Dilema del Prisionero 4.2, Tabla del Juego de Deadlock 4.3, Tabla de Descoordinación 4.4a y Tabla del Juego de Platonia 4.4b), donde el agente 0 puede seleccionar entre acciones de fila y el agente 1 entre acciones de columna y obtener una recompensa de $(R_{jugador0}, R_{jugador1})$.

\	Agente 1		
Agente 0	\	C	D
	C	(a ; b)	(c ; d)
	D	(e ; f)	(g ; h)

CUADRO 4.1: Representación matricial general de un juego con dos jugadores y dos estrategias.

\	Agente 1		
Agente 0	\	C	D
	C	(6.6 ; 6.6)	(0 ; 10)
	D	(10 ; 0)	(3.3 ; 3.3)

(A) Versión 1.

\	Agente 1		
Agente 0	\	C	D
	C	(5 ; 5)	(-10 ; 30)
	D	(30 ; -10)	(-5 ; -5)

(B) Versión 2.

CUADRO 4.2: Representación de la matriz de pagos del dilema del prisionero (e>a>g>c y d>b>h>f).

Para estudiar juegos cuánticos seguimos utilizando el protocolo *EWL* [51] para 2 agentes. Como se mencionó anteriormente, el primer paso es asignar un estado

\	Agente 1		
Agente 0	\	C	D
	C	(6.6 ; 6.6)	(10 ; 0)
	D	(0 ; 10)	(3.3 ; 3.3)

(A) Versión 1.

\	Agente 1		
Agente 0	\	C	D
	C	(5 ; 5)	(30 ; -10)
	D	(-10 ; 30)	(-5 ; -5)

(B) Versión 2.

CUADRO 4.3: Representación de la matriz de pagos del juego de deadlock (c>a>g>e y f>b>h>d).

\	Agente 1		
Agente 0	\	R	L
	R	(10 ; 0)	(0 ; 10)
	L	(0 ; 10)	(10 ; 0)

(A) Juego de descoordinación.

\	Agente 1		
Agente 0	\	R	L
	R	(0 ; 0)	(0 ; 10)
	L	(10 ; 0)	(0 ; 0)

(B) Juego egoísta.

CUADRO 4.4: Representación matricial de pagos de otros juegos útiles.

cuántico a cada una de las estrategias posibles. En el caso de dos estrategias, por ejemplo, en el Dilema del Prisionero, $C \rightarrow |0\rangle$ y $D \rightarrow |1\rangle$. El segundo paso es crear un circuito cuántico donde a cada agente se le asigna un qubit que comienza en el estado $|0\rangle$. El tercer paso es crear un estado entrelazado entre todos los agentes. Esto se hace aplicando el operador de entrelazamiento $J = cos(\frac{\gamma}{2})\mathbb{I}^{\otimes N} + isin(\frac{\gamma}{2}) * \sigma_x^{\otimes N}$, como se mostró ya en la Figura 3.2, donde $\mathbb{I}$ es la matriz identidad, σ_x la puerta Pauli X, $N = 2$ representa el número de agentes y γ un valor que determina la cantidad de entrelazamiento, siendo $\gamma = 0$ sin entrelazamiento en absoluto y $\gamma = \frac{\pi}{2}$ el entrelazamiento máximo.

En el cuarto paso, cada agente elige su estrategia más adecuada de manera individual e independiente. Esto se hace modificando el estado de su propio qubit localmente. Para hacer esto, cada agente aplica una o más puertas cuánticas de un qubit, modificando el estado de su qubit. Una puerta general de un qubit [13] es una matriz unitaria que puede representarse como:

$$U(\theta, \phi, \lambda) = \begin{pmatrix} \cos(\frac{\theta}{2}) & -e^{i\lambda}\sin(\frac{\theta}{2}) \\ e^{i\phi}\sin(\frac{\theta}{2}) & e^{i(\phi+\lambda)}\cos(\frac{\theta}{2}) \end{pmatrix} \tag{4.1}$$

Ya podemos destacar el hecho de que mientras los agentes clásicos solo tienen 2 posibles estrategias puras (por ejemplo, cooperar o defectar), los agentes cuánticos tienen un número infinito de estrategias puras, es decir, cualquier combinación de valor real para los tres parámetros θ, ϕ y λ. El quinto paso es aplicar el operador $J^\dagger$ (transpuesto conjugado de J) después de las estrategias de los agentes. Finalmente, el sexto paso consiste en medir el estado de los qubits para leer las salidas clásicas del circuito y, por lo tanto, la acción final de cada agente. Las lecturas se utilizan como entradas de la matriz de pagos para determinar las recompensas de los agentes.

Una última cosa a añadir es el hecho de que vamos a reemplazar la puerta general de un qubit de tres parámetros $U(\theta, \phi, \lambda)$ por tres puertas cuánticas de rotación de un parámetro $R_X(\varphi_1)$, $R_Y(\varphi_2)$ y $R_X(\varphi_3)$, con $R_X(\varphi) = exp(-i\frac{\varphi}{2}X) = \begin{pmatrix} cos(\frac{\varphi}{2}) & -isin(\frac{\varphi}{2}) \\ -isin(\frac{\varphi}{2}) & cos(\frac{\varphi}{2}) \end{pmatrix}$ y $R_Y(\varphi) = exp(-i\frac{\varphi}{2}Y) = \begin{pmatrix} cos(\frac{\varphi}{2}) & -sin(\frac{\varphi}{2}) \\ sin(\frac{\varphi}{2}) & cos(\frac{\varphi}{2}) \end{pmatrix}$. Esto es posible sin perder generalidad ya que $U(\theta, \phi, \lambda) = e^{i\alpha}R_{\hat{n}}(\beta)R_{\hat{m}}(\gamma)R_{\hat{n}}(\delta)$ [13]. Dicho esto, el circuito de la Figura 3.2 se convierte en el de la Figura 3.5.

Modelo de aprendizaje

Esta sección presenta un nuevo método para aprender estrategias mixtas en juegos cuánticos. Los resultados se centran en juegos con dos agentes, pero el método es extensible para N agentes. La principal dificultad al calcular estrategias mixtas en juegos cuánticos es que las posibles estrategias puras ya son infinitas (cualquier valor real para θ, ϕ y λ). Por lo tanto, una estrategia mixta se obtiene a través de una función de densidad de probabilidad sobre estas tres variables continuas.

Algoritmo de aprendizaje - Primero, los valores de θ, ϕ y λ se restringen a cualquier valor real entre $[0; 2\pi)$ (ya que son ángulos). Por lo tanto, una estrategia mixta se obtiene a través de una función de densidad de probabilidad (PDF) de tres variables. La idea principal es usar Q-learning [35], discretizando el espacio de estrategias, para aprender el valor aproximado de cada estrategia y construir una PDF con él. Los agentes seleccionan sus estrategias en cada iteración mediante el muestreo de la PDF. Reciben una recompensa, actualizan sus propias tablas Q y PDFs con esta nueva información y continúan con la siguiente iteración.

Tres dificultades surgen cuando los agentes intentan aprender sus estrategias mixtas en juegos cuánticos. El primer obstáculo se presenta debido a que, en el proceso de aprendizaje en juegos, la recompensa de cada agente depende no solo de la estrategia que elija, sino también de las estrategias de otros agentes. En otras palabras, la retroalimentación que un agente usa para actualizar el valor Q de una estrategia variará dependiendo de las estrategias que los otros agentes hayan seleccionado.

El segundo problema proviene de la diferencia en la naturaleza entre las estrategias clásicas y cuánticas. Mientras las estrategias clásicas puras siempre devuelven resultados determinísticos, las estrategias cuánticas puras pueden llevar a resultados no deterministas. Por ejemplo, en la versión 4.2a del Dilema del Prisionero, si el jugador0 selecciona la estrategia $S_0 = (\theta_0, \phi_0, \lambda_0) = (\pi, 0, 0)$ y el jugador1 también selecciona la estrategia $S_1 = (\theta_1, \phi_1, \lambda_1) = (\pi, 0, 0)$, el estado cuántico antes de medir va a ser $|\psi_{out}\rangle = J^\dagger(U(\pi,0,0) \otimes U(\pi,0,0))J|00\rangle = \frac{|00\rangle+|01\rangle+|10\rangle+|11\rangle}{2}$. Esto significa que las acciones finales van a ser (cooperar, cooperar) con probabilidad 0.25, (cooperar, traicionar) con probabilidad 0.25, (traicionar, cooperar) con probabilidad 0.25 y (traicionar, traicionar) con probabilidad 0.25. Por lo tanto, incluso si ambos agentes repiten sus estrategias puras, las recompensas para el jugador0 (y el jugador1) serán 6,6 con probabilidad 0.25, 0,0 con probabilidad 0.25, 10 con probabilidad 0.25 y 3,3 con probabilidad 0.25. Esta propiedad no determinista de las estrategias cuánticas puras hace que el aprendizaje de estrategias cuánticas mixtas sea incluso más lento.

La tercera dificultad se debe al hecho de que, al tratar con estrategias mixtas, aprender los valores Q no es suficiente. Los agentes necesitan ser capaces de transformar correctamente esa información en una función de densidad de probabilidad que no los haga ni demasiado codiciosos ni demasiado indecisos. Un comportamiento codicioso hará que los agentes intenten obtener más recompensas utilizando el valor estimado actual de las estrategias y no aprendiendo lo suficiente del entorno, este concepto se llama explotación y puede causar convergencia a mínimos locales. Por otro lado, un comportamiento indeciso hará que los agentes se enfoquen principalmente en mejorar su conocimiento sobre cada acción en lugar de obtener más recompensas, esto se llama exploración y puede causar que no haya convergencia en absoluto.

En este contexto, los agentes comienzan en la primera iteración inicializando su tabla Q, la cual, idealmente, convergerá a una representación cuantitativa de cuán buena es cada estrategia. Luego, establecen en cero un contador que indica cuántas

veces cada estrategia fue seleccionada. Este contador se utilizará para actualizar la tabla Q; cuantas más veces un agente haya seleccionado una estrategia, menor será la influencia que tendrá una nueva recompensa obtenida al actualizar su valor Q. Cabe aclarar que la primera estrategia de cada jugador es seleccionada al azar. Después de que todos los agentes tengan una estrategia, juegan el juego y obtienen la recompensa correspondiente según su matriz de pagos de 2×2.

En la siguiente iteración, los agentes usan sus estrategias seleccionadas y las recompensas recibidas para actualizar el contador: $count(s)_t = count(s)_{t-1} + 1$, y el valor Q: $Q_t(s) = Q_{t-1}(s) + (\frac{1}{count_t(s)}) * (R_{t-1} - Q_{t-1}(s))$. Ahora, los agentes tienen que seleccionar una nueva estrategia. Para lograr esto, van a convertir los valores Q en probabilidades y crear una PDF de la cual van a muestrear para seleccionar la siguiente estrategia. La probabilidad de cada estrategia se definirá mediante la función softmax: $p_t(s) = \frac{e^{\frac{Q_t(s)}{T}}}{\sum_{\text{all s}} e^{\frac{Q_t(s)}{T}}}$, donde T se denomina parámetro de temperatura. Para temperaturas altas ($T \to \infty$), todas las estrategias tienen casi la misma probabilidad y cuanto más baja es la temperatura, más afectan los valores Q a la probabilidad. Para una temperatura baja ($T \to 0$), la probabilidad de la estrategia con la recompensa esperada más alta tiende a 1.

Hay dos últimos conceptos que vale la pena mencionar antes de pasar a estudiar los resultados del algoritmo. Para evitar converger a máximos locales, añadimos un factor ϵ-codicioso. En cada iteración, cada agente genera un número real aleatorio entre 0 y 1 y si este número es menor que ϵ, el agente selecciona su estrategia al azar en lugar de muestrear de la PDF. Por último, el factor T de la función softmax comenzará desde un valor inicial alto T_o (priorizando la exploración) y luego disminuirá a un valor final bajo T_f (priorizando la explotación) siguiendo la ecuación: $T = T_f + (T_o - T_f)e^{-\frac{t}{\tau}}$, siendo τ un factor de tasa de decaimiento.

En el Algoritmo 3, es posible observar una descripción completa del método de aprendizaje utilizando pseudocódigo.

Modelo descentralizado - El sistema de aprendizaje es descentralizado, cada agente toma decisiones autónomas locales dirigidas a sus objetivos individuales que posiblemente entren en conflicto con los de otros agentes. Aprenderán individual e independientemente, sin comunicación entre ellos ni una influencia ordenadora central de un sistema centralizado que ejerza control directamente sobre los componentes de nivel inferior del sistema.

Además, los agentes no tendrán información perfecta sobre el entorno. Específicamente, los agentes no conocerán la matriz de pagos del juego que están jugando actualmente, no sabrán la estrategia o la recompensa que otros agentes están recibiendo en ningún momento, e incluso no saben contra cuántos agentes están jugando. Su *única* retroalimentación es la recompensa que reciben después de seleccionar una estrategia.

Tanto el modelo descentralizado como el de información imperfecta para dos agentes se pueden observar en la Figura 4.1. Los agentes seleccionan una estrategia $S_X = [\varphi_{X1}, \varphi_{X2}, \varphi_{X3}]$ y la envían al dispositivo cuántico. Las dos estrategias de los agentes funcionan como 6 parámetros (φ_{A1}, φ_{A2}, φ_{A3}, φ_{B1}, φ_{B2} y φ_{B3}) en un circuito cuántico parametrizado. Se ejecuta el circuito cuántico y las lecturas se asignan a sus acciones clásicas correspondientes (las que corresponden al primer paso del protocolo *EWL*). Estas acciones se utilizan para jugar el juego en curso y las recompensas se envían a sus agentes. Los agentes usan su recompensa para ajustar su estrategia (de acuerdo con el Algoritmo 3), seleccionan una nueva estrategia y el circuito cuántico parametrizado se ejecuta nuevamente.

Algorithm 3 Agentes que aprenden estrategias mixtas en juegos cuánticos

Require: $N \geq 2$
  $N \leftarrow 2$ ▷ Numero de agentes
  $T_o \leftarrow 4$ ▷ Temperatura inicial
  $T_f \leftarrow 0{,}125$ ▷ Temperatura final
  $\tau \leftarrow 30000$ ▷ Tasa de decaimiento
  $\epsilon \leftarrow 0{,}01$ ▷ Factor ϵ-codicioso
  for $t = 0$ **to** 200000 **do** ▷ Número de iteraciones
    $T \leftarrow T_f + (T_o - T_f)e^{-\frac{t}{\tau}}$ ▷ Actualizar temperatura
    for $n = 0$ **to** $N - 1$ **do** ▷ Para todos los agentes
      if $t = 0$ **then** ▷ Primer paso
        $count_n \leftarrow 0$ ▷ Inicializar el contador de estrategias
        $Q_n \leftarrow Q_0$ ▷ Inicializar la Q-table
        $strategy[n] \leftarrow strategies(randint)$ ▷ Estrategia aleatoria para jugador n
      else
        $count_n(s) \leftarrow count_n(s) + 1$ ▷ Actualizar contador para estrategia s
        $Q_n(s) \leftarrow Q_n(s) + (R[n] - Q_n(s))/count_n(s)$ ▷ Actualizar Q-table
        if random(1) $< \epsilon$ **then** ▷ Si no acuta codiciosamente
          $strategy[n] \leftarrow strategies(randint)$ ▷ Retornar acción
        end if
        for s **in** estrategias **do** ▷ Para todas las estrategias
          $p(s) \leftarrow \frac{e^{\frac{Q(s)}{T}}}{\sum_s e^{\frac{Q(s)}{T}}}$ ▷ Función de densidad de probabilidad
        end for
        str $\leftarrow$ sample from p ▷ Seleccionar una estrategia muestreando PDF
        $strategy[n] \leftarrow strategies(str)$ ▷ Retornar la estrategia seleccionada
      end if
    end for
    $R \leftarrow GAME(strategy)$ ▷ Obtener recompensas de todos los agentes
  end for

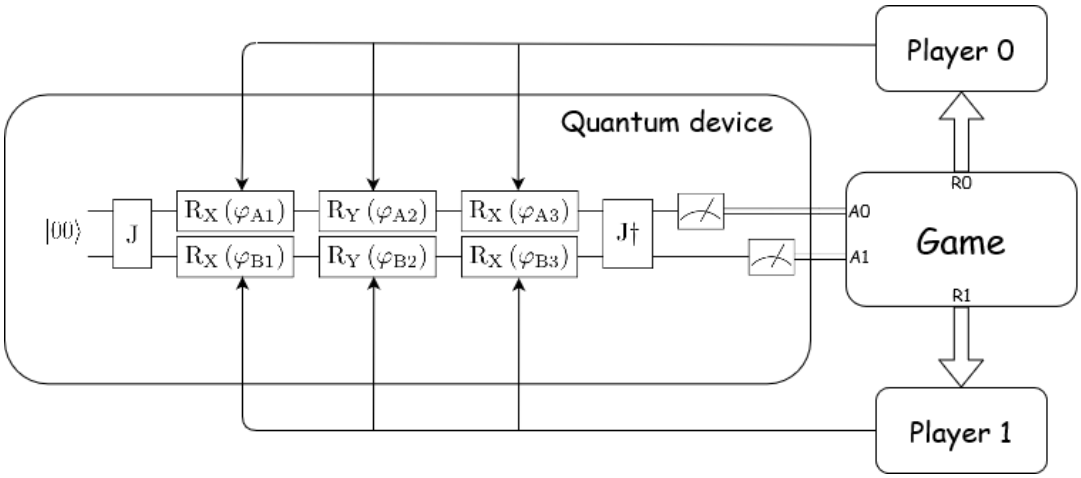


FIGURA 4.1: Modelo de dos jugadores aprendiendo estrategias mixtas utilizando circuitos cuánticos parametrizados.

Resultados

En esta sección, procederemos a aplicar el Algoritmo 3 al entorno clásico de los juegos presentados en la sección anterior para verificar su comportamiento. Los

equilibrios obtenidos se compararán con los equilibrios de Nash bien conocidos de estos juegos. Más tarde, se aplicará el Algoritmo 3 a la versión cuántica de los juegos y se verificará si los equilibrios cambiaron y, en caso afirmativo, determinando si dichos cambios representan mejoras o desmejoras.

En la primera parte de la sección de resultados, los agentes cuánticos estarán máximamente entrelazados en el momento en que apliquen sus puertas. Sin embargo, luego, se analizará el impacto del entrelazamiento en el rendimiento de los juegos cuánticos. De manera similar, en los primeros resultados, se asumirá que los circuitos cuánticos son ideales en términos de ausencia de ruido, para luego introducir ruido de despolarización en dichos circuitos. Finalmente, se procederá a caracterizar cómo esto sesgará el comportamiento de los agentes.

Performance de juegos clásicos vs cuánticos en condiciones ideales - Lógicamente es posible encontrar equilibrios de la versión clásica de los juegos presentados anteriormente aplicando el Algoritmo 3. En este caso, los agentes solo tienen dos estrategias (0 o 1, C o D, R o L, ...), por lo que las estrategias mixtas se definen por p_0 y p_1, donde $p_0 + p_1 = 1$. En la columna de equilibrios clásicos de la Tabla 4.5, es evidente que tanto los equilibrios de Nash conocidos como los equilibrios obtenidos por los agentes aplicando el Algoritmo 3 coinciden aproximadamente.

Juegos	Equilibrios clásicos		Equilibrios cuánticos	
	Nash	Obtenidos	Tendencia	Obtenidos
El dilema del Prisionero v1	[3.3 ; 3.3]	[3.316 ; 3.316]	[5 ; 5]	[4.968 ; 4.962]
El dilema del Prisionero v2	[-5 ; -5]	[-4.837 ; -4.857]	[10 ; 10]	[10.142 ; 9.618]
Juego de Deadlock v1	[6.6 ; 6.6]	[6.581 ; 6.585]	[5 ; 5]	[4.979 ; 4.964]
Juego de Deadlock v2	[5 ; 5]	[5.046 ; 5.055]	[10 ; 10]	[9.652 ; 9.464]
Juego de Descoordinación	[5 ; 5]	[5.001 ; 4.999]	[5 ; 5]	[4.987 ; 5.013]
Juego del Egoísmo	[0 ; 0]	[0.148 ; 0.094]	[5 ; 5]	[4.952 ; 4.946]

CUADRO 4.5: Rendimiento de los juegos clásicos y cuánticos después de aplicar el algoritmo 3 (recompensas promedio de los últimos 50000 valores).

Es importante destacar el hecho de que todas las estrategias de los agentes convergen en estrategias puras, con la excepción de aquellos que juegan el juego de discoordinación ($p_0 = 0{,}5$ y $p_1 = 0{,}5$). Esto verifica el comportamiento predicho por los equilibrios de Nash conocidos en todos los casos.

