Boolean Matrix Decomposition for Label Space Dimension Reduction: Method, Framework and Applications

Lin Liu¹, Lin Tang², *

¹ School of Information Yunnan Normal University Kunming, China.
² Key Laboratory of Educational Informatization for Nationalities Ministry of Education Yunnan Normal University Kunming, China.

liulinrachel@163.com, *Corresponding author e-mail: maitanweng2@163.com

Abstract. Recently, the large number of labels challenges the traditional multi-label classification methods in various application scenes. We introduce a framework of multi-label classification based on Boolean matrix decomposition for giving various multi-label classifiers the ability to predict labels in a high dimension label space. The BMD module in this framework satisfies two conditions of ‘the column use condition’ and exacts BMD, which can reduce the burden on training and predicting task of multi-label classifier to some extent, and the predicted result of multi-label classifier can be restored to original label space by simply Boolean multiplying with matrix. Experimental results on yeast datasets demonstrate that our framework can work particularly well on datasets with a large number of labels and obtain a better predicting accuracy. In summary, the methods discussed in this paper constitute important basic for utilizing more multi-label classifiers in high label dimension space, which is the main contribution of this paper.

1. Introduction

As an important established method for analysis and preprocessing data, matrix decomposition has become a thoroughly explored direction in data analysis. For matrix decomposition, a matrix is able to be represented as the product of two matrices, which are the concise representations of data. For example, the well-known singular value decomposition (SVD) and nonnegative factorization (NMF) have been applied in the field of image processing, DNA analysis, label space dimension reduction for multi-label classification and so on[1]. To deal with binary data in data mining, there also have recently been intensive research activities in Boolean matrix decomposition (BMD, called Boolean matrix factorization), including not only single algorithm but also its applications in many different guises.

The goal of BMD is to represent a given binary matrix as the product of two or more factor matrices. Given an $m \times n$ Boolean matrix $A \in \{0,1\}^{m \times n}$, BMD aims to find two matrices $C \in \{0,1\}^{m \times k}$ and $X \in \{0,1\}^{k \times n}$ such that the difference $\|A - C \circ X\|$ under some norm $L$ is minimized with a given $k$ or as small a $k$ as possible. The operator $\circ$ is the Boolean matrix product, which is equivalent to normal matrix multiplication with addition defined as $1 + 1 = 1$. The minimum possible $k$ is called the Boolean rank of $Y$. If a condition of $\|A - C \circ X\|_L = 0$ under any norm $L$ is satisfied, then this BMD is called an exact BMD. BMD is typically a tough problem, with the basic problem being NP-hard.

In BMD described above, the matrix $C$ can be viewed as concept matrix, which represents a set of meaningful concepts, and the second matrix $X$ can be expressed as a union of a subset of the concepts.
In the area of text mining, for instance, binary matrix \( A \) assumed to be the “documents” and “words” matrix, representing the documents-words mapping. Then matrix \( C \) and \( X \) represent the document-topics mapping and topic-words mapping respectively, then topics are the meaningful concepts.

For classification problem, the ability of concepts representation of BMD also offers a powerful tool for data analysis. If the attribute space is represented as the column of matrix \( A \), and rows are instances to be classified. Then BMD can switch the original attributes space to the lower dimension features space [2]. However, the feature value always is real-valued data. In addition to, there are many classification scenes with many labels in multi-label classification [3, 4]. For example, there were already 5 billion photos and more than 20 million unique tags in Flickr. While many of these tags may be redundant, it has been suggested that humans can still recognize between 10,000 and 100,000 unique object classes. The Dmoz data set, which is constructed by crawling webpages from the Open Directory Project, also has more than 30,000 labels. Nonetheless, it is generally known that the large number of labels will cause a long training time of multi-label classifier. Especially, several multi-label algorithms couldn’t handle classification problem with many labels, for instance, binary relevance method can’t be used in multi-label classification that includes \( 10^7 \) labels [5]. Meanwhile, previous research has shown that the classification performance of discriminative models such as SVMs will fall dramatically in the face of many labels dataset [6]. Aiming at these application scenarios, several recent approaches called label space dimension reduction (LSDR) have been proposed to address the multi-label classification problem with many labels. These kinds of methods aim to obtain the lower dimension representation of label space. A natural idea of LSDR is matrix decomposition, such as SVD have provided a powerful tool in many LSDR studies. Obviously, compared with real-valued data matrix decomposition, BMD is more suitable for LSDR which accounts for the Boolean data type of the class labels. Therefore, our attention in this paper is multi-label classification problem with many labels, and we delve more into the issues pertaining to this.

