

Combining Instance-Based Learning and Logistic Regression for Multilabel Classification*

Weiwei Cheng and Eyke Hüllermeier

Department of Mathematics and Computer Science
University of Marburg, Germany
{cheng,eyke}@mathematik.uni-marburg.de

Abstract. Multilabel classification is an extension of conventional classification in which a single instance can be associated with multiple labels. Recent research has shown that, just like for conventional classification, instance-based learning algorithms relying on the nearest neighbor estimation principle can be used quite successfully in this context. However, since hitherto existing algorithms do not take correlations and interdependencies between labels into account, their potential has not yet been fully exploited. In this paper, we propose a new approach to multilabel classification, which is based on a framework that unifies instance-based learning and logistic regression, comprising both methods as special cases. This approach allows one to capture interdependencies between labels and, moreover, to combine model-based and similarity-based inference for multilabel classification. As will be shown by experimental studies, our approach is able to improve predictive accuracy in terms of several evaluation criteria for multilabel prediction.

Reference

1. Cheng, W., Hüllermeier, E.: Combining Instance-Based Learning and Logistic Regression for Multilabel Classification. *Machine Learning* (2009) DOI: 10.1007/s10994-009-5127-5

* This is an extended abstract of an article published in the machine learning journal [1].