Semantic Rule Learning from Internet of Things Data

Erkan Karabulut [1], Paul Groth [1], and Victoria Degeler [1]

[1] University of Amsterdam, The Netherlands
{e.karabulut,p.t.groth,v.o.degeler}@uva.nl

Association Rule Mining (ARM) is the task of discovering commonalities in data in the form of logical implications. ARM is used in the Internet of Things (IoT) for different tasks including monitoring and decision-making. However, existing methods give limited consideration to IoT-specific requirements such as heterogeneity and volume. Moreover, they do not utilize important static domain-specific description data about IoT systems which is often represented as a knowledge graph with an underlying domain ontology [1].

We propose a novel pipeline that integrates dynamic sensor data and knowledge graphs [2, 3] to learn semantic association rules. Sensor data is enriched by the corresponding semantic properties from the knowledge graph before the rule learning process. An example rule based on sensor data only looks as follows: *'if sensor1 measures a value in range R, then sensor2 must measure a value in range R2'.* This rule can only be applied to *sensor1* and *sensor2*. In contrast, rules learned from sensor data and knowledge graphs are more contextual as seen in the following example in the water network domain: *'if a water flow sensor placed in a pipe P1 with diameter* $\geq$ *A1 measures a value in range R, then a water pressure sensor placed in a junction J1 connected to P1 measures a value in range R2'.* The semantic association rule is no longer about individual sensors. Instead, it describes a certain context that a sensor is placed in and therefore is generically applicable to multiple sensors and also more explainable.

Furthermore, we propose an Autoencoder-based ARM method as part of the proposed pipeline to address the high volume of IoT data which can lead to a high number of rules that are resource-intensive to process. The Autoencoder-based ARM learns a neural representation of the semantically enriched sensor data, and then our algorithm described in [2] extracts rules from the neural representation. Extensive evaluations on 3 IoT datasets from 2 domains in comparison to 8 baselines show that ARM on both static and dynamic IoT data results in more generically applicable rules while the Autoencoder-based ARM approach can learn a concise set of high-quality association rules with full coverage over the datasets.

References

[1] E. Karabulut, S. F. Pileggi, P. Groth, and V. Degeler, "Ontologies in digital twins: A systematic literature review," *Future Generation Computer Systems*, vol. 153, pp. 442–456, 2024.

[2] E. Karabulut, V. Degeler, and P. Groth, "Ae semrl: Learning semantic association rules with autoencoders," *arXiv preprint arXiv:2403.18133*, 2024.

[3] E. Karabulut, V. Degeler, and P. Groth, "Semantic association rule learning from time series data and knowledge graphs," in *Proceedings of the 2nd International Workshop on Semantic Industrial Information Modelling (SemIIM 2023) co-located with 22nd International Semantic Web Conference (ISWC 2023)*, pp. 1–7, 2023.