Kardan Journal of Engineering and Technalogy (KJET)

ISSN: 2706-7815 (Print and Online), Journal homepage: kjet.kardan.edu.af

Technique Recommendation to Recover Missing Labels by Performance Analysis of Multi-Label Machine Learning Algorithms

Sayed Mortaza Kazemi, Zahidullah Sherzad, Muhammad Haleem, and Nasratullah Zarif

To cite this article: Kazemi, S. M. Sherzad, Z. Haleem, M. Zarif, N. (2022), Technique recommendation to recover missing labels by performance analysis of Multi-Label Machine Learning Algorithms, *Kardan Journal of Engineering and Technology, 4 (1), 21-28.* DOI: 10.31841/KJET.2022.25

To link to this article: http://dx.doi.org/10.31841/KJET.2022.25

Published online: 25 December 2022

Article

Technique recommendation to recover missing labels by performance analysis of Multi-Label Machine Learning Algorithms

Kardan Journal of Engineering and Technology 4 (1) 21–28
©2022 Kardan University
Kardan Publications
Kabul, Afghanistan
DOI: 10.31841/KJET.2022.25
https://kardan.edu.af/journals/CurrentIssue.aspx?j=KJET

Received: 27 July 22
Revised: 20 August 22
Accepted: 25 Oct 22
Published: 25 Dec 22

Sayed Mortaza Kazemi
Zahidullah Sherzad
Muhammad Haleem
Nasratullah Zarif

Abstract

With today's technological growth Machine learning is revalued as a tool that provides the ability in automating systems. The machines learn from experience based on the available datasets with no intervention from humans. Here in this research, we consider a variant of the classification, multi-label learning where each instance belongs to more than one label simultaneously. In contrast to other classification tasks, there are a number of challenges, the most significant is to figure out the class when the labels are absent. We compared various multi-label methods from algorithm adaptation and problem transformation perspectives. We applied the algorithms to eight datasets. Problem transformation algorithms are applied to analyze label correlation in a positive sense and then both in a positive and negative sense. We contribute by recommending the technique to address the issue of missing labels by either relying solely on label correlation or by combining it with data locality. Experimental results on multiple benchmark data sets demonstrate that MLLCRS-ML outperforms other cutting-edge techniques.

Introduction

Multi-label learning was developed as a result of research into the text classification problem, in which each document may belong to multiple predetermined subjects at the same time. The goal of multi-label learning is to predict the label sets of previously unseen examples by assessing training instances with known labels. The training set consists of cases, each of which is connected with a set of labels. For instance, with image annotation, a road and an automobile can be tagged on a picture. Multi-label learning has rapidly advanced in recent works in a variety of study fields, including bio-informatics [1], text categorization [2], image annotation [3], and video annotation [4]. Numerous algorithms have been put out to address multi-label classification issues. By limiting each instance to have a single label, traditional two-class, and multi-class problems can both be transformed into multi-label ones. On the other hand, multi-label problems' inherent generality makes learning them more challenging. Multi-label learning strategies can be categorized into two groups using the well-known taxonomy shown in [5]. Approaches to Problem Transformation (PT) and Approaches to Algorithm Adaptation (AA).

1.1 Problem Transformation

It solves a multi-label problem intuitively and divides it into numerous independent binary classification questions (one per category). However, this method does not take into account the relationships between the various labels of each instance, therefore a system of this type may have limited expressive potential. One of the representative algorithms of problem transformation techniques, the Binary Relevance (BR)[]approach, treats each label as a separate binary (one-vs-rest) classification issue. In each binary classification problem, examples that belong to a particular class are viewed as positive examples, whereas instances that don't belong to that class are viewed negatively.

1.2 Algorithm adaptation

These techniques aim to adapt existing single-label classification algorithms to multi-label classification problems, it includes Rank-SVM [6], ML-kNN [7] which are created by tweaking the k-nearest neighbor approach, and others. The fundamental goal of algorithm adaptation techniques is to modify the single-label learning algorithms that are already in use in order to directly address the multi-label learning issue.

