Similarity learning for fuzzy rough rule induction

Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression [1].

FRS make use of a similarity relation, which expresses the degree to which two elements in a dataset are related. Instead of using a normal, fixed similarity relation, it is possible to apply similarity learning or distance metric learning (DML)[2] to learn the optimal relation from the data. This has been successfully applied to neighborhoud-based machine learning methods, such as k-nearest neighbours.

Rule induction is a machine learning model which learns rules from the data with the aim of e.g. predicting the class of new samples. One such technique which makes use of FRS theory, is QuickRules [3], which creates rules in a greedy way.

The goal of this thesis is to design and conduct an extensive experimental study on a collection of benchmark datasets, in order to evaluate the predictive and descriptive potential of similarity learning in the context of rule induction.

The specific research questions that the thesis should address are:

• How can we apply DML to rule induction?

• Which DML methods perform best in this context, what are their strengths and weaknesses w.r.t. predictive and descriptive performance?

• Can we improve these methods or create our own technique which outperforms them?

• Are there specific kinds of data (e.g., imbalanced data, high-dimensional data, multi-label data, numerical vs categorical data, ...) on which the approach performs better or worse?

Prior knowledge of fuzzy rough sets is not necessary, and a Python implementations of QuickRules will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

References

[1] S. Vluymans, et al. “Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey.” Fundamentae Informaticae 142.1-4 (2015): 53-86.
[2] J. L. Suárez, S. García, F. Herrera, A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges, Neurocomputing 425 (2021) 300–322.
[3] Richard Jensen, Chris Cornelis, and Qiang Shen. Hybrid fuzzy-rough rule induction and feature selection. pages 1151 – 1156, 09 2009.

Experimental evaluation of new fuzzy rough set models for machine learning tasks

Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression. [1]

A prominent example is fuzzy rough nearest neighbour classification (FRNN), which is a lazy learner that associates an upper and a lower approximation with each decision class and classifies test instances according to their membership in these. This has a transparent interpretation: upper approximation membership encodes to what extent a test instance is similar to the training instances of a class, and so possibly belongs to this class, whereas lower approximation membership encodes to what extent a test instance is not similar to the training instances of other classes and so necessarily belongs to this class.

One of the most promising FRS models uses Ordered Weighted Averaging (OWA) operators in the calculation of the upper and lower approximations. Recently, a new class of fuzzy rough set models have been developed at Ghent University, based on fuzzy quantification models. [2]

The goal of this thesis is to design and conduct an extensive experimental study on a collection of benchmark datasets, in order to evaluate the machine learning potential of the newly proposed FRS models.

The specific research questions that the thesis should address are:

- Are the new fuzzy rough set models based on fuzzy quantification an improvement on the existing OWA FRS model for feature selection?
- What are the strengths and weaknesses of each FRS model?
- Are there specific kinds of data (e.g., imbalanced data, high-dimensional data, multi-label data, numerical vs categorical data, ...) on which the approach performs better or worse?

Prior knowledge of fuzzy rough sets is not necessary, and Python implementations of the different FRS models will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

References

[1] Vluymans, Sarah, et al. "Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey." Fundam. Informaticae 142.1-4 (2015): 53-86.

[2] Theerens, Adnan, and Chris Cornelis. "Fuzzy Rough Sets Based on Fuzzy Quantification." arXiv preprint arXiv:2212.04327 (2022).