

Redes neuronales clásicas-cuánticas híbridas para reconocimiento de partículas supersimétricas

Castro García José Luis¹, Ojeda Guillen Didier¹

¹ESCOM-IPN, Ciudad de México, México

Teléfono (55) 5729-6000 Ext. 52027 E-mail: jcastrog1702@alumno.ipn.mx

Resumen — La física de partículas requiere clasificar una gran cantidad de datos experimentales para descubrir nuevas partículas exóticas. Para esta tarea de clasificación se utilizan diferentes técnicas, dentro de las cuales se encuentra el aprendizaje automático. El presente trabajo propone un modelo de aprendizaje automático cuántico híbrido para el reconocimiento de partículas supersimétricas.

Palabras Clave – Aprendizaje automático, computación cuántica, física de partículas.

Abstract — Particle physics requires sorting through a large amount of experimental data to discover new exotic particles. Different techniques are used for this classification task, among which is machine learning. The present work proposes a hybrid quantum machine learning model for the recognition of supersymmetric particles.

Keywords — Machine learning, quantum computing, particle physics.

I. INTRODUCCIÓN

La física de partículas cuenta con una gran variedad de problemas que requieren el manejo de conjuntos de datos a gran escala. Para gestionar tales datos se han empleado técnicas clásicas de aprendizaje automático y aprendizaje profundo para acelerar el descubrimiento de nuevas partículas o eventos exóticos.

Los métodos recientes de aprendizaje automático cuántico (QML) tienen el potencial de aplicar estas técnicas en problemas de reconocimiento de partículas debido a su capacidad de procesar datos escasos, común en los conjuntos de datos de partículas supersimétricas.

Sin embargo, las técnicas comunes de QML no son muy factibles con los dispositivos cuántico ruidoso de escala intermedia (NISQ) con los que contamos: las máquinas de vectores de soporte cuántica (QSVM) y el clasificador cuántico variacional (VQC) requieren una cantidad de qubits equivalentes al número de características presentes en el conjunto de datos, por lo que tradicionalmente ha sido necesario utilizar algoritmos de preprocesamiento como análisis de componentes principales (PCA) antes de poder utilizar el algoritmo QML [1].

Nuevas propuestas proponen usar circuitos variacionales unidos a arquitecturas híbridas de aprendizaje automático para crear nuevas técnicas de QML. En el trabajo presentado

por Davis Arthur *et al.* [2] se implementa una red neuronal híbrida usando circuitos variacionales pequeños como neuronas dentro de la arquitectura, asegurándose que cada neurona reciba una entrada válida a partir de escalamientos dentro de la red. Otro método lo encontramos en el trabajo de Samuel Yen-Chein *et al.* [3], donde se presenta un modelo híbrido que combina las redes de tensores y los circuitos cuánticos variacionales que entrena, usando la red de tensores como un compresor de datos para, posteriormente, alimentar el circuito VQC y realizar la clasificación binaria.

En este trabajo desarrollaremos una arquitectura híbrida similar a ambos trabajos para realizar la clasificación de partículas supersimétricas: utilizaremos arquitecturas basadas en redes neuronales para crear un espacio representativo de nuestros datos mientras que introduciremos circuitos variacionales cuánticos para crear un nuevo espacio de características reducido y alimentar, finalmente, una red neuronal poco profunda para realizar la clasificación. Esto nos permitirá usar algunas ideas probadas ampliamente en redes neuronales mientras que aprovechamos la ventaja cuántica de encontrar relaciones entre los datos imposibles de generar clásicamente.

II. METODOLOGÍA

El primer problema de usar algoritmos QML en datos de naturaleza no cuántica es la manera de codificar los datos en un circuito cuántico con un número limitado de qubits. Teóricamente, se requiere una gran cantidad de compuertas cuánticas para hacer que los algoritmos QML puedan aprovechar la ventaja cuántica.

Tradicionalmente, ha sido necesario una compresión de los datos usando PCA; no obstante, otra alternativa es el uso de *Autoencoders* [4] para esta tarea, que es una técnica de análisis de componentes principales no lineal que utiliza redes neuronales artificiales. Sin embargo, sigue presente la desventaja de que la fase de preprocesamiento y la fase de clasificación usando métodos cuánticos se llevan a cabo por separado. Por consiguiente, es necesario idear un esquema que integre ambas para tener un entrenamiento continuo.