2. Related Work

The existing, well-known algorithms have three groups based on the label correlation. The first group is without label correlation [7]. The second group shows a pairwise double correlation while the third group has a label-wise cross correlation. The second group under the taxonomy [8] the first, second, and high orders. Here the label correlations are disregarded by first-order algorithms. The representative techniques are ML-kNN and BR. The pair-wise correlations are utilized by second-order methodologies. Examples are MLLCRS [9], IMLSF [10] and etc. The correlations that come above second-order correlations have the advantages of high-order approaches. The representative techniques primarily consist of CC [11] and CCPU [12]. Based on the reviewed literature to the best of our knowledge no contribution exists to identify and recover the missing labels. Hence our comparison analysis recommends our contributed technique. With solid arguments from the performance analysis.

2.1 Existing Multi-label Algorithms in Brief

2.1.1 ML-KNN

Ml-knn (A Lazy Learning Approach to Multi-Label Learning) [7]: This study presents the Mlknn multi-label lazy learning strategy, which evolved from the classical k-Nearest Neighbor (kNN) algorithm. In the training set, the k nearest neighbors of each unseen instance are initially determined. Then, depending on statistical data obtained from these surrounding instances' label sets, i.e. the label assigned for the unseen instance is decided using the maximum a posteriori (MAP) principle and the number of surrounding instances belonging to each potential class. Experiments on three various multi-label learning issues from the real world.

2.1.2 Multi-Label Learning with Label-Specific Features (LIFT) [8]

This study examines an alternative method for learning from multi-label data that uses label-specific features to aid in class label differentiation. In light of this, the multi-label learning with label-specific features method known as LIFT, which is simple yet effective, is suggested. By performing a clustering analysis on each label's positive and negative examples, LIFT first builds features specific to each label before doing training and testing by querying the clustering results.

2.1.3 *Learning Label-Specific Features for Multi-Label Classification [13]*

A well-known paradigm for multi-label classification is binary relevance (BR). For each label, it breaks down the multi-label classification problem into a bi- nary (one vs. rest) classification subproblem. Although the BR approach is a clear and easy method for multi-label categorization, it nevertheless has a number of shortcomings. First, label correlations are not taken into account. Second, the problem of class imbalance may affect each binary classifier. Third, for data sets with numerous labels, it may become computationally prohibitive. By making use of label correlations between labels and performing label space dimension reduction, a number of solutions have been put forth to address these issues. Inconsistency, another possible BR flaw, is frequently disregarded by researchers when they create multi-label studies. Certain label correlations may be shared only by a limited local collection of instances in real-world applications, but others may be essential globally. Additionally, only incomplete labels are frequently discovered, making label correlations much more difficult to employ. That is, assessing label correlations gets harder when many labels are missing. GLOCAL [14] is a novel multi-label strategy that addresses both full-label and missing-label cases by merging global and local label correlations, learning a latent label representation, and optimizing label manifolds.

2.1.4 *Multi-Label Learning with Global and Local Label Correlation*

Existing methods either presuppose that the label correlations are local and shared only by a small subset of the data, or that they are global and shared by all instances. In real-world applications, some label correlations may be shared only by a small local set of examples, whereas other label correlations may be important globally. Furthermore, only incomplete labels are commonly discovered, making use of label correlations far more difficult. That is, when many labels are missing, estimating label correlations becomes difficult. In this paper, we describe GLOCAL [14], a unique multi-label technique that tackles both full-label and missing-label scenarios by combining global and local label correlations, learning a latent label representation, and optimizing label manifolds.

2.1.5 *Multi-label classification by exploiting local positive and negative pairwise label correlation*

By utilizing Local Positive and Negative Pairwise Label Correlations, or LPLC [15], the authors of this work offer a straightforward and efficient Bayesian model for multi-label classification. The positive and negative label correlations of each ground truth label for each training example are found during the train- ing phase. For each test example, the k nearest neighbors and the related positive and negative pairwise label correlations are found first in the test stage. The posterior probability, which is based on the label distribution, the local positive and negative pairwise label correlations represented in the k nearest neighbors, is then maximized to create predictions.