2. The framework of multi-label classification based on BMD

2.1. LSDR approaches

Given an instance-labels mapping matrix \( A \), whose elements is 0 or 1. If \( A \) is a high-dimensional vector in label space, then it’s difficult to train classifier on original matrix \( A \). The aim of LSDR is using a compression step to transform the original high dimension label space to a lower dimension space. Then classifiers can be trained on this lower dimension space, which can reduce the computation burden of classifier. Obviously, many existing method is able to apply in this dimension reduction process, such as SVD, PCA and NMF.

A first attempt of LSDR is employed by Hsu et al. [7], which projects the higher-dimensional label vector using compressed sensing. Subsequently, many variants have been developed along this line, such as principal component analysis [9], canonical correlation analysis [8] and other singular value decompositions [10, 11]. A common characteristic of these studies is that they all reduce the possibly large number of labels to a more manageable set of transformed labels. Yet, a major limitation is that the transformed labels which lost the meaning of original labels, we called this kind of methods ‘LSDR based on transformed labels’.

Instead of using label transformation, Balasubramanian et al. [12] proposed to train only a small subset of the labels. Since this subset come from the original labels, their learning problems will not be made more difficult. Afterwards, Bi et al. [3] proposed a Column Subset Selection for Multi-Label (CSS-ML) to select exactly \( k \) representative labels based on a randomized sampling procedure, then learned \( k \) classifiers for these selected labels. We called this kind of methods’ LSDR based on label subset’.

2.2. BMD for LSDR

In connection with data mining, the seminal work by Miettinen et al. [13, 14] was a catalyst to ignite a wave of interest in BMD and its applications to data mining [15, 16, 17, 18, 19]. Meanwhile, there had been some related work prior to that, e.g., [20, 21], in combinatorics research. BMD also has applications
in such areas as educational testing [22] and role Mining [23, 24], as well as in more traditional data analysis. Miettinen et al. [14] designed a simple greedy algorithm, called Asso, to solve the discrete basis problem. Drineas et al. [25, 26] introduced CX-decompositions, where a given matrix $M$ is decomposed into two matrices $C$ and $X$, with the condition that the columns of $C$ must be a subset of the columns of $M$. Then the ‘column use condition’ is imposed. Obviously, the ‘column use condition’ is consistent with the idea of ‘LSDR based on label subset’.

In reference [27], authors introduced a multi-label classifier based on Asso. Work by Belohlavek et al. [16, 17] addresses exact as well as approximate BMD, which proposed two heuristic algorithms, named GreConD and GreEss. However, they also do not impose the ‘column use condition’. In reference [28], Sun et al. first prove an exact formula for the Boolean matrix $J$ such that $TMMJ = \otimes$ holds for exact BMD problem. Since minimizing $k$ is NP-hard in exact BMD problem, they proposed two heuristic algorithms for finding suboptimal but good decomposition and measured the performance of their algorithms on several real datasets in comparison with other representative heuristic algorithms for BMD. Their algorithms impose the ‘column use condition’ and run very fast.

In fact, both BMD and singular value decomposition (SVD) belong to matrix factorization. Although SVD provides a powerful tool in many LSDR studies, compared with BMD, the existing researches about LSDR have their limitations respectively [29]: transferring original matrix to a new labels space suffers from the lack of interpretability; selecting columns from original labels only can predict a subset of original labels. Obviously, these two limitations can be released by an exact BMD which is imposed by ‘the column use condition’: ‘the column use condition’ retains the interpretability of low dimension labels space; the exact BMD can restore low dimension predicted label matrix to the original label matrix by matrix $X$. Therefore, we present a multi-label classification framework based on BMD, according with ‘the column use condition’ and extract BMD simultaneously.

