A Novel Online Stacked Ensemble for Multi-Label Stream Classification

Alican Büyükçakır[1], Hamed Bonab[2] and Fazli Can[1]

[1]: Bilkent Information Retrieval Group, Computer Engineering Department, Bilkent University.
[2]: College of Information and Computer Sciences, University of Massachusetts Amherst.

CIKM 2018

Introduction

Multi-label stream classification (MLSC) is a supervised learning problem where each instance in the data stream is classified into one or more pre-defined sets of labels. It involves the difficulties of data streams (e.g. **(1)** data can be seen only once, **(2)** finite time and memory against possibly infinite data stream, **(3)** concept drifts) and multi-label classification (e.g. **(1)** 2^L (exponential) number of labelings for a data, **(2)** learning dependencies between labels). MLSC is used in diverse areas including text tagging and classification, aviation safety predictions, news labeling, and bioinformatics.

Ensembles in MLSC. Ensembles are commonly used in data streams due to their capability of improving the accuracies of their individual weak components, and adapting to the changes (new concepts) in the distribution of the data over time (drifts).

Related Work

Two main ways of tackling multi-label problems [1] are **(1)** Problem Transformation, where data is transformed to be compatible with single-label methods such as decision trees and KNN, **(2)** Algorithm Adaptation where the algorithms are modified to be compatible to multi-label data such as modifying split criterion of a decision tree.

Ensembles. In MLSC, there are *component-agnostic* ensembles that can work with any existing multi-label algorithm, and *component-sensitive* ones that use parameters that can only work with specific multi-label methods.

- **Online Bagging-based**: component-agnostic with static weighting. (e.g. EBR, ECC, EPS, E$_B$RT, ...)
- **ADWIN Bagging-based**: component-agnostic with explicit drift detection. (e.g. E$_a$BR, E$_a$CC, E$_a$PS, E$_a$HT$_{PS}$, ...)
- **Component-Sensitive**: dynamic weighting but cannot work with any method. (e.g. MW, SWMEC, SMART)

We propose **GOOWE-ML** that is component-agnostic with dynamic weighting.

Proposed Method

GOOWE-ML (**G**eometrically **O**ptimum **O**nline **W**eighted **E**nsemble for **M**ulti-**L**abel Classification) is a batch-incremental and evolving (dynamic size) ensemble that assigns weights to its components using a geometric interpretation of label space [2] and solving a linear least squares problem at the end of each batch [3].

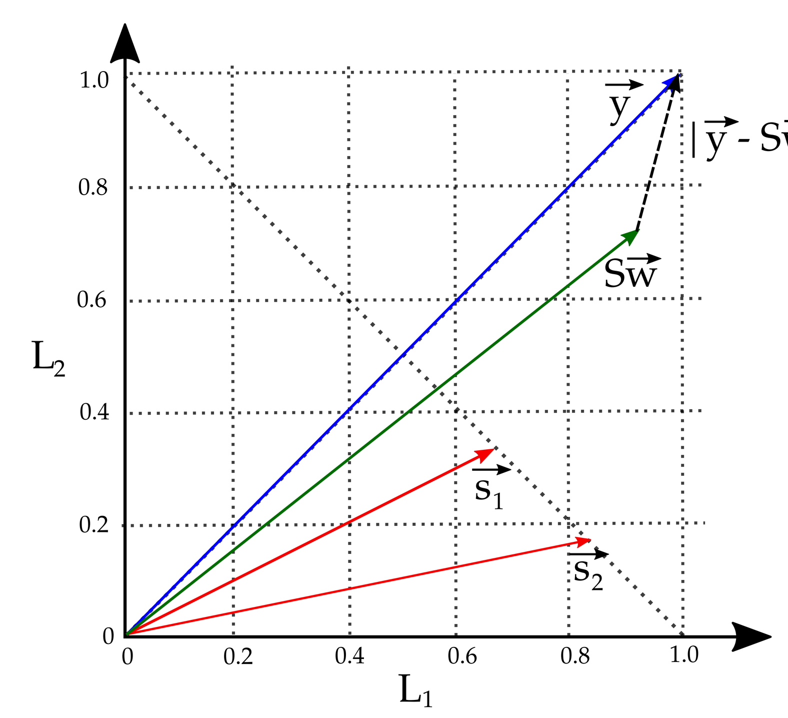

Figure 1: Representation in label space. Relevance scores of the components (S_1 and S_2, red) are combined into the prediction of the ensemble ($S\vec{w}$, green). The optimal vector ($\vec{y}$, blue) is generated from the ground truth. The distance between $\vec{y}$ and $S\vec{w}$ is minimized.