2.1.6 *Improving multi-label classification with missing labels by learning label-specific features*

Current multi-label learning algorithms typically rely on an identical data representation made up of all the characteristics used to distinguish between all the labels, and they assume that all of the class labels are contained in every training sample. However, with multi-label learning, each class label may be characterized by certain unique properties of its own, and for some practical applications, just a portion of each example's label set can be recovered. LSML [16] is proposed in this paper as a unique technique for learning

Label-Specific features for multi-label classification with Missing Labels. First, a new supplemental label matrix is added to the incomplete label matrix by learning high-order label correlations. The multi-label classifier is then built simultaneously using the learned high-order label correlations and a label-specific data representation for each class label that has been learned.

2.1.7 Joint label-specific features and label correlation for multi-label learning with a missing label

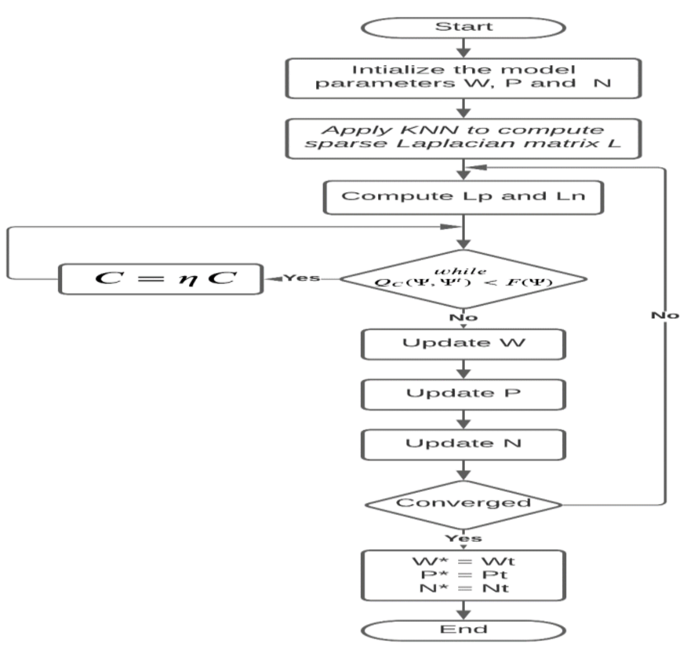
The fact that some features in the original feature space may decide class labels is ignored by existing multi-label learning classification techniques. Ad- ditionally, for some practical applications, only a partial label of each instance can be obtained. As a result, we suggest a unique technique for multi-label learning with missing data called combined Label-Specific features and Label Correlation. To address the aforementioned issues, Label (LSLC-ML) [17] and its optimized version were used. After selecting label-specific features, a multi-label classification task can be modeled by combining the label-specific feature selections, missing labels, and positive and negative label correlations. First, a missing label can be recovered by the learned positive and negative label correlations from the incomplete training data sets.

3. Proposed Solution

3.1 Multi-label classification with Missing Labels using Label Correlation and Robust Structural Learning

This study proposes a unique multi-label classifier called Multi-label classification with Missing Labels utilizing Label Correlation and Robust Structural Learning. They also present a unified learning method that addresses the mentioned problem (MLLCRS-ML). In order to recover the missing labels, the proposed classifier takes into account label-specific features in addition to making use of the structural property of the data and pairwise label correlation

3.2 Flowchart of the proposed solution



Algorithm 1: Linear MLLCRS-ML

Input: $X \in R^{n \times m}$, $Y \in \{-1, 1\}^{n \times l}$

Output: W ; P ; N; 1

Initialization

Model Parameters W ; P ; N;

