






Embeddings de grafos de conocimiento basados en algoritmos cuánticos variacionales

Martin Santesteban^{1,2} , Patricio Bruno² , Santiago Cifuentes^{1,2} , Guido Bellomo² , and Gustavo Martin Bosyk² 

¹ Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales.
Departamento de Computación. Buenos Aires, Argentina

² CONICET-Universidad de Buenos Aires. Instituto de Ciencias de la Computación (ICC). Buenos Aires, Argentina
martin.p.santesteban@gmail.com, pdbruno@gmail.com,
santiagocifuentes66@hotmail.com, gbellomo@icc.fcen.uba.ar,
gbosyk@icc.fcen.uba.ar

Resumen Se propone un esquema híbrido cuántico-clásico para generar *embeddings* en grafos de conocimiento mediante algoritmos variacionales (VQAs). El modelo optimiza representaciones cuánticas de entidades y relaciones a través de circuitos parametrizados, combinando procesamiento cuántico (preparación de estados y medición) con optimización clásica. Se analizan diversas arquitecturas de circuitos cuánticos, estableciendo un marco unificado.

Palabras claves: Grafos de conocimiento, Embeddings, Algoritmos cuánticos variacionales

Knowledge graph embeddings based on variational quantum algorithms

Abstract A hybrid quantum-classical framework is proposed to generate embeddings in knowledge graphs using variational quantum algorithms (VQAs). The model optimizes quantum representations of entities and relations via parameterized circuits, combining quantum processing (state preparation and measurement) with classical optimization. Various quantum circuit architectures are analyzed, establishing a unified framework.

Keywords: Knowledge graphs, Embeddings, Variational quantum algorithms

1 Introducción

Los *Knowledge Graphs* (KGs) (Hogan et al., 2021) cobran relevancia cuando las conexiones entre datos son tan importantes como sus valores intrínsecos,

encontrando aplicaciones desde redes sociales (Fan, 2012) hasta Web Semántica (Arenas and Pérez, 2011) y procedencia de datos (Anand et al., 2010).

Matemáticamente, un KG es un multigrafo dirigido y etiquetado, $KG = (E, R, F)$, donde E es el conjunto de nodos (entidades), R el conjunto de etiquetas (relaciones) y $F \subseteq E \times R \times E$ el conjunto de hechos (o tripletas) (h, r, t) , que indican que la entidad origen (cabeza) $h \in E$ se relaciona con la entidad destino (cola) $t \in E$ mediante la relación $r \in R$.

Sobre estos grafos se aplican procesos de inferencia como data completion (Rossi et al., 2021), clustering (Saeedi et al., 2018) o link prediction (Kumar et al., 2020), abordados mediante técnicas de embeddings (Dai et al., 2020; Ge et al., 2024). Estas técnicas representan cada entidad $e \in E$ y relación $r \in R$ en un espacio vectorial de baja dimensión, buscando que embeddings similares reflejen semánticas estructurales comunes. La validez de una tripleta (h, r, t) se estima con una función de score $\delta_{hrt} : E \times R \times E \rightarrow \mathbb{R}$, optimizando los embeddings sobre un conjunto de entrenamiento $\mathcal{D} = \{(h, r, t), y_{hrt}\}$ —donde y_{hrt} indica si la tripleta pertenece al KG— usando una función de pérdida $\mathcal{L}(\delta_{hrt}, y_{hrt}; \Theta)$, con Θ los parámetros del modelo. Ejemplos de este enfoque son RESCAL (Nickel et al., 2011), TransE (Asmara et al., 2023) y RotatE (Sun et al., 2019).

En el ámbito cuántico, los modelos operan sobre vectores en espacios de Hilbert. Recientemente, se propusieron VQAs para calcular embeddings (Kurokawa et al., 2022; Ma et al., 2019), aunque con scores dependientes de la arquitectura del circuito. En este contexto, presentamos qCUERO (Quantum Classical Universal Embeddings for Relations and Objects), un esquema general y adaptable para embeddings de KGs mediante VQAs.

2 qCUERO: Modelos Cuánticos de Embeddings

El esquema qCUERO define embeddings cuánticos para entidades en un espacio de Hilbert de dimensión 2^n (con n qubits) tal que $\forall e \in E$, $|e\rangle = V(\vec{\theta}_e)H^{\otimes n}|0\rangle^{\otimes n}$, donde $H^{\otimes n}$ denota compuertas Hadamard aplicadas en paralelo y $V(\vec{\theta}_e)$ es un ansatz parametrizado con $\vec{\theta}_e \in \mathbb{R}^{d_e}$. Para cada relación $r \in R$, se define una unitaria $U(\vec{\theta}_r)$ con $\vec{\theta}_r \in \mathbb{R}^{d_r}$. Tanto el número de qubits n como las dimensiones de los vectores de parámetros d_e y d_r son hiperparámetros del modelo. La consistencia para una tripleta $(h, r, t) \in E$ se impone mediante $U(\vec{\theta}_r)|h\rangle \approx |t\rangle$.

