The Journal of Open Source Software

NiaARM: A minimalistic framework for Numerical Association Rule Mining

Žiga Stupan [1¶] **and Iztok Fister Jr.** [1]

1 University of Maribor, Faculty of Electrical Engineering and Computer Science ¶ Corresponding author

DOI: 10.21105/joss.04448

Software

- Review
- Repository
- Archive

Editor: Fabian Scheipl

Reviewers:

- @diegomcarvalho
- @fabian-s

Submitted: 27 April 2022
Published: 29 September 2022

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

Association Rule Mining (ARM) is a data mining method intended for discovering relations between attributes in transaction databases in the form of implications (Agrawal & Srikant, 1994; Fister Jr. & Fister, 2020). Traditional approaches, such as the Apriori algorithm (Agrawal & Srikant, 1994) or ECLAT (Zaki, 2000), require the attributes in the database to be discretized. This can result in the incorporation of noise into data, and potentially the obtained associations may not reveal the story fully (Varol Altay & Alatas, 2020). On the contrary, Numerical Association Rule Mining (NARM) is an extension of ARM that allows handling numerical attributes without discretization (Fister Jr. et al., 2021; Kaushik et al., 2020). Thus, an algorithm can operate directly, not only with categorical but also with numerical attributes concurrently. Interestingly, most NARM algorithms are based on stochastic population-based nature-inspired algorithms, which proved to be very efficient in searching for association rules (Alatas et al., 2008; Kaushik et al., 2021).

The NiaARM framework is an extended implementation of the ARM-DE algorithm (Fister et al., 2018; Fister Jr. et al., 2021), where Numerical Association Rule Mining is modeled as a single objective, continuous optimization problem, where the fitness is a weighted sum of the support and confidence of the built rule. The approach is extended by allowing the use of any optimization algorithm from the related NiaPy framework (Vrbančič et al., 2018) and having the option to select various interest measures and their corresponding weights for the fitness function.

The flow of the NiaARM framework is shown in Figure 1. Users have the option to construct a dataset either from a CSV file or a Pandas Dataframe. The dataset is then used to build the optimization problem, along with user selected interest measures to be used in the computation of the fitness function. Then the optimization problem can be solved using any algorithm in the NiaPy library to mine association rules from the dataset. The rules can be exported to a CSV file, analyzed statistically, or visualized using the visualization methods implemented in the framework, such as the hill slopes method (Fister et al., 2020). A simple command-line interface for mining rules is also provided.

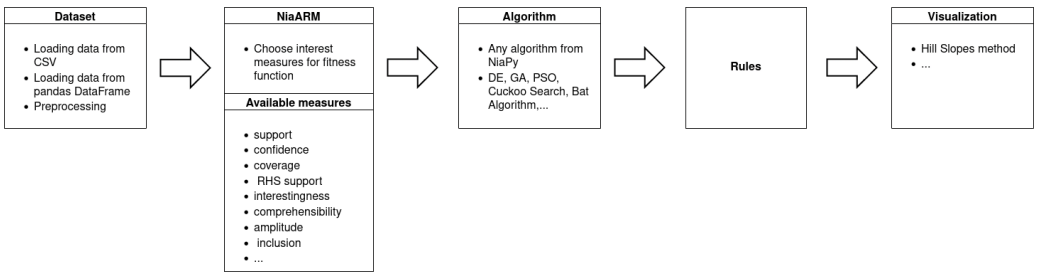

Figure 1: NiaARM flow.

Statement of need

Numerical Association Rule Mining plays a vital role in the data revolution era (Telikani et al., 2020). Several research papers that present NARM methods exist, but universal software where all primary tasks of NARM, i.e., preprocessing, searching for association rules, and visualization, is lacking. The NiaARM framework provides users with methods that allow them to preprocess their data, implement several interest measures, and powerful visualization techniques. In a nutshell, the benefits of the NiaARM framework are:

1. A simple way to mine association rules on numerical, categorical, or mixed attribute-type datasets.
2. Combined with the NiaPy library, it allows testing out the proposed approach using arbitrary nature-inspired algorithms.
3. A vast collection of implemented popular interest measures to measure the mined rules' quality.
4. Powerful visualization methods.
5. A simple command-line interface for easier handling with the proposed tool.

