Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Multi-Label Rule Learning

Project Manager
Michael Rapp

Principal Investigator
Prof. Dr. Johannes Fürnkranz

Project Term
2019 - 2020

Clusters
Lichtenberg Cluster Darmstadt

Software
Python

Additional Software
Cython

University
Technische Universität Darmstadt

Introduction

This project aimed at solving multi-label classification problems using rule learning algorithms. Multi-label classification has received a lot of attention in the recent machine learning literature and is nowadays used in applications as diverse as music categorization, semantic scene classification, or protein function classification. Multi-label classification becomes particularly challenging when it comes to discovering hidden dependencies between labels. As rules provide a natural form of expressing such dependencies, they are a natural choice to capture label correlations. The Lichtenberg cluster was used to evaluate the predictive accuracy of the developed algorithms and existing baselines using synthetic and realworld benchmark data sets.

Methods

We utilized the well-known gradient boosting framework to develop a novel rule learning algorithm that is able to optimize different types of multivariate loss functions (decomposable or non-decomposable) and therefore can flexibly be tailored to different multi-label evaluation measures that are commonly used to assess the quality of predictions in multi-label classification.

Results

As shown by the experiments that have been conducted on the Lichtenberg cluster, our rule learning method can successfully be used to optimize decomposable and non-decomposable loss functions. It has further been shown that it is able to outperform

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

conventional state-of-theart boosting methods on data sets of moderate size.

Discussion

Existing boosting-based approaches to multi-label classification are usually limited to the use of decomposable loss function, which allows for a straight-forward adaption of gradient boosting from binary to multi-label classification. However, capturing dependencies between labels has become a key motivation for research in this machine learning domain. To address such use cases, we proposed a generalization of the gradient boosting framework that supports the minimization of non-decomposable loss function. In the future, it could could serve as a basis for developing new methods specifically tailored to this particular type of loss functions. As an instantiation of the framework, we presented a novel rule learning algorithm that has been shown to achieve state-of-the-art performance. A drawback of the proposed algorithm are the computational costs that come with the use of large data sets with many labels. To compensate for this, we plan to investigate approximations that exploit the sparsity in the label space in future work.

Publications

Rapp, M.; Loza Mencía, E.; Fürnkranz, J.; Nguyen, V.; Hüllermeier, E.: Learning Gradient Boosted Multi-label Classification Rules, in: Proceedings of the European Conference of Machine Learning (ECML2020), Springer, 2020 https://arxiv.org/abs/2006.13346

Last Update: 2022-04-22 12:10