# Multilabel classification using heterogeneous ensemble of multi-label classifiers

Muhammad Atif Tahir , Josef Kittler , Ahmed Bouridane

a Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, UK

b School of Computing, Engineering and Information Sciences, University of Northumbria, Newcastle upon Tyne NE2 1XE, UK

Josu Maiora

# 1. Introduction

- A conventional multi-class classification system assigns each instance x a single label l from a set of disjoint labels L.

- In this paper each instance is to be assigned to a subset of labels $Y \subseteq L$. This problem is known as multi-label learning.

- They include highly imbalanced training sets, as very limited data is available for some labels, and capturing correlation among classes.

- In this paper, we focus on highly imbalanced data distributions using ensemble of multi-label classifiers.

# 1. Introduction

- Ensemble techniques are becoming increasingly important → they improve the accuracy with highly imbalanced data.

- Ensembles can be homogeneous (every base classifier using the same algorithm), or heterogeneous (different algorithms).

- The aim of this paper is to use **heterogeneous** ensembles of **multi-label** learners to improve the performance.

# 1. Introduction

- The proposed ensemble multilabel learning approach (EML)[1] is applied to six publicly available multi-label data sets

# 2. Related work

- Multi-label classification
  - Problem transformation methods
  - Algorithm adaptation methods
- RaKEL (Tsoumakas and Vlahavas, 2007),
- Calibrated Label Ranking (CLR) (Furnkranz et al., 2008),
- Multi-label KNN (MLKNN) (Zhang and Zhou, 2007),
- Instance Based Logistic Regression (IBLR) (Chengand Hullermeier, 2009) and
- Ensemble of Classifier Chains (ECC)(Read et al., 2009).

# 3. Ensemble of multi-label classifiers (EML)

- Let X denote a set of instances and let Y = {1, 2, . . . ,N} be a set of labels

$$S = \{(x_1, y_1), \ldots, (x_m, y_m)\}$$

- The goal is to design a multi-label classifier H that predicts a set of labels for an unseen example.

- Ensemble of multi-label classifiers train q multi-label classifiers H1,H2, . . . ,Hq. For an unseen instance x, each kth individual model (of q models) produces an N-dimensional vector Pk = [p1k,p2k,. . . ,pNk], where the value pbk is the probability of the bth class label assigned by classifier k being correct.

# 3. Ensemble of multi-label classifiers (EML)

- MEAN, MAX, MIN are the simplest and most popular way to combine the scores of classifiers

# 3. Ensemble of multi-label classifiers (EML)

1. Average of probabilities ($EML_A$)

$$\mu_b(x) = \frac{1}{q}\sum_{k=1}^{q} p_{bk}(x)$$

# 3. Ensemble of multi-label classifiers (EML)

## 2. Average of probabilities and threshold selection via multi-labelled-ness ($EML_T$)

- Properly adjusting the decision thresholds (instead of the traditional value of 0.5) can improve the performance of a multi-label classifier.
- Let $X_T$ be the training set and $X_S$ the test set. A threshold t is then selected using Eq. (2) to choose the final predicted multi-label set Z.

$$t = \arg \min_{\{t \in 0.00, 0.001, \ldots, 1.00\}} |LCard(X_T) - LCard(H_t(X_s))|$$

$$LCard(X) = \frac{\sum_{i=1}^{|X|} |E_i|}{|X|}$$

where Ei is the actual set of labels for the training set and a predicted set of labels under threshold t for the test set

# 3. Ensemble of multi-label classifiers (EML)

## 3. Static weighting by N-Fold Cross Validation ($EML_S$)

In static weighting, the weights for each classifier are computed in the training phase. In this paper, the weights for each classifier are learnt via N-Fold Cross Validation (N = 5)

# 3. Ensemble of multi-label classifiers (EML)

## 4. Dynamic weighting using Dudani rule ($EML_D$)

- A weighted k-NN rule is proposed for classifying new patterns
- The main idea is to weight a neighbor with smaller distance more heavily than the one with a greater distance

$$w_j = \begin{cases} \frac{d_k - d_j}{d_k - d_1} & \text{if } d_k \neq d_1 \\ 1 & \text{otherwise} \end{cases}$$

- For multi-label classifier

$$weight(M_j) = \begin{cases} \frac{d_q - d_j}{d_q - d_1} & \text{if } d_q \neq d_1 \\ 1 & \text{otherwise} \end{cases}$$

- w(Mj) is the weight of multi-label classifier j for instance x

# 3. Ensemble of multi-label classifiers (EML)

## 5. Dynamic weighting using Shepard rule (EMLP)

- Shepard:

  ‘‘the relevance of a previous stimulus for the generalization to a new stimulus is an exponentially decreasing function of its distance in psychological space’’.

