1 Introduction

1.1 Resumen:

The introduction section of the paper discusses the importance and challenges of relational learning in machine learning, particularly in the context of graph-based data representation. It highlights the two main approaches to relational learning - latent feature (connectionist) approach and graph pattern-based (symbolic) approach. It also points out the limitations of existing query systems based on graph isomorphisms and their computational complexity, as well as the lack of efficient query expansion operations. The paper aims to present a novel graph query framework that addresses these limitations and enables controlled complexity in relational learning methods through atomic operations and polynomial time evaluation of cyclic patterns.

1.2 Evaluación:

Evaluation Criteria:

- Motivation
- Novelty
- Clarity
- Grammar and Style
- Typos and Errors

Evaluation Levels:

- YES
- Can be improved
- Must be Improved
- Not Applicable

Evaluation Justification and Examples from the Evaluated Section:

Motivation:

The section clearly explains the study's significance and relevance. The problem of relational learning is well justified, with specific examples provided of its applications in different domains, such as social network classification, protein characterisation, and toxic effect identification.

Improvement:

Suggest ways to strengthen the motivation, such as using data or references to highlight the problem's importance.

Novelty:

The section clearly describes the proposed approach's novelty or originality. It differentiates itself from existing work by addressing two fundamental problems in relational learning: computational complexity arising from relational queries and the lack of robust and general frameworks for symbolic relational learning methods.

Improvement:

Suggest ways to emphasize the novelty, such as explicitly comparing with related work and highlighting unique contributions.

Clarity:

The section is well-written and easy to understand. It uses appropriate terminology and avoids ambiguity. The text flows smoothly, making it comprehensible for readers who are familiar with relational learning concepts.

Improvement:

Suggest ways to improve clarity, such as restructuring complex sentences, defining technical terms, and using illustrative examples.

Grammar and Style:

The section is free of grammatical and stylistic errors. The language used is appropriate for an academic setting.

Improvement:

No specific grammatical corrections or stylistic improvements are needed.

Typos and Errors:

There are no typos or other errors in the section.

Improvement:

No specific corrections are required.

Overall, the Introduction section meets all evaluation criteria and is well-written, clear, and engaging for readers interested in relational learning concepts.

2 Related work

2.1 Resumen:

The Related work section of the paper discusses various approaches to graph pattern matching and relational learning. It highlights two types of relational learning models: latent feature approach and graph-pattern based approach, with a focus on the latter. The section briefly reviews ILP, TILDE, MRDTL, Selection Graphs, and DT-GBI as representative methods for graph-pattern based relational learning. The paper aims to develop an approach that supports learning from general subgraphs as base cases and is capable of executing cyclic queries, which sets it apart from the previously discussed techniques.

2.2 Evaluación:

Motivation:
Evaluation Level: YES
Evaluation justification and exemples from the evaluated section: The section clearly explains the study's significance and relevance by providing an overview of existing approaches to relational querying. It highlights the limitations of these methods, such as their inability to evaluate non-existence of elements, lack of support for cyclic patterns, and insufficiency when dealing with general subgraphs.

Improvement: The motivation could be strengthened by providing specific examples or statistics that demonstrate the impact of these limitations on real-world applications.

Novelty:
Evaluation Level: YES
Evaluation justification and exemples from the evaluated section: The section clearly describes the proposed approach's novelty or originality by comparing it with existing work in the field of graph pattern matching. It emphasizes that their proposal supports learning from general subgraphs as base cases, which distinguishes it from other methods.

Improvement: The novelty could be further emphasized by explicitly stating the unique contributions and advantages of the proposed approach compared to other related works.

Clarity:
Evaluation Level: Can be improved
Evaluation justification and exemples from the evaluated section: Although the section is generally well-written, some complex sentences may need restructuring for better comprehension. For instance, "While systems for querying based on graph isomorphism present NP complexity, those based on simulations present polynomial complexity" could be rephrased to improve clarity.

Improvement: Suggest ways to improve clarity by restructuring complex sentences and using illustrative examples when necessary.

Grammar and Style:
Evaluation Level: Can be improved

Evaluation justification and exemples from the evaluated section: The section contains a few grammatical and stylistic errors, such as "As stated above, there are two fundamentally different types of relational learning models" lacking an article before "types." Additionally, the use of acronyms like "ILP" should be accompanied by their full form in parentheses for better understanding.

Improvement: Suggest specific grammatical corrections and stylistic improvements, such as using more concise and precise language throughout the section.

Typos and Errors:
Evaluation Level: Can be improved
Evaluation justification and exemples from the evaluated section: There are a few typos and errors in the text, like "graphlog" instead of "GraphLog," which should be corrected for accuracy.

Improvement: Suggest specific corrections for typos and other errors found in the text.