Las recompensas en el equilibrio y el comportamiento de aprendizaje de los agentes que aprenden a jugar juegos cuánticos usando el Algoritmo 3 se muestran en la Tabla 4.5 y en la Figura 4.2, respectivamente. Todos los valores en la Tabla 4.5 corresponden al promedio de los últimos 50,000, es decir, después de que se haya realizado la mayor parte de la fase de exploración. Es posible observar cómo en cuatro (Tablas 4.2a, 4.2b, 4.3b y 4.4b) de los seis juegos, el rendimiento cuántico es superior al clásico. Además, hay un caso (Tabla 4.4a) donde las recompensas son iguales y uno donde el caso clásico (Tabla 4.3a) supera al cuántico. Las recompensas promedio trazadas en la Figura 4.2 tambien se calcularon tomando el valor medio de una ventana de las últimas 50,000 recompensas, cuando fue posible:

$$recompensa\_promedio(t) = \begin{cases} promedio(recompensas(0:t)) & \text{si } t < 50{,}000 \\ promedio(recompensas(t - 50{,}000 : t)) & \text{si } t >= 50{,}000 \end{cases} \tag{4.2}$$

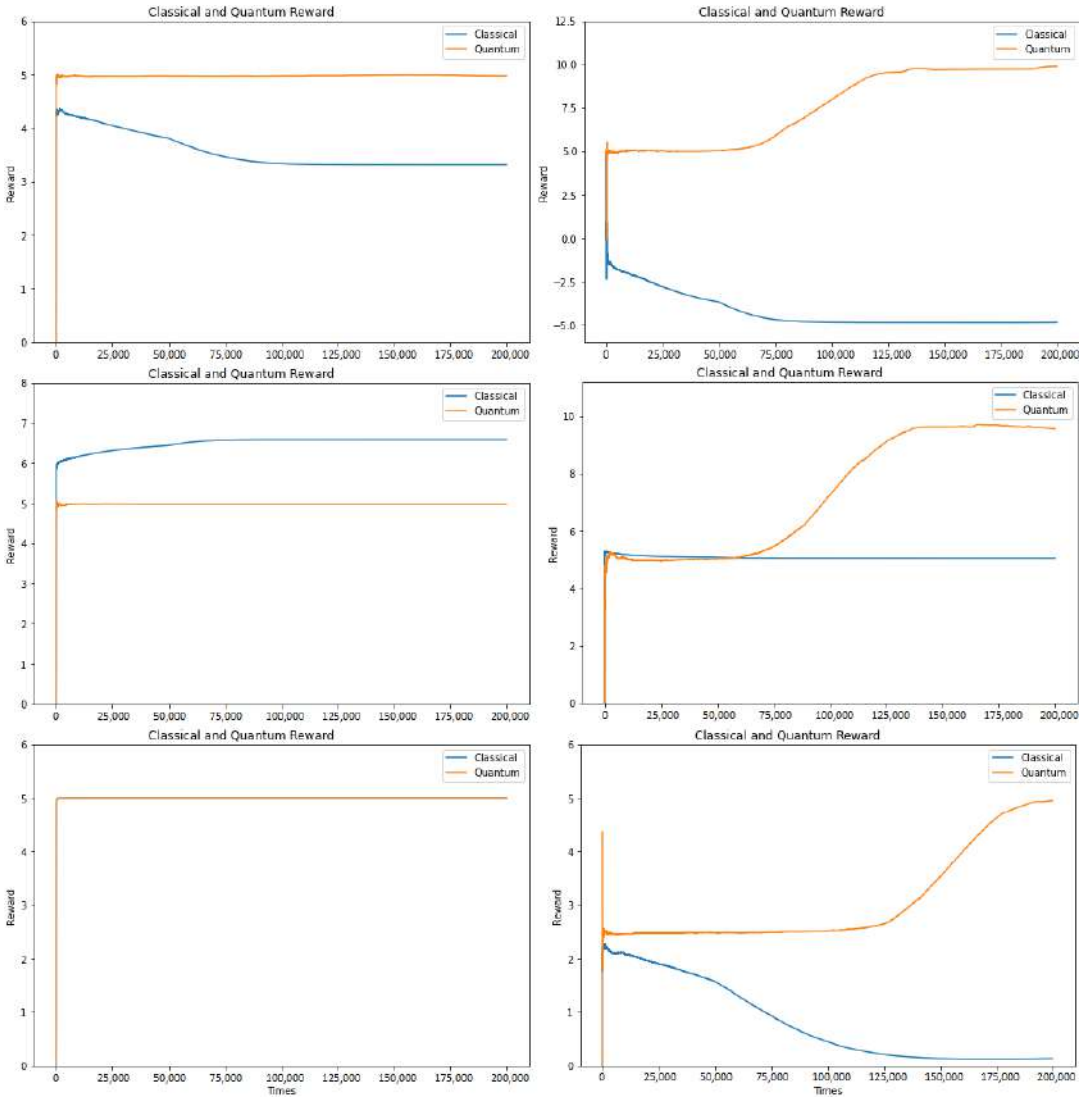


FIGURA 4.2: Recompensa promedio (desde una ventana de los últimos 50.000 valores) para los agentes que aprenden a jugar. Juegos de izquierda a derecha. Primera fila: Dilema del Prisionero v1 y Dilema del Prisionero v2. Segunda fila: Juego Deadlock v1 y Juego Deadlock v2. Tercera fila: Juego de Discoordinación y Juego Egoísta.

Es interesante observar que en ambas versiones del Dilema del Prisionero, los agentes cuánticos de alguna manera superan el famoso dilema y obtienen una recompensa más alta que el equilibrio de Nash clásico ($[5;5]$ en lugar de $[3{,}3;3{,}3]$ y $[10;10]$ en lugar de $[-5;-5]$). Además, en la versión 2 del Dilema del Prisionero (Tabla 4.2b), el resultado es incluso más alto que la eficiencia de Pareto clásica ($[10;10]$ en lugar de $[5;5]$). Esta segunda condición se cumplirá siempre que $[\frac{c+e}{2};\frac{d+f}{2}] > [a;b]$ en la Tabla 4.1.

De manera similar, en el juego de Deadlock, que es un juego que no presenta ningún dilema en su configuración clásica, sí lo hace en el mundo cuántico. Los agentes clásicos tienen una estrategia de cooperación mutua que es tanto un equilibrio de Nash como una eficiencia de Pareto. Sin embargo, cuando los agentes pasan al juego cuántico, pierden ese privilegio. Si se cumple la condición $[\frac{c+e}{2};\frac{d+f}{2}] > [a;b]$, esto es una buena noticia, ya que los agentes cuánticos aún logran tener un nuevo equilibrio con una recompensa incluso más alta que la de los agentes clásicos ($[10;10]$ en

lugar de $[5;5]$). De lo contrario, las recompensas de los agentes clásicos terminarán superando a las cuánticas ($[6{,}6;6{,}6]$ versus $[5;5]$).

Los últimos dos juegos son: el juego de discoordinación y el juego de Platonia. El primero no presenta ningún cambio cuando se cuantiza, lo cual es en sí mismo una particularidad. El segundo, por otro lado, es el juego que presenta la mayor diferencia. Los agentes clásicos siempre convergen a una estrategia pura (L), que les da la menor recompensa posible ($[g;h] = [0;0]$), mientras que los agentes cuánticos convergen a una estrategia que les permite obtener la recompensa más alta posible, $[\frac{c+e}{2}; \frac{d+f}{2}] = [5;5]$.

Dependencia del entrelazamiento - Hasta ahora, hemos adoptado un valor de $\gamma = \frac{\pi}{2}$ en $J = \cos(\frac{\gamma}{2})\mathbb{I}^{\otimes N} + i\sin(\frac{\gamma}{2}) * \sigma_x^{\otimes N}$. Esto asume un entrelazamiento máximo entre agentes. Se sabe que el entrelazamiento es el recurso que permite ventajas en los juegos cuánticos sobre los clásicos [129]. Sin embargo, lograr esta situación de entrelazamiento máximo a veces es costoso de producir. Por lo tanto, en la Figura 4.3 es posible visualizar cómo varía la relación entre las recompensas cuánticas y clásicas en equilibrio en función de γ (siendo $\gamma = 0$ y $\gamma = \frac{\pi}{2}$ sin entrelazamiento y entrelazamiento máximo respectivamente) para tres juegos con comportamientos diferentes.

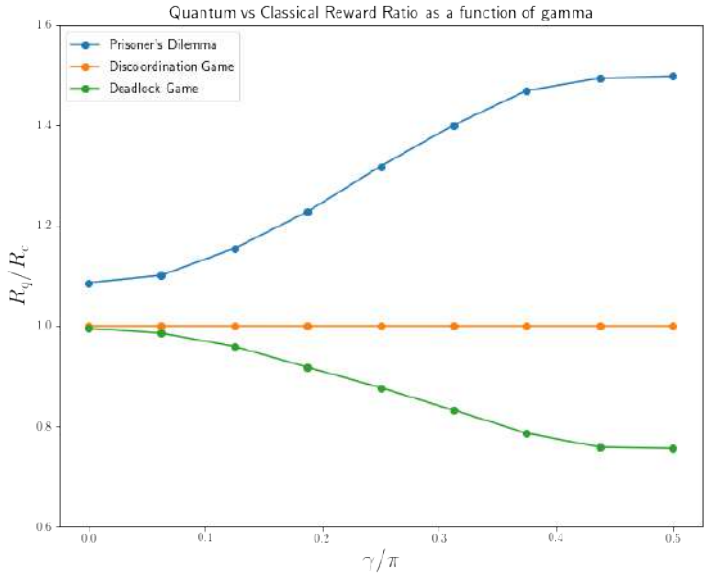


FIGURA 4.3: Relación de recompensa cuántica versus clásica en función del factor de entrelazamiento γ.

Para empezar, cuando $\gamma = 0$, es decir, sin entrelazamiento, los agentes de todos los juegos cuánticos tienen los mismos equilibrios que los agentes clásicos, $\frac{R_q}{R_c} \simeq 1$. Este comportamiento confirma que las diferencias entre los juegos clásicos y cuánticos provienen del entrelazamiento y no del hecho de que su espacio de estrategia se haya extendido de dos estrategias puras a un número infinito de estrategias puras. En el juego cuántico con $\gamma = 0$, los agentes todavía pueden manipular sus qubits con cualquier compuerta cuántica del conjunto infinito de ellas, tienen el mismo espacio de estrategia mixta con o sin entrelazamiento. Por lo tanto, entrelazamiento significa cuántico. Sin entrelazamiento significa clásico, incluso si se les permite manipular qubits.

A partir de ahí, en la Figura 4.3, a medida que aumentamos el valor de γ, podemos ver cómo la recompensa de los agentes se acerca progresivamente al valor previamente calculado con entrelazamiento máximo. Para el Dilema del Prisionero

v1, $\frac{R_q}{R_c} = \frac{4{,}968+4{,}962}{3{,}316+3{,}316} \simeq 1{,}5$. Para el juego de la discoordinación, $\frac{R_q}{R_c} = \frac{4{,}987+5{,}013}{5{,}001+4{,}999} = 1$. Para el juego del deadlock v1, $\frac{R_q}{R_c} = \frac{4{,}979+4{,}964}{6{,}581+6{,}585} \simeq 0{,}75$. Para concluir, podemos decir que para notar una ventaja (o desventaja) en los juegos cuánticos, no es necesario superar un cierto umbral de nivel de entrelazamiento. Cuanto mayor sea el nivel de entrelazamiento entre los agentes, por pequeño que sea, mayores serán las ventajas (o desventajas) que experimentarán los agentes.

Dependencia del ruido - Un análisis similar pero para un fenómeno diferente se puede realizar examinando el ruido cuántico. En las secciones anteriores, asumimos que, después de la aplicación del operador J, todos los agentes podían aplicar sus compuertas a sus qubits en un canal ideal. Esta condición también es difícil de cumplir, por lo que procederemos a modelar esta situación de manera más cuidadosa.

El modelo de ruido utilizado será el canal de depolarización. El estado del sistema cuántico después de este ruido es: $\varepsilon(\rho) = \frac{\lambda I}{2} + (1-\lambda)\rho = (1-\lambda)\rho + \frac{\lambda}{3}(X\rho X + Y\rho Y + Z\rho Z)$, siendo $\rho = |\psi\rangle\langle\psi|$ la matriz de densidad del estado cuántico antes de que se aplique el ruido. La forma de modelar esto (Figura 4.4) es añadiendo una cuarta compuerta después de $R_X(\varphi_1)R_Y(\varphi_2)R_X(\varphi_3)$, esta compuerta será seleccionada aleatoriamente siguiendo las probabilidades:

$$U4 = \begin{cases} I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \text{with } p = (1-\lambda) \\ X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \text{with } p = \frac{\lambda}{3} \\ Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} & \text{with } p = \frac{\lambda}{3} \\ Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} & \text{with } p = \frac{\lambda}{3} \end{cases} \tag{4.3}$$

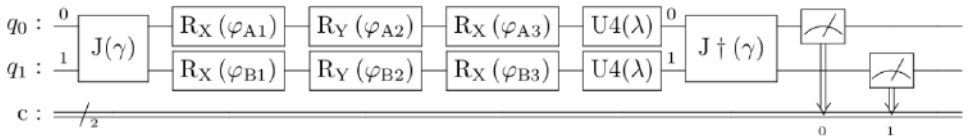


FIGURA 4.4: Un modelo completo del protocolo EWL teniendo en cuenta el factor de entrelazamiento y el ruido del canal despolarizante.

Esto significa que el estado cuántico en cada canal permanecerá intacto con una probabilidad de $(1-\lambda)$ y será modificado con una probabilidad de λ. Para modificar el estado cuántico del canal, las compuertas X, Y o Z serán aplicadas con la misma probabilidad. En la Figura 4.5 es posible visualizar cómo varía el rendimiento de dos juegos (Dilema del Prisionero v2 y Juego de Platonia) en función de λ.

En la Figura 4.5, es claro cómo el ruido afecta el rendimiento y el comportamiento de los agentes cuánticos. Las recompensas comienzan a disminuir rápidamente incluso con pequeños aumentos en el valor de λ. Además, para un valor de $\lambda \geq 0{,}4$, el proceso de aprendizaje de los agentes se ve tan afectado por el ruido que no pueden hacer nada mejor que aleatorizar entre todas las estrategias. Para el Dilema del Prisionero v2 (Tabla 4.2b), $R_q \simeq \frac{5-10+30-5}{4} = 5$. Para el juego de Platonia (Tabla 4.4b), $R_q \simeq \frac{0+10+0+0}{4} = 2{,}5$.

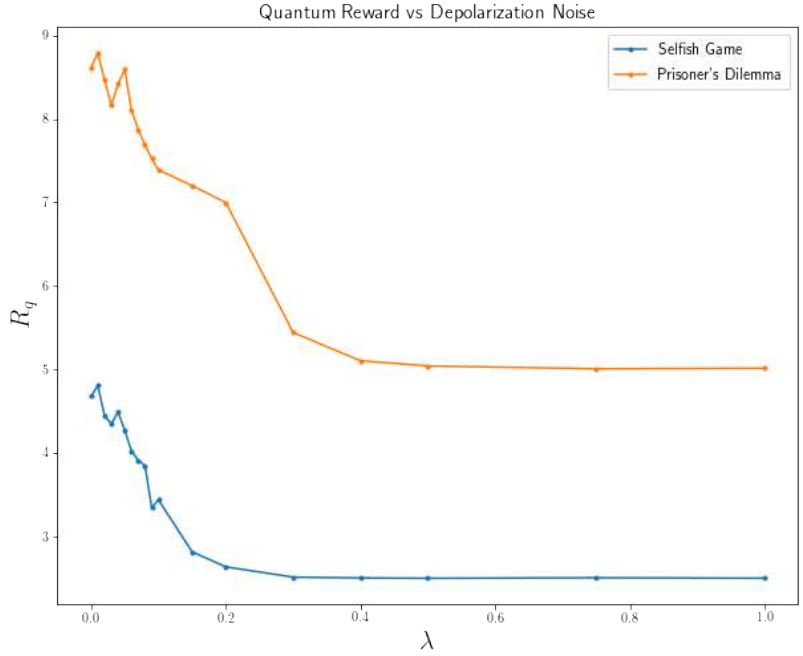


FIGURA 4.5: Recompensa de los jugadores cuánticos en función del parámetro λ para el modelo de canal despolarizante.

Una forma de interpretar esto es que los agentes están aprendiendo de una fuente corrupta. A veces, seleccionan una estrategia y reciben una recompensa correspondiente a otra estrategia. Esto causa que los agentes se comporten irracionalmente, lo que, a su vez, hace que aprender del otro agente sea también más difícil. Todo esto significa que si no se puede garantizar un sistema cuántico con poco ruido, el aprendizaje de los agentes se verá notablemente afectado.

Visualización de estrategias mixtas - Otra forma de entender el comportamiento de los agentes es visualizando sus estrategias. Las estrategias mixtas están representadas por una función de densidad de probabilidad sobre el espacio de estrategias. En los juegos cuánticos, el espacio de estrategias está representado por tres variables continuas, cada una de ellas representando los ángulos de las puertas de rotación cuántica $R_X(\varphi_1)$, $R_Y(\varphi_2)$ y $R_X(\varphi_3)$. En la Figura 4.6, es posible observar diferentes estrategias aprendidas por agentes en juegos cuánticos bajo distintas configuraciones. Cada punto en la PDF corresponde a una estrategia diferente $S = (\varphi_1, \varphi_2, \varphi_3)$ y tiene asignado un valor entre 0 y 1 que corresponde a la probabilidad de que esa estrategia sea seleccionada por el agente ($\sum_i P_{S_i} = 1$).

Cada punto está representado con un tamaño y un color proporcional a su valor. Cuanto mayor es el valor de la probabilidad, mayor es su tamaño y más amarillo su color. Es posible observar cómo, en algunos casos, las estrategias mixtas convergen a una estrategia pura, $P_{S_i} = 1$ y $\forall j \neq i \; P_j = 0$ (Figura 4.6, abajo a la derecha). Sin embargo, en la mayoría de los casos, las estrategias mixtas convergen en una distribución particular que puede o no tener una forma interesante.

Conclusión

En la culminación de la exploración de estrategias mixtas en juegos cuánticos con información imperfecta, se ha diseñado un algoritmo que no solo se destaca por su descentralización, sino que también capacita a los agentes para aprender dentro

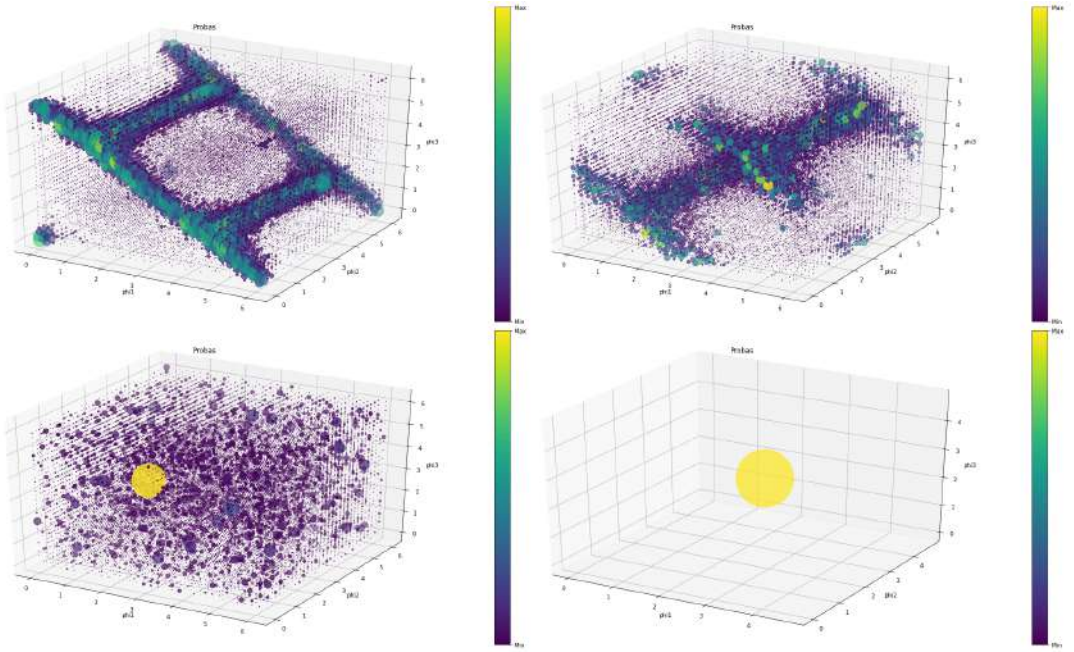


FIGURA 4.6: Función de Densidad de Probabilidad sobre las tres variables $(\varphi_1, \varphi_2, \varphi_3)$ que representan diferentes estrategias cuánticas mixtas. **Arriba a la izquierda**: juego de deadlock ($\gamma = 0$ y $\lambda = 0$). **Arriba a la derecha**: dilema del prisionero ($\gamma = 0$ y $\lambda = 0$). **Abajo izquierda**: juego de discoordinación ($\gamma = \frac{\pi}{2}$ y $\lambda = 0$). **Abajo a la derecha**: juego egoísta ($\gamma = \frac{\pi}{2}$ y $\lambda = 0$).

de un marco de información incompleta. Este algoritmo representa un avance significativo en la búsqueda sistemática de equilibrios en una variedad de juegos, permitiendo una comparación directa entre las recompensas obtenidas en las versiones cuánticas y clásicas de los juegos. La eficacia del algoritmo se ha corroborado a través de la coincidencia de los equilibrios obtenidos para las versiones clásicas de los juegos con sus respectivos equilibrios de Nash, proporcionando así una validación robusta de su rendimiento.