Optimal Weighting

Solving the linear least square problem $\min_{\vec{w}} ||\vec{y} - S\vec{w}||_2^2$ for each batch is equivalent to solving $Aw = d$ where matrix A has elements:

$$a_{qk}^i = \sum_{i=1}^{n}\sum_{j=1}^{L} S_{qj}^i S_{kj}^i \quad (1 \le q, k \le K)$$

and vector d has elements:

$$d_q^i = \sum_{i=1}^{n}\sum_{j=1}^{L} y_j^i S_{qj}^i \quad (1 \le q \le K)$$

After predicting for each instance, using relevance scores of each component and the ground truth vector, A and d are filled. When the data chunk is full, **(1)** the linear system $Aw = d$ is solved to obtain w, **(2)** the component *with the lowest weight* is replaced by an empty classifier, and **(3)** the new classifier as well as the existing ones are trained from the chunk.

The updated weights do not depend on the previous values in w, they only depend on the performance of the components on the latest chunk. This allows the ensemble to capture sudden changes in the distribution of the data.

Experimental Setup

We generated 4 different GOOWE-ML-based ensembles using different problem transformations (and Hoeffding Tree classifiers after transformations):

1. **GOBR**: Ensemble using Binary Relevance
2. **GOCC**: Ensemble using Classifier Chains
3. **GOPS**: Ensemble using Pruned Sets
4. **GORT**: Ensemble using iSOUP Regression Trees

We compared our method against 7 existing methods [4] (EBR, ECC, EPS, E$_B$RT, E$_a$BR, E$_a$CC, E$_a$PS) on 7 datasets.

We evaluated the models on the following metrics: Exact Match, Hamming Score, Accuracy, Example-based and Label-based {Precision, Recall and F1 Score}, Time and Memory Consumption. Afterwards, we applied Friedman test with Nemenyi post-hoc analysis ($\alpha = 0.05$) for statistical significance.