$$weight(M_j) = e^{-\alpha d_j^{\beta}}$$

  α y β constants

# 4. Experiments

- Datasets: six multi-label datasets from a variety of domains
- Features: publicly available feature vectors are used for all datasets
- Evaluation measures: Hamming Loss, Accuracy, F1, and Classification Accuracy from the example-based category, and Micro/Macro F1/AUC from the label-based group. Additionally, we use One-error, Coverage, Ranking Loss and Average Precision from the ranking-based group

**Table 1**
Standard and multilabel statistics for the data sets used in the experiments.

| Datasets | Domain | Samples | Features | Labels | LCard |
|---|---|---|---|---|---|
| Enron | Text | 1702 | 1001 | 53 | 3.38 |
| Medical | Text | 978 | 1449 | 45 | 1.25 |
| Scene | Vision | 2407 | 294 | 6 | 1.07 |
| Pascal07 | Vision | 9963 | 500 | 20 | 1.44 |
| Yeast | Biology | 2417 | 103 | 14 | 4.24 |
| Emotions | Music | 593 | 72 | 6 | 1.87 |

# 4. Experiments

- Benchmark methods

**Table 4**
Comparison of the proposed ensemble method (EML) with the state-of-the-art multi-label classifiers for emotions.

| | MLkNN | IBLR | RAkEL | CLR | ECC | $EML_A$ | $EML_T$ | $EML_S$ | $EML_P$ | $EML_D$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Hamming Loss ↓ | 0.204 | 0.201 | 0.231 | 0.205 | 0.204 | **0.185**$^{*}$ | 0.199 | 0.200 | 0.199 | 0.200 |
| Accuracy ↑ | 0.512 | 0.523 | 0.482 | 0.522 | 0.564 | **0.579**$^{*}$ | 0.568 | 0.567 | 0.568 | 0.568 |
| Fmeasure ↑ | 0.625 | 0.630 | 0.599 | 0.635 | 0.681$^{+}$ | 0.694 | 0.701$^{+}$ | 0.700$^{+}$ | **0.702**$^{*}$ | 0.699$^{+}$ |
| ClassAcc ↑ | 0.261 | 0.288 | 0.227 | 0.267 | 0.304 | **0.326**$^{*}$ | 0.272 | 0.272 | 0.271 | 0.272 |
| Micro $F_1$ ↑ | 0.644 | 0.656 | 0.616 | 0.660 | 0.676 | **0.697**$^{*}$ | 0.680 | 0.680 | 0.680 | 0.679 |
| Macro $F_1$ ↑ | 0.608 | 0.632 | 0.603 | 0.647 | 0.663 | **0.673**$^{*}$ | 0.656 | 0.656 | 0.657 | 0.656 |
| Micro AUC ↑ | 0.844 | 0.851 | 0.811 | 0.844 | 0.828 | **0.871**$^{*}$ | 0.855 | 0.855 | 0.855 | 0.851 |
| Macro AUC ↑ | 0.820 | 0.832 | 0.793 | 0.827 | 0.821 | **0.855**$^{*}$ | 0.848 | 0.848 | 0.849 | 0.845 |
| One-error ↓ | 0.284 | 0.279 | 0.327 | 0.271 | 0.275 | 0.249$^{+}$ | 0.249$^{+}$ | **0.248**$^{*}$ | 0.251$^{+}$ | 0.254$^{+}$ |
| Coverage ↓ | 1.83 | 1.77 | 2.02 | 1.73 | 1.90 | **1.68**$^{*}$ | **1.68**$^{*}$ | **1.68**$^{*}$ | **1.68**$^{*}$ | 1.694 |
| Ranking Loss ↓ | 0.170 | 0.164 | 0.205 | 0.154 | 0.181 | **0.143**$^{*}$ | **0.143**$^{*}$ | **0.143**$^{*}$ | 0.144$^{+}$ | 0.147 |
| AvgPrecision ↑ | 0.791 | 0.798 | 0.762 | 0.807 | 0.795 | **0.818**$^{*}$ | **0.818**$^{*}$ | **0.818**$^{*}$ | 0.817$^{+}$ | 0.815 |
| # Wins (Ind) | 0/12 | 3/12 | 0/12 | 4/12 | 5/12 | – | – | – | – | |
| # Wins (All) | 0/12 | 0/12 | 0/12 | 0/12 | 0/12 | 10/12 | 3/12 | 4/12 | 2/12 | 0/12 |

# 4. Conclusions

- Heterogeneous ensemble of multi-label learners is proposed to simultaneously tackle both **class imbalance and class correlation problems**
- Ensemble methods are well-known for **overcoming over-fitting** problems and improving the performance of individual classifiers
- It has been shown that the presented approach provides a **very accurate and efficient solution** when compared with the state-of-the-art multi-label methods