El uso del algoritmo para identificar equilibrios en juegos cuánticos reveló que las estrategias mixtas tienden a converger hacia distribuciones particulares, las cuales pueden o no presentar configuraciones interesantes. Este hallazgo subraya la complejidad inherente a la cuantificación de juegos y la necesidad de herramientas sofisticadas como la propuesta para navegar por este espacio expandido de estrategias. La capacidad de adaptarse y aprender en entornos cuánticos, donde la información puede ser fragmentaria o incierta, es crucial para el avance de la teoría de juegos cuántica y su aplicación en escenarios del mundo real, como mercados cuánticos y redes de comunicación cuánticas. La investigación futura se beneficiará de la exploración de estos algoritmos en sistemas con ruido y entrelazamiento no ideal, ampliando aún más el entendimiento de la toma de decisiones estratégicas en el ámbito cuántico.

4.2.2. QESRL: Exploración del aprendizaje por refuerzo egoísta para juegos cuánticos repetitivos

Introducción

En la vanguardia de la teoría de juegos cuántica, el artículo 'QESRL: Exploring Selfish Reinforcement Learning for Repeated Quantum Games' representa un avance significativo en la comprensión y aplicación de estrategias de aprendizaje por refuerzo en contextos cuánticos. Este trabajo se distingue de la sección al introducir un algoritmo unificado de aprendizaje que se aplica tanto a juegos clásicos como cuánticos de suma no cero teniendo el cuenta la distribución de las recompensas entre los agentes. Basándose en el algoritmo de 'Exploración Egoísta de Aprendizaje por Refuerzo' (ESRL) [128] previamente propuesto en juegos clásicos, este enfoque se extiende para manejar juegos cuánticos con información imperfecta. La singularidad de este algoritmo radica en su capacidad para permitir a los agentes explorar y aprender estrategias periódicas en juegos cuánticos, aprovechando la cuantización de los juegos para descubrir resultados más justos.

El algoritmo QESRL, en su esencia, es una fusión innovadora de la teoría de juegos, la mecánica cuántica y el aprendizaje por refuerzo. Esta confluencia de disciplinas permite una comprensión más profunda de las interacciones estratégicas, revelando la compleja trama que subyace a la optimización de comportamientos de agentes en diversos escenarios. Un aspecto central de este estudio es la evaluación de la equidad en la distribución de recompensas entre los agentes. A través de la implementación del algoritmo QESRL, se observa una notable diferencia en la justicia de la distribución de recompensas en comparación con algoritos de aprendizaje por refuerzo cuánticos (QRL) tradicionales. Este enfoque mantiene la adaptabilidad y la dinámica de convergencia del algoritmo adquiriendo la capacidad de equilibrar las recompensas individuales y colectivas, abordando así el desafío de distribuir recompensas de manera equitativa entre los agentes en escenarios multiagente. La capacidad del algoritmo QESRL para navegar en el intrincado espacio donde las interacciones estratégicas se encuentran con los fenómenos cuánticos promete remodelar nuestra comprensión de los comportamientos cooperativos y competitivos.

Modelo clásico ESRL

ESRL (Aprendizaje por Refuerzo Egoísta Exploratorio) [128] es un algoritmo basado en la teoría de autómatas de aprendizaje, diseñado para juegos repetidos estocásticos y de suma no-cero. Los agentes ESRL son independientes y actualizan una distribución probabilística sobre las acciones basadas en las recompensas recibidas. Los agentes alternan entre fases de exploración, priorizando la optimización individual, y fases de sincronización, coordinando objetivos sociales. Una vez transcurrido un período fijo de tiempo, donde los agentes han estado alternando entre exploración y sincronización, comienzan a explotar sus estrategias preferidas aprendidas durante las fases anteriores. En particular, ESRL se adapta a diferentes formas de juegos sin conocimiento previo. Este enfoque innovador armoniza a los agentes que comparten información muy limitada, utilizando aprendizaje autónomo y optimización colaborativa dentro de un marco dinámico. En las siguientes subsecciones, explicaremos generalmente las fases de exploración, sincronización y explotación. Sin embargo, en la siguiente sección daremos una descripción más profunda de ellas al definir el algoritmo QESRL.

Fase de exploración - Durante la fase de exploración del algoritmo ESRL, los agentes operan de manera autónoma como aprendices autodirigidos tradicionales. Empleando algoritmos de autómatas de aprendizaje por refuerzo, los agentes ajustan sus probabilidades de acción basadas en los resultados de sus acciones y las recompensas recibidas. Esta fase lleva a la convergencia de los agentes hacia equilibrios puros de Nash conjuntos. Este proceso de aprendizaje autocontenido permite a los agentes adaptarse a varios escenarios de juego sin depender del conocimiento previo.

Fase de sincronización - En la fase de sincronización, el enfoque cambia de la optimización individual a la toma de decisiones colaborativa. Los agentes evalúan la solución conjunta alcanzada durante la fase de exploración, considerando su relevancia para sus objetivos individuales. Esta fase introduce una comunicación limitada entre los agentes, permitiéndoles compartir su rendimiento y contribuir a la evaluación colectiva de la calidad del equilibrio. A través de esta cooperación, el algoritmo ESRL identifica equilibrios de Nash óptimos en Pareto, explora nuevos atractores de solución y combina experiencias de aprendizaje individuales para lograr resultados refinados y beneficiosos colectivamente.

Fase de explotación - En esta fase final, los agentes aprovechan las estrategias aprendidas en las fases anteriores de exploración y sincronización. Después de un tiempo suficiente, todos los agentes comienzan a implementar la política periódica que han aprendido coordinando su comportamiento y alternando entre las diferentes acciones conjuntas seleccionadas por cada agente. Esta política periódica se mantendrá hasta el final del episodio.

Modelo cuántico ESRL

El algoritmo QESRL adapta el algoritmo ESRL para ser aplicado a modelos de juegos cuánticos como los presentados en [7] y descritos en la sección 4.2.1. Las siguientes tablas representan todos los juegos que se estudiarán en el resto de la sección: (Batalla de los Sexos 4.6b, Dilema del Prisionero 4.6c y Juego de Platonia 4.6d), donde el agente 0 puede seleccionar entre acciones de fila y el agente 1 entre acciones de columna y obtener una recompensa de $(R_{jugador0}, R_{jugador1})$.

\	Agente 1		
Agente 0	\	A	B
	A	(a ; b)	(c ; d)
	B	(e ; f)	(g ; h)

(A) Representación matricial general de un juego.

\	Agente 1		
Agente 0	\	A	B
	A	(10 ; 5)	(2 ; 2)
	B	(0 ; 0)	(5 ; 10)

(B) Matriz de pagos de la Batalla de los Sexos.

\	Agente 1		
Agente 0	\	A	B
	A	(6.6 ; 6.6)	(0 ; 10)
	B	(10 ; 0)	(3.3 ; 3.3)

(C) Matriz de resultados del Dilema del Prisionero.

\	Agente 1		
Agente 0	\	A	B
	A	(0 ; 0)	(0 ; 10)
	B	(10 ; 0)	(0 ; 0)

(D) Matriz de pagos del juego Platonia.

Descripción del algoritmo QESRL - Para adaptar el ESRL a los juegos cuánticos, debemos ser capaces de aprender una función de densidad de probabilidad sobre las estrategias disponibles para los agentes en los juegos cuánticos. Cada estrategia cuántica está representada por 3 ángulos: φ_1, φ_2 y φ_3, los cuales podrían estar en el rango de $[0 : 2\pi)$. Como se describió anteriormente en [7], discretizaremos el espacio de estrategias y crearemos un vector donde cada elemento corresponde a un conjunto de ángulos: $S_i = [\varphi_1; \varphi_2; \varphi_3]$. Por lo tanto, este vector mapeará el espacio

de estrategias con la PDF actualizada en cada iteración del algoritmo. El diagrama de la Figura 4.1, que representaba visualmente el modelo de aprendizaje del artículo [7], se corresponde también con el modelo de aprendizaje del algoritmo QESRL con dos agentes. La descripción completa de las fases de exploración y sincronización del QESRL, adaptadas del ESRL clásico, se describe en el resto de esta sección.

En el algoritmo 4 es posible observar el pseudocódigo de la fase de exploración del QESRL, donde: α es la tasa de aprendizaje. PDF es un vector que representa la función de densidad de probabilidad en todas las acciones disponibles del agente. $Actions$ es un vector con todas las acciones seleccionadas por todos los agentes en una iteración. $Rewards$ es un vector con todas las recompensas recibidas de cada agente en una iteración. $dones$ es un vector con una bandera que indica si cada agente ha convergido.

Algorithm 4 Fase de exploración del algoritmo QESRL.

Require: Definición de la matriz del Juego
Require: N ▷ Número de agentes
Require: A ▷ Número de acciones
  $\alpha \leftarrow 0{,}001$ ▷ Tasa de aprendizaje
  $t_{max} \leftarrow 100000$ ▷ Número de iteraciones
  # *Inicialización*
  for $i = 1$ **to** N **do** ▷ Para cada agente
    Creación de agente i ▷ Inicializar PDF uniformemente
    $Actions[i] \leftarrow sample\_action(PDF)$ ▷ Muestrear PDF para determinar acción
  end for
  $Rewards \leftarrow Game(Actions)$ ▷ Recompensas de todos los agentes
  # *Loop principal*
  while $(!done\ or\ t_{max})$ **do**
    for $i = 1$ **to** N **do** ▷ Para cada agente
      $PDF \leftarrow (1 - \alpha * Rewards[i]) * PDF$ ▷ Actualizar de PDF
      $PDF[Actions[i]] \leftarrow PDF[Actions[i]] + \alpha * Rewards[i]$ ▷ Actualizar de PDF
      $Actions[i] \leftarrow sample\_action(PDF)$ ▷ Muestrear PDF
      $dones[i] \leftarrow done\_checking(PDF)$ ▷ Comprobar convergencia de agentes
    end for
    $Rewards \leftarrow Game(Actions)$ ▷ Recompensas de todos los agentes
    $done \leftarrow AND(dones)$ ▷ Comprobar convergencia de agentes
  end while

En el algoritmo 5 es posible observar el pseudocódigo de la fase de sincronización del QESRL, donde: $Actions$ es un vector con todas las acciones finales de cada agente. $Rewards$ es un vector con todas las recompensas finales para cada agente. Finalmente, $Hist$ es una matriz de N filas y 2 columnas. Cada fila, representando un agente diferente, tiene 2 columnas: la primera representa la acción conjunta preferida de cada agente en equilibrio y la segunda su recompensa correspondiente.

Resultados

Esta sección presenta dos estudios realizados utilizando el algoritmo QESRL. Comenzaremos comparando el rendimiento de dos agentes jugando tres juegos diferentes en su versión clásica versus cuántica. Luego se estudia cómo el rendimiento y la equidad de los agentes que juegan un juego particular crecen en función del número de agentes.

Algorithm 5 Fase de sincronización del algoritmo QESRL.

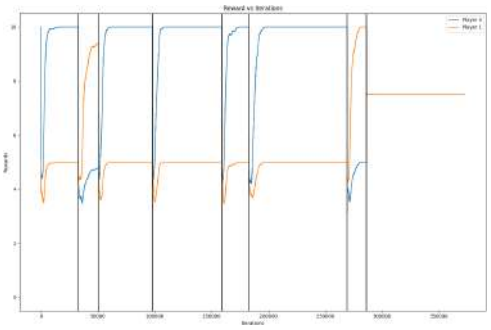
Require: Hist: Una lista donde cada agente almacenará su equilibrio preferido y su respectiva recompensa.
for $i = 1$ **to** N **do** ▷ Para cada agente
  if $Rewards[i] > Hist[i][1]$ **then** ▷ Si la recompensa recibida supera la preferida
    $Hist[i][0] \leftarrow Actions$ ▷ Actualizar equilibrio favorito
    $Hist[i][1] \leftarrow Rewards[i]$ ▷ Actualizar recompensa correspondiente
  end if
end for

Juegos clásicos versus juegos cuanticos - Los resultados de agentes utilizando el algoritmo QESRL para el Juego de la Batalla de los Sexos, el Dilema del Prisionero y el Juego de Platonia, tanto en sus versiones clásicas como cuánticas, se presentan en la Figura 4.7. Hay dos consideraciones importantes a tener en cuenta antes de continuar con los resultados: 1) las líneas verticales negras en los gráficos representan los cambios en las fases del algoritmo, siendo la última fase siempre la fase de explotación, que permanecerá indefinidamente; 2) la diferencia entre los juegos clásicos y cuánticos se establece modificando el parámetro γ del operador $J(\gamma)$, donde $\gamma = 0$ significa juego clásico y $\gamma = \frac{\pi}{2}$ juego cuántico. La tasa de aprendizaje se estableció en $\alpha = 0{,}001$ para todos los casos.

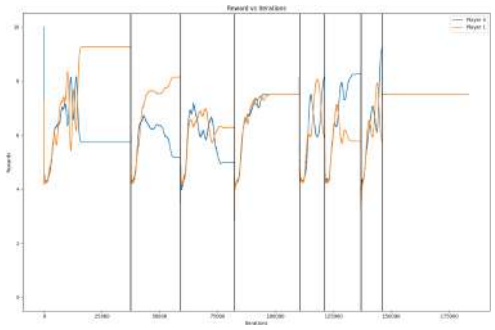
El juego de la Batalla de los Sexos (BoS) ilustra los desafíos de coordinación, ya que dos agentes eligen entre las acciones 'A' y 'B' para lograr un objetivo común, como se muestra en la tabla 4.6b. Las diferencias en las preferencias de recompensas subrayan la complejidad de las decisiones, resaltando los desafíos para alcanzar un consenso en medio de prioridades distintas. En la figura 4.7a es posible observar cómo dos agentes jugando la versión clásica del BoS convergen a una recompensa de $R_{1,2} = 7{,}5$. En particular, estos resultados confirman que la versión clásica de QESRL produce los mismos resultados que el algoritmo ESRL reportado en [128]. El juego BoS tiene tres equilibrios de Nash con recompensas: $[R_1, R_2] = [10, 5]$ (puro), $[R_1, R_2] = [5, 10]$ (puro), y $[R_1, R_2] = [3{,}84, 3{,}84]$ (mixto). Esto significa que las recompensas en el equilibrio alcanzado por el algoritmo ESRL (y QESRL) ($[R_1, R_2] = [7{,}5, 7{,}5]$) son más altas que el otro equilibrio con recompensas distribuidas equitativamente ($[R_1, R_2] = [3{,}84, 3{,}84]$) y tienen las mismas recompensas totales que los otros dos equilibrios ($[R_1, R_2] = [10, 5]$ y $[R_1, R_2] = [5, 10]$) pero ahora con recompensas distribuidas equitativamente. La combinación de rendimiento máximo y equidad es lo que hace que estos algoritmos sean tan fascinantes.

En la Figura 4.7b se puede ver cómo las recompensas de los dos agentes usando QESRL con entrelazamiento cuántico convergen exitosamente a $R_{1,2} = 7{,}5$. Aunque es posible observar que la dinámica de las recompensas de los agentes antes de la convergencia es más complicada que en la versión clásica (debido al entrelazamiento), ambos agentes logran una recompensa idéntica a la versión clásica.

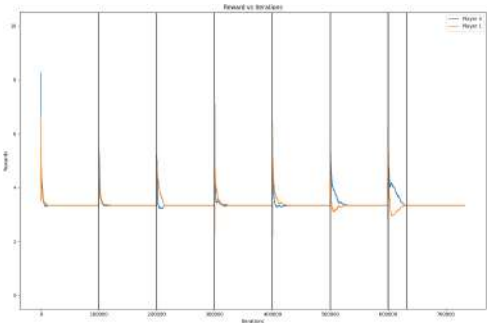
El Dilema del Prisionero (PD) ejemplifica los conflictos de cooperación. Los agentes eligen cooperar o traicionar, destacando la tensión entre el interés propio y el beneficio mutuo, ilustrando los desafíos de incentivar la cooperación con motivos en conflicto como se muestra en la Tabla 4.6c. La Figura 4.7c muestra la convergencia de la QESRL clásica cuando los agentes juegan el Dilema del Prisionero. Es importante destacar que el algoritmo QESRL sin entrelazamiento converge al equilibrio de Nash clásico, $[R_1, R_2] = [3{,}33, 3{,}33]$. Dado que la versión clásica del PD solo tiene un equilibrio de Nash, coincide con el equilibrio del algoritmo QESRL, verificando nuevamente su correcto funcionamiento.



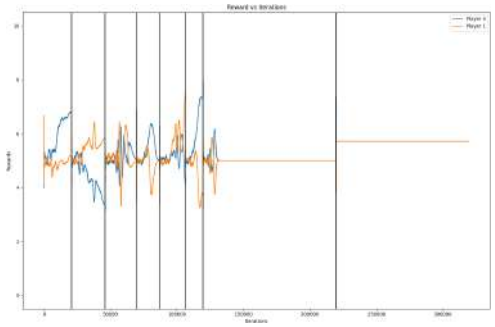
(A) Batalla de los sexos clásica.
Recompensas QESRL: Jugador 0 = 7.5000. Jugador 1 = 7.5000.



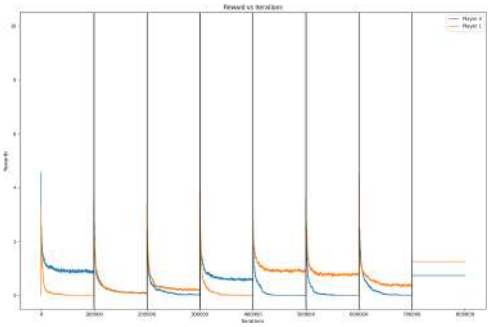
(B) Batalla de los sexos cuántica.
Recompensas QESRL: Jugador 0 = 7.5000. Jugador 1 = 7.5000.



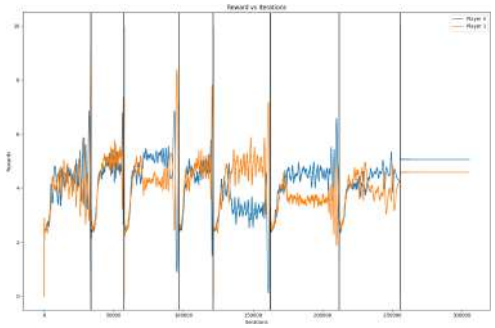
(C) El dilema del prisionero clásico.
Recompensas QESRL: Jugador 0 = 3.3333. Jugador 1 = 3.3333.



(D) El dilema del prisionero cuántico.
Recompensas QESRL: Jugador 0 = 5.7113. Jugador 1 = 5.7113.



(E) Juego de Platonia clásico.
Recompensas QESRL: Jugador 0 = 0.7322. Jugador 1 = 1.2500.



(F) Juego de Platonia cuántico.
Recompensas QESRL: Jugador 0 = 5.0537. Jugador 1 = 4.5802.

FIGURA 4.7: Agentes que utilizan el algoritmo QESRL en juegos Clásicos y Cuánticos con 2 jugadores.

Las recompensas de los agentes jugando la versión cuántica del dilema del prisionero se pueden observar en la Figura 4.7d. En ese gráfico, es posible visualizar la correcta convergencia del algoritmo y confirmar que las recompensas de los agentes usando entrelazamiento $[R_1, R_2] = [5{,}71, 5{,}71]$ (juego cuántico) son más altas que las de aquellos que no lo usan $[R_1, R_2] = [3{,}33, 3{,}33]$ (juego clásico).

Finalmente, el Juego de Platonia (PG) involucra a N participantes en busca de una recompensa sustancial. Un único individuo que reclame el premio lo obtiene todo, mientras que múltiples o ninguno resulta en ninguna recompensa para nadie,

como se muestra en la tabla 4.6d. Este escenario revela los desafíos de la toma de decisiones cooperativa. En las Figuras 4.7e y 4.7f es posible observar las recompensas de los agentes jugando el mismo juego (PG) usando QESRL con o sin entrelazamiento cuántico. En este juego en particular, podemos observar tanto la exitosa convergencia del algoritmo como la clara ventaja de la configuración cuántica versus la clásica: $[R_1, R_2] = [5{,}05, 4{,}58]$ vs. $[R_1, R_2] = [0{,}73, 1{,}25]$, respectivamente.