Results of Prequential Evaluation

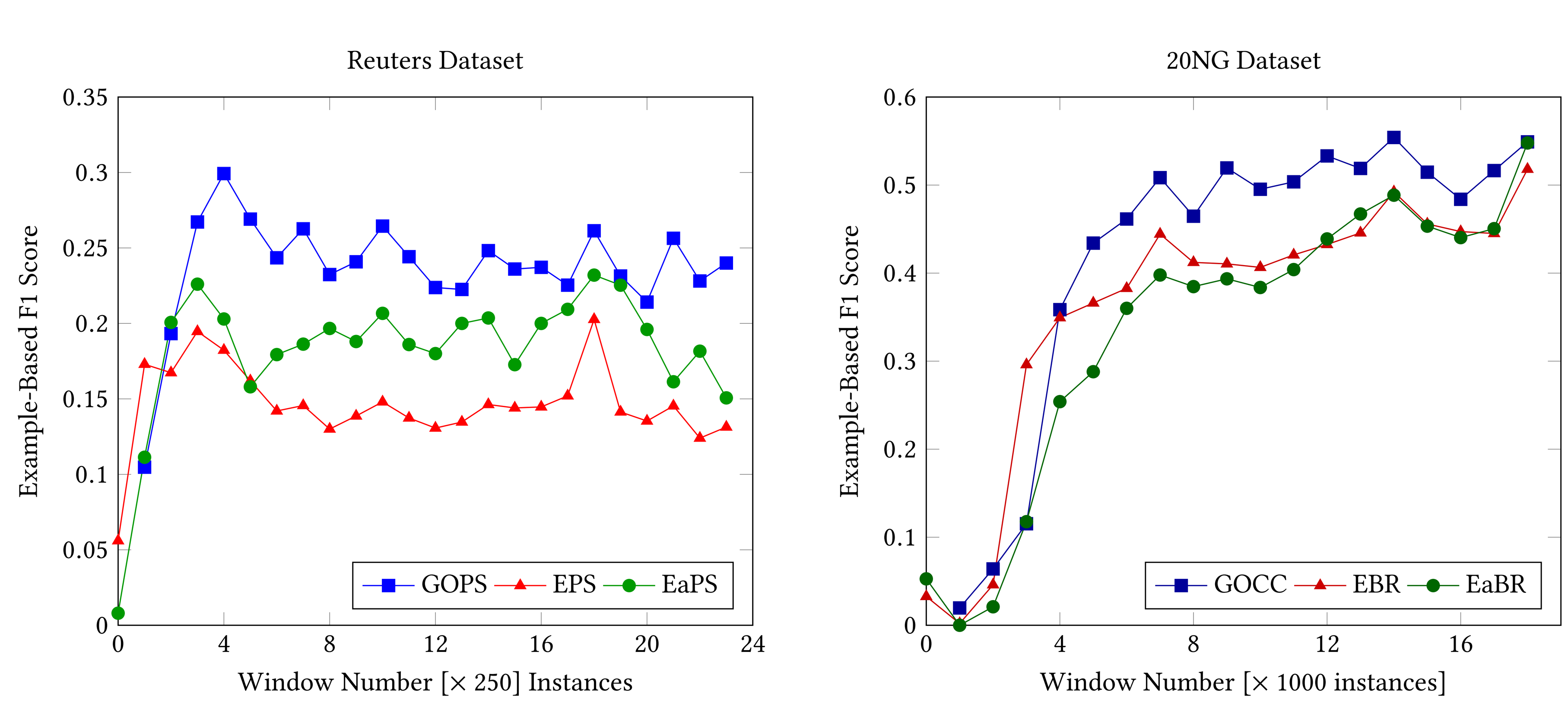

Figure 2: Prequential Evaluation of Models: Example-Based F1 Score for Reuters and 20NG datasets. Best performing models among GOOWE-ML, Online Bagging and ADWIN Bagging are picked.

Overall Results

Example-based Accuracy and Micro-averaged F1 Score are reported in the below table. For more results, visit the repository and see the paper.

Table 1: Overall Results (Acc_{ex} and $F1_{micro}$)

	20NG	Yeast	Ohsumed	Slashdot	Reuters	IMDB	TMC7	Avg. Rank
(d) Example-Based Accuracy (Acc_{ex}) ↑								**Avg. Rank**
GOBR	0.239	0.508	0.184	0.106	0.040	0.164	0.457	4.57
GOCC	**0.391**	**0.509**	**0.277**	0.025	0.120	0.138	0.515	**3.00**
GOPS	0.137	0.504	0.211	**0.299**	0.160	**0.204**	0.327	3.29
GOBRT	0.115	0.454	0.178	0.107	0.040	0.164	0.298	6.71
EBR	0.352	0.502	0.191	0.020	0.098	0.055	0.520	4.29
ECC	0.337	0.493	0.180	0.018	0.093	0.012	0.511	6.14
EPS	0.094	0.460	0.180	0.260	0.143	0.105	0.246	6.29
EBRT	0.100	0.372	0.049	0.001	0.000	0.001	0.007	10.57
EaBR	0.330	0.502	0.169	0.016	0.056	0.024	**0.529**	6.14
EaCC	0.152	0.495	0.004	0.018	0.004	0.001	0.516	7.71
EaPS	0.108	0.455	0.170	0.250	**0.179**	0.083	0.290	6.43
(b) Micro-Averaged F1 Score ($F1_{micro}$) ↑								**Avg. Rank**
GOBR	0.237	0.638	0.291	0.187	0.076	0.276	0.584	4.86
GOCC	**0.516**	**0.640**	**0.410**	0.050	0.196	0.228	0.634	**2.71**
GOPS	0.206	0.629	0.298	**0.315**	**0.210**	**0.314**	0.447	3.43
GOBRT	0.153	0.598	0.270	0.187	0.077	0.277	0.439	6.57
EBR	0.499	0.631	0.294	0.041	0.141	0.099	0.638	4.29
ECC	0.486	0.625	0.280	0.037	0.134	0.025	0.631	6.14
EPS	0.115	0.584	0.216	0.286	0.162	0.138	0.342	7.00
EBRT	0.174	0.519	0.076	0.001	0.000	0.001	0.008	10.58
EaBR	0.477	0.632	0.266	0.033	0.081	0.041	**0.640**	5.71
EaCC	0.262	0.627	0.007	0.037	0.007	0.002	0.632	7.71
EaPS	0.180	0.580	0.205	0.278	0.200	0.118	0.378	6.71

Conclusion

We proposed a novel ensembling scheme for multi-label classification task in data streams. We leveraged geometric interpretation of label space to assign optimal weights for the components in the ensemble. Our method improved the predictive performance of existing multi-label methods, and surpassed the state-of-the-art in multi-label ensemble models.

References

[1] Min-Ling Zhang and Zhi-Hua Zhou.
A review on multi-label learning algorithms.
IEEE TKDE, 26(8):1819–1837, 2014.

[2] Shengli Wu and Fabio Crestani.
A geometric framework for data fusion in information retrieval.
Information Systems, 50(Supp. C):20 – 35, 2015.

[3] Hamed Bonab and Fazli Can.
GOOWE: Geometrically Optimum and Online-Weighted Ensemble classifier for evolving data streams.
ACM TKDD, 12(2):25, 2018.

[4] Jesse Read, Albert Bifet, Geoff Holmes, and Bernhard Pfahringer.
Scalable and efficient multi-label classification for evolving data streams.
Machine Learning, 88(1-2):243–272, 2012.

Contact and Reproducibility

- Email: alicanbuyukcakir@bilkent.edu.tr
- Project Repository: http://www.github.com/abuyukcakir/gooweml

University of Massachusetts Amherst