Las redes neuronales clásicas-cuánticas híbridas (HQCNN) son redes compuestas por bloques de funciones basados en redes neuronales clásicas y circuitos variacionales compuestos entre sí en la topología de un gráfico dirigido. En este trabajo, usaremos una arquitectura que consta de una fase de preprocesamiento compuesta de dos fases usando redes

neuronales con una fase cuántica entre estas dos fases (Figura 1).



Fig. 1. Modelo HQCNN utilizado. Los nodos marcados por DNN son arquitecturas de redes neuronales clásicas, mientras que el nodo QNN es un nodo con circuitos variacionales.

La primera fase, que llamamos **codificador**, se encargará de mapear los datos de entrada, que consta de 18 características, a un espacio de 20 dimensiones cuyos valores se encuentran en el rango de $[0, \pi]$ que usaremos como parámetros para el circuito cuántico. La siguiente fase, que llamamos **circuito variacional**, contará con 20 compuertas paramétricas actuando sobre un estado cuántico de 10 qubits en estado superpuesto. Finalmente, tenemos la última fase que consta de una red neuronal que llamaremos **clasificador**, donde se recibe un vector real de dimensión 10 (los valores esperados obtenidos del circuito cuántico) que utilizaremos como clasificador final.

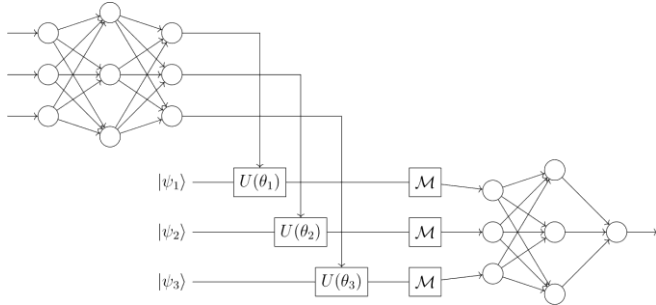


Fig. 2. Arquitectura HQCNN utilizada durante el proyecto.

A. Codificador

El mapa de características de un circuito cuántico variacional juega un papel importante en la expresividad del modelo [5]. Los datos deben codificarse de tal manera que cada posible entrada tenga una codificación qubit única antes de pasar al ansatz del modelo.

En la mayoría de los circuitos variacionales propuestos [6], los mapas de codificación interactúan con los datos clásicos mediante compuertas de rotación de Pauli aplicadas a un registro de qubits. Usar este tipo de compuertas trae una serie de ventajas, en particular, si restringimos cada característica al intervalo $[0, \pi]$, podemos crear un mapa de características que garantiza que cada entrada única tendrá una codificación cuántica única sin requerir una gran cantidad de puertas cuánticas.

Para realizar esta codificación, usamos una red neuronal para que aprenda un mapeo antes de alimentar el circuito cuántico. La red neuronal de codificación consta de sólo 3 capas que se describen en la Tabla I.

TABLA I.
ARQUITECTURA DEL CODIFICADOR PARA HQCNN.

Tipo de capa	Dimensión de salida	No. de parámetros
Lineal densamente conectada (ELU).	36	684
Lineal densamente conectada (sigmoide).	20	740
Funcional ($x \rightarrow \pi x$)	20	-

B. Circuito variacional

Para el circuito variacional se utilizó el ansatz de amplitudes reales. Originalmente, el ansatz consiste en aplicar una compuerta R_y en cada qubit seguido de compuertas CX aplicadas a cada combinación posible de qubits en el estado cuántico. Finalmente, se aplica otra compuerta R_y en cada qubit.

En la práctica, notamos que el modelo tenía mejor rendimiento si modificábamos ligeramente el circuito: utilizamos compuertas R_z en la primera capa del circuito en lugar de compuertas R_y ; además, usamos circuitos de entrelazamiento entre los pares de qubits contiguos en lugar de solo compuertas CX entre todos los qubits. Estas modificaciones reducen el número de compuertas del orden de $\mathcal{O}(q^2)$ a $\mathcal{O}(q)$, siendo q el número de qubits. Este circuito se muestra en la Figura 3, donde se muestra que la medición se realiza respecto a la base Z .