La Fig. 1 ilustra el flujo híbrido cuántico-clásico: (i) preparación cuántica de entidades y relaciones y cálculo del score (en rojo), (ii) cálculo clásico de \mathcal{L}^1 (en azul), y (iii) optimización clásica de parámetros. Para obtener ventajas frente a métodos clásicos, los ansatz deben ser circuitos poco profundos (complejidad lineal en n), de forma que la complejidad por época sea $O(|\mathcal{D}|\text{poly}(\log d)/\epsilon^2)$, donde $|\mathcal{D}|$ es el tamaño del conjunto de entrenamiento, $d = 2^n$, y $1/\epsilon^2$ es el número de mediciones requeridas para alcanzar una precisión ϵ .

El modelo qCUERO tiene la capacidad de expresar relaciones *simétricas* y *antisimétricas* junto con la *inversión* y *composición de relaciones*. Entre los

¹ Aunque la función de pérdida podría obtenerse cuánticamente si se logra escribir como el valor medio de algún observable.

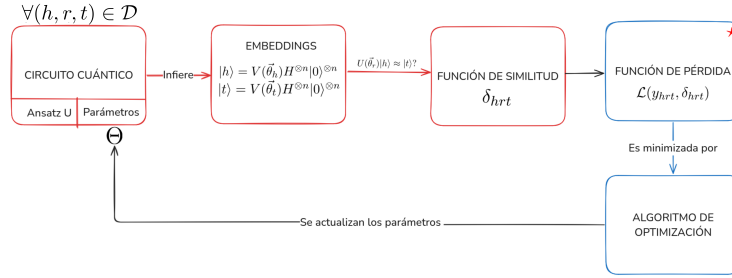


Fig. 1: Flujo híbrido cuántico-clásico para entrenar un modelo de embeddings cuántico. Etapas cuánticas en color rojo; etapas clásicas en color azul.

embeddings clásicos, RotaE posee una capacidad de expresividad equivalente (Sun et al., 2019).

Teorema. *qCUERO es capaz de modelar relaciones simétricas/antisimétricas, inversas y compuestas: (i) simetría: $U(\vec{\theta}_r)U(\vec{\theta}_r) = I$, (ii) anti-simetría: $U(\vec{\theta}_r)U(\vec{\theta}_r) \neq I$, (iii) inversión: $U(\vec{\theta}_{r_1})U(\vec{\theta}_{r_2}) = I$, y (iv) composición: $U(\theta_{r_1}) = U(\theta_{r_2})U(\theta_{r_3})$, donde I es la identidad.*

Existen dos propuestas recientes de VQAs (Kurokawa et al., 2022; Ma et al., 2019) que se enmarcan en el esquema de la Fig. 1. La diferencia principal radica en el score y en la cantidad de qubits empleados.

En (Ma et al., 2019) se introduce el fQCE (Fully Parametrized Quantum Circuit Embedding), que utiliza $n + 1$ qubits. En n qubits se codifica la acción de la relación r sobre la cabeza h y la cola t mediante unitarias $U_1 = U(\vec{\theta}_r)U(\vec{\theta}_h)H^{\otimes n}$ y $U_2 = U(\vec{\theta}_t)H^{\otimes n}$, respectivamente (Fig. 2a). La medición del qubit adicional del switch test gate (Chamorro-Posada and Garcia-Escartin, 2023) se emplea para obtener el score: $\delta_{hrt}^{fQCE} = \Re(\langle t|U(\vec{\theta}_r)|h\rangle) \in [-1, 1]$.

Por otro lado, (Kurokawa et al., 2022) utiliza un modelo con $2n + 1$ qubits. En n qubits se codifica la acción de r sobre h mediante $U_1 = U(\vec{\theta}_r)U(\vec{\theta}_h)H^{\otimes n}$ y en otros n qubits se codifica la cola t como $U_2 = U(\vec{\theta}_t)H^{\otimes n}$ (Fig. 2b). La medición del qubit adicional del swap test (Buhrman et al., 2001) se usa para obtener el score: $\delta_{hrt}^{ku} = |\langle t|U(\theta_r)|h\rangle|^2 \in [0, 1]$.

En ambos modelos, la complejidad de calcular el del score es $O(\text{poly}(\log d)/\epsilon^2)$. Además, se minimiza la función de pérdida del error cuadrático medio, $= \min_{\Theta} \frac{1}{|\mathcal{D}|} \sum_{hrt} (\delta_{hrt} - y_{hrt})^2$ con $\Theta = \{(\vec{\theta}_e, \vec{\theta}_r)\}$. La complejidad total por época es $O(|\mathcal{D}|\text{poly}(\log d)/\epsilon^2)$.

3 Comentarios finales

Se presentó el esquema qCUERO para abordar el embedding de KG mediante VQAs, demostrando su capacidad para modelar relaciones simétricas, antisimétricas, inversas y compuestas, con una expresividad comparable a RotaE.