3 Relational machine learning

3.1 Resumen:

The Relational Machine Learning section of the manuscript presents a technique for acquiring relational classifiers on graph data sets using the framework introduced. This is achieved through a top-down decision tree induction that employs graph queries as test tools in the internal nodes. The process involves determining which refinement set generates the maximum information gain while separating the training set and applying it to the initial query. Examples of relational learning are provided, including node classification problems using a social network toy and character species classification in the Star Wars toy graph.

3.2 Evaluación:

Motivation:
Evaluation Level: Can be improved
Evaluation justification and exemples from the evaluated section: The section clearly explains the study's significance and relevance. However, it could benefit from providing specific examples of problems and wider impacts to justify the importance further.
Improvement suggestion: Provide a few real-world applications or challenges where relational machine learning can be applied, such as fraud detection in financial transactions or identifying relationships between proteins in biological networks.

Novelty:
Evaluation Level: YES
Evaluation justification and exemples from the evaluated section: The section clearly describes the proposed approach's novelty or originality by explaining how it leverages graph query framework for relational learning. It differentiates itself from existing work by using a top-down decision tree induction with graph queries as test tools, resulting in queries that define classes within the graph dataset.

Clarity:
Evaluation Level: Can be improved
Evaluation justification and exemples from the evaluated section: The section is well-written but could benefit from restructuring some complex sentences for better comprehension. For example, the sentence "The procedure for tree learning is standard" might be clearer if it was rephrased to describe the process more explicitly.
Improvement suggestion: Provide a brief overview of the relational decision tree learning process and then explain each step in detail.

Grammar and Style:
Evaluation Level: Can be improved
Evaluation justification and exemples from the evaluated section: The section has some grammatical errors, such as "a top-down decision tree induction will be conducted" instead of "a top-down decision tree induction is conducted." It also lacks concise language in places.
Improvement suggestion: Correct all grammatical and stylistic errors, and use more concise and

precise language throughout the section.

Typos and Errors:
Evaluation Level: Can be improved
Evaluation justification and exemples from the evaluated section: The section contains some typos and errors that should be corrected for accuracy, such as "" instead of "S" in the equation.
Improvement suggestion: Identify all typos and other errors and correct them accordingly.

Overall evaluation:
Evaluation Level: Can be improved
The section provides a clear overview of the proposed approach but could benefit from improvements in clarity, grammar, style, and accuracy to enhance its quality and readability.

4 Conclusions and future work

4.1 Resumen:

The paper presents a novel graph query framework that allows for polynomial cyclic assessment of queries and refinements based on atomic operations. It fulfills essential requirements such as consistent grammar, support for subgraph evaluation, and automated query construction via refinements. The framework can be implemented on hypergraph data and offers potential for future research in developing automated methods to generate refinement sets. Patterns obtained from the graph learning procedure can serve as features for other machine learning methods, and it is concluded that effective techniques for matching graph patterns and learning symbolic relationships are feasible. Future work includes investigating probabilistic amalgamation of queries, evaluating additional machine learning algorithms alongside this query framework, and exploring opportunities for relational learning.

4.2 Evaluación:

Evaluation Criteria: Evaluation Level, Evaluation justification and exemples from the evaluated section.

Motivation: YES
The section clearly explains the study's significance and relevance by highlighting the problem of exponential complexity in graph isomorphism-based query systems and the need for a novel framework that enables polynomial time evaluation of cyclic patterns.

Improvement: The motivation could be further strengthened by providing specific examples or real-world scenarios where this novel framework would be particularly useful.

Novelty: YES
The section describes the proposed approach's originality and differentiates it from existing work by emphasizing its ability to assess subgraphs beyond individual nodes, support cyclic queries within polynomial time, and utilize a consistent grammar for both queries and evaluated structures.

Improvement: The novelty could be further emphasized by providing more explicit comparisons with related work and highlighting the unique contributions of the proposed framework.

Clarity: YES
The section is well-written and easy to understand, using appropriate terminology and avoiding ambiguity. It clearly explains the main contribution of the paper and demonstrates its capabilities through experimentation in relational learning procedures.

Improvement: There are no specific suggestions for improvement in terms of clarity.

Grammar and Style: YES
The section is free of grammatical and stylistic errors, using language appropriate for an academic setting.

Improvement: There are no specific suggestions for improvement in terms of grammar and style.

Typos and Errors: YES
The section is free of typos and other errors.

Improvement: There are no specific suggestions for improvement in terms of typos and errors.

Overall, the "Conclusions and future work" section meets all evaluation criteria and provides a clear understanding of the study's significance, originality, and contributions.

Acknowledgements

Proyecto PID2019-109152G financiado por MCIN/AEI/10.13039/501100011033

DISARM project - Grant n. PDC2021-121197, and the HORUS project - Grant n. PID2021-126359OB-I00 funded by MCIN/AEI/310.13039/501100011033 and by the "European Union NextGenerationEU/PRTR"

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.