1 Introduction

1.1 Resumen:

The introduction to the paper titled "Logical-Mathematical Foundations of a Graph Query Framework for Relational Learning" outlines the importance and potential applications of relational learning methods, which take into account the connections between data. The two main approaches to relational learning are presented: the latent feature or connectionist approach and the graph pattern-based approach or symbolic approach. The paper focuses on addressing the challenges faced by the graph pattern-based approach, including computational complexity arising from relational queries and the lack of robust and general frameworks for symbolic relational learning methods. A novel graph query framework is proposed to solve these problems, with a focus on formalizing an efficient graph query system that enables controlled complexity and stepwise pattern expansion using well-defined operations. The paper's structure is outlined, with sections covering related research, the novel graph query framework, its implementation for relational machine learning, and conclusions and future research directions.

1.2 Evaluación:

Evaluation Criteria: Clarity

Evaluation Level: Can be improved

Evaluation Justification: The introduction section of the paper is generally clear, providing a good overview of the topic and the motivation for the research. However, it could benefit from some improvements in clarity. For instance, the first sentence could be rephrased to make it more understandable for readers who are not familiar with machine learning concepts. Additionally, the transition between discussing different types of relational learning methods and introducing the novel graph query framework could be smoother.

Example from the text: "In relational learning, relationships between objects are also taken into account during the learning process, and data is represented as a graph composed of nodes (entities) and links (relationships), both with possible associated properties." This sentence could be rephrased to make it clearer for readers who may not have prior knowledge in this area.

Improvement: To improve clarity, the authors could consider using more concise language, restructuring complex sentences, and providing examples or illustrations to help explain key concepts. Additionally, they could clarify the transition between discussing different types of relational learning methods and introducing their novel approach.

2 Related work

2.1 Resumen:

The 'Related Work' section of the paper discusses various techniques for executing relational queries and approaches to relational learning. The common approach is Graph Pattern Matching, which has been researched for over three decades. However, this work focuses on the graph-pattern based approach for relational learning. Techniques like Inductive Logic Programming (ILP), Multi-relational Decision Tree Learning (MRDTL), and Graph-Based Induction of Decision Trees (DT-GBI) are mentioned as examples of pattern-based approaches to relational learning. However, these methods have limitations such as lack of support for atomic operations or inability to construct cyclic patterns. The paper's proposal aims to overcome these limitations by supporting learning from general subgraphs and executing cyclic queries.

2.2 Evaluación:

Evaluation Criteria: Clarity, Novelty

Evaluation Level: Can be improved

Evaluation Justification: The related work section provides a clear overview of the existing approaches in the field of relational learning. However, it can be improved by emphasizing more on the specific contributions and differences between the proposed method and the existing ones. For instance, while discussing Multi-relational decision tree learning (MRDTL) and Graph-Based Induction of Decision Trees (DT-GBI), there is a need to explain how these methods differ from the proposed approach and why the proposed approach is better suited for certain tasks.

Example from the text: "Multi-relational decision tree learning (MRDTL [11]) is a learning algorithm for relationships and is supported by Selection Graphs [15], a graph representation of SQL queries that selects records from a relational database based on certain constraints. Selection graphs enable atomic operations to enhance queries, but they lack the ability to distinguish between query elements that constitute the query result and those that relate to objects that should or should not be linked to the query result. Consequently, queries performed using selection graphs yield records that satisfy the given selection graph conditions, but cannot identify subgraphs. The refinement operations presented on the selection graphs are: adding positive conditions, adding negative conditions, adding present edges and opening nodes, adding absent edges and closing nodes. This set of operations does not allow for the construction of cyclic patterns."

Improvement: In addition to describing the limitations of existing methods, it would be helpful to explain how the proposed approach addresses these limitations and provides unique capabilities. For instance, in the case of Selection Graphs, one could explain how the proposed method can identify subgraphs and enable the construction of cyclic patterns through its atomic operations.

3 Relational machine learning

3.1 Resumen:

In this section, the authors describe how to use the graph query framework for relational machine learning. They propose a pattern search technique based on information gain to identify typical patterns for each subgraph class in a labeled subgraph set within a graph dataset. To achieve this, they conduct top-down decision tree induction using graph queries as test tools at internal nodes. The best refinement sets are identified during the tree construction process, resulting in queries that define classes within the graph dataset.