Performance versus Equidad - Ahora vamos a analizar cómo escala el algoritmo QESRL cuando el número de agentes en el juego aumenta. El Juego de Platonia se puede extender fácilmente a N agentes utilizando la siguiente tabla de recompensas 4.7:

Acciones		Resto de los N-1 agentes.	
		Todos seleccionan 0	Al menos uno selecciona 1
Agente i	0	0	0
	1	10	0

CUADRO 4.7: Recompensas del Agente i para el Juego Platonia de N jugadores.

Los juegos se jugarán en un entorno cuántico. La tasa de aprendizaje se establecerá en $\alpha = 0{,}001$ y el número máximo de iteraciones antes de detener la simulación cuando no se logra la convergencia será de $t_{max} = 100000$ en todos los casos. Esta última especificación es muy condicional, especialmente para juegos más grandes. A medida que aumenta el número de agentes, también aumenta el número de iteraciones necesarias para alcanzar el equilibrio. Por lo tanto, se espera observar una disminución en el rendimiento de los agentes a medida que crece el número de agentes. Sin embargo, adoptaremos este valor porque el tiempo de simulación aumenta exponencialmente a medida que crece el número de agentes. Esto se debe a dos razones: 1) más agentes de aprendizaje significa actualizar más funciones de densidad de probabilidad en cada iteración, y 2) el tamaño del circuito cuántico a simular en cada iteración también aumenta. De lo contrario, la simulación se volvería exponencialmente más costosa, requiriendo exponencialmente más iteraciones con tiempos de ejecución significativamente más prolongados en cada iteración. Además, mantener un número fijo de iteraciones máximas asegura consistencia en el análisis.

Dicho esto, en la Figura 4.8 (izquierda) es posible observar cómo evoluciona el rendimiento del QRL puro (lo que significa solo la fase de exploración del algoritmo QESRL) y el propio QESRL a medida que aumenta el número de agentes. Es importante destacar que el rendimiento de los agentes es comparable para ambos algoritmos. Sin embargo, cuando miramos la equidad en la distribución de recompensas entre los agentes, los resultados son notablemente diferentes.

El coeficiente de Gini [130], una medida esencial en la teoría de juegos, evalúa la equidad en la distribución de recompensas entre los agentes. Varía entre 0 y 1, reflejando desde un equilibrio perfecto hasta un desequilibrio total. Calculado a partir de la curva de Lorenz, cuantifica la disparidad entre la curva y la línea de equidad ideal, proporcionando perspectivas sobre la equidad dentro de contextos matemáticos y físicos. La fórmula del coeficiente de Gini para un conjunto de N recompensas r_i es $G = \frac{2\sum_{i=1}^{N} i r_i}{N\sum_{i=1}^{N} r_i} - \frac{N+1}{N}$. En la Figura 4.8 (derecha) es posible observar la equidad entre agentes $(1 - G)$ en función del número de agentes. En este gráfico, se puede observar que aunque el rendimiento de los agentes es casi el mismo entre el QRL puro y el QESRL, la equidad en la distribución de recompensas entre los agentes es significativamente mayor para el QESRL.

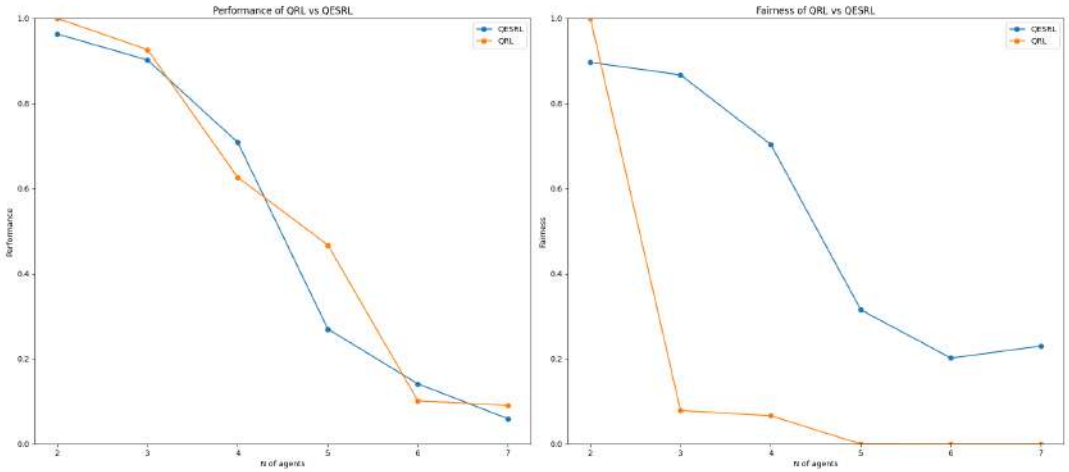


FIGURA 4.8: Rendimiento y equidad de las recompensas de N agentes que juegan el Juego Platonia.

Conclusión

La investigación presentada culmina con conclusiones significativas que marcan un avance en el campo del aprendizaje por refuerzo en juegos cuánticos. En primer lugar, el estudio destaca la eficacia del algoritmo QESRL no solo en términos de rendimiento y adaptabilidad, sino también en su capacidad para fomentar un entorno de juego más justo y equitativo. Esta característica del QESRL, que promueve una distribución más equilibrada de las recompensas entre los agentes, representa un paso importante hacia la creación de sistemas de aprendizaje por refuerzo que no solo buscan la optimización de las recompensas, sino también la equidad en el proceso de aprendizaje. Este enfoque en la justicia, especialmente en juegos cuánticos donde las interacciones pueden ser extremadamente complejas, es crucial para el desarrollo de aplicaciones prácticas que requieren una consideración cuidadosa de la equidad entre los participantes.

En segundo lugar, las conclusiones del estudio resaltan la importancia de la escalabilidad y la robustez del algoritmo QESRL. A diferencia de los enfoques tradicionales, el QESRL demuestra ser efectivo no solo en juegos simples, sino también en escenarios más complejos con un mayor número de agentes. Esta capacidad para mantener un alto rendimiento y justicia en entornos más complejos y dinámicos abre nuevas posibilidades para la aplicación de algoritmos de aprendizaje por refuerzo en una variedad de contextos cuánticos. Así, el estudio proporciona una base sólida para futuras investigaciones y desarrollos en el campo del aprendizaje por refuerzo cuántico, enfatizando la necesidad de considerar la justicia y la equidad como componentes esenciales en el diseño de estos algoritmos.

4.3. Maximizar recompensas locales en juegos cuánticos de múltiples agentes mediante estrategias de aprendizaje basadas en gradientes

Introducción

En la evolución de la investigación sobre juegos cuánticos, [8] representa un avance significativo, integrando estrategias de aprendizaje basadas en gradientes en el contexto de juegos cuánticos con múltiples agentes. Este enfoque se distingue de las secciones previas, 4.2.1 y 4.2.2, al enfocarse en la maximización de recompensas locales a través de técnicas de aprendizaje adaptativas y más sofisticadas, basadas en la estimación del gradiente de la función de recompensa.

Esta sección introduce una metodología novedosa que combina la teoría de juegos cuánticos con estrategias de aprendizaje automático, específicamente el aprendizaje basado en gradientes. Esta combinación permite una exploración más profunda de cómo los agentes pueden optimizar sus estrategias en entornos cuánticos, especialmente en situaciones donde múltiples agentes interactúan y compiten por recompensas. A diferencia de los anteriores, este estudio pone un énfasis particular en la eficiencia y efectividad del aprendizaje en entornos con poco ruido cuántico, profundizando sobre nuevas vías para comprender la interacción entre la mecánica cuántica y los algoritmos de aprendizaje automático. Este trabajo establece un marco teórico y práctico para investigar cómo los agentes pueden aprender y adaptarse de manera óptima en juegos cuánticos complejos, marcando otro paso adelante en la integración de la teoría cuántica y el aprendizaje automático.

Descripción del Modelo

Este trabajo se basa en juegos con N agentes que tienen dos estrategias puras cada uno. En este caso, las recompensas se pueden definir con una matriz de pagos con 2^N filas (una para cada combinación de acciones conjuntas) donde cada fila tiene N columnas (una para cada agente), cada columna representa la recompensa para cada agente dada una combinación de la acción conjunta. En la Tabla 4.8, es posible observar una representación general de una matriz de pagos para dos agentes con dos estrategias donde A0 y A1 son las estrategias del agente A; B0 y B1 son las estrategias del agente B; Ra1, Ra2, Ra3 y Ra4 son las recompensas del agente A; Rb1, Rb2, Rb3 y Rb4 son las recompensas del agente B. En consecuencia, una matriz de pagos general para tres y cuatro agentes se puede encontrar en las Tablas 4.9 y 4.10, respectivamente.

Estrategias	Recompensas
(A0, B0)	(Ra1, Rb1)
(A0, B1)	(Ra2, Rb2)
(A1, B0)	(Ra3, Rb3)
(A1, B1)	(Ra4, Rb4)

CUADRO 4.8: Matriz general de pagos para 2 jugadores y 2 estrategias.

A lo largo de la sección se consideran los siguientes juegos de N agentes con dos estrategias: el *juego de las minorías* [131], que premia a los agentes que seleccionan una estrategia que fue elegida por menos del 50 % de los agentes. El *juego de Platonia* [132], que recompensa a un agente si elige la estrategia 1 y todos los demás

Estrategias	Recompensas
(A0, B0, C0)	(Ra1, Rb1, Rc1)
(A0, B0, C1)	(Ra2, Rb2, Rc2)
(A0, B1, C0)	(Ra3, Rb3, Rc3)
(A0, B1, C1)	(Ra4, Rb4, Rc4)
(A1, B0, C0)	(Ra5, Rb5, Rc5)
(A1, B0, C1)	(Ra6, Rb6, Rc6)
(A1, B1, C0)	(Ra7, Rb7, Rc7)
(A1, B1, C1)	(Ra8, Rb8, Rc8)

CUADRO 4.9: Matriz general de pagos para 3 jugadores y 2 estrategias.

Estrategias	Recompensas
(A0, B0, C0, D0)	(Ra1, Rb1, Rc1, Rd1)
(A0, B0, C0, D1)	(Ra2, Rb2, Rc2, Rd2)
(A0, B0, C1, D0)	(Ra3, Rb3, Rc3, Rd3)
(A0, B0, C1, D1)	(Ra4, Rb4, Rc4, Rd4)
(A0, B1, C0, D0)	(Ra5, Rb5, Rc5, Rd5)
(A0, B1, C0, D1)	(Ra6, Rb6, Rc6, Rd6)
(A0, B1, C1, D0)	(Ra7, Rb7, Rc7, Rd7)
(A0, B1, C1, D1)	(Ra8, Rb8, Rc8, Rd8)
(A1, B0, C0, D0)	(Ra9, Rb9, Rc9, Rd9)
(A1, B0, C0, D1)	(Ra10, Rb10, Rc10, Rd10)
(A1, B0, C1, D0)	(Ra11, Rb11, Rc11, Rd11)
(A1, B0, C1, D1)	(Ra12, Rb12, Rc12, Rd12)
(A1, B1, C0, D0)	(Ra13, Rb13, Rc13, Rd13)
(A1, B1, C0, D1)	(Ra14, Rb14, Rc14, Rd14)
(A1, B1, C1, D0)	(Ra15, Rb15, Rc15, Rd15)
(A1, B1, C1, D1)	(Ra16, Rb16, Rc16, Rd16)

CUADRO 4.10: Matriz general de pagos para 4 jugadores y 2 estrategias.

agentes eligen 0. El *dilema de la cena* [133], que es una extensión para N agentes del bien conocido dilema del prisionero. Finalmente, el *dilema del voluntario* [134], que es una extensión para N agentes del famoso juego del gallina. Todos estos juegos se expandirán más en la subsección de resultados, sin embargo, ahora podemos observar una matriz de pagos de tres agentes de estos juegos como ejemplo (aunque todos son extensibles a N agentes): el problema de las minorías en la Tabla 4.11, el juego de Platonia en la Tabla 4.12, el dilema despiadado en la Tabla 4.13 y el dilema del voluntario en la Tabla 4.14.

En el resto de las sección modelaremos los juegos cuánticos siguiendo el modelo *EWL* [51] presentados secciones anteriores pero extendido a N agentes en [55].

Algoritmo de aprendizaje - Es momento de definir el algoritmo empleado por los agentes para aprender acciones óptimas en los juegos cuánticos presentados. El algoritmo opera de forma descentralizada, con cada agente intentando maximizar únicamente su propia recompensa a largo plazo. Aunque los agentes no comparten ninguna información con sus oponentes sobre el juego en el que participan o las recompensas que reciben, tienen que compartir sus acciones con otros participantes.

Estrategias	Recompensas
(0, 0, 0)	(0, 0, 0)
(0, 0, 1)	(0, 0, 10)
(0, 1, 0)	(0, 10, 0)
(0, 1, 1)	(10, 0, 0)
(1, 0, 0)	(10, 0, 0)
(1, 0, 1)	(0, 10, 0)
(1, 1, 0)	(0, 0, 10)
(1, 1, 1)	(0, 0, 0)

CUADRO 4.11: Matriz de pagos del juego de las minorías para 3 jugadores.

Estrategias	Recompensas
(0, 0, 0)	(0, 0, 0)
(0, 0, 1)	(0, 0, 10)
(0, 1, 0)	(0, 10, 0)
(0, 1, 1)	(0, 0, 0)
(1, 0, 0)	(10, 0, 0)
(1, 0, 1)	(0, 0, 0)
(1, 1, 0)	(0, 0, 0)
(1, 1, 1)	(0, 0, 0)

CUADRO 4.12: Matriz de pagos del dilema de Platonia para 3 jugadores.

Estrategias	Recompensas
(0, 0, 0)	(6.66, 6.66, 6.66)
(0, 0, 1)	(3.33, 3.33, 10.0)
(0, 1, 0)	(3.33, 10.0, 3.33)
(0, 1, 1)	(0.00, 6.66, 6.66)
(1, 0, 0)	(10.0, 3.33, 3.33)
(1, 0, 1)	(6.66, 0.00, 6.66)
(1, 1, 0)	(6.66, 6.66, 0.00)
(1, 1, 1)	(3.33, 3.33, 3.33)

CUADRO 4.13: Matriz de pagos de dilemas sin escrúpulos para 3 jugadores.

Antes de adentrarnos en la descripción del algoritmo, es importante definir algunos conceptos cruciales. Primero, la estrategia de cada agente se representa como el producto matricial de tres compuertas de rotación cuántica dependientes de parámetros (acciones): $u_i = R_X(\varphi_{i3}) \times R_Y(\varphi_{i2}) \times R_X(\varphi_{i1})$. Segundo, la estrategia conjunta de todos los agentes se representa como el producto tensorial de Kronecker de todas las acciones individuales de los agentes. Para el caso de dos agentes A y B, esto se expresa como $U = u_A \otimes u_B = (R_X(\varphi_{A3}) \times R_Y(\varphi_{A2}) \times R_X(\varphi_{A1})) \otimes (R_X(\varphi_{B3}) \times R_Y(\varphi_{B2}) \times R_X(\varphi_{B1}))$. Este formalismo permite encapsular las estrategias colectivas de los agentes en el juego cuántico.

El estado cuántico de salida del circuito cuántico se calcula entonces como $|\psi_{out}\rangle = J^\dagger \times U \times J \times |00\rangle$. Por último, la probabilidad de que cada estado clásico posible se mida al final del circuito cuántico se determina mediante la regla de Born $prob(\varphi_{A1}, \varphi_{A2}, \varphi_{A3}, \varphi_{B1}, \varphi_{B2}, \varphi_{B3}) =$

Estrategias	Recompensas
(0, 0, 0)	(0, 0, 0)
(0, 0, 1)	(0, 0, 1)
(0, 1, 0)	(0, 1, 0)
(0, 1, 1)	(0, 1, 1)
(1, 0, 0)	(1, 0, 0)
(1, 0, 1)	(1, 0, 1)
(1, 1, 0)	(1, 1, 0)
(1, 1, 1)	(-10, -10, -10)

CUADRO 4.14: Matriz de pagos del dilema de los voluntarios para 3 jugadores.

$|\psi_{out}|^2 = \begin{bmatrix} p_{00} \\ p_{01} \\ p_{10} \\ p_{11} \end{bmatrix}$. Esta ecuación relaciona directamente las estrategias de los agentes (ángulos) con las probabilidades de que cada acción (*0* o *1*) sea seleccionada en un juego cuántico. Todos estos conceptos son fácilmente extensibles a un caso general que involucre a N agentes.

Todos los juegos con N agentes y dos acciones, como se mencionó anteriormente, pueden representarse mediante una matriz de 2^N filas y N columnas. Cada fila corresponde a una combinación específica de acciones elegidas por los agentes, mientras que cada columna representa la recompensa recibida por cada agente. Esta representación matricial sirve como un resumen directo entre todas las recompensas asignadas en el juego y cada combinación posible de acciones. Como ejemplo ilustrativo, un juego genérico de dos agentes con dos acciones puede representarse por la siguiente matriz, donde los pares $[Rai, Rbj]$ denotan las combinaciones de recompensas para el agente A y el agente B en respuesta a sus posibles acciones, respectivamente, $juego = \begin{bmatrix} Ra1 & Rb1 \\ Ra2 & Rb2 \\ Ra3 & Rb3 \\ Ra4 & Rb4 \end{bmatrix}$.

Utilizando los conceptos definidos anteriormente, se puede calcular un vector de recompensas. Esto se logra multiplicando el vector que representa la probabilidad de que cada acción sea seleccionada por la matriz del juego. La representación matemática de esta operación se da por $recompensa(\varphi_{A1}, \varphi_{A2}, \varphi_{A3}, \varphi_{B1}, \varphi_{B2}, \varphi_{B3}) = prob.transpuesta() * juego$. En este cálculo, la matriz *recomensa* está compuesta por dos filas, la primera representa la recompensa asignada al Agente A: $recompensa_a = p_{00} * Ra1 + p_{01} * Ra2 + p_{10} * Ra3 + p_{11} * Ra4$. Esto es una suma ponderada de las recompensas para el Agente A, con pesos correspondientes a las probabilidades de las posibles combinaciones de acciones. De manera similar, la segunda fila corresponde a la recompensa para el Agente B: $recompensa_b = p_{00} * Rb1 + p_{01} * Rb2 + p_{10} * Rb3 + p_{11} * Rb4$. Así, el marco propuesto permite una clara relación entre las recompensas esperadas para cada agente dadas sus estrategias.

Ahora es posible describir el algoritmo o policy empleado por los agentes para actualizar sus estrategias basadas en las recompensas recibidas en cada iteración. Aquí, el término policy se refiere a una función utilizada por el agente para seleccionar acciones dada la recompensa instantanea recibida [45]. Cada agente comienza con un conjunto de acciones seleccionadas al azar y utiliza el algoritmo Adam [135], un prominente método de optimización basado en gradientes en el dominio del

aprendizaje automático, que ajusta adaptativamente las tasas de aprendizaje para las acciones. Sin embargo, antes de continuar, es necesario definir un aspecto crítico: el método por el cual cada agente calcula el gradiente de la función de recompensa con respecto a sus acciones.

