Efficient and effective automatic feature discovery

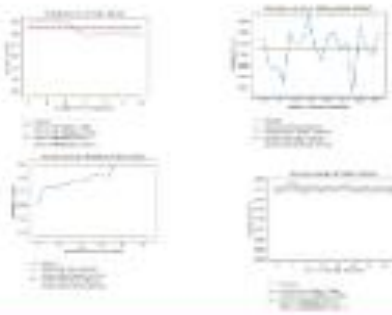
2022

Andreas Benedict Bien

How to make automatic feature discovery efficient and effective?

Experiments

Main Results



Methodology

Conclusion

The heuristic performs well. In 3/4 datasets it performs equally or better than all then all other experiments, with very K values less than 5.

Future work

Collect more data on the datasets for higher precision on further analysis:
A collection of data thousands of datasets instead of hundreds could allow for better analysis of the covariance between statistical characteristic scores, and the independence of the likelihoods..

Pair with machine learning:
The heuristic itself could be finetuned using a reinforcement learning algorithm, or possibly used as an input to train a machine learning model that predicts the utility of the feature.

Pair with other heuristics:
Pairing the model with heuristics that take into account the statistical values rather than just the rankings might synergize well with a heuristic that only considers ranking of features, as it is the heuristic I developed is less precise when all features similar statistical characteristics.