![Figure 1. The framework of multi-label classification based on BMD](image)

2.3. framework descriptions

The basic idea of our framework of multi-label classification based on BMD(MLC-EBMD): at first, the feature matrix $F_{\text{max}} \in \mathbb{R}^{n \times f}$ and functional label matrix $A_{\text{max}} \in \{0,1\}^{m \times n}$ of multi-label classification dataset are input to our framework; an exact BMD module decompose the label matrix $A_{\text{max}} \in \{0,1\}^{m \times n}$ into two matrix: $C_{\text{max}} \in \{0,1\}^{m \times k}$ and $X_{\text{max}} \in \{0,1\}^{k \times n}$; $F_{\text{max}}$ and $C_{\text{max}}$ are combined to form a matrix $D_{\text{max}}$ that is the training dataset for multi-label classifier; then, the trained classifier predicts labels for annotated instances with feature vector $p \in \mathbb{R}^{n \times f}$, the predicted label vector is $l \in \{0,1\}^{m \times 1}$; at last, the predicted label vector $l \in \{0,1\}^{m \times 1}$ can be restored to the vector of original label space $l' \in \{0,1\}^{n \times 1}$ by $l' = l \ast X_{\text{max}}$.

As shown in Figure 1, the advantage of this framework is that the matrix $C$ obtained by an exact BMD can reduce the burden on training and predicting task of multi-label classifier to some extent, and the predicted result of multi-label classifier can be restored to original label space by simply Boolean multiplying with matrix $X$. In our framework, as the exact BMD and multi-label classifier are two separate modules, this framework is able to adopt multiple classifiers. The label space dimension reduction performed by an exact BMD can not only reserve the meaning of lower dimension label space, but also would not decrease the accuracy of classification. Intuitively, the exact BMD algorithm is the most important and core module in this framework, which can be employed by any existing BMD algorithms. Specifically, the prediction results on new label space can be easily transformed into the
original label space by Boolean matrix multiplication with a label-correlated matrix obtained from the decomposition.

3. Experimental results
In order to test and verify prediction effect of MLC-EBMD, we utilize protein function prediction dataset Saccharomyces cerevisiae (S.C) for multi-label classification, which is proposed in reference [30]. We mainly use one of the S.C datasets called D1, whose feature vector includes amino acid frequency ratios, molecular weight, sequence length and hydrophobicity, etc. To estimate the influence of different label characteristics on classification results, D1 dataset is divided into three parts by categories of GO terms: BP, CC, and MF. The evaluation criterions employed in this paper is same as reference [30], which won’t be described again.

In our experiments, MLKNN (Multi-label K-nearest neighbor, MLKNN) [31] is taken as multi-label classifier of our framework, and Remove-Smallest algorithm is taken as BMD module of framework. MLC-BMaD (Multi-Label Classification using Boolean Matrix Decomposition) [27] is the comparison framework. We test them on the same datasets. Although MLC-BMaD is similar to our framework, which also includes the processes of label space dimension reducing and restoring, the BMD algorithm adopted in MLC-BMaD is not an exact BMD algorithm. What’s more, ten-fold cross-validation is performed on this data set, and the experimental results are shown in Figure 2.

![Figure 2. The performance comparison of MLC-EBMD and MLC-BMaD for three datasets](image)

Meanwhile, in order to estimate the label restoring performance of exact BMD algorithm in PFP-BMD, Table 1 shows the comparison of original label space and reduced label space on accuracy of multi-label classification, where the values showed in bold are the best results in comparing algorithms.