Cada agente estima el valor del gradiente recalculando su propia recompensa pero modificando una de sus propias acciones por una pequeña diferencia de ϵ. Por ejemplo, el Agente A, poseyendo información sobre el juego en curso y las acciones tomadas por otros agentes, puede calcular el gradiente con respecto a cada una de sus acciones de la siguiente manera:

$$\nabla\varphi_{A1} = \frac{reward_a(\varphi_{A1}+\varepsilon, \varphi_{A2}, \varphi_{A3}, \varphi_{B1}, \varphi_{B2}, \varphi_{B3}) - reward_a(\varphi_{A1}, \varphi_{A2}, \varphi_{A3}, \varphi_{B1}, \varphi_{B2}, \varphi_{B3})}{\varepsilon}$$

$$\nabla\varphi_{A2} = \frac{reward_a(\varphi_{A1}, \varphi_{A2}+\varepsilon, \varphi_{A3}, \varphi_{B1}, \varphi_{B2}, \varphi_{B3}) - reward_a(\varphi_{A1}, \varphi_{A2}, \varphi_{A3}, \varphi_{B1}, \varphi_{B2}, \varphi_{B3})}{\varepsilon}$$

$$\nabla\varphi_{A3} = \frac{reward_a(\varphi_{A1}, \varphi_{A2}, \varphi_{A3}+\varepsilon, \varphi_{B1}, \varphi_{B2}, \varphi_{B3}) - reward_a(\varphi_{A1}, \varphi_{A2}, \varphi_{A3}, \varphi_{B1}, \varphi_{B2}, \varphi_{B3})}{\varepsilon}$$

Al calcular los gradientes, las acciones de los agentes podrían ajustarse directamente en la dirección de aumentar las recompensas utilizando ascenso de gradiente estocástico (SGD) de manera directa como en (4.4), donde α es la tasa de aprendizaje, es decir, la velocidad a la que un modelo se ajusta a nuevos comentarios. Los agentes también podrían incorporar un promedio móvil de los últimos gradientes para dar pasos más grandes en áreas más planas y pasos más pequeños en áreas más empinadas, como en RMSProp.

$$\begin{cases} \varphi_{A1}^{t+1} = \varphi_{A1}^{t} + \alpha * \nabla\varphi_{A1} \\ \varphi_{A2}^{t+1} = \varphi_{A2}^{t} + \alpha * \nabla\varphi_{A2} \\ \varphi_{A3}^{t+1} = \varphi_{A3}^{t} + \alpha * \nabla\varphi_{A3} \end{cases} \tag{4.4}$$

Sin embargo, el algoritmo Adam supera a SGD y RMSProp al ofrecer tasas de aprendizaje adaptativas que se ajustan para cada acción individualmente para navegar por la función de recompensa, empleando promedios móviles (como en RMSProp), pero también incorporando el cálculo de la estimación de momentos. Si bien este algoritmo de aprendizaje resulta extremadamente más eficiente, requiere que los agente realicen algunos cálculos adicionales, que se definen a continuación:

$$\begin{cases} m_{A1}^{t+1} = \beta_1 * m_{A1}^{t} + (1-\beta_1) * \nabla\varphi_{A1} \\ m_{A2}^{t+1} = \beta_1 * m_{A2}^{t} + (1-\beta_1) * \nabla\varphi_{A2} \\ m_{A3}^{t+1} = \beta_1 * m_{A3}^{t} + (1-\beta_1) * \nabla\varphi_{A3} \\ v_{A1}^{t+1} = \beta_2 * v_{A1}^{t} + (1-\beta_2) * \nabla\varphi_{A1}^2 \\ v_{A2}^{t+1} = \beta_2 * v_{A2}^{t} + (1-\beta_2) * \nabla\varphi_{A2}^2 \\ v_{A3}^{t+1} = \beta_2 * v_{A3}^{t} + (1-\beta_2) * \nabla\varphi_{A3}^2 \end{cases}$$

$$\begin{cases} \hat{m}_{A1} = m_{A1}^{t+1} / (1-\beta_1^{t+1}) \\ \hat{m}_{A2} = m_{A2}^{t+1} / (1-\beta_1^{t+1}) \\ \hat{m}_{A3} = m_{A3}^{t+1} / (1-\beta_1^{t+1}) \\ \hat{v}_{A1} = v_{A1}^{t+1} / (1-\beta_2^{t+1}) \\ \hat{v}_{A2} = v_{A2}^{t+1} / (1-\beta_2^{t+1}) \\ \hat{v}_{A3} = v_{A3}^{t+1} / (1-\beta_2^{t+1}) \end{cases}$$

$$\begin{cases} \varphi_{A1}^{t+1} = \varphi_{A1}^{t} + (\alpha_1 * \hat{m}_{A1})/(\sqrt{\hat{v}_{A1}} + \alpha_2) \\ \varphi_{A2}^{t+1} = \varphi_{A2}^{t} + (\alpha_1 * \hat{m}_{A2})/(\sqrt{\hat{v}_{A2}} + \alpha_2) \\ \varphi_{A3}^{t+1} = \varphi_{A3}^{t} + (\alpha_1 * \hat{m}_{A3})/(\sqrt{\hat{v}_{A3}} + \alpha_2) \end{cases}$$

donde α_1 determina el tamaño del paso en cada iteración. α_2 es un número muy pequeño para prevenir cualquier división por cero. β_1 se utiliza para el promedio de decaimiento exponencial de gradientes pasados. β_2 se utiliza para el promedio de decaimiento exponencial de gradientes cuadrados pasados. Finalmente, m es una estimación del primer momento (la media) de los gradientes y v es una estimación del segundo momento (la varianza) de los gradientes. Para obtener más información sobre el algoritmo Adam, por favor consulte la bibliografía [135].

El método de cálculo del gradiente y el algoritmo Adam implican varios hiperparámetros. Estos valores han sido estudiados meticulosamente, resultando en la identificación de un conjunto óptimo de valores. Estos valores óptimos se adoptan para el resto del artículo, asegurando la consistencia y resultados óptimos en análisis y discusiones subsiguientes.

$$\begin{cases} \varepsilon = 1 \times 10^8 \\ \alpha_1 = 0{,}0001 \\ \alpha_2 = 1 \times 10^8 \\ \beta_1 = 0{,}9 \\ \beta_2 = 0{,}999 \end{cases}$$

Finalmente, un diagrama de bloques que muestra el sistema integral que incluye tanto a los agentes que implementan el algoritmo de aprendizaje como al entorno del juego cuántico se puede encontrar en la Figura 4.9.

Resultados

Esta sección presenta los resultados obtenidos al aplicar el algoritmo previamente descrito a los juegos de suma no nula mencionados anteriormente. Se proporcionan representaciones gráficas de las recompensas para cada agente a lo largo de 1,000,000 de iteraciones para dos escenarios:

1. Juegos sin entrelazamiento, logrados con un valor de $\gamma = 0$ en el operador J. Este escenario corresponde a la versión clásica del juego.

2. Juegos con entrelazamiento, logrados con un valor de $\gamma = \frac{\pi}{2}$ en el operador J, que corresponde a la versión completamente cuántica del juego.

Los resultados se presentan de esta manera para lograr representar dos objetivos: 1) verificar la correcta funcionalidad del algoritmo (el escenario clásico debería comportarse como lo predice la teoría de juegos clásica), 2) comparar la dinámica y el rendimiento de los agentes entre los casos clásicos y cuánticos. Esta comparación puede revelar cualquier cambio notable en el comportamiento o la estrategia que surja en la transición de un marco clásico a uno cuántico.

Este análisis comienza con el juego de las minorías. La Figura 4.10 ilustra la evolución de las recompensas de los agentes tanto para los casos clásicos (izquierda) como cuánticos (derecha), específicamente para juegos que involucran tres, cuatro y cinco agentes. Cabe destacar que el juego de las minorías no tiene una interpretación significativa para un escenario de dos agentes.

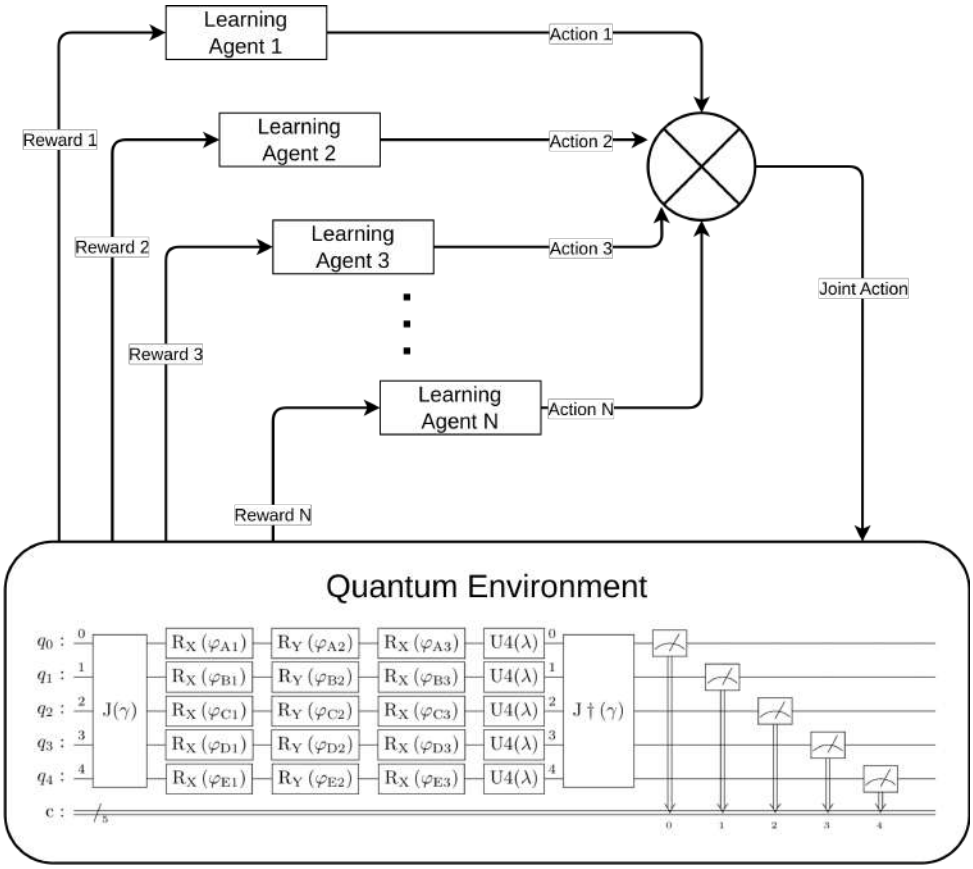
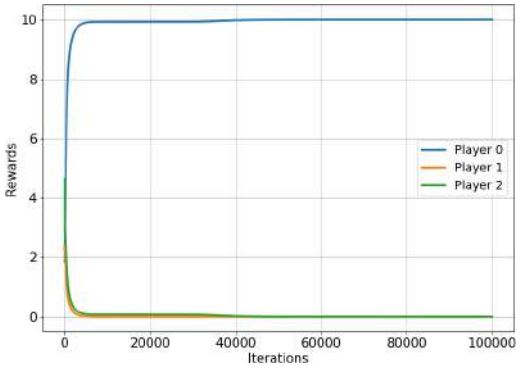


FIGURA 4.9: Modelo completo del sistema: aprendizaje de agentes + entorno cuántico.

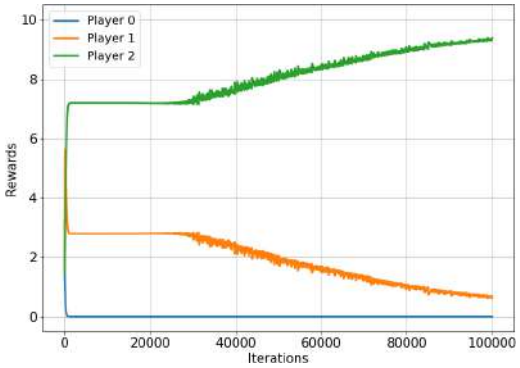
Los resultados son, de hecho, convincentes, aunque una descripción más específica del juego de las minorías es beneficiosa para nuestro análisis. Tal como se mencionó anteriormente, el juego de la minoría para N agentes ofrece a cada agente dos opciones (0 y 1), luego, se otorga una recompensa de $R_N = 10$ a los agentes que elijan la opción minoritaria, que se define como cualquier selección realizada por menos del 50 % de los agentes [131].

La Figura 4.10 muestra que las recompensas promedio para el juego que involucra a tres y cinco agentes son casi idénticas en ambos casos, clásico y cuántico ($R_3 = \frac{10*1}{3} = 3{,}333$ y $R_5 = \frac{10*2}{5} = 3{,}999$ versus $R_3 = 3{,}333$ y $R_5 = 3{,}999$, respectivamente), exactamente como los equilibrios de Nash clásicos y cuánticos calculados teóricamente en [55]. No solo los valores son casi los mismos, sino que también convergen hacia un estado de equilibrio óptimo: en un juego de tres agentes, un solo agente puede estar en la minoría, mientras que en un juego de cinco agentes, dos agentes es la máxima cantidad que constituye una minoría. Este equilibrio óptimo denota un estado de balance donde las recompensas se maximizan, subrayando la exitosa implementación del algoritmo tanto en contextos clásicos como cuánticos.

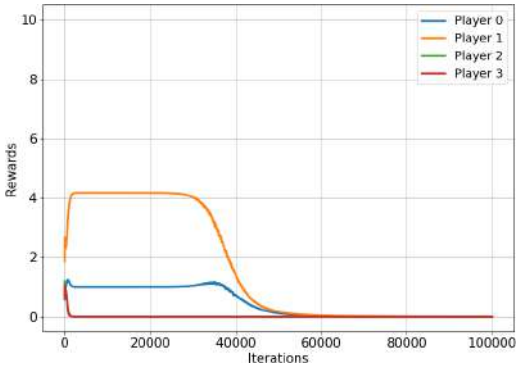
Sin embargo, para el juego de cuatro agentes, el caso clásico converge a una recompensa de $R_3 \approx 0$ para todos los agentes, mientras que el caso cuántico no. Los agentes en el juego cuántico logran aprender un conjunto de estrategias, asegurando que cada agente gane con una probabilidad $p \approx 1/4$, también como se predijo en [55]. Esto resulta en que cada agente reciba una recompensa promedio de aproximadamente $R_4 = \frac{10*1}{4} \approx 2{,}2451$, demostrando una clara ventaja del juego cuántico sobre el enfoque clásico en este escenario particular.



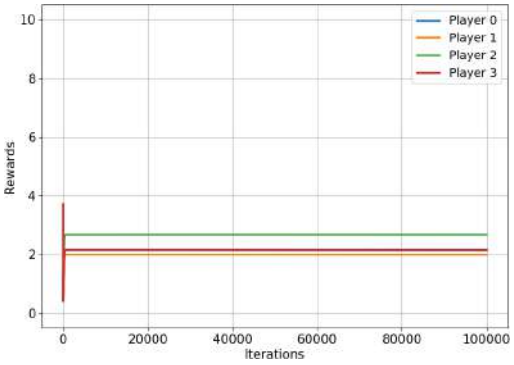
(A) Clásico $N = 3$. $\bar{R} = 3,3333$.



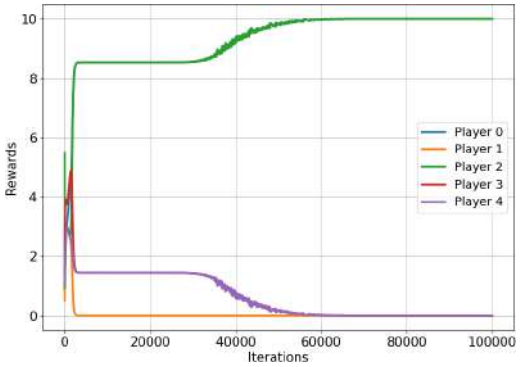
(B) Cuántico $N = 3$. $\bar{R} = 3,3333$.



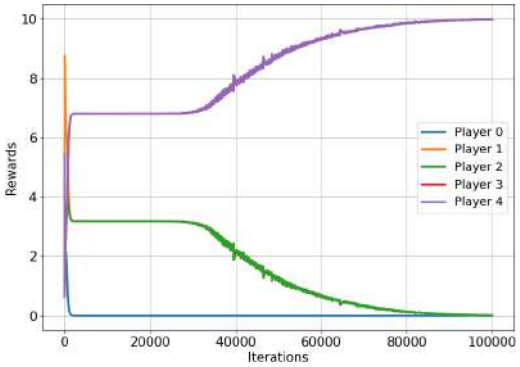
(C) Clásico $N = 4$. $\bar{R} = 0,0007$.



(D) Cuántico $N = 4$. $\bar{R} = 2,2451$.



(E) Clásico $N = 5$. $\bar{R} = 3,9999$.



(F) Cuántico $N = 5$. $\bar{R} = 3,9999$.

FIGURA 4.10: Aprendizaje de juegos de las minorías clásicos versus cuántico para 3, 4 y 5 jugadores. (a): Recompensa promedio $= 3,3333$. (b): Recompensa promedio $= 3,3333$. (c): Recompensa promedio $= 0,0007$. (d): Recompensa promedio $= 2,2451$. (e): Recompensa promedio $= 3,9999$. (f): Recompensa promedio $= 3,9999$.

Ahora dirijamos nuestra atención a la Figura 4.11, que muestra la evolución de las recompensas de los agentes en el juego de Platonia tanto para configuraciones clásicas (izquierda) como cuánticas (derecha), para dos, tres, cuatro y cinco agentes. En este juego, se otorga una recompensa de $R_N = 10$ solo a un agente que seleccione la acción 1, si y solo si todos los demás agentes han elegido la opción 0, como se muestra en la Tabla 4.15.

Agente i \Otros (N-1) Agentes	Todos 0	Algún 1
0	0	0
1	10	0

CUADRO 4.15: Matriz de pagos para el Juego Platonia con N agentes.

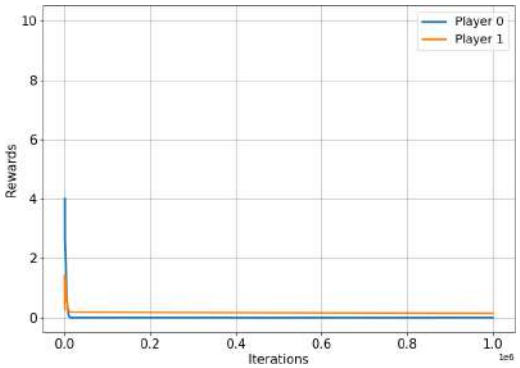
Los resultados de este juego son ciertamente más llamativos. Todos los agentes en el juego clásico convergen (independientemente del número de agentes, N) a su equilibrio de Nash clásico, resultando en una recompensa total de $R_N \approx 0$ para todos los agentes. A pesar de ser predecible, este resultado resultó en el peor rendimiento posible. En marcado contraste, la configuración cuántica muestra el poder del entrelazamiento, ya que los agentes logran consistentemente una recompensa total significativamente más alta, independientemente del número de agentes involucrados ($R_2 = 9{,}66$, $R_3 = 9{,}99$, $R_4 = 4{,}21$ y $R_5 = 9{,}58$). Este notable resultado resalta la eficacia del algoritmo y el potencial de las estrategias cuánticas en juegos multijugador complejos, que pueden ofrecer ventajas sobre sus contrapartes clásicas.

La Figura 4.12 muestra la evolución de las recompensas para los agentes tanto en los escenarios clásicos (izquierda) como cuánticos (derecha) del dilema de la cena. Este juego involucra a dos, tres, cuatro y cinco agentes y es una extensión de N agentes del conocido dilema del prisionero. El dilema de la cena modela una situación en la que un grupo, acordando dividir el costo total de una comida en partes iguales, debe elegir individualmente entre un plato menos caro o uno más caro. Aunque elegir el plato más caro puede parecer ventajoso para cada individuo, ya que si eligen el plato más barato terminarán pagando partes de los platos caros de otros comensales, esta decisión colectiva finalmente conduce a un resultado financiero general menos óptimo (todos pagando un precio muy alto) [133]. En la Tabla 4.16, presentamos una matriz de pagos para la versión de dos agentes del dilema de la cena, que también sirve como representación del renombrado dilema del prisionero.

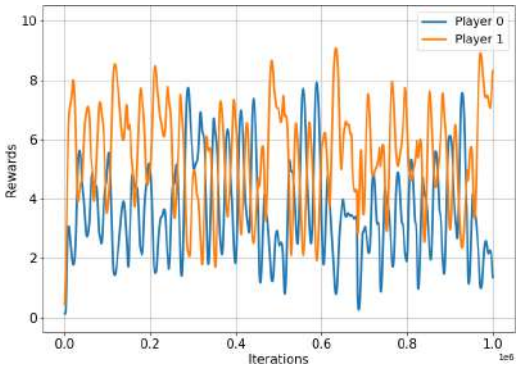
A \B	0	1
0	(6.66; 6.66)	(0.00; 10.0)
1	(10.0; 0.00)	(3.33; 3.33)

CUADRO 4.16: Matriz de pagos para el Juego sin escrúpulos con 2 jugadores o Dilema del prisionero.

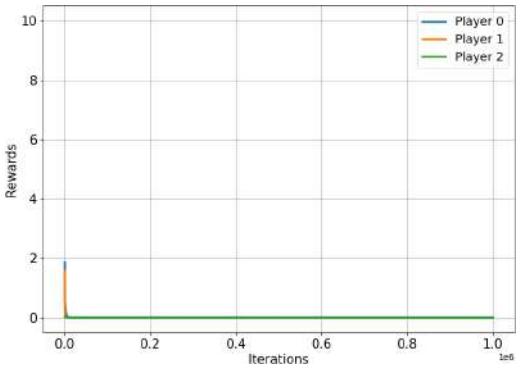
Como se observa en la Figura 4.12, las recompensas de los agentes clásicos siempre (para cada N) convergen rápidamente al valor de $R_N \approx 3{,}33$. Este valor representa el equilibrio de Nash clásico esperado y corresponde a la acción egoísta que produce la recompensa mínima cuando es elegida por todos los agentes. Sin embargo, el caso cuántico presenta otro contraste alentador. En el caso de dos agentes, ambas recompensas convergen a un valor de $R_2 \approx 6{,}66$, correspondiente al valor ideal asociado con la cooperación de ambos agentes. Por otro lado, para tres, cuatro y cinco agentes, las recompensas de los agentes cuánticos no convergen completamente; en cambio, fluctúan en torno a alcanzar los valores finales de



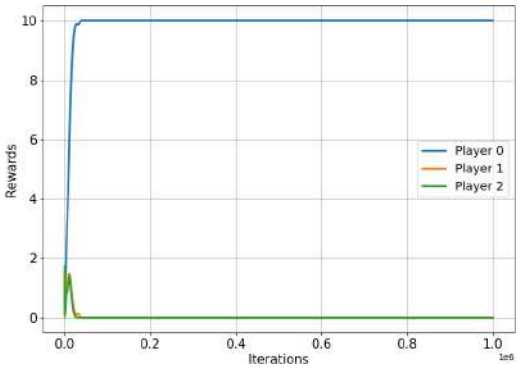
(A) Clásico $N = 2$. $\bar{R} = 0,1445$.