The authors then provide examples of node classification problems and classify more intricate structures like the Star Wars toy graph. They demonstrate how to use the query framework and refinement sets to perform relational learning by identifying distinctive patterns for each node type or species, which can be used to directly assess nodes and clarify future classifications. The resulting decision trees accurately assign types to all nodes in the graph by exploiting relational information from the network.

Overall, this section presents a method for using the graph query framework for relational machine learning that involves identifying characteristic patterns of subgraph classes through top-down decision tree induction with graph queries as test tools at internal nodes. The authors provide practical examples to illustrate the process of performing relational learning by utilizing the query framework and refinement sets to classify nodes or other structures in a graph dataset.

3.2 Evaluación:

Evaluation Criteria: Clarity, Evaluation Level: Can be improved, Evaluation Justification: The section could benefit from more explicit explanations of the methodology and its significance. An example is the use of "Q" without any explanation or definition. It would improve clarity if the terms were defined in a glossary at the beginning of the paper or as they are introduced in the text.

4 Conclusions and future work

4.1 Resumen:

The paper proposes a novel framework for graph queries that allows polynomial cyclic assessment of queries and refinements based on atomic operations. The system uses a consistent grammar for both queries and evaluated structures, supports the assessment of subgraphs beyond individual nodes, and offers controlled and automated query construction via refinements. The framework evaluates existence/non-existence of paths and nodes in a graph instead of demanding isomorphism, thus enabling polynomial time evaluation of cyclic patterns. The proof-of-concept implementation has been demonstrated through experimentation in relational learning procedures, showing that interesting patterns can be extracted from relational data. Although the presented query definition utilizes binary graph data sets, it can be implemented on hypergraph data as well. Future research will focus on developing automated methods to generate refinement sets based on a given learning task and the specific characteristics of the graph dataset. Additionally, patterns obtained from the graph learning procedure can serve as features in other machine learning methods, enabling non-relational machine learning methods to learn from them.

4.2 Evaluación:

Evaluation Criteria: Clarity, Novelty

Evaluation Levels:
- YES: The criterion is fully met in the provided section.
- Can be improved: The criterion is partially met but could be strengthened.

Evaluation Justification and Examples from the Evaluated Section:

Clarity: Partially met. While the text does provide a clear overview of the framework's capabilities, it lacks specific examples to illustrate these capabilities.

Novelty: Fully met. The text clearly describes the novelty of the approach by differentiating it from existing work (e.g., "Graph isomorphism-based query systems exhibit exponential complexity when presented with cyclic queries"). It also emphasizes its unique contributions, such as evaluating cyclic patterns in polynomial time and supporting top-down learning techniques through refinement sets.

Specific Suggestions for Improvement:

Clarity: Add specific examples to illustrate the framework's capabilities, such as instances of successful pattern extraction or instances where the framework outperformed other methods.

Novelty: Emphasize the novelty further by explicitly comparing with related work and highlighting unique contributions in more detail.

Acknowledgements

References

References

Almagro-Blanco, P., & Sancho-Caparrini, F. (2017). Generalized graph pattern matching. *CoRR*, *abs/1708.03734*. URL: `http://arxiv.org/abs/1708.03734`. arXiv:1708.03734.

Barceló, P., Libkin, L., & Reutter, J. L. (2011). Querying graph patterns. In *Proceedings of the Thirtieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* PODS '11 (pp. 199–210). New York, NY, USA: ACM. URL: `http://doi.acm.org/10.1145/1989284.1989307`. doi:10.1145/1989284.1989307.

Blockeel, H., & Raedt, L. D. (1998). Top-down induction of first-order logical decision trees. *Artificial Intelligence*, *101*, 285–297. URL: `http://www.sciencedirect.com/science/article/pii/S0004370298000344`. doi:10.1016/S0004-3702(98)00034-4.

Bonifati, A. Fletcher, G.Voigt, H.Yakovets, N.Jagadish, H. V. (2018 Querying Graphs Morgan & Claypool Publishers

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 2787–2795). Curran Associates, Inc. URL: `http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf`.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? workforce implications. *Science*, *358*, 1530–1534.

Camacho, R., Pereira, M., Costa, V. S., Fonseca, N. A., Adriano, C., Simões, C. J., & Brito, R. M. (2011). A relational learning approach to structure-activity relationships in drug design toxicity studies. *Journal of integrative bioinformatics*, *8*, 176–194.