Trade-off parameters: λ_1 ; λ_2 ; λ_3 ; λ_4 ; λ_5 ; λ_6 ; a ;

iteration parameters: β_1 ; β_2 ; t ;

relative parameters: L_p ; L_n ; C ;

2 Calculate the Laplacian matrix L;

3 Repeat

Calculate L_p and L_n ;

While $Q_C(\Psi, \Psi^t) < F(\Psi)$ do

$C = \eta C$ where η is the search step size;

Calculate W^t ;

$Gtw \leftarrow W t - 1/c\ \nabla W f\ (W t, Pt, Nt)$

$W_{t+1} \leftarrow prox_{\lambda 1/c}\ (G^t{}_w)$; by

Calculate P^t

$GtP \leftarrow Pt - 1/c\ \_\nabla Pf\ (Wt, Pt, Nt)$;

$P_t{+}_1 \leftarrow prox_{\lambda 2/c}\ (G^t{}_p)$; (28)

Calculate N^t by (30)

$GtN \leftarrow Nt - 1/c\ \nabla Nf\ (Wt, Pt, Nt)$

$N_t{+}_1 \leftarrow prox_{\lambda 3/c}\ (G^t{}_n)$;

$B_{t+1} = (1+sqrt(4\beta_t{}^2+1)/2)$

$t \leftarrow t+1$

Until MaxIteration Reached

$W_* = W_t$; $P^* = P_t$; $N^* = N_t$;

Return W^*, P^*, N^*

Where W is model weight matrix, P is positive correlation and N is negative correlation matrix

Algorithm 2: Model Prediction

Input: Model parameter W and test data D_t;

Output: Predicted Labels Y_{pred};

1 Calculate $S_c = D_t W$

2 $Y_{pred} = \mathrm{sign}(S_c - \tau)$ 3 Return Y_{pred}

4. Result Analysis

Here, we draw attention to a few factors that may help to explain their actions. Only the local positive and negative label correlations are looked at by LPLC. While LLSF takes label correlations into account as well, LIFT solely takes label-specific features into account. MLKNN, LPLC, LIFT, and LLSF, however, are unable to address the issue of missing labels. Glocal does take label correlations and incomplete labels into account, but it is unable to choose features that are specific to a given label. While LSML and LSLC take into account label correlations, label-specific features, and the missing label problem, they only use label correlations to address the issue. Our suggested algorithm first determines positive and negative label correlations from incomplete training data sets, then addresses the missing label issue by utilizing both the learned label correlation and the locality of the data, and finally models the multi-label classification issue by repeatedly choosing label-specific features.

a. Result table and graphs:

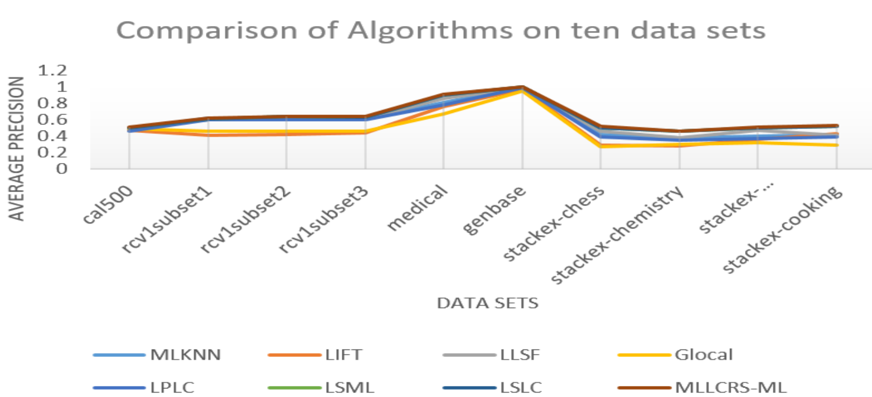


Figure 1: Comparison result of eight algorithms based on Average precision evaluation metric.