To the authors' knowledge, NiaARM is one of only three publicly available software solutions that implement any form of numerical association rule mining, the other two being KEEL (Alcalá-Fdez et al., 2009) and uARMSolver (Fister & Fister Jr, 2020). KEEL is a software tool used to assess evolutionary algorithms for machine learning problems of various kinds such as regression, classification, unsupervised learning, etc. It's a GUI application written in Java primarily intended for research and educational purposes. Although its scope is much wider it also includes some popular algorithms for numerical association rule mining including GAR, GENAR and MODENAR. The uARMSolver framework, written in C++, also implements the ARM-DE algorithm. Comparatively, NiaARM offers better ease of use, the ability to use arbitrary nature-inspired algorithms from the NiaPy framework (uARMSolver only implements DE and PSO), and the ability to optimize using more interest measures.

References

Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the 20th International Conference on Very Large Data Bases*, 487–499. ISBN: 9781558601536

Alatas, B., Akin, E., & Karci, A. (2008). MODENAR: Multi-objective differential evolution algorithm for mining numeric association rules. *Applied Soft Computing*, *8*(1), 646–656. https://doi.org/10.1016/j.asoc.2007.05.003

Alcalá-Fdez, J., Sanchez, L., Garcia, S., Jesus, M. J. del, Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., & others. (2009). KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, *13*(3), 307–318. https://doi.org/10.1007/s00500-008-0323-y

Fister, I., Fister, D., Iglesias, A., Galvez, A., Osaba, E., Del Ser, J., & Fister, I. (2020). Visualization of Numerical Association Rules by Hill Slopes. In C. Analide, P. Novais, D. Camacho, & H. Yin (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2020* (pp. 101–111). Springer International Publishing. https://doi.org/10.1007/978-3-030-62362-3_10

Fister, I., & Fister Jr, I. (2020). *uARMSolver: A framework for Association Rule Mining*. arXiv. https://doi.org/10.48550/arXiv.2010.10884

Fister, I., Iglesias, A., Galvez, A., Del Ser, J., Osaba, E., & Fister, I. (2018). Differential Evolution for Association Rule Mining Using Categorical and Numerical Attributes. In H. Yin, D. Camacho, P. Novais, & A. J. Tallón-Ballesteros (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2018* (pp. 79–88). Springer International Publishing. https://doi.org/10.1007/978-3-030-03493-1_9

Fister Jr., I., & Fister, I. (2020). A brief overview of swarm intelligence-based algorithms for numerical association rule mining. *arXiv:2010.15524 [Cs]*. https://doi.org/10.48550/ARXIV.2010.15524

Fister Jr., I., Podgorelec, V., & Fister, I. (2021). Improved Nature-Inspired Algorithms for Numeric Association Rule Mining. In P. Vasant, I. Zelinka, & G.-W. Weber (Eds.), *Intelligent Computing and Optimization* (pp. 187–195). Springer International Publishing. https://doi.org/10.1007/978-3-030-68154-8_19

Kaushik, M., Sharma, R., Peious, S. A., Shahin, M., Ben Yahia, S., & Draheim, D. (2020). On the potential of numerical association rule mining. *International Conference on Future Data and Security Engineering*, 3–20. https://doi.org/10.1007/978-981-33-4370-2_1

Kaushik, M., Sharma, R., Peious, S. A., Shahin, M., Yahia, S. B., & Draheim, D. (2021). A systematic assessment of numerical association rule mining methods. *SN Computer Science*, *2*(5), 1–13. https://doi.org/10.1007/s42979-021-00725-2

Telikani, A., Gandomi, A. H., & Shahbahrami, A. (2020). A survey of evolutionary computation for association rule mining. *Information Sciences*, *524*, 318–352. https://doi.org/10.1016/j.ins.2020.02.073

Varol Altay, E., & Alatas, B. (2020). Performance analysis of multi-objective artificial intelligence optimization algorithms in numerical association rule mining. *Journal of Ambient Intelligence and Humanized Computing*, *11*(8), 3449–3469. https://doi.org/10.1007/s12652-019-01540-7

Vrbančič, G., Brezočnik, L., Mlakar, U., Fister, D., & Fister, I. (2018). NiaPy: Python microframework for building nature-inspired algorithms. *Journal of Open Source Software*, *3*(23), 613. https://doi.org/10.21105/joss.00613

Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, *12*(3), 372–390. https://doi.org/10.1109/69.846291