Chang, K.-W., Yih, S. W.-t., Yang, B., & Meek, C. (2014). Typed tensor decomposition of knowledge bases for relation extraction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. ACL – Association for Computational Linguistics. URL: `https://www.microsoft.com/en-us/research/publication/typed-tensor-decomposition-of-knowledge-bases-for-relation-extraction/`.

Consens, M. P., & Mendelzon, A. O. (1990). Graphlog: A visual formalism for real life recursion. In *Proceedings of the Ninth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems* PODS '90 (pp. 404–416). New York, NY, USA: ACM. URL: `http://doi.acm.org/10.1145/298514.298591`. doi:10.1145/298514.298591.

Cook, S. A. (1971). The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing* STOC '71 (pp. 151–158). New York, NY, USA: ACM. URL: `http://doi.acm.org/10.1145/800157.805047`. doi:`10.1145/800157.805047`.

Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., & Zhang, W. (2014). Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* KDD '14 (pp. 601–610). New York, NY, USA: ACM. URL: `http://doi.acm.org/10.1145/2623330.2623623`. doi:`10.1145/2623330.2623623`.

Fan, W., Li, J., Ma, S., Tang, N., Wu, Y., & Wu, Y. (2010). Graph pattern matching: From intractable to polynomial time. *Proc. VLDB Endow.*, *3*, 264–275. URL: `http://dx.doi.org/10.14778/1920841.1920878`. doi:`10.14778/1920841.1920878`.

Gallagher, B. (2006). Matching structure and semantics: A survey on graph-based pattern matching. *AAAI FS*, *6*, 45–53.

García-Jiménez, B., Pons, T., Sanchis, A., & Valencia, A. (2014). Predicting protein relationships to human pathways through a relational learning approach based on simple sequence features. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *11*, 753–765. doi:`10.1109/TCBB.2014.2318730`.

Geamsakul, W., Matsuda, T., Yoshida, T., Motoda, H., & Washio, T. (2003). Classifier construction by graph-based induction for graph-structured data. In K.-Y. Whang, J. Jeon, K. Shim, & J. Srivastava (Eds.), *Advances in Knowledge Discovery and Data Mining: 7th Pacific-Asia Conference, PAKDD 2003, Seoul, Korea, April 30 – May 2, 2003 Proceedings* (pp. 52–62). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: `http://dx.doi.org/10.1007/3-540-36175-8_6`. doi:`10.1007/3-540-36175-8_6`.

Gupta, S. (2015). *Neo4j Essentials*. Community experience distilled. Packt Publishing. URL: `https://books.google.es/books?id=WJ7NBgAAQBAJ`.

Luc De Raedt, Sebastijan Dumančić, Robin Manhaeve, Giuseppe Marra. *From Statistical Relational to Neural-Symbolic Artificial Intelligence*. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI'20)*, 2021. ISBN: 9780999241165. Article No.: 688. Pages: 8. Yokohama, Japan.

Henzinger, M. R., Henzinger, T. A., & Kopke, P. W. (1995). Computing simulations on finite and infinite graphs. In *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on* (pp. 453–462). IEEE.

Jacob, Y., Denoyer, L., & Gallinari, P. (2014). Learning latent representations of nodes for classifying in heterogeneous social networks. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining* WSDM '14 (pp. 373–382). New York, NY, USA: ACM. URL: `http://doi.acm.org/10.1145/2556195.2556225`. doi:`10.1145/2556195.2556225`.

Jiang, J. Q. (2011). Learning protein functions from bi-relational graph of proteins and function annotations. In T. M. Przytycka, & M.-F. Sagot (Eds.), *Algorithms in Bioinformatics* (pp. 128–138). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*, 3.

Karp, R. M. (1975). On the computational complexity of combinatorial problems. *Networks*, *5*, 45–68.

Knobbe, A. J., Siebes, A., Wallen, D. V. D., & Syllogic B., V. (1999). Multi-relational decision tree induction. In *In Proceedings of PKDD' 99, Prague, Czech Republic, Septembre* (pp. 378–383). Springer.

Latouche, P., & Rossi, F. (2015). Graphs in machine learning: an introduction.

Leiva, H. A., Gadia, S., & Dobbs, D. (2002). Mrdtl: A multi-relational decision tree learning algorithm. In *Proceedings of the 13th International Conference on Inductive Logic Programming (ILP 2003* (pp. 38–56). Springer-Verlag.