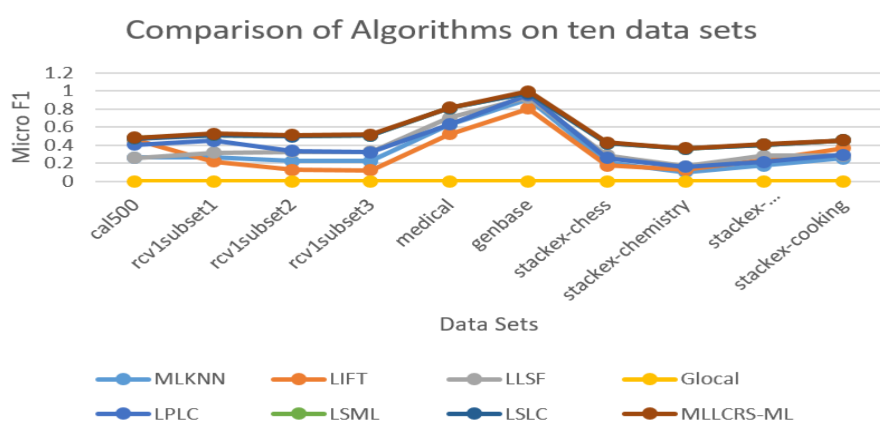


Figure 2: Comparison result of eight algorithms based on Micro F1 evaluation metric.

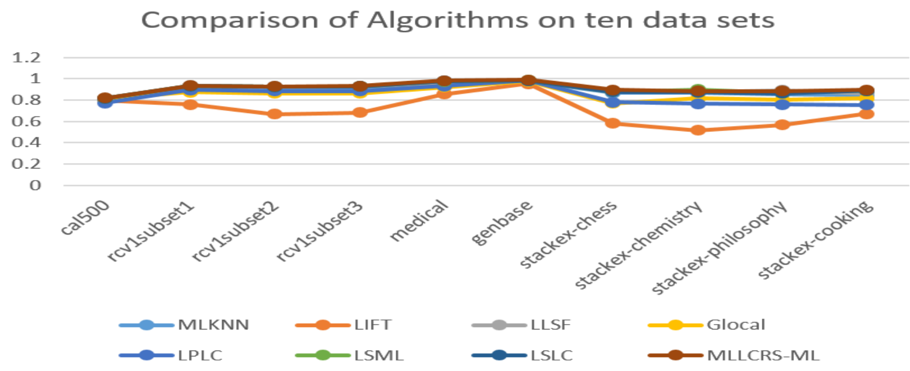


Figure 3: Comparison result of eight algorithms based on AUC evaluation metric.

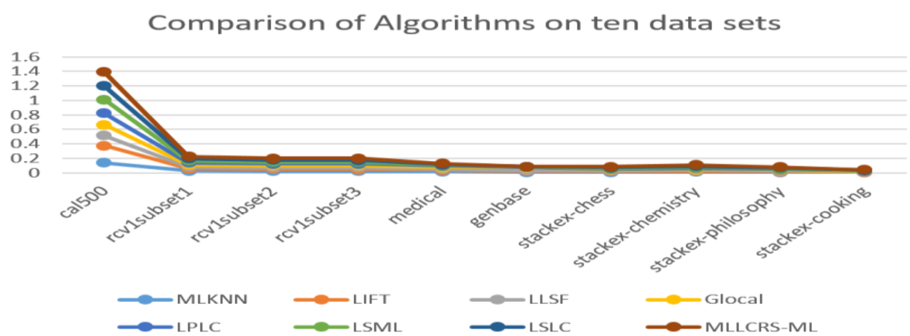


Figure 4: Comparison result of eight algorithms based on hamming Loss evaluation metric.

5. Conclusion

We compared eight different multi-label algorithms in this paper, and MLLCAS-ML performed better than all other algorithms because it fills the missing label matrix by utilizing both label correlations (both positive and negative label correlations) and a structural property of data called locality of data, which ensures that similar data instances will have similar class labels. This technique considers label correlation and label-specific features to improve classifier accuracy.

