Classifying multiple categories through label and feature grouping

Kwanit Gupta (B19EE046) , Girish Pandey (M22CS056)

Abstract

Managing cases with several labels at once is the goal of multi-label classification (MLC), a difficult task in machine learning.

Effective feature selection and dimensionality reduction strategies are necessary to enhance the performance and efficiency of MLC models when dealing with the high-dimensional and noisy data often associated with MLC jobs.

To make use of the label-specific features and the label correlations in multi-label datasets, we offer a unique MLC framework that combines feature selection with label clustering.

Methodology

During data abstraction , the LIBSVM file was transformed to generate 2 sparse representation each for feature and label.

we split it into train and test data , for training we have used various
 Multilabel containers like Binary Relevance , classifier chain ,etc. with several
combinations of Base estimators like SVC , GaussianNB ,etc.
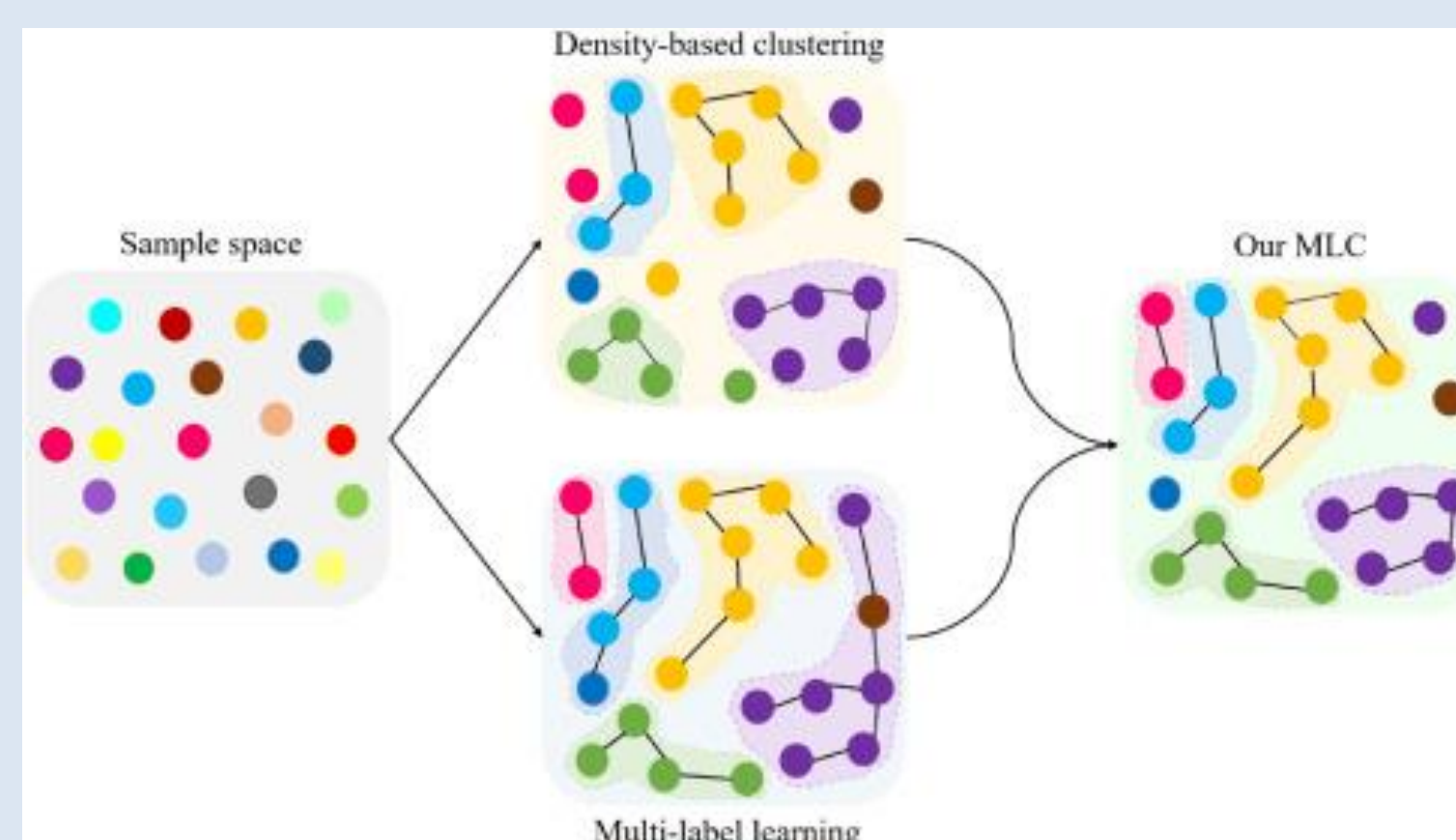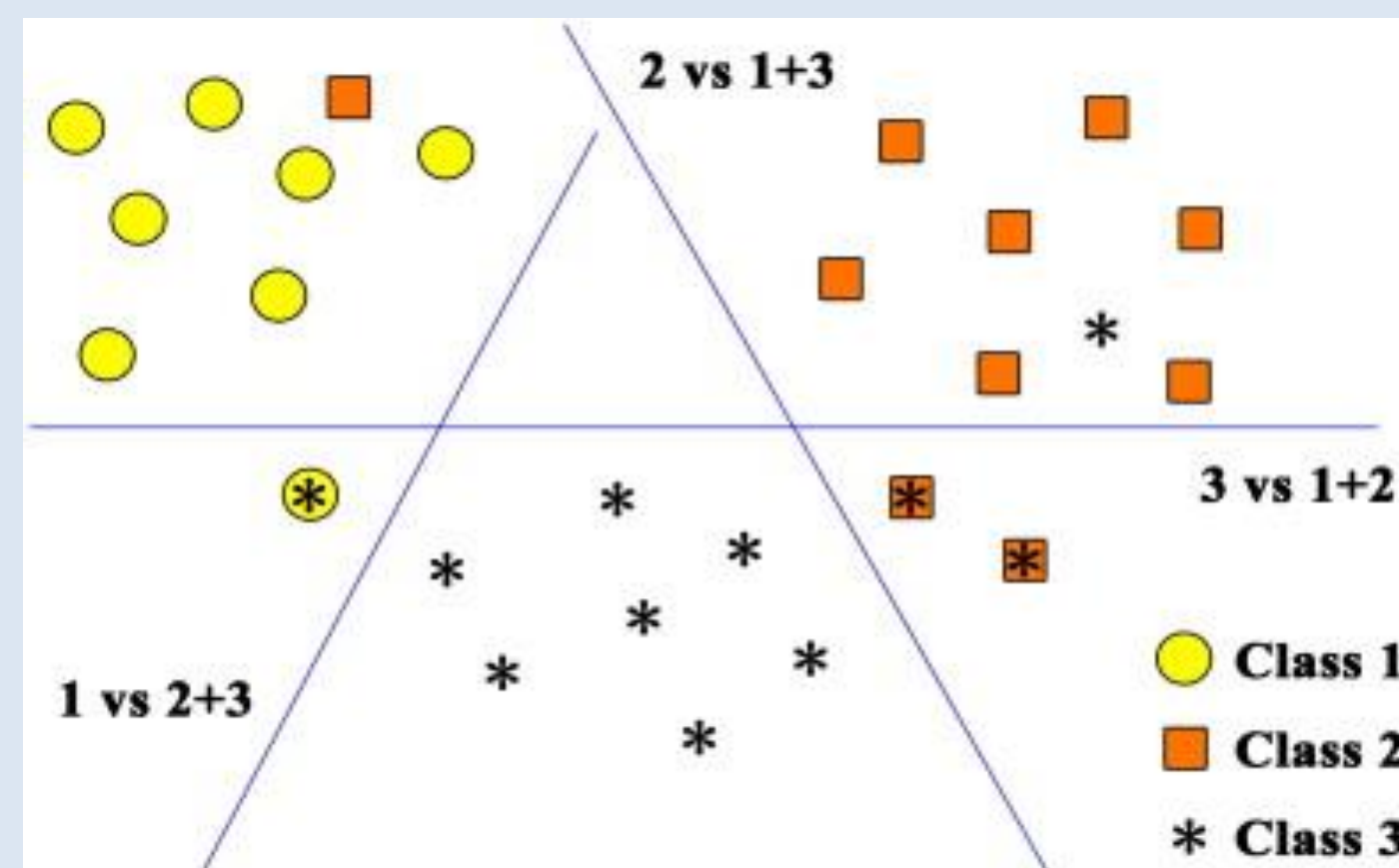
Result

Approaches	Train Time (s)	Predict Time (s)
Binary Relevance	60 – 1683	1 – 412
Classifier Chains	49 – 1415	1 – 431
Label Power Set	1 – 564	0.1 – 117
RakelD	16 – 1964	4 – 404

Table 1: Time Comparison with Approaches

Approaches	Virtual Usage (MiB)	Residual Usage (MiB)
Binary Relevance	3325 – 5402	2119 – 4198
Classifier Chains	3271 – 3394	2092 – 2184
Label Power Set	3412 – 4641	2165 – 3376
RakelD	3604	2339

Table 3: Memory Comparison with Approaches

Future work

Developing efficient and scalable online multilabel classification methods that can adapt to changing data distributions over time.

Developing transfer learning methods that can effectively leverage the shared structure of the label space and the data distribution.

Contribution

Coding and Reporting done by both