**Title:** Classifier Chains for Multi-label Classification

**Author:** Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank

Department of Computer Science

The University of Waikato

Hamilton, New Zealand

**Year:** 2009

Published in:

**Binary Relevance Method(BM):**

BM transforms any multi-label problem into one binary problem for each label. Hence this method trains |L| binary classifiers C1, · · · , C|L|. Each classifier Cj is responsible for predicting the 0/1 association for each corresponding label lj ∈ L.

BM is mentioned throughout the literature but consistently sidelined on the grounds of its assumption of label independence. That is to say, during its transformation process, BM ignores label correlations that exist in the training data. The argument is that, due to this information loss, BM’s predicted label sets are likely to contain either too many or too few labels, or labels that would never co-occur in practice.

**Label combination method, or label power-set (CM):** The basis of this method is to combine entire label sets into atomic (single) labels to form a single-label problem for which the set of possible single labels represents all distinct label subsets in the original multi- label representation. Each (x,S) is transformed into (x,l) where l is the atomic label representing a distinct label subset. Overcomes the ignorance of label independence.

**Sidenote:**

The consensus view in the literature is that it is crucial to take into account label correlations during the classification process. However as the size of multi-label datasets grows, most methods struggle with the exponential growth in the number of possible correlations. Consequently, these methods are able to be more accurate on small datasets, but are not as applicable to larger datasets. This necessarily restricts their usefulness as many multi-label contexts involve large numbers of examples and labels.

**Binary Method Advantages:**

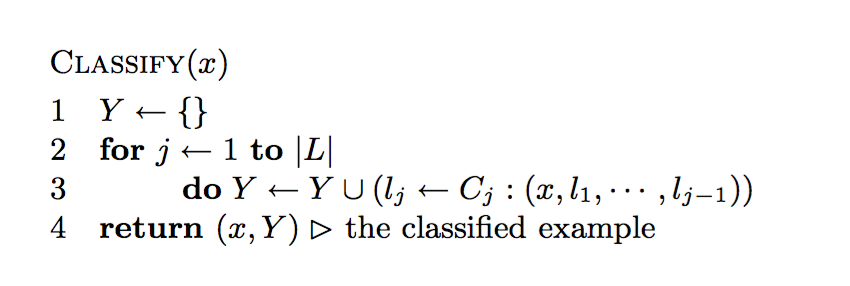
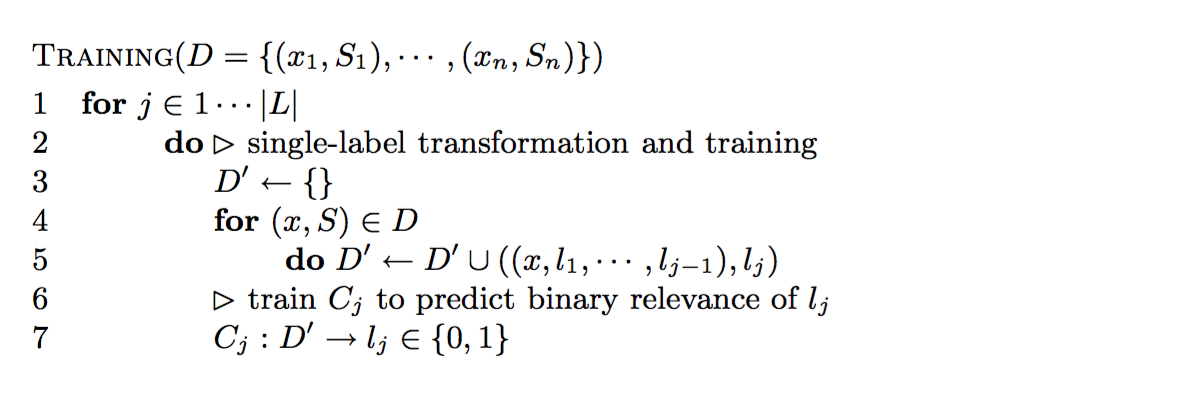
* Although it is mentioned to have many disadvantages because of its assumption of label independence, BM is suitable for cases where new data may be irrelevant to known data or where label relationships may change over the test data; even the label set L may be altered dynamically. - making BM ideal for active learning and data stream scenarios.

* However the most important and widely relevant advantage of BM is its low computational complexity compared to other methods. Given a constant number of examples, BM scales linearly with the size of the known label set L.

