Mining Exceptional Models with Rank Correlation

Lennart Downar[1]

[1] *Technische Universtität Dortmund, Germany*

ABSTRACT

Subgroup Discovery (SD) is a Local Pattern Mining task, that classically tries to find subgroups of a given dataset with unusual distributions for a specific target attribute. *Exceptional Model Mining* (EMM) has been proposed as a generalization of SD by Leman et al. [1] and was thoroughly introduced by Duivesteijn [2]. It allows for multiple target attributes and more complex models to be applied to possible subgroups (e.g., Regression, Correlation or Bayesian Networks). In the EMM framework, the attributes of the dataset are partitioned into two: one part (the descriptors) is used to define subgroups on, and one part (the targets) is used to evaluate subgroups on. The concept of interest in subgroups is captured by learning, from (a subset of) the dataset, a model fitted on the targets.

The general development of a an EMM Model Class consists of defining a quality measure to gauge the exceptionality of a subgroup. In the existing Correlation Model Class for EMM, the exceptionality of a subgroup is expressed by the correlation between two targets. The employed correlation coefficient is the standard Pearson correlation: dividing the covariance of the targets by the product of their standard deviations. However, implicitly, this assumes that the two targets are normally distributed. In many real-life situations, this assumption is unjustified.

We have taken the existing Correlation Model Class and investigated, whether the theoretical limitations could be lifted by applying so-called rank correlation coefficients, which promise to detect monotone relationships instead of just linear ones and are in general non-parametric. We found that Spearman's rank correlation can substitute Pearson correlation and due to its mathematical background allows for broader applications. Additionally Kendall's rank correlation provides an alternative correlation-based quality measure.

References

[1] D. Leman, A. Feelders, and A. Knobbe, *Exceptional model mining*, ECML PKDD'08, vol. 5212, pp. 1–16, 2008.

[2] W. Duivesteijn, *Exceptional Model Mining*, PhD thesis, Leiden University, 2013.

[3] L. Downar, *A Rank Correlation Model Class for Exceptional Model Mining*, Bachelor thesis, TU Dortmund, to appear