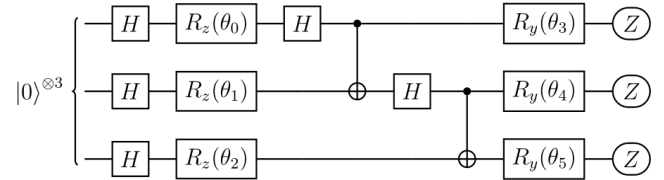


Fig. 3. Circuito variacional de amplitudes reales utilizado con 3 qubits utilizado en el modelo HQCNN.

C. Clasificador

Como se puede observar en la Figura 2, la entrada de la última red neuronal es un vector con los valores esperados del circuito variacional. Como su nombre lo indica, la tarea de esta red es simplemente realizar la clasificación con las características generadas por el circuito cuántico; para ello utilizamos una red neuronal poco profunda que consta de dos capas ocultas con 64 y 128 neuronas respectivamente, ambas usando $ReLU$ como función de activación. La capa de salida consta solamente de una neurona con activación sigmoidea. Esta arquitectura se describe en la Tabla II.

TABLA II.
ARQUITECTURA DEL CLASIFICADOR PARA HQCNN.

Tipo de capa	Dimensión de salida	No. de parámetros
Lineal densamente conectada (ReLU).	64	704
Lineal densamente conectada (ReLU).	128	8320
Lineal densamente conectada (sigmoide).	1	129

D. Conjunto de datos

El conjunto de datos para este trabajo se obtuvo del repositorio de UCI [7], extraídos del trabajo realizado por Baldi *et al.* [8]. El conjunto de datos original cuenta con 5,000,000 de ejemplos, de los cuales el 46% pertenecen a clase de partículas supersimétricas (SUSY). Sin embargo, a lo largo de este trabajo, se utilizó un subconjunto de 10,000 muestras con las mismas características estadísticas que el conjunto original.

La tarea de clasificación presente en el conjunto de datos es la de distinguir entre un evento de producción de partículas supersimétricas contra un evento de fondo con el mismo nivel de detección (Figura 4).

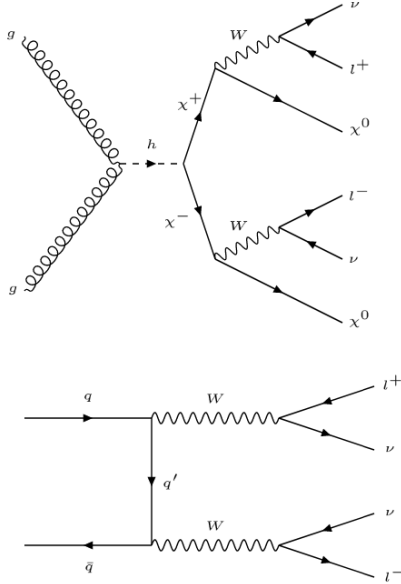


Fig. 4. Arriba: diagrama de Feynman que describe el proceso de desintegración de partículas supersimétricas χ^\pm en partículas supersimétricas neutras χ^0 , leptones cargados l^\pm y neutrinos ν . Abajo: proceso de fondo en el cual dos bosones W se descomponen en dos leptones cargados y dos neutrinos.

El proceso supersimétrico consta de dos partículas supersimétricas cargadas eléctricamente χ^\pm , que se desintegran en bosones W y una partícula supersimétrica eléctricamente neutra χ^0 . A su vez, los bosones W se desintegran en leptones l y neutrinos ν . La descomposición

final de la partícula supersimétrica está compuesta por dos leptones cargados ll y el momento perdido arrastrado por las partículas que no son detectables ($\chi^0\chi^0\nu\nu$).

El proceso de fondo que tiene el mismo nivel de detectabilidad es la producción de pares de bosones W , que se desintegran en leptones cargados l y neutrinos invisibles ν . Ambos eventos contienen dos leptones finales y una pérdida considerable de impulso.

III. RESULTADOS

A. Entrenamiento

Para el entrenamiento utilizamos una muestra de 8000 datos usando una máquina con AMD Ryzen 5 2600X Six-Core con una tarjeta gráfica NVIDIA GeForce RTX 3060.

