Minor Revision based on Reviewers' Comments

Paper: Learning Semantic Association Rules from Internet of Things Data

We would like to thank the reviewers for their time and constructive feedback! We have addressed the reviewer's comments and answered their questions as described below. In our revised **paper**, we used blue font for the revised parts.

For the rest of this document:

Blue tone: refers to the feedback given by the reviewers.
Green tone: describes our actions and answers.

Reviewer 1 - Andreas Martin

The paper presents a technically sound and well-organized study on semantic association rule mining in IoT, combining static knowledge graphs with dynamic sensor data. The proposed Autoencoder-based Neurosymbolic ARM method effectively reduces the number of rules while maintaining full data coverage. The experimental evaluation is thorough and provides strong support for the approach.

Strengths:
- Novel and well-justified integration of semantic and sensor data for rule mining.
- Comprehensive evaluation across diverse ARM methods and datasets.
- Clear and relevant discussion of scalability, execution time, and potential extensions.

Issues Addressed in the Revision:
- The manuscript now conforms to the journal's formatting and layout requirements.
- The methodology and pipeline are more clearly explained, improving accessibility.
- The authors have added a discussion on model generalizability.
- The section on real-world scalability has been expanded.
- Practical implications, including integration into digital twins, are now explicitly addressed.
- The revision effectively resolves the initial concerns and enhances the clarity and applicability of the work.

We would like to thank the reviewer for their time and feedback, and for accepting our paper.

Reviewer 2 - Marvin Schiller

Sorry for the late review of the revision. I very much appreciate the inclusion of the running example. This makes the whole procedure much clearer. In particular, I noted that classes are used like categorical/nominal data ranges (exclusive values per feature), not like classes generally in OWL/RDFS ontologies (per default, these are not exclusive, but on the contrary, often part of a class hierarchy).

This is correct. We further clarified it in our paper and provided an explicit definition of the term *generically applicable* as having high data support and data coverage, as shown in ***Experimental Setting 1***.

Particular thanks also for the highlighting of the changes and the detailed discussion of the reviews.

We would like to thank the reviewer for their time and constructive feedback, which improved our paper significantly.

One remark just to make sure I understand correctly:
- In Equation 3 and 4, the "c_i" appear to be used to specify the count of different values in the range of a feature f_i, correct?

This is correct. We have now clarified this point explicitly before Equation 3 in the revised manuscript.

When looking through the bibliography, I found some items where the references are not up to journal standard. Examples are:

* Goldberg DE (2013) Genetic algorithms. pearson education India. --> Besides the typo, is it possible that this should rather refer to a textbook called "Genetic Algorithms in Search, Optimization, and Machine Learning" published by Addison Wesley?

We corrected the reference to "Goldberg DE (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Professional."

* Gruber T (1993) What is an ontology. --> This reference indirectly "points" (via google scholar) to a web page that has already disappeared.

We changed this reference to the same Author's another 1993 paper with ontology definition that is available online: "Gruber TR (1993) A translation approach to portable ontology specifications. Knowledge acquisition 5(2): 199–220."

* Kennedy J and Eberhart R; and Khedr et al: "IEEE" was formatted as "Ieee" and "ieee"

Fixed.

Typographic remarks:

- "lost function" typo on p.5 is still present (should be "loss function")
- p. 4 To enhance readibility, better write "e(e_k) = (v_i, v_j) or e(e_k) = (v_j ,v_i) --> Otherwise the lone appearance of "(v_j ,v_i)" is confusing
- p. 6 "a Neurosymbolic approach" --> lowercase the "n"
- p. 8 "input to the trained Autoencoder represents consequent" -> Add an article before "consequent", .e.g. "the consequent"

The typographic mistakes mentioned above are corrected.

Reviewer 3 - Anonymous Reviewer

The author's response to the reviews were really helpful in understanding the paper. A few suggestions:

We would like to thank the reviewer for their time and constructive feedback, which improved our paper significantly.

(1) The major hypothesis in the paper is that using an under-complete AE helped in extracting only the significant rules. It would be really nice if this hypothesis can be shown empirically by experimenting with an over-complete AE and showing that the rules extracted from under complete-AE is a sub-set of the rules extracted from over-complete AE.

We agree with the reviewer that different AE architectures, including an over-complete AE, can also be investigated to test their strength in learning a concise set of rules. In the ***Conclusion and Future Work*** section, we listed the evaluation of different AE architectures, as well as other neural network architectures, for their capability to capture strong associations between data features. This also includes supporting AE's neural representation with a graph neural network-based representation to better capture semantic relations in an IoT knowledge graph.

(2) It would be nice to see the relationship between the rules extracted by the proposed method and the baselines. Are the rules extracted by the proposed method a subset of the rules extracted by the baseline? Or, are these rules not extracted by the baselines but is identified only by the proposed method? If so, could you present coverage or support of the rules exclusively extracted by the proposed method separately? A similar presentation is useful when comparing under-complete vs over-complete AE.

We agree that understanding the relationship between rules extracted by different methods can offer a useful perspective.

In our case, a direct one-to-one comparison is challenging due to the different nature of the methods: optimization-based approaches generate numerical intervals that are not directly comparable to the discretized rules produced by any other method, and exhaustive methods can indeed recover all rules found by our method—given sufficiently low support and confidence thresholds—but at the cost of rule explosion and impractical runtimes (as shown in ***Experimental Setting 2***).

Our neurosymbolic method instead prioritizes conciseness, aiming to extract a smaller set of high-quality, semantically meaningful rules, while having full data coverage. This design makes average rule quality a more appropriate evaluation metric than one-to-one comparison. As such, we follow the common ARM practice of evaluating based on aggregate quality measures, which we report in ***Experimental Setting 2***.

We have discussed these considerations in the ***Discussion*** section ("*Neurosymbolic methods can help learning a concise set of high-quality rules*") and agree that a deeper rule-level comparison (e.g., support

or coverage of exclusive rules) is an interesting direction. We also note in the ***Conclusion and Future Work*** section that downstream task evaluation—e.g., rule-list classifiers trained from our extracted rules—can serve as an additional validation of their practical relevance, and we plan to pursue this in future work.