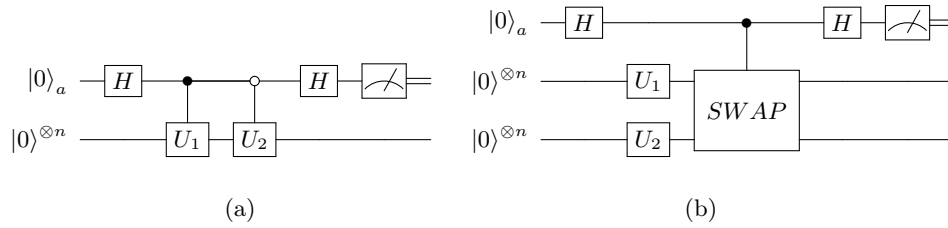


Fig. 2: Propuestas de VQAs para embedding de KGs. (a) Circuito con Switch SWAP test. (b) Circuito con SWAP test.

Asimismo, se mostró que las propuestas de (Ma et al., 2019) y (Kurokawa et al., 2022) se enmarcan en este esquema.

Como líneas futuras se plantea: (i) validar el método en escenarios reales con diversos grafos de conocimiento; (ii) explorar nuevas arquitecturas cuánticas para optimizar la eficiencia del embedding; (iii) investigar si el entrenamiento en superposición cuántica reduce exponencialmente la complejidad (pasando de $|\mathcal{D}|$ a $\log |\mathcal{D}|$); y (iv) analizar el impacto del ruido cuántico para desarrollar estrategias de mitigación que permitan su implementación en hardware real.

Agradecimientos

GMB reconoce el financiamiento de los proyectos PIBAA-0718 y PIP-0135 del CONICET (Argentina). MS agradece la Beca de Iniciación en Investigación Científica otorgada por la FCEyN-UBA y LambdaClass.

Referencias

- Anand, M. K., Bowers, S., & Ludäscher, B. (2010). Techniques for efficiently querying scientific workflow provenance graphs. *EDBT*, 10(2010), 287–298.
- Arenas, M., & Pérez, J. (2011). Querying semantic web data with sparql. *Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART Symp. on Principles of Database Systems*, 305–316.
- Asmara, S. M., Sahabudin, N. A., Ismail, N. S. N., & Sabri, I. A. A. (2023). A review of knowledge graph embedding methods of transe, transh and transr for missing links. *2023 IEEE 8th Int. Conf. On Software Engineering and Computer Systems (ICSECS)*, 470–475.
- Buhrman, H., Cleve, R., Watrous, J., & de Wolf, R. (2001). Quantum fingerprinting. *Phys. Rev. Lett.*, 87, 167902.
- Chamorro-Posada, P., & Garcia-Escartin, J. C. (2023). The switch test for discriminating quantum evolutions. *Journal of Physics A: Mathematical and Theoretical*, 56(35), 355301.

- Dai, Y., Wang, S., Xiong, N. N., & Guo, W. (2020). A survey on knowledge graph embedding: Approaches, applications and benchmarks. *Electronics*, 9(5), 750.
- Fan, W. (2012). Graph pattern matching revised for social network analysis. *Proceedings of the 15th Int. Conf. on Database Theory*, 8–21.
- Ge, X., Wang, Y. C., Wang, B., Kuo, C.-C. J., et al. (2024). Knowledge graph embedding: An overview. *APSIPA Trans. on Signal and Inf. Proc.*, 13(1).
- Hogan, A., Blomqvist, E., Cochez, M., D’amato, C., Melo, G. D., et al. (2021). Knowledge graphs. *ACM Comput. Surv.*, 54(4).
- Kumar, A., Singh, S. S., Singh, K., & Biswas, B. (2020). Link prediction techniques, applications, and performance: A survey. *Physica A: Statistical Mechanics and its Applications*, 553, 124289.
- Kurokawa, M., Giri, P. R., & Saito, K. (2022). Evaluating variational quantum circuit designs for knowledge graph completion. *2022 IEEE Int. Conf. on Quantum Computing and Engineering (QCE)*, 777–778.
- Ma, Y., Tresp, V., Zhao, L., & Wang, Y. (2019). Variational quantum circuit model for knowledge graph embedding. *Adv. Quantum Tech.*, 2(7-8), 1800078.
- Nickel, M., Tresp, V., Kriegel, H.-P., et al. (2011). A three-way model for collective learning on multi-relational data. *Icml*, 11(10.5555), 3104482–3104584.
- Rossi, A., Barbosa, D., Firmani, D., Matinata, A., & Merialdo, P. (2021). Knowledge graph embedding for link prediction: A comparative analysis. *ACM Trans. on Knowledge Discovery from Data (TKDD)*, 15(2), 1–49.
- Saeedi, A., Peukert, E., & Rahm, E. (2018). Using link features for entity clustering in knowledge graphs. *European Semantic Web Conf.*, 576–592.
- Sun, Z., Deng, Z.-H., Nie, J.-Y., & Tang, J. (2019). Rotate: Knowledge graph embedding by relational rotation in complex space. *Int. Conf. on Learning Representations*.