Debido a la utilización de redes neuronales, fue necesario entrenar los modelos híbridos usando backpropagation; para calcular el gradiente del circuito variacional se utilizó el método de gradiente adjunto [9]. Cabe mencionar que este método para calcular el gradiente no puede implementarse en una computadora cuántica; sin embargo, este método es útil cuando se realizan experimentos de sistemas fácilmente simulables de forma clásica, obteniendo un alto rendimiento incluso en circuitos profundos con muchos parámetros.

En la Tabla III se muestran los hiperparámetros utilizados en el entrenamiento del modelo. Originalmente se pensaba entrenar durante 100 épocas usando lotes de 256 datos; sin embargo, el uso de *early-stopping* con una tolerancia de 3 épocas limitó el entrenamiento a 64 épocas. Para la optimización se utilizó el algoritmo de ADAM [10], usando todos los parámetros por defecto excepto por la tasa de aprendizaje.

TABLA III.
PARÁMETROS UTILIZADOS EN EL ENTRENAMIENTO DEL MODELO HQCNN.

Atributo	Valor
Tamaño de datos de entrenamiento	8000
Tamaño de datos de prueba	2000
Función de pérdida	Entropía cruzada
Optimizador	ADAM
Tasa de aprendizaje	0.0001

B. Métricas de desempeño

El modelo HQCNN logró superar a las redes neuronales profundas (DNN). En el entrenamiento de los modelos notamos una estabilidad en la evolución de la función de pérdida que incluso supera a los algoritmos clásicos, al igual que el *Valor Predictivo Positivo* (PPV), como se puede observar en la Figura 5.

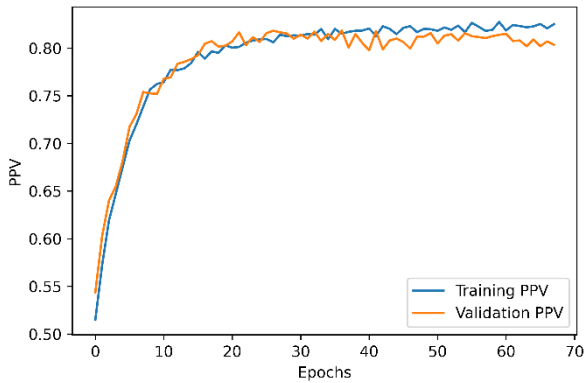
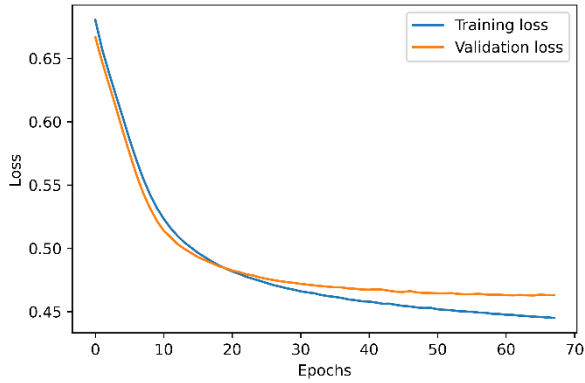


Fig. 5. Circuito variacional de amplitudes reales utilizado con 3 qubits utilizado en el modelo HQCNN.

Como se observa en la Tabla IV, HQCNN ha superado en ambas métricas a las redes neuronales profundas entrenadas con la misma cantidad de datos y usando entropía cruzada. En este caso consideramos PPV y la significancia media aproximada (AMS) debido a que son las métricas más usadas en física de partículas [11].

TABLA IV.
RESULTADOS OBTENIDOS POR HQCNN RESPECTO A REDES NEURONALES.

Modelo	PPV	AMS
HQCNN	0.788	11.690σ
DNN	0.755	10.374σ

IV. DISCUSIÓN

Aunque no se utilizaron circuitos cuánticos muy complejos, el método HQCNN logró superar a las redes neuronales usando circuitos poco complejos y redes neuronales que no son profundas. Esto permitió superar los problemas observados en los métodos de QSVM y VQC.