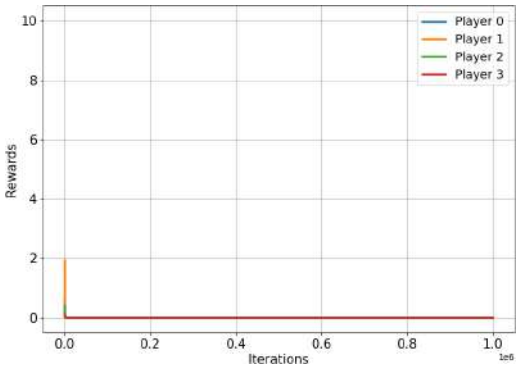
(B) Cuántico $N = 2$. $\bar{R} = 9,6604$.



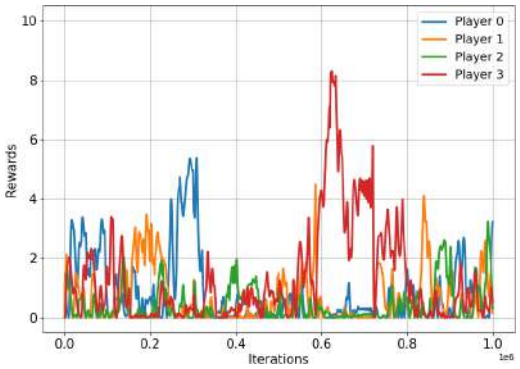
(C) Clásico $N = 3$. $\bar{R} = 0,0090$.



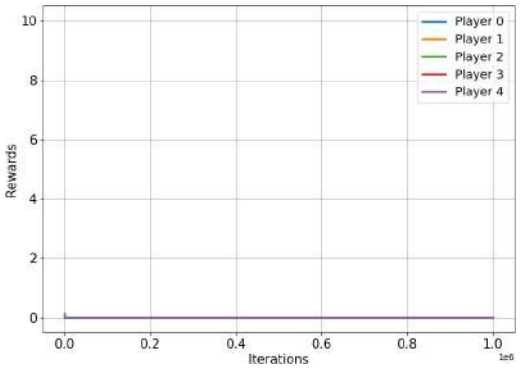
(D) Cuántico $N = 3$. $\bar{R} = 9,9999$.



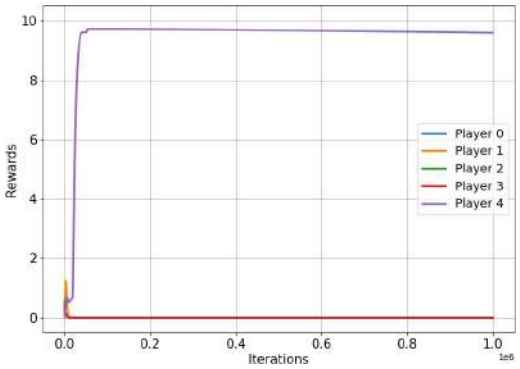
(E) Clásico $N = 4$. $\bar{R} = 0,0008$.



(F) Cuántico $N = 4$. $\bar{R} = 4,2078$.

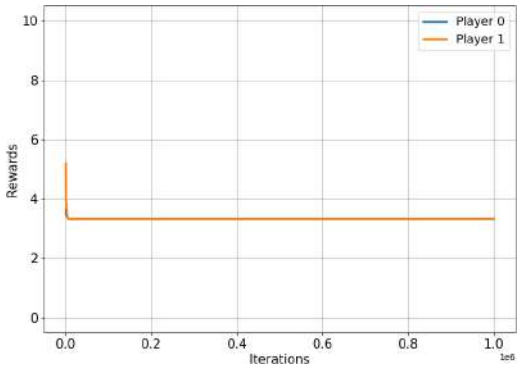


(G) Clásico $N = 5$. $\bar{R} = 0,0001$.

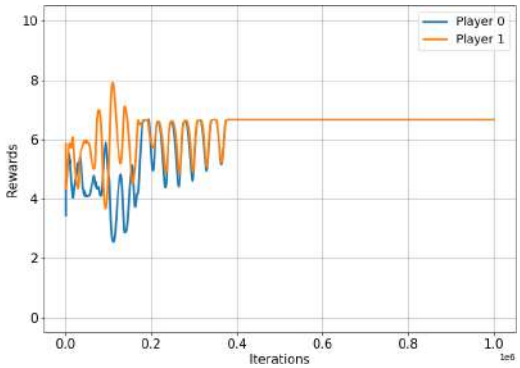


(H) Cuántico $N = 5$. $\bar{R} = 9,5895$.

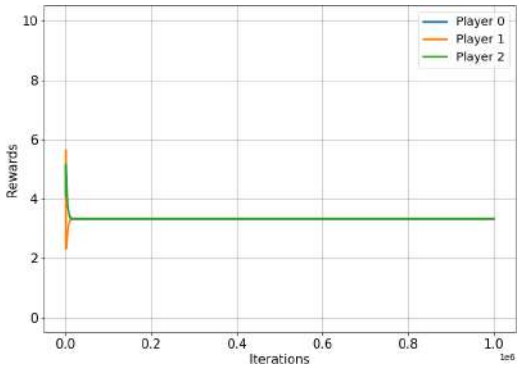
FIGURA 4.11: Aprendizaje del Juego Platonia clásico versus cuántico para 2, 3, 4 y 5 jugadores. (a): Recompensa total $= 0,1445$. (b) Recompensa total $= 9,6604$. (c) Recompensa total $= 0,0090$. (d) Recompensa total $= 9,9999$. (e) Recompensa total $= 0,0008$. (f) Recompensa total $= 4,2078$. (g) Recompensa total $= 0,0001$. (h) Recompensa total $= 9,5895$.



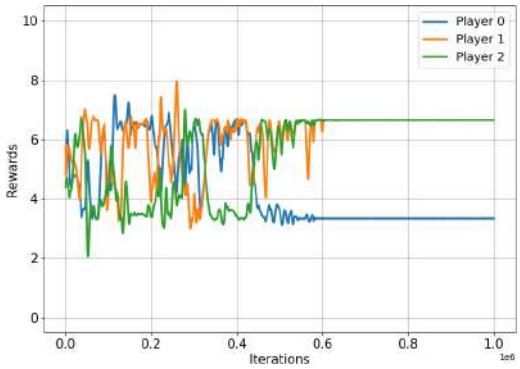
(A) Clásico $N = 2$. $\bar{R} = 3{,}3333$.



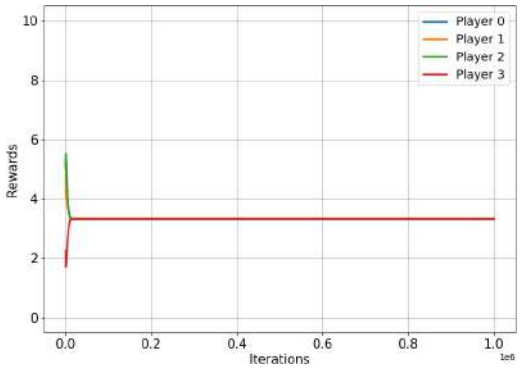
(B) Cuántico $N = 2$. $\bar{R} = 6{,}6666$.



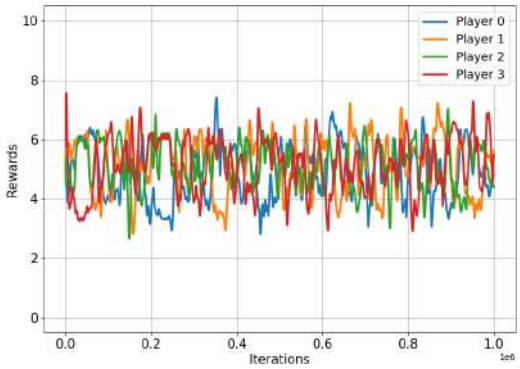
(C) Clásico $N = 3$. $\bar{R} = 3{,}3333$.



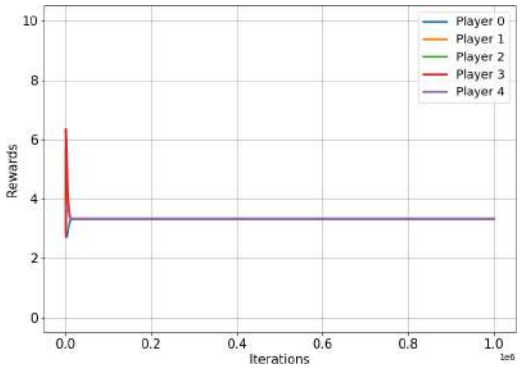
(D) Cuántico $N = 3$. $\bar{R} = 5{,}5525$.



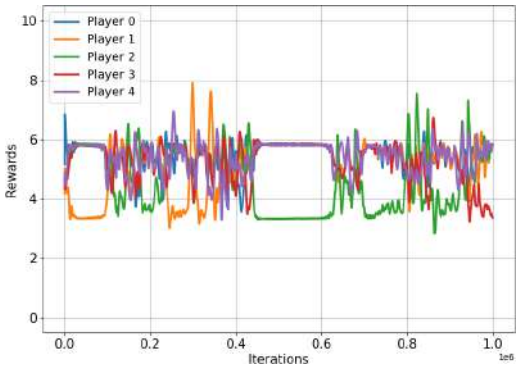
(E) Clásico $N = 4$. $\bar{R} = 3{,}3333$.



(F) Cuántico $N = 4$. $\bar{R} = 5{,}0717$.



(G) Clásico $N = 5$. $\bar{R} = 3{,}3333$.



(H) Cuántico $N = 5$. $\bar{R} = 5{,}3244$.

FIGURA 4.12: Aprendizaje del juegos sin escrúpulos clásico versus cuántico para 2, 3, 4 y 5 jugadores. (a) Recompensa promedio = 3,3333. (b) Recompensa promedio = 6,6666. (c) Recompensa promedio = 3,3333. (d) Recompensa promedio = 5,5525. (e) Recompensa promedio = 3,3333. (f) Recompensa promedio = 5,0717. (g) Recompensa promedio = 3,3333. (h) Recompensa promedio = 5,3244.

$R_N = [R_3, R_4, R_5] = [5{,}55, 5{,}07, 5{,}32]$. Aunque este valor no es el más alto posible, es significativamente mayor que los $R_N \approx 3{,}33$ obtenidos en el caso clásico.

Finalmente, nos enfocamos en el juego del voluntario, como se muestra en la Figura 4.13. El juego del voluntario es una extensión de N agentes del juego del gallina, que encarna escenarios en los que todo un grupo se beneficia de un sacrificio menor de al menos un individuo, pero todos sufren si nadie se sacrifica [134]. La matriz de pagos para el juego del dilema del voluntario de dos agentes se ilustra en la Tabla 4.17, representando al mismo tiempo el famoso juego del gallina.

A \B	0	1
0	(0; 0)	(-1, 1)
1	(1; -1)	(-10; -10)

CUADRO 4.17: Matriz de pagos para el Dilema del voluntario con 2 jugadores o Juego de la gallina.

En este escenario final, observamos que los comportamientos de ambos casos, el clásico y el cuántico, son casi idénticos para cada N. Las recompensas convergen de manera consistente al valor esperado del equilibrio de Nash clásico $R_N = 1/N$ en ambos escenarios. Este resultado confirma que las ventajas cuánticas no siempre pueden estar presentes y que nuestro algoritmo propuesto es una herramienta valiosa para verificar de manera eficiente y sistemática estos comportamientos.

Efectos del ruido

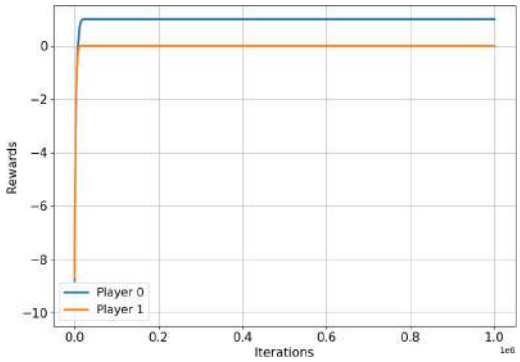
Esta sección examina cómo el ruido inherente en las computadoras cuánticas podría afectar el rendimiento del algoritmo de aprendizaje. Este análisis es un primer paso para comprender las implicaciones prácticas de implementar tales algoritmos en entornos computacionales cuánticos reales, donde el ruido es un factor inevitable.

En la sección anterior, asumimos que, después de la aplicación del operador J, todos los agentes podrían aplicar sus puertas a sus qubits en un canal ideal, sin ningún tipo de ruido que afecte la vulnerabilidad del sistema cuántico. Dada la dificultad de cumplir con tal condición, se hace necesario emprender un enfoque de modelado para tener en cuenta esta situación con precisión.

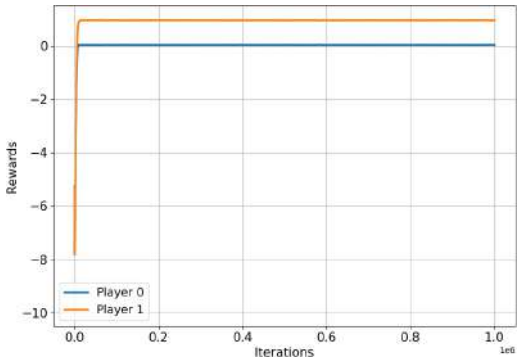
El modelo de ruido elegido para nuestro estudio es el canal de depolarización, principalmente porque modela efectivamente tanto los errores de bit-flip como los de phase-flip. Si bien el modelo de ruido del canal de depolarización captura las características esenciales de numerosos procesos de ruido en el mundo real, es un modelo simplificado y que no representa ningún dispositivo cuánticos específicos. El estado del sistema cuántico de un qubit después de este ruido es $\varepsilon(\rho) = (1-\lambda)\rho + \frac{\lambda I}{2} = (1-\lambda)\rho + \frac{\lambda}{3}(X\rho X + Y\rho Y + Z\rho Z)$, con $\rho = |\psi\rangle\langle\psi|$ siendo la matriz de densidad del estado cuántico antes de que se aplique el ruido.

La manera de modelar esto es agregando una cuarta puerta después de $R_X(\varphi_1)$, $R_Y(\varphi_2)$ y $R_X(\varphi_3)$; esta compuerta se seleccionará al azar siguiendo las probabilidades, tal como se explicó en la ecuación 4.3 y en la figura 4.4. Esto significa que el estado cuántico en cada canal permanecerá intacto con una probabilidad de $(1-\lambda)$ y será modificado con una probabilidad de λ. Para modificar el estado cuántico del canal, las puertas X, Y o Z se aplicarán con la misma probabilidad.

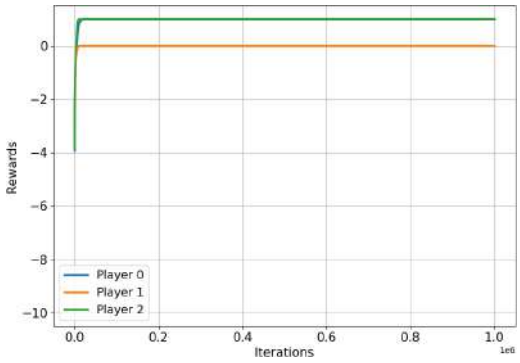
Para este análisis, nuestro enfoque estará en el juego de Platonia en su variante cuántica ($\gamma = \frac{\pi}{2}$), ya que este juego manifiesta la ventaja más significativa al contrastar los modelos clásicos y cuánticos. Las Figuras 4.14 y 4.15 ilustran la variación en la recompensa promedio de los agentes en función del ruido cuántico (λ y $\log(\lambda)$,



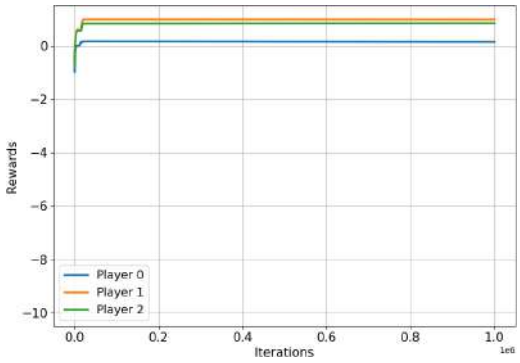
(A) Clásico $N = 2$. $\bar{R} = 0{,}4999$.



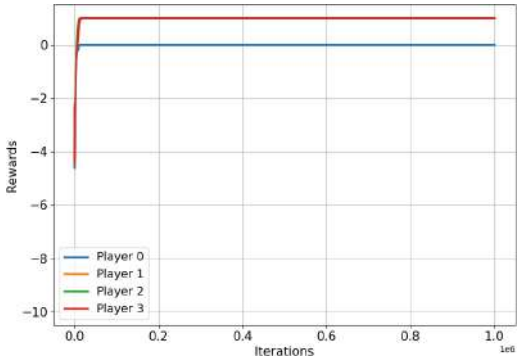
(B) Cuántico $N = 2$. $\bar{R} = 0{,}4999$.



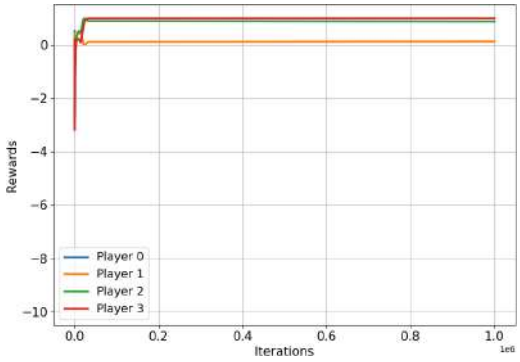
(C) Clásico $N = 3$. $\bar{R} = 0{,}6666$.



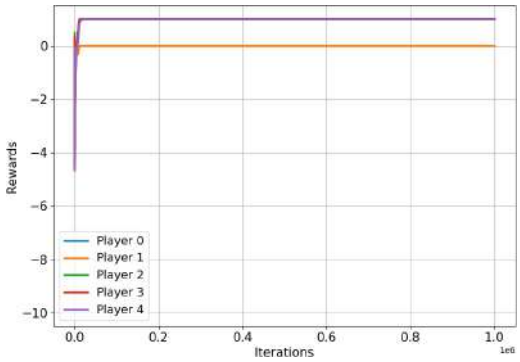
(D) Cuántico $N = 3$. $\bar{R} = 0{,}6606$.



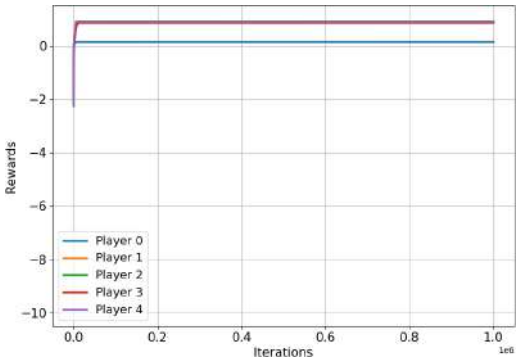
(E) Clásico $N = 4$. $\bar{R} = 0{,}7499$.



(F) Cuántico $N = 4$. $\bar{R} = 0{,}7164$.



(G) Clásico $N = 5$. $\bar{R} = 0{,}7999$.



(H) Cuántico $N = 5$. $\bar{R} = 0{,}7390$.

FIGURA 4.13: Aprendizaje del Juego del Voluntario clásico versus cuántico para 2, 3, 4 y 5 jugadores. (a) Recompensa promedio = 0,4999. (b) Recompensa promedio = 0,4999. (c) Recompensa promedio = 0,6666. (d) Recompensa promedio = 0,6606. (e) Recompensa promedio = 0,7499. (f) Recompensa promedio = 0,7164. (g) Recompensa promedio = 0,7999. (h) Recompensa promedio = 0,7390.

respectivamente, y la línea roja punteada representa el sistema sin ruido, $\lambda = 0$) inherente en el circuito cuántico del modelo EWL para dos, tres, cuatro, cinco, seis y siete agentes.