**Classifier Chain (CC)**:

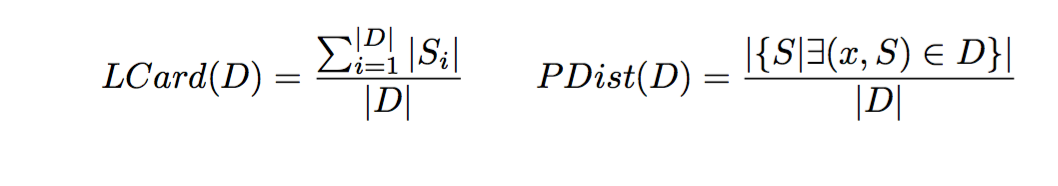
This method benefits from the complexity of the BM while overcoming the disadvantage of label independence of BM. The Classifier Chain model (CC) involves |L| binary classifiers as in BM.

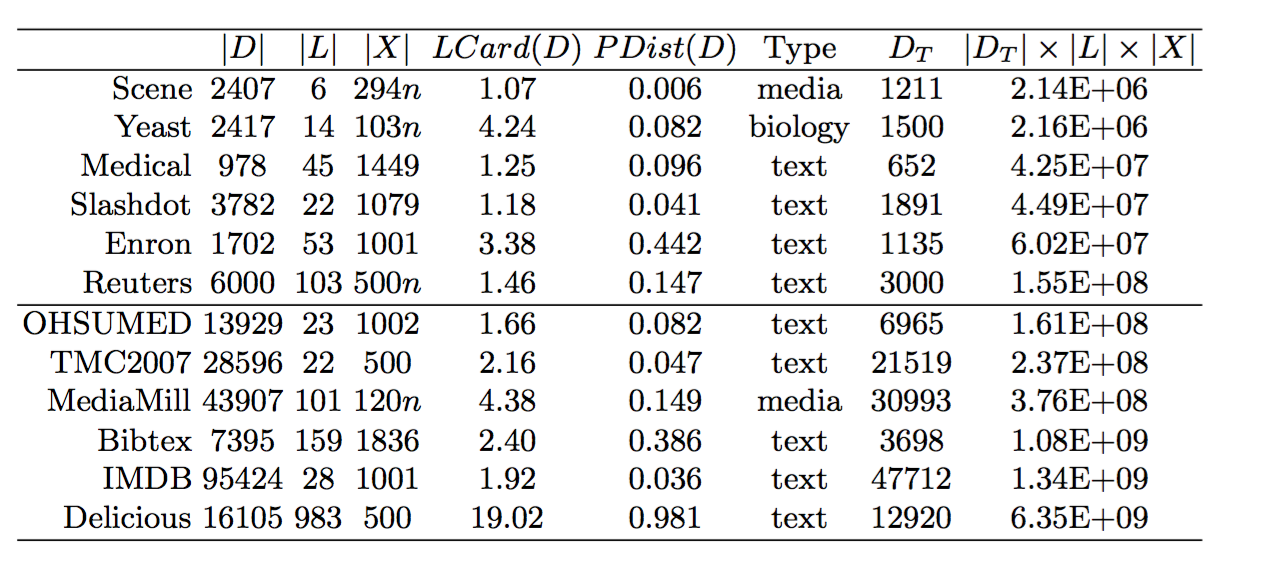
Hence a chain C1 , · · · , C|L| of binary classifiers is formed. Each classifier Cj in the chain is responsible for learning and predicting the binary association of label lj given the feature space, augmented by all prior binary relevance predictions in the chain l1, · · · , lj−1. The classification process begins at C1 and propagates along the chain: C1 determines P r(l1 |x) and every following classifier C2 · · · C|L| predicts Pr(lj|xi,l1,...,lj−1).



**Datasets**

Table below displays datasets from a variety of domains and their associated statistics. Label Cardinality (LCard) is a standard measure of “multi-labelledness”. It is simply the average number of labels relevant to each instance. The Proportion of Distinct label combinations (PDist) is simply the number of distinct label subsets relative to the total number of examples:





In total 12 datasets are used, with dimensions ranging from 6 to 983 labels, and from less than 1,000 examples to almost 100,000. The datasets are roughly ordered by complexity (|DT | × |L| × |X|) and divided between regular and large sizes. Included are two new real-world multi-label text collections: Slashdot, which is collected from http://slashdot.org, and IMDB from http://imdb.org (data obtained from http://www.imdb.com/interfaces#plain). All datasets and further information about them can be found at various sources1.

1 http://www.cs.waikato.ac.nz~/jmr30/#datasets and http://mlkd.csd.auth. gr/multilabel.html#Datasets

.