1 Introduction

1.1 Resumen:

The introduction section of the research paper provides an overview of relational learning and its two basic approaches: latent feature or connectionist approach, and graph pattern-based approach. The authors highlight that the latter has been less successful due to computational complexity issues arising from relational queries and lack of robust and general frameworks for this type of symbolic relational learning methods.

The paper proposes a novel graph query framework that aims to solve these two fundamental problems by providing efficient pattern matching with controlled complexity and stepwise pattern expansion using well-defined operations. The authors state that their study focuses on formalizing an efficient graph query system and defining a set of operations for refining queries, but does not conduct extensive analysis or comparison in terms of performance or efficiency compared to other methods.

The paper is structured as follows: Section 1 provides an overview of related research; Section 2 introduces the novel graph query framework with its main definitions and properties that guarantee its utility; representative query examples are presented along with computational complexity analysis. Finally, section 3 describes implementation details for performing relational machine learning using this new approach while section 4 summarizes conclusions drawn from this investigation including potential avenues for future research

1.2 Evaluación:

Evaluation Criteria:

2 Relational machine learning

2.1 Resumen:

In this section, the authors introduce a top-down decision tree induction method to obtain characteristic patterns of subgraph classes using their graph query framework. They propose leveraging information gain as a criterion for selecting refinement sets during the tree construction process.

The training set consists of pairs (S, y), where S is a subgraph and y represents its associated class. The decision trees derived from this approach are not predominantly binary like those in the literature.

The authors provide examples to demonstrate relational learning using their query framework and refinement sets on small social network data, which accurately assign types (User A, User B, or Item) to all nodes in the graph by exploiting relational information from the network. They also show how they can classify characters' species in a Star Wars toy graph based on each character node and its corresponding specie property as training datasets. The leaf patterns of these trees characterize each species: human characters are born friends of Luke, while droids are unborn friends of Luke, wookies are those born in Kashyyk, etc.

Overall, this section presents an approach to relational machine learning using their graph query framework and refinement sets based on information gain as a criterion for selecting refinement sets during the tree construction process. They provide examples illustrating how they can classify nodes or character species in toy graphs accurately by exploiting relational information from those datasets through decision trees generated with this methodology.

2.2 Evaluación:

Evaluation Criteria:

- Motivation

Clarity: The section clearly explains the study's significance and relevance. The problem's importance and its wider impacts are justified. Specific examples from the text include "The objective is to classify the nodes based on the patterns extracted from the dataset." and "By utilizing each character node in the Star Wars toy graph (Figure 5) and the corresponding specie property as a training dataset, the relational decision tree shown in Figure 6 categorizes and explains each character's species in the graph."

Improvement: The motivation could be further strengthened by providing more concrete examples that demonstrate the practical applications of the proposed approach.

- Novelty

Originality: The section clearly describes the proposed approach's novelty or originality. It differentiates itself from existing work by introducing a relational decision tree algorithm specifically designed for graph data sets, utilizing refinement sets and queries to extract characteristic patterns.

Improvement: The novelty could be emphasized by explicitly comparing it with related works in the literature and highlighting its unique contributions more prominently.

- Clarity

Comprehension: The section is well-written and easy to understand. It uses appropriate terminology and avoids ambiguity. Specific examples from the text include "The training set, , consists of pairs (S, y), where S denotes a subgraph of G and y represents its associated class." and "Every node n in the resulting decision tree is linked to: - a subset of the training set: \subseteq ,"

Improvement: The clarity could be improved by restructuring complex sentences, defining technical terms more explicitly, and using illustrative examples.

- Grammar and Style

Correctness: The section is mostly free of grammatical and stylistic errors; however, some minor corrections are needed for a few instances to enhance the overall language usage in an academic setting.

Examples from the text include "Negative nodes/edges are identified with a cross, while nodes with predicate (v,S):= v S are larger and white in hue."

Improvement: Specific grammatical corrections and stylistic improvements could be suggested, such as using more concise and precise language.

- Typos and Errors

Accuracy: The section is mostly free of typos and other errors; however, a few minor corrections are needed to ensure accuracy in the text.

Examples from the text include "The leaf patterns of the tree characterize each species: human characters are born friends of Luke, while droids are unborn friends of Luke, wookies are those born in Kashyyk, etc."

Improvement: Specific corrections for typos and other errors should be suggested.

Evaluation Levels:

- Motivation: Can be improved

- Novelty: YES

- Clarity: Must be Improved

- Grammar and Style: Can be improved

- Typos and Errors: Must be Improved

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.