Por un lado, para los escenarios que involucran dos, tres, cuatro y cinco agentes, se observa un resultado esperado: a medida que aumenta el ruido en el circuito cuántico, el rendimiento de los agentes se deteriora. Esto es previsible ya que la retroalimentación recibida por los agentes para actualizar sus estrategias, utilizando el algoritmo propuesto, contiene progresivamente más ruido a medida que aumenta λ. En consecuencia, la eficacia de aprendizaje de los agentes disminuye con el aumento de λ, lo que lleva a una disminución en su rendimiento general. Todos los agentes alcanzan una recompensa promedio máxima cuando $\lambda = 0$, con el caso ideal de no tener ruido.

Por otro lado, se observa un comportamiento particular cuando el número de agentes en el juego cuántico aumenta a seis y siete agentes. Se observa que una leve presencia de ruido en el circuito cuántico ofrece ventajas al algoritmo en términos de maximización de la recompensa promedio de los agentes. En ambos casos, los juegos que involucran a seis y siete agentes producen la mayor recompensa promedio para un valor no nulo de λ ($\lambda = 0{,}0078125$ y $\lambda = 0{,}00390625$, respectivamente).

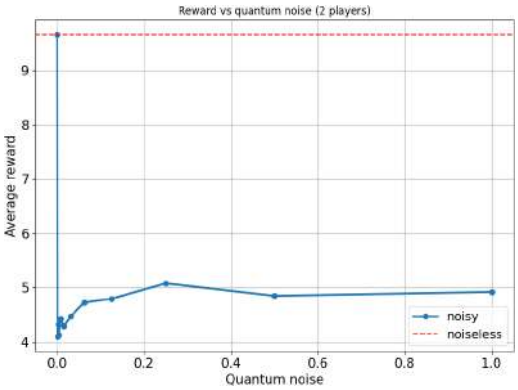
Es importante destacar que podría haber un caso trivial en el que el rendimiento de los agentes mejore de manera no decreciente a medida que aumenta la cantidad de ruido en el circuito cuántico. Es decir, en varios juegos, un agente que actúa completamente al azar (máximo ruido) logra una recompensa mayor que un agente racional (sin ruido). Sin embargo, tal comportamiento no sería particularmente notable. Por ejemplo, esto se observa en el dilema clásico del prisionero (Tabla 4.16), donde el comportamiento racional converge en una recompensa de $R = 3{,}33$, mientras que un comportamiento completamente aleatorio produce una recompensa de $R = \frac{0+3{,}33+6{,}66+10}{4} = 4$.

En nuestro estudio, una pequeña cantidad de ruido es beneficiosa, pero el ruido excesivo sigue siendo desventajoso. Este comportamiento no trivial es relevante ya que cualquier sistema cuántico no aislado tendrá ruido intrínseco que podría ser aprovechado potencialmente. Esto es particularmente notable en el contexto actual, donde todas las computadoras cuánticas contemporáneas inevitablemente poseen un nivel significativo de ruido. Por lo tanto, descubrir aplicaciones donde dicho ruido pueda ser aprovechado para mejorar el rendimiento del sistema es de suma relevancia.

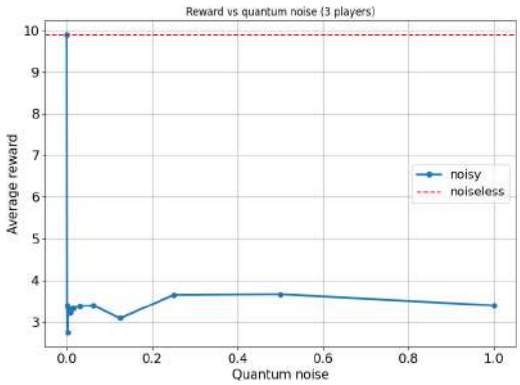
Una explicación plausible para este fenómeno podría derivarse del hecho de que cuando aumenta el número de agentes, la función de recompensa se vuelve más compleja y, si el algoritmo opera en un entorno sin ruido, puede quedar atrapado en máximos locales. Sin embargo, una cantidad minúscula de ruido en los circuitos cuánticos podría ser suficiente para sacar el sistema de estos máximos locales, permitiendo la exploración de otras acciones con recompensas más altas. Es importante que el nivel de ruido sea lo suficientemente bajo para no dañar el proceso de aprendizaje general y, por lo tanto, el rendimiento del sistema.

Conclusión

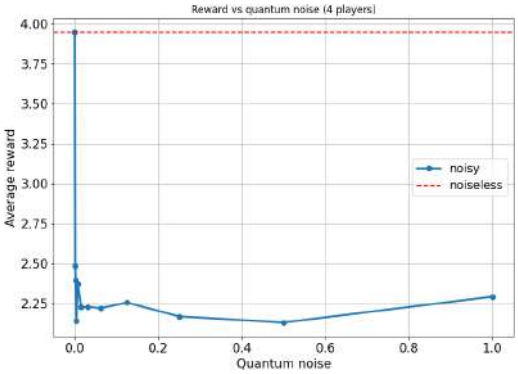
Esta investigación ha explorado las dinámicas intrincadas de los juegos cuánticos que involucran múltiples agentes, arrojando luz sobre las complejas relaciones entre el ruido cuántico, el rendimiento del sistema y la eficacia del aprendizaje. Se ha demostrado que las estrategias basadas en gradientes en un entorno multiagente



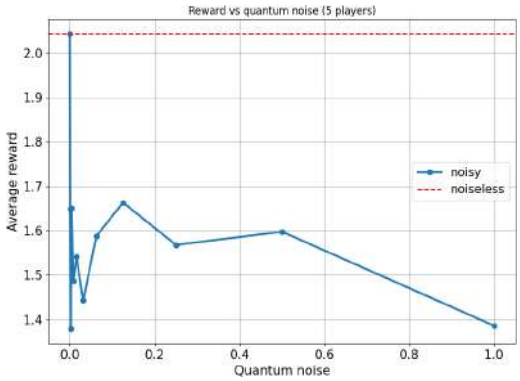
(A) Cuántico N = 2.



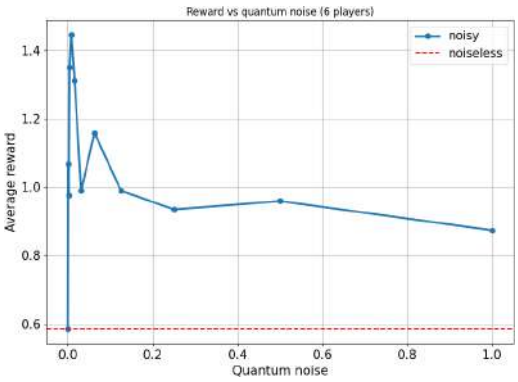
(B) Cuántico N = 3.



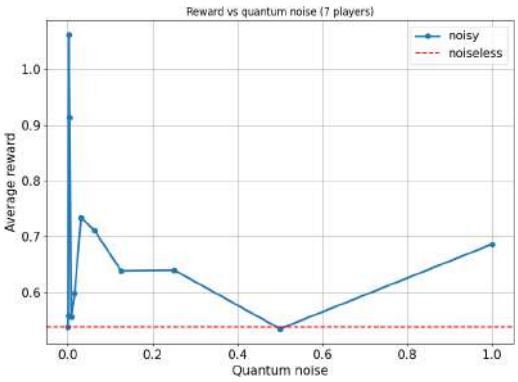
(C) Cuántico N = 4.



(D) Cuántico N = 5.

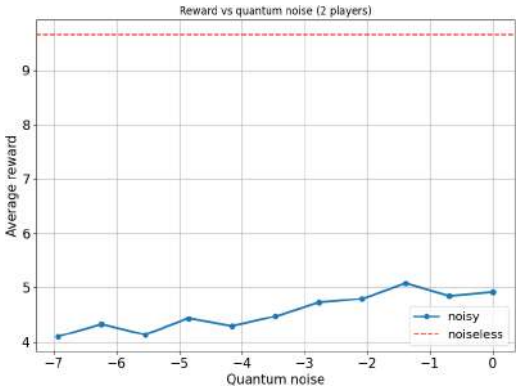


(E) Cuántico N = 6.

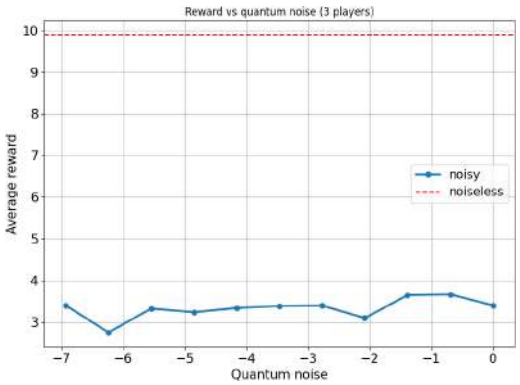


(F) Cuántico N = 7.

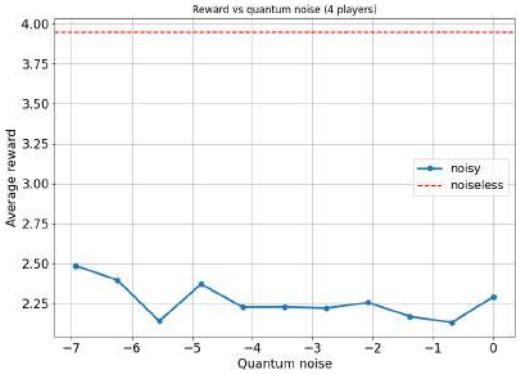
FIGURA 4.14: Recompensa promedio del juego Platonia de N jugadores versus ruido cuántico representada como λ (para $\lambda = [0, \frac{1}{1024}, \frac{1}{512}, \frac{1}{256}, \frac{1}{128}, \frac{1}{64}, \frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1]$). (a) Recompensa promedio máxima = 9.882 por ruido cuántico para λ = 0. (b) Recompensa promedio máxima = 9.971 para ruido cuántico para λ = 0. (c) Recompensa promedio máxima = 4.319 para ruido cuántico para λ = 0. (d) Recompensa promedio máxima = 6.451 para ruido cuántico para λ = 0. (e) Recompensa promedio máxima = 2.221 para ruido cuántico para λ = 0.0078125. (f) Recompensa promedio máxima = 1.686 para ruido cuántico por λ = 0.00390625.



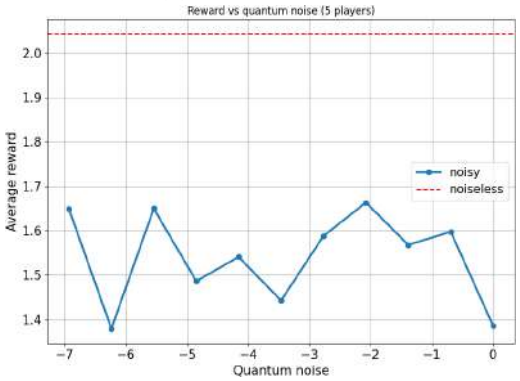
(A) Cuántico N = 2.



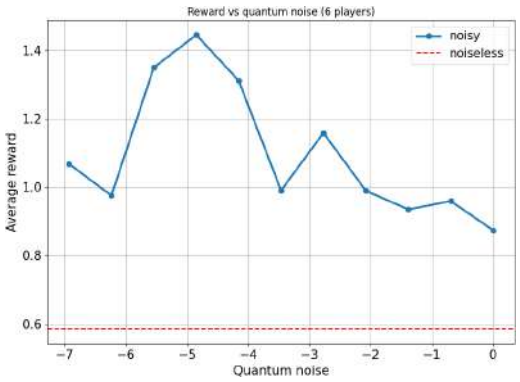
(B) Cuántico N = 3.



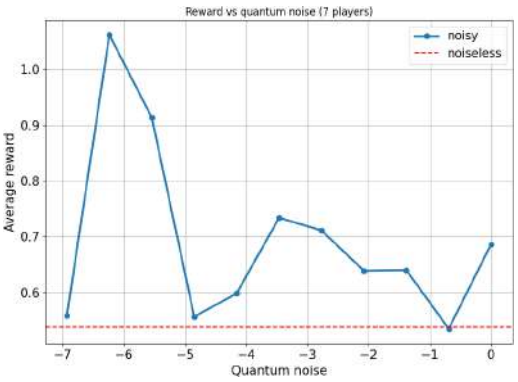
(C) Cuántico N = 4.



(D) Cuántico N = 5.



(E) Cuántico N = 6.



(F) Cuántico N = 7.

FIGURA 4.15: Recompensa promedio del juego Platonia de N jugadores versus ruido cuántico representada como $\log(\lambda)$ (para $\lambda = [0, \frac{1}{1024}, \frac{1}{512}, \frac{1}{256}, \frac{1}{128}, \frac{1}{64}, \frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1]$). (a) Recompensa promedio máxima = 9.882 por ruido cuántico para $\lambda = 0$. (b) Recompensa promedio máxima = 9.971 para ruido cuántico para $\lambda = 0$. (c) Recompensa promedio máxima = 4.319 para ruido cuántico para $\lambda = 0$. (d) Recompensa promedio máxima = 6.451 para ruido cuántico para $\lambda = 0$. (e) Recompensa promedio máxima = 2.221 para ruido cuántico para $\lambda = 0.0078125$. (f) Recompensa promedio máxima = 1.686 para ruido cuántico por $\lambda = 0.00390625$.

presentan características intrigantes cuando se toman en cuenta la mecánica cuántica.

El algoritmo de aprendizaje cuántico propuesto ha mostrado un rendimiento robusto en diversas configuraciones de juego. Además, la investigación de los efectos del ruido cuántico reveló un comportamiento único que podría tener implicaciones de gran alcance. Como era de esperar, el aumento del ruido causó que el rendimiento de los agentes en juegos con pocos participantes se deteriorara. Sin embargo, sorprendentemente, una pequeña cantidad de ruido pareció ser ventajosa en juegos con seis y siete agentes, facilitando la salida de máximos locales y permitiendo la exploración de acciones más óptimas. Este descubrimiento estimula la discusión sobre el posible aprovechamiento del ruido del sistema, un desafío común en las computadoras cuánticas contemporáneas, para mejorar el rendimiento.

Las percepciones de este estudio tienen implicaciones sustanciales para la aplicación práctica de la computación cuántica. En el campo de la inteligencia artificial, donde las mejoras cuánticas podrían optimizar la exploración de soluciones más eficientes, o en las redes de comunicación, donde los principios de los juegos cuánticos podrían permitir el desarrollo de sistemas de comunicación más robustos. La exploración de juegos cuánticos con aprendizaje multiagente en esta investigación apenas ha arañado la superficie. Hay una gran cantidad de dinámicas complejas y fenómenos fascinantes por explorar en este campo emergente.

Capítulo 5

Conclusión

5.1. Trabajos Futuros

A partir de la experiencia adquirida en esta tesis doctoral, se plantean diversas direcciones futuras de investigación que pueden construirse sobre las bases de este trabajo. Entre ellas, se contempla la posibilidad de desarrollar un modelo más complejo para redes de comunicación que refleje con mayor precisión la realidad de la congestión, incorporando parámetros avanzados como diferentes topologías de red, capacidades de canal variadas y modelos de los medios físicos. En la misma dirección, se sugiere la exploración de la aplicación de algoritmos descentralizados de aprendizaje por refuerzo en el protocolo de ruteo de la sección 3.2, contrastando con el enfoque centralizado de la sección 3.3 que busca minimizar la latencia total de la red.

Otras áreas de interés incluyen la incorporación de algoritmos de corrección de errores en los agentes para mitigar las pérdidas de rendimiento causadas por el ruido cuántico, además de la caracterización del rendimiento de los agentes ante cambios en el juego que están jugando a lo largo de un proceso de aprendizaje prolongado. Finalmente, se propone la aplicación de algoritmos de aprendizaje cuántico para la optimización de estrategias en entornos cuánticos, una vía muy poco explorada que podría aprovechar las capacidades únicas de la computación cuántica para mejorar la eficiencia del aprendizaje en juegos cuánticos.

5.2. Conclusiones finales

Esta tesis doctoral ha navegado por la confluencia de la computación cuántica, la teoría de juegos y el aprendizaje por refuerzo, desvelando un panorama innovador para el modelado de redes de comunicación y sistemas multi-agente. Hemos demostrado cómo la integración de estas disciplinas permite superar desafíos de congestión en redes y mejora las estrategias de ruteo, extendiendo las capacidades más allá de los métodos clásicos. A través de la exploración de propiedades cuánticas como la superposición y el entrelazamiento, esta investigación ha ampliado el espacio estratégico, permitiendo optimizaciones que eran inaccesibles dentro del marco de juegos clásicos y el aprendizaje tradicional.

Los avances presentados en esta tesis destacan: a) la creación de un modelo basado en la teoría de juegos que representa el problema de la congestión y evidencia las ventajas de utilizar juegos cuánticos versus juegos clásicos, b) el diseño de un protocolo de ruteo basado en juegos cuánticos y el análisis de su rendimiento bajo condiciones ideales, entrelazamiento parcial y ruido cuántico, c) la incorporación de algoritmos de aprendizaje por refuerzo en el protocolo de ruteo para agregar la capacidad de autoadaptación en en tiempo real en función del estado de la red,

d) el desarrollo de un algoritmo de aprendizaje por refuerzo multi-agente descentralizado para juegos cuánticos con información imperfecta donde los agentes pueden aprender tanto estrategias puras como mixtas, e) la integración de la capacidad de distribuir equitativamente las recompensas entre los agentes en el algoritmo de aprendizaje multi-agente, incluso cuando los agentes intentan maximizar únicamente su propio rendimiento, f) el diseño de un algoritmo de aprendizaje para sistemas multi-agente que incorpora la estimación del gradiente de la función de recompensa para buscar las estrategias óptimas. El éxito de estos algoritmos en la mejora de la gestión de congestiones y en la optimización de las recompensas subraya el potencial de los enfoques cuánticos sobre los clásicos, marcando un hito en el desarrollo de redes de comunicación y sistemas inteligentes.

Este trabajo representa un avance significativo en la aplicación práctica de la teoría de juegos cuántica, abriendo nuevas vías para la optimización de las telecomunicaciones y el aprendizaje automático. Al desentrañar las complejidades de estas interacciones estratégicas en diversos dominios, hemos sentado las bases para futuras investigaciones en sistemas distribuidos autoadaptativos, donde los algoritmos cuánticos pueden jugar un papel crucial en la superación de desafíos contemporáneos y en el establecimiento de nuevos paradigmas tanto científicos como tecnológicos.

Bibliografía

[1] Agustin Silva y Claudio. Gonzalez. «SoC FPGA implementation of Hopfield Neural Network for solving the Shortest Path Problem». En: *Congreso Argentino de Sistemas Embebidos*. Vol. 11. http://www.sase.com.ar/case/ediciones/case2021/. 2021, págs. 100-103.

[2] Agustin Silva, Omar G Zabaleta y Claudio Gonzalez. «Aceleracion de simulación de circuitos cuánticos parametrizados en SoC FPGA». En: *Congreso Argentino de Sistemas Embebidos*. Vol. 12. http://www.sase.com.ar/case/ediciones/case-2022/. 2022, págs. 44-46.

[3] Agustin Silva, Omar G Zabaleta y Constancio M Arizmendi. «Quantum Game Theory approach for data network routing: a solution for the congestion problem». En: *Journal of Physics: Conference Series*. Vol. 2207. 1. IOP Publishing. 2022, pág. 012034.

[4] Agustin Silva, Omar G Zabaleta y Constancio M Arizmendi. «Mitigation of Routing Congestion on Data Networks: A Quantum Game Theory Approach». En: *Quantum Reports* 4.2 (2022), págs. 135-147.

[5] John Preskill. «Quantum Computing in the NISQ era and beyond». En: *Quantum* 2 (2018), pág. 79.

[6] Agustin Silva, Omar G Zabaleta y Constancio M Arizmendi. «Learning-based Protocol for Routing in Quantum Networks». En: *IFAC-PapersOnLine* 55.40 (2022), págs. 211-216.

[7] Agustin Silva, Omar G Zabaleta y Constancio M Arizmendi. «Learning Mixed Strategies in Quantum Games with Imperfect Information». En: *Quantum Reports* 4.4 (2022), págs. 462-475.

[8] Agustin Silva, Omar Gustavo Zabaleta y Constancio Miguel Arizmendi. «Maximizing Local Rewards on Multi-Agent Quantum Games through Gradient-Based Learning Strategies». En: *Entropy* 25.11 (2023), pág. 1484.

[9] A Cicuttin et al. «Looking for suitable rules for true random number generation with asynchronous cellular automata». En: *Nonlinear Dynamics* (2022), págs. 1-12.

[10] Werner Florian Samayoa et al. «HyperFPGA: An Experimental Testbed for Heterogeneous Supercomputing». En: (2023).

[11] Ethan Bernstein y Umesh Vazirani. «Quantum complexity theory». En: *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*. 1993, págs. 11-20.

[12] Eric W Weisstein. «Number field sieve». En: *https://mathworld. wolfram. com/* (2000).