References

[1] A. Clare, R. D. King, Knowledge discovery in multi-label phenotype data, in: European Conference on Principles of Data Mining and Knowledge Discovery, *Springer*, 2001, pp. 42–53.

[2] A. K. McCallum, Multi-label text classification with a mixture model trained by em, in: AAAI 99 Workshop on Text Learning, *Citeseer*, 1999, pp. 1–7.

[3] M. R. Boutell, J. Luo, X. Shen, C. M. Brown, Learning multi-label scene classification, *Pattern Recognition* 37 (9) (2004) 1757–1771.

[4] F. Kang, R. Jin, R. Sukthankar, Correlated label propagation with application to multi-label learning, in: 2006 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'06), Vol. 2, IEEE, 2006, pp. 1719–1726.

[5] G. Tsoumakas, I. Katakis, I. Vlahavas, *Mining multi-label data, in: Data Mining and Knowledge Discovery Handbook*, Springer, 2009, pp. 667–685.

[6] A. Jiang, C. Wang, Y. Zhu, Calibrated rank-svm for multi-label image categorization, in: 2008 *IEEE International Joint Conference on Neural Networks* (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 1450–1455.

[7] M.-L. Zhang, Z.-H. Zhou, Ml-knn: A lazy learning approach to multi-label learning, *Pattern Recognition* 40 (7) (2007) 2038–2048.

[8] M.-L. Zhang, L. Wu, Lift: Multi-label learning with label-specific features, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (1) (2014) 107–120.

[9] R. Rastogi, S. Mortaza, Multi-label classification with missing labels using label correlation and robust structural learning, *Knowledge-Based Systems* 229 (2021) 107336.

[10] R. Rastogi, S. Mortaza, Imbalance multi-label data learning with label specific features, *Neurocomputing* (2022).

[11] J. Read, B. Pfahringer, G. Holmes, E. Frank, Classifier chains for multi-label classification, *Machine Learning* 85 (3) (2011) 333–359.

[12] P. Teisseyre, Classifier chains for positive unlabelled multi-label learning, *Knowledge-Based Systems* 213 (2021) 106709–106725.

[13] J. Huang, G. Li, Q. Huang, X. Wu, Learning label-specific features and class-dependent labels for multi-label classification, *IEEE Transactions on Knowledge and Data Engineering* 28 (12) (2016) 3309–3323.

[14] Y. Zhu, J. T. Kwok, Z.-H. Zhou, Multi-label learning with global and local label correlation, *IEEE Transactions on Knowledge and Data Engineering* 30 (6) (2017) 1081–1094.

[15] J. Huang, G. Li, S. Wang, Z. Xue, Q. Huang, Multi-label classification by exploiting local positive and negative pairwise label correlation, *Neurocom- Puting* 257 (2017) 164–174.

[16] J. Huang, F. Qin, X. Zheng, Z. Cheng, Z. Yuan, W. Zhang, Q. Huang, Improving multi-label classification with missing labels by learning label- specific features, *Information Sciences* 492 (2019) 124–146.

[17] Z. Cheng, Z. Zeng, Joint label-specific features and label correlation for multi-label learning with missing label, *Applied Intelligence* 50 (11) (2020) 4029–4049.

About the Authors

Mr. Sayed Mortaza Kazemi, *Lecturer, Faculty of Computer Science, Kardan University, Kabul Afghanistan.* <*m.kazemi@kardan.edu.af*>
Mr. Zahidullah Sherzad, *Researcher, Faculty of Computer Science, Mannheim University of applied sciences, Mannheim, Germany.* <*z.sherzad@hs-mannheim.de*>
Mr. Muhammad Haleem, *Assistant Professor, Faculty of Computer Science, Kardan University, Kabul Afghanistan.* <*m.haleem@kardan.edu.af*>
Mr. Nasratullah Zarif, *Lecturer, Department of Computer Science, Faculty of Computer Science, Kardan University, Kabul Afghanistan.* <*n.zarif@kardan.edu.af*>