Liu, Jing; Guo, Zhongwen; Sun, Zhongwei; Liu, Shiyong; Wang, Xupeng

Summary: In large-scale multi-label classification, excessive computational complexity has severely restricted the application of non-linear kernel support vector machines (SVMs). Therefore, we proposed fast multi-label SVM classification algorithm using divide-and-conquer strategy (MLDC-SVM), which firstly uses the binary relevance problem transformation strategy to transform the multi-label classification problem into multiple binary classification problems. After that, each binary classification problem can be solved by using the improved binary SVM classification algorithm using divide-and-conquer strategy. The improvement is reflected in the use of DEC (Different Error Cost) method to overcome the label data imbalance issue. Finally, it achieves fast multi-label classification by integrating the solutions of each binary classification problems. This algorithm is superior to other fast multi-label classification algorithms in training speed, testing speed and testing performance. Experimental results on two large-scale multi-label data sets show that the training and testing speed of MLDC-SVM classification algorithm is the fastest. Meanwhile, the testing performance of MLDC-SVM classification algorithm is similar to that of ML-LIBSVM classification algorithm, which is better than that of other fast multi-label classification algorithms.

MSC:
68T05 Learning and adaptive systems in artificial intelligence

Keywords:
multi-label classification; support vector machine; non-linear kernel; divide-and-conquer strategy; label data imbalance issue; different error cost method

Full Text: DOI