[13] Michael A Nielsen e Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[14] John F Nash Jr. «Equilibrium points in n-person games». En: *Proceedings of the national academy of sciences* 36.1 (1950), págs. 48-49.

[15] Tim Roughgarden. «Algorithmic game theory». En: *Communications of the ACM* 53.7 (2010), págs. 78-86.

[16] David Easley, Jon Kleinberg et al. «Networks, crowds, and markets». En: *Cambridge Books* (2012).

[17] Amar Prakash Azad, Eitan Altman y Rachid El-Azouzi. «Routing games: From egoism to altruism». En: *8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*. IEEE. 2010, págs. 528-537.

[18] Richard J La y Venkat Anantharam. «Optimal routing control: Repeated game approach». En: *IEEE transactions on automatic control* 47.3 (2002), págs. 437-450.

[19] Zhu Han, Zhu Ji y KJ Ray Liu. «Dynamic distributed rate control for wireless networks by optimal cartel maintenance strategy». En: *IEEE Global Telecommunications Conference, 2004. GLOBECOM'04.* Vol. 6. IEEE. 2004, págs. 3454-3458.

[20] Fabio Milan, Juan José Jaramillo y R Srikant. «Achieving cooperation in multihop wireless networks of selfish nodes». En: *Proceeding from the 2006 workshop on Game theory for communications and networks*. 2006, 3-es.

[21] Allen B MacKenzie y Stephen B Wicker. «Stability of multipacket slotted aloha with selfish users and perfect information». En: *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*. Vol. 3. IEEE. 2003, págs. 1583-1590.

[22] Eitan Altman et al. «Non-cooperative forwarding in ad-hoc networks». En: *NETWORKING 2005. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems: 4th International IFIP-TC6 Networking Conference, Waterloo, Canada, May 2-6, 2005. Proceedings 4*. Springer. 2005, págs. 486-498.

[23] Márk Félegyházi, Levente Buttyán y Jean-Pierre Hubaux. «Equilibrium analysis of packet forwarding strategies in wireless ad hoc networks–the static case». En: *Personal Wireless Communications: IFIP-TC6 8th International Conference, PWC 2003, Venice, Italy, September 23-25, 2003. Proceedings 8*. Springer. 2003, págs. 776-789.

[24] Vikram Srinivasan et al. «Cooperation in wireless ad hoc networks». En: *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*. Vol. 2. IEEE. 2003, págs. 808-817.

[25] Pietro Michiardi y Refik Molva. «A game theoretical approach to evaluate cooperation enforcement mechanisms in mobile ad hoc networks». En: *WiOpt'03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*. 2003, 4-pages.

[26] Pietro Michiardi y Refik Molva. «Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks». En: *Advanced Communications and Multimedia Security: IFIP TC6/TC11 Sixth Joint Working Conference on Communications and Multimedia Security September 26–27, 2002, Portorož, Slovenia*. Springer. 2002, págs. 107-121.

[27] Sonja Buchegger y Jean-Yves Le Boudec. «Performance analysis of the CONFIDANT protocol». En: *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*. 2002, págs. 226-236.

[28] Eitan Altman, Yezekael Hayel e Hisao Kameda. «Evolutionary dynamics and potential games in non-cooperative routing». En: *2007 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops*. IEEE. 2007, págs. 1-5.

[29] Engin Zeydan et al. «Bottleneck throughput maximization for correlated data routing: a game theoretic approach». En: *2010 44th Annual Conference on Information Sciences and Systems (CISS)*. IEEE. 2010, págs. 1-6.

[30] S-K Ng y Winston Khoon Guan Seah. «Game-theoretic model for collaborative protocols in selfish, tariff-free, multihop wireless networks». En: *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*. IEEE. 2008, págs. 216-220.

[31] See-Kee Ng y Winston KG Seah. «Game-theoretic approach for improving cooperation in wireless multihop networks». En: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 40.3 (2010), págs. 559-574.

[32] Petteri Nurmi. «Modelling routing in wireless ad hoc networks with dynamic Bayesian games». En: *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.* IEEE. 2004, págs. 63-70.

[33] Zhu Han. *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge university press, 2012.

[34] Zhu Han et al. *Game theory for next generation wireless and communication networks: Modeling, analysis, and design*. Cambridge University Press, 2019.

[35] Richard S Sutton y Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[36] Ming Tan. «Multi-agent reinforcement learning: Independent vs. cooperative agents». En: *Proceedings of the tenth international conference on machine learning*. 1993, págs. 330-337.

[37] Caroline Claus y Craig Boutilier. «The dynamics of reinforcement learning in cooperative multiagent systems». En: *AAAI/IAAI* 1998.746-752 (1998), pág. 2.

[38] Eduardo Rodrigues Gomes y Ryszard Kowalczyk. «Dynamic analysis of multiagent Q-learning with ε-greedy exploration». En: *Proceedings of the 26th annual international conference on machine learning*. 2009, págs. 369-376.

[39] Michael Wunder, Michael L Littman y Monica Babes. «Classes of multiagent q-learning dynamics with epsilon-greedy exploration». En: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010, págs. 1167-1174.

[40] Ardeshir Kianercy y Aram Galstyan. «Dynamics of Boltzmann Q learning in two-player two-action games». En: *Physical Review E* 85.4 (2012), pág. 041145.

[41] Wolfram Barfuss, Jonathan F Donges y Jürgen Kurths. «Deterministic limit of temporal difference reinforcement learning for stochastic games». En: *Physical Review E* 99.4 (2019), pág. 043305.

[42] Shuyue Hu, Chin-wing Leung y Ho-fung Leung. «Modelling the dynamics of multiagent q-learning in repeated symmetric games: a mean field theoretic approach». En: *Advances in Neural Information Processing Systems* 32 (2019).

[43] Stefanos Leonardos y Georgios Piliouras. «Exploration-exploitation in multiagent learning: Catastrophe theory meets game theory». En: *Artificial Intelligence* 304 (2022), pág. 103653.

[44] Georgios Papoudakis et al. «Benchmarking multi-agent deep reinforcement learning algorithms in cooperative tasks». En: *arXiv preprint arXiv:2006.07869* (2020).

[45] Stefano V Albrecht, Filippos Christianos y Lukas Schäfer. «Multi-agent reinforcement learning: Foundations and modern approaches». En: *Massachusetts Institute of Technology: Cambridge, MA, USA* (2023).

[46] John Von Neumann y Oskar Morgenstern. «Theory of games and economic behavior, 2nd rev». En: (1947).

[47] David A Meyer. «Quantum strategies». En: *Physical Review Letters* 82.5 (1999), pág. 1052.

[48] Kazuki Ikeda y Shoto Aoki. «Theory of quantum games and quantum economic behavior». En: *Quantum Information Processing* 21.1 (2022), pág. 27.

[49] Edwin Ho et al. «Game Theory in defence applications: A review». En: *Sensors* 22.3 (2022), pág. 1032.

[50] Magda M Madbouly, Yasser F Mokhtar y Saad M Darwish. «Quantum game application to recovery problem in mobile database». En: *Symmetry* 13.11 (2021), pág. 1984.

[51] Jens Eisert, Martin Wilkens y Maciej Lewenstein. «Quantum games and quantum strategies». En: *Physical Review Letters* 83.15 (1999), pág. 3077.

[52] Simon C Benjamin y Patrick M Hayden. «Comment on "Quantum Games and Quantum Strategies"». En: *Physical Review Letters* 87.6 (2001), pág. 069801.

[53] Jens Eisert y Martin Wilkens. «Quantum games». En: *Journal of Modern Optics* 47.14-15 (2000), págs. 2543-2556.

[54] Luca Marinatto y Tullio Weber. «A quantum approach to static games of complete information». En: *Physics Letters A* 272.5-6 (2000), págs. 291-303.

[55] Simon C Benjamin y Patrick M Hayden. «Multiplayer quantum games». En: *Physical Review A* 64.3 (2001), pág. 030301.

[56] Neil F Johnson. «Playing a quantum game with a corrupted source». En: *Physical Review A* 63.2 (2001), pág. 020302.

[57] A Iqbal y AH Toor. «Evolutionarily stable strategies in quantum games». En: *Physics Letters A* 280.5-6 (2001), págs. 249-256.

[58] Adrian P Flitney y Derek Abbott. «Quantum version of the Monty Hall problem». En: *Physical Review A* 65.6 (2002), pág. 062318.

[59] A Iqbal y AH Toor. «Quantum mechanics gives stability to a Nash equilibrium». En: *Physical Review A* 65.2 (2002), pág. 022306.

[60] Jiangfeng Du et al. «Entanglement enhanced multiplayer quantum games». En: *Physics Letters A* 302.5-6 (2002), págs. 229-233.

[61] Adrian P Flitney y Derek Abbott. «Quantum games with decoherence». En: *Journal of Physics A: Mathematical and General* 38.2 (2004), pág. 449.

[62] Azhar Iqbal y Stefan Weigert. «Quantum correlation games». En: *Journal of Physics A: Mathematical and General* 37.22 (2004), pág. 5873.

[63] Jing-Ling Chen, Leong Chuan Kwek y Choo Hiap Oh. «Noisy quantum game». En: *Physical Review A* 65.5 (2002), pág. 052320.

[64] Hong Guo, Juheng Zhang y Gary J Koehler. «A survey of quantum games». En: *Decision Support Systems* 46.1 (2008), págs. 318-332.

[65] Faisal Shah Khan et al. «Quantum games: a review of the history, current state, and interpretation». En: *Quantum Information Processing* 17 (2018), págs. 1-42.

[66] Junichi Shimamura et al. «Quantum and classical correlations between players in game theory». En: *International Journal of Quantum Information* 2.01 (2004), págs. 79-89.

[67] P Zhang et al. «Optical realization of quantum gambling machine». En: *Europhysics Letters* 82.3 (2008), pág. 30002.

[68] WF Balthazar et al. «Experimental realization of the quantum duel game using linear optical circuits». En: *Journal of Physics B: Atomic, Molecular and Optical Physics* 48.16 (2015), pág. 165505.

[69] ARC Pinheiro et al. «Vector vortex implementation of a quantum game». En: *JOSA B* 30.12 (2013), págs. 3210-3214.

[70] Robert Prevedel et al. «Experimental realization of a quantum game on a one-way quantum computer». En: *New Journal of Physics* 9.6 (2007), pág. 205.

[71] JB Altepeter et al. «Experimental realization of a multi-player quantum game». En: *Nonlinear Optics: Materials, Fundamentals and Applications*. Optica Publishing Group. 2009, PDNTuA2.

[72] Christian Schmid et al. «Experimental implementation of a four-player quantum game». En: *New Journal of Physics* 12.6 (2010), pág. 063031.

[73] WM Reisman. «Arms and Influence. By Thomas C. Schelling. New Haven and London: Yale university Press, 1966. pp. ix, 293. Index. 7,50.». En: *American Journal of International Law* 61.2 (1967), págs. 625-626.

[74] Omar Gustavo Zabaleta, Juan Pablo Barrangú y Constancio M Arizmendi. «Quantum game application to spectrum scarcity problems». En: *Physica A: Statistical Mechanics and its Applications* 466 (2017), págs. 455-461.

[75] OG Zabaleta y CM Arizmendi. «Evolutionary quantum minority game: A wireless network application». En: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.7 (2018).

[76] Damien Challet e Y-C Zhang. «Emergence of cooperation and organization in an evolutionary game». En: *Physica A: Statistical Mechanics and its Applications* 246.3-4 (1997), págs. 407-418.

[77] Adrian P Flitney y Lloyd CL Hollenberg. «Multiplayer quantum minority game with decoherence». En: *Quantum Information & Computation* 7.1 (2007), págs. 111-126.

[78] Neal Solmeyer, Ricky Dixon y Radhakrishnan Balu. «Quantum routing games». En: *Journal of Physics A: Mathematical and Theoretical* 51.45 (2018), pág. 455304.

[79] Tim Roughgarden. «On the severity of Braess's paradox: Designing networks for selfish users is hard». En: *Journal of Computer and System Sciences* 72.5 (2006), págs. 922-953.

[80] Indrakshi Dey et al. «Quantum Game Theory meets Quantum Networks». En: *arXiv preprint arXiv:2306.08928* (2023).

[81] Shantom Kumar Borah et al. «Analysis of Adversarial Jamming From a Quantum Game Theoretic Perspective». En: *IEEE Systems Journal* 17.1 (2022), págs. 881-891.

[82] Faisal Shah Khan y Ning Bao. «Quantum Prisoner's Dilemma and high frequency trading on the quantum cloud». En: *Frontiers in Artificial Intelligence* 4 (2021), pág. 769392.

[83] Kyriakos Lotidis, Panayotis Mertikopoulos y Nicholas Bambos. «Learning in quantum games». En: *arXiv preprint arXiv:2302.02333* (2023).

[84] Lei Cui et al. «Coherent feedback control for linear quantum systems and two-strategy evolutionary game theory». En: (2017).

[85] Pan-Yang Su. «Analyzing Quantum Cryptography and Communication based on Quantum Game Theory». En: *Quantum* 2 (), pág. 14.

[86] Olivier Ezratty. «Perspective on superconducting qubit quantum computing». En: *The European Physical Journal A* 59.5 (2023), pág. 94.

[87] Sergey Bravyi et al. «The future of quantum computing with superconducting qubits». En: *Journal of Applied Physics* 132.16 (2022).

[88] John Bardeen, Leon N Cooper y J Robert Schrieffer. «Microscopic theory of superconductivity». En: *Physical Review* 106.1 (1957), pág. 162.

[89] John Bardeen, Leon N Cooper y John Robert Schrieffer. «Theory of superconductivity». En: *Physical review* 108.5 (1957), pág. 1175.

[90] Philip W Anderson y John M Rowell. «Probable observation of the Josephson superconducting tunneling effect». En: *Physical Review Letters* 10.6 (1963), pág. 230.

[91] John M Martinis, Michel H Devoret y John Clarke. «Energy-level quantization in the zero-voltage state of a current-biased Josephson junction». En: *Physical review letters* 55.15 (1985), pág. 1543.

[92] Philip Krantz et al. «A quantum engineer's guide to superconducting qubits». En: *Applied physics reviews* 6.2 (2019).

[93] Hans Mooij. «Superconducting quantum bits». En: *Physics world* 17.12 (2004), pág. 29.

[94] Agustin Di Paolo et al. «Extensible circuit-QED architecture via amplitude- and frequency-variable microwaves». En: *arXiv preprint arXiv:2204.08098* (2022).

[95] Jonas Larson y Th K Mavrogordatos. «The Jaynes-Cummings model and its descendants». En: *arXiv preprint arXiv:2202.00330* (2022).

[96] Xiu Gu et al. «Microwave photonics with superconducting quantum circuits». En: *Physics Reports* 718 (2017), págs. 1-102.

[97] Yilun Xu et al. «Radio frequency mixing modules for superconducting qubit room temperature control systems». En: *Review of Scientific Instruments* 92.7 (2021).

[98] Sunmi Kim et al. «Enhanced coherence of all-nitride superconducting qubits epitaxially grown on silicon substrate». En: *Communications Materials* 2.1 (2021), pág. 98.

[99] Abhinandan Antony et al. «Miniaturizing transmon qubits using van der Waals materials». En: *Nano letters* 21.23 (2021), págs. 10122-10126.

[100] Julia Zotova et al. «Compact Superconducting Microwave Resonators Based on Al-AlO x-Al Capacitors». En: *Physical Review Applied* 19.4 (2023), pág. 044067.

[101] Rebecca Hicks et al. «Active readout-error mitigation». En: *Physical Review A* 105.1 (2022), pág. 012419.

[102] Martin Beisel et al. «Configurable readout error mitigation in quantum workflows». En: *Electronics* 11.19 (2022), pág. 2983.

[103] Jay M Gambetta et al. «Investigating surface loss effects in superconducting transmon qubits». En: *IEEE Transactions on Applied Superconductivity* 27.1 (2016), págs. 1-5.

[104] A Romanenko et al. «Three-dimensional superconducting resonators at T< 20 mK with photon lifetimes up to τ= 2 s». En: *Physical Review Applied* 13.3 (2020), pág. 034032.

[105] Alexander Opremcak et al. «High-fidelity measurement of a superconducting qubit using an on-chip microwave photon counter». En: *Physical Review X* 11.1 (2021), pág. 011027.

[106] Jules Tilly et al. «The variational quantum eigensolver: a review of methods and best practices». En: *Physics Reports* 986 (2022), págs. 1-128.

[107] Leo Zhou et al. «Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices». En: *Physical Review X* 10.2 (2020), pág. 021067.

[108] Maria Schuld y Francesco Petruccione. *Machine learning with quantum computers*. Springer, 2021.

[109] Olivier Ezratty. «Understanding Quantum Technologies 2023». En: *arXiv preprint arXiv:2111.15352* (2023).

[110] Marco Fellous-Asiani et al. «Optimizing resource efficiencies for scalable full-stack quantum computers». En: *PRX Quantum* 4.4 (2023), pág. 040319.

[111] Alexia Auffeves. «Quantum technologies need a quantum energy initiative». En: *PRX Quantum* 3.2 (2022), pág. 020101.

[112] David Awschalom. *From Long-distance Entanglement to Building a Nationwide Quantum Internet: Report of the DOE Quantum Internet Blueprint Workshop*. Inf. téc. Brookhaven National Lab.(BNL), Upton, NY (United States), 2020.

[113] SLN Hermans et al. «Qubit teleportation between non-neighbouring nodes in a quantum network». En: *Nature* 605.7911 (2022), págs. 663-668.

[114] Yu-Ao Chen et al. «An integrated space-to-ground quantum communication network over 4,600 kilometres». En: *Nature* 589.7841 (2021), págs. 214-219.

[115] Stephanie Wehner, David Elkouss y Ronald Hanson. «Quantum internet: A vision for the road ahead». En: *Science* 362.6412 (2018), eaam9288.

[116] Antonio Manzalini. «Quantum communications in future networks and services». En: *Quantum Reports* 2.1 (2020), págs. 221-232.

[117] Koji Azuma et al. «Quantum repeaters: From quantum networks to the quantum internet». En: *Reviews of Modern Physics* 95.4 (2023), pág. 045006.

[118] Edgar N Gilbert. «Random graphs». En: *The Annals of Mathematical Statistics* 30.4 (1959), págs. 1141-1144.

[119] Mohammad Reza Jabbarpour et al. «Applications of computational intelligence in vehicle traffic congestion problem: a survey». En: *Soft Computing* 22.7 (2018), págs. 2299-2320.

[120] Ruijin Ding et al. «Packet routing against network congestion: A deep multi-agent reinforcement learning approach». En: *2020 International Conference on Computing, Networking and Communications (ICNC)*. IEEE. 2020, págs. 932-937.

[121] Michael A Nielsen e Isaac Chuang. *Quantum computation and quantum information*. 2002. American Association of Physics Teachers.

[122] «IBM Quantum Experience. Available online: http://www.research.ibm.com/quantum». En: *(accessed on 10 Jun 2021)*.

[123] Jason Brownlee. «Statistical methods for machine learning». En: *Machine Learning Mastery.* (2020).

[124] Yaodong Yang y Jun Wang. «An overview of multi-agent reinforcement learning from game theoretical perspective». En: *arXiv preprint arXiv:2011.00583* (2020).

[125] Drew Fudenberg y David K Levine. «Whither game theory? Towards a theory of learning in games». En: *Journal of Economic Perspectives* 30.4 (2016), págs. 151-170.

[126] George W Brown. «Iterative solution of games by fictitious play». En: *Act. Anal. Prod Allocation* 13.1 (1951), pág. 374.

[127] James Hannan. «Approximation to Bayes risk in repeated play». En: *Contributions to the Theory of Games* 3 (1957), págs. 97-139.

[128] Katja Verbeeck et al. «Exploring selfish reinforcement learning in repeated games with stochastic rewards». En: *Autonomous Agents and Multi-Agent Systems* 14 (2007), págs. 239-269.

[129] Jiangfeng Du et al. «Entanglement playing a dominating role in quantum games». En: *Physics Letters A* 289.1-2 (2001), págs. 9-15.

[130] Robert Dorfman. «A formula for the Gini coefficient». En: *The review of economics and statistics* (1979), págs. 146-149.

[131] Damien Challet, Matteo Marsili y Yi-Cheng Zhang. *Minority games: interacting agents in financial markets*. OUP Oxford, 2004.

[132] Douglas R Hofstadter. *Metamagical themas: Questing for the essence of mind and pattern*. Hachette UK, 2008.

[133] Natalie S Glance y Bernardo A Huberman. «The dynamics of social dilemmas». En: *Scientific American* 270.3 (1994), págs. 76-81.

[134] Andreas Diekmann. «Volunteer's dilemma». En: *Journal of conflict resolution* 29.4 (1985), págs. 605-610.

[135] Diederik P Kingma y Jimmy Ba. «Adam: A method for stochastic optimization». En: *arXiv preprint arXiv:1412.6980* (2014).