Milner, R. (1989). *Communication and Concurrency.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Nguyen, P. C., Ohara, K., Motoda, H., & Washio, T. (2005). Cl-gbi: A novel approach for extracting typical patterns from graph-structured data. In T. B. Ho, D. Cheung, & H. Liu (Eds.), *Advances in Knowledge Discovery and Data Mining: 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam, May 18-20, 2005. Proceedings* (pp. 639–649). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: `http://dx.doi.org/10.1007/11430919_74`. doi:`10.1007/11430919_74`.

Wenfei Fan. *Graph Pattern Matching Revised for Social Network Analysis.* In *Proceedings of the 15th International Conference on Database Theory (ICDT '12)*, 2012. ISBN: 9781450307918. Publisher: Association for Computing Machinery. Address: New York, NY, USA. Pages: 8–21. Num. Pages: 14. Location: Berlin, Germany. DOI: 10.1145/2274576.2274578.

Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Juncheng Liu, Bryan Hooi. *NodeAug: Semi-Supervised Node Classification with Data Augmentation.* En *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20)*, 2020, páginas 207–217. DOI: `https://doi.org/10.1145/3394486.3403063`.

Seyed Mehran Kazemi and David Poole. *RelNN: A Deep Neural Model for Relational Learning.* In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, University of British Columbia, Vancouver, Canada, 2018. Email: `smkazemi@cs.ubc.ca`, `poole@cs.ubc.ca`.

Maria Leonor Pacheco and Dan Goldwasser. *Modeling Content and Context with Deep Relational Learning. Transactions of the Association for Computational Linguistics*, 2021; 9, 100–119. DOI: `https://doi.org/10.1162/tacl_a_00357`.

K. Ahmed, A. Altaf, N.S.M. Jamail, F. Iqbal, R. Latif. *ADAL-NN: Anomaly Detection and Localization Using Deep Relational Learning in Distributed Systems. Applied Sciences*, 2023, 13, 7297. DOI: `https://doi.org/10.3390/app13127297`.

Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Maosong Sun. *Graph Neural Networks: A Review of Methods and Applications. AI open*, 2020, 1, 57-81.

Lingfei Wu, Peng Cui, Jian Pei, Liang Zhao, Xiaojie Guo. *Graph Neural Networks: Foundation, Frontiers and Applications*. En *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, 2022, páginas 4840–4841. DOI: `https://doi.org/10.1145/3534678.3542609`.

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, *104*, 11–33.

Namkyeong L., Dongmin H., Gyoung S. Na, Sungwon K., Junseok L. and Chanyoung P. (2023). Conditional Graph Information Bottleneck for Molecular Relational Learning.

Plotkin, G. (1972). Automatic methods of inductive inference .

van Rest, O., Hong, S., Kim, J., Meng, X., & Chafi, H. (2016). Pgql: a property graph query language. In *Proceedings of the Fourth International Workshop on Graph Data Management Experiences and Systems* (p. 7). ACM.

Reutter, J. L. (2013). *Graph Patterns: Structure, Query Answering and Applications in Schema Mappings and Formal Language Theory*. Ph.D. thesis Laboratory for Foundations of Computer Science School of Informatics University of Edinburgh.

Segaran, T., Evans, C., Taylor, J., Toby, S., Colin, E., & Jamie, T. (2009). *Programming the Semantic Web*. (1st ed.). O'Reilly Media, Inc.

Tang, L., & Liu, H. (2009). Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 817–826). ACM.

Zou, L., Chen, L., & Özsu, M. T. (2009). Distance-join: Pattern match query in a large graph database. *Proc. VLDB Endow.*, *2*, 886–897. URL: `http://dx.doi.org/10.14778/1687627.1687727`. doi:`10.14778/1687627.1687727`.

Shuai Ma, Yang Cao, Wenfei Fan, Jinpeng Huai, Tianyu Wo. *Strong Simulation: Capturing Topology in Graph Pattern Matching. ACM Trans. Database Syst.*, January 2014, Volume 39, Number 1, Article No. 4. ISSN: 0362-5915. Publisher: Association for Computing Machinery. Address: New York, NY, USA. DOI: 10.1145/2528937. Keywords: dual simulation, graph simulation, data locality, Strong simulation, subgraph isomorphism.