A diferencia de VQC y QSVM, este método escala bien en tiempo de ejecución, ya que, a diferencia de los métodos

anteriormente mencionados donde entre más características más tiempo de entrenamiento tomaba. Esto nos obligaba a utilizar métodos de compresión de datos, por lo que se perdía mucha información antes de incluso pasar a la fase de clasificación.

Por último, en trabajos anteriores, se ha encontrado que los algoritmos clásicos logran obtener buenos resultados si se utilizan funciones de pérdida basadas en significancia estadística [12] (que indirectamente maximiza PPV); por lo que cabría preguntarse qué tanto mejorarían los algoritmos cuánticos con este tipo de funciones de pérdida.

V. CONCLUSIONES

Los detectores de física de partículas recopilan miles de datos cada segundo; sin embargo, sólo una pequeña fracción de estos datos son realmente de interés. Es aquí donde entra en juego las nuevas técnicas de aprendizaje automático, que se han convertido en un estándar en el análisis de datos de física de partículas.

Como se observó, los HQCNN sobrepasó en rendimiento a las redes neuronales. Este método podría permitir usar métodos de QML para entrenar modelos más robustos en comparación a otros algoritmos como QSVM y VQC ya que es capaz de utilizar la ventaja cuántica sin requerir una gran cantidad de qubits, como era necesario con los algoritmos cuánticos anteriormente mencionados, donde el número de qubits es igual al número de características.

Se espera que este tipo de algoritmos híbridos sea capaz de superar a los algoritmos de aprendizaje automático cuando los datos tengan una naturaleza cuántica o demasiado compleja; sin embargo, antes es necesario tener un acceso libre a las computadoras cuánticas, debido a que, durante el desarrollo del trabajo, se utilizaron simuladores sin ruido cuántico.

APÉNDICE

El código utilizado para este trabajo puede encontrarse en forma de Notebooks en la plataforma de Google Colab. Los enlaces se muestran a continuación:

HQCNN - notebook (Red neuronal clásica-cuántica híbrida).
DNN - notebook (Red neuronal profunda).

AGRADECIMIENTOS

Los autores agradecen al comité organizador de la RNFM por la oportunidad de presentar este trabajo, a la Secretaría de Investigación y Posgrado (SIP) por su apoyo económico en el proyecto 20220405 y el apoyo económico mediante el programa BEIFI.

REFERENCIAS

- [1] Wu, Sau Lan, et al. "Application of quantum machine learning using the quantum variational classifier method to high energy physics analysis at the LHC on IBM quantum computer simulator and hardware with 10 qubits." *Journal of Physics G: Nuclear and Particle Physics* 48.12 (2021).
- [2] Arthur, Davis. "A hybrid quantum-classical neural network architecture for binary classification." arXiv preprint arXiv:2201.01820 (2022).
- [3] Chen, Samuel Yen-Chi, et al. "Hybrid quantum-classical classifier based on tensor network and variational quantum circuit." arXiv preprint arXiv:2011.14651 (2020).
- [4] Kramer, Mark A. "Nonlinear principal component analysis using autoassociative neural networks." *AIChE journal* 37.2 (1991).
- [5] Schuld, Maria, Ryan Sweke, and Johannes Jakob Meyer. "Effect of data encoding on the expressive power of variational quantum-machine-learning models." *Physical Review A* 103.3 (2021).
- [6] Sim, Sukin, Peter D. Johnson, and Alán Aspuru-Guzik. "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms." *Advanced Quantum Technologies* 2.12 (2019).
- [7] Dua, Dheeru and Graff, Casey. "UCI Machine Learning Repository." Irvine, CA: University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/m>. (2019).
- [8] Baldi, Pierre, Peter Sadowski, and Daniel Whiteson. "Searching for exotic particles in high-energy physics with deep learning." *Nature communications* 5.1 (2014).
- [9] Chen, Ricky TQ, et al. "Neural ordinary differential equations." *Advances in neural information processing systems* 31 (2018).
- [10] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
- [11] Adam-Bourdarios, Claire, et al. "The Higgs boson machine learning challenge." *NIPS 2014 workshop on high-energy physics and machine learning*. PMLR, (2015).
- [12] Elwood, Adam, and Dirk Krücker. "Direct optimisation of the discovery significance when training neural networks to search for new physics in particle colliders." arXiv preprint arXiv:1806.00322 (2018).