From Table 1, we can observe that, after restoring to original label space, the AUPRC value of PFP-BMD has improved on three datasets. As AUPRC considers the importance of each functional label to be equal, it illustrates that PFP-BMD is able to greatly improve the overall accuracy of protein function prediction. Especially for D1-CC dataset, it is also worth noting that there are improvements on three evaluation criteria. The reason may be that the BMD of D1-CC label matrix reduces the value of ‘Cardinality’ remarkably, so it illustrates PFP-BMD is suitable for multi-label classification data with high label-correlated degree.

| Datasets | AU(PRC) | AUPRC | AUPRC |
|----------|---------|-------|-------|
| D1-BP    | 0.3161  | 0.1758| 0.2106|
|          | 0.2884  | 0.2432| 0.3464|
| D1-MF    | 0.3429  | 0.2467| 0.2551|
|          | 0.3040  | 0.3169| 0.2420|
| D1-CC    | 0.3958  | 0.1741| 0.2573|
|          | 0.4952  | 0.2075| 0.4404|

4. Conclusion
To solve the problem of multi-label classification with many labels, a framework based on Boolean matrix decomposition is presented for label space dimension reduction. As the BMD module satisfies two conditions of ‘column use condition’ and exact decomposition, the lower dimension label space generated by BMD module can identify and represent original labels in a compact manner, and their
predicted results are multiplied with a matrix obtained from BMD module to obtain the final full predicting results. Our experimental results demonstrate that this framework can obtain a better predicting accuracy compared with other LSDR methods. MLC-EBMD allows various multi-label classifiers to be applied in multi-label classification, which can lay the foundation for improvement of classification accuracy.

Acknowledgments
The research is supported by a National Nature Science Fund Project (61562093) and Starting Foundation for Doctoral Research of Yunnan Normal University (2017ZB013).

References
[1] Vavasis, S(2010). On the complexity of nonnegative matrix factorization. SIAM J. Optim. 20, 1364–1377 .
[2] J. Outrata(2010). Boolean factor analysis for data preprocessing in machine learning, in: Proc. ICMLA, 899–902 .
[3] Bi W, Kwok J T Y(2013). Efficient Multi-label Classification with ManyLabels[C]//ICML (3). 405-413.
[4] Li L, Wang H(2016). Towards Label Imbalance in Multi-label Classification with Many Labels. arXiv preprint arXiv:1604.01304.
[5] Liu T Y, Yang Y, Wan H(2015). Support vector machines classification with a very large-scale taxonomy. Acm Sigkdd Explorations Newsletter, 7(1):36-43.
[6] Rubin T N, Chambers A, Smyth P, et al. Statistical topic models for multi-label document classification. Machine Learning, 2012, 88(1):157-208.
[7] Hsu D J, Kakade S, Langford J, et al(2009). Multi-label prediction via compressed sensing[C]//NIPS. 22: 772-780.
[8] Tai F, Lin H T(2012). Multilabel classification with principal label space transformation. Neural Computation, 24(9): 2508-2542.
[9] Zhang Y, Schneider J G(2011). Multi-Label Output Codes using Canonical Correlation Analysis[C]//AISTATS. 873-882.
[10] Chen Y N, Lin H T(2012). Feature-aware label space dimension reduction for multi-label classification[C]//Advances in Neural Information Processing Systems. 1529-1537.
[11] Houfeng Wang L L L Z(2014). Muli-label Text Categorization with Hidden Components.
[12] Balasubramaniam K, Lebanon G(2012). The landmark selection method for multiple output prediction. arXiv preprint arXiv:1206.6479.
[13] Miettinen P(2009). Matrix decomposition methods for data mining: Computational complexity and algorithms.
[14] Miettinen P, Mieliikäinen T, Gionis A(2008). The discrete basis problem. IEEE Transactions on Knowledge and Data Engineering., 20(10): 1348-1362.
[15] Barnes T(2010). Novel derivation and application of skill matrices: The q-matrix method. Handbook on educational data mining, 159-172.
[16] Belohlavek R, Trnecka M(2015). From-below approximations in Boolean matrix factorization: Geometry and new algorithm. Journal of Computer and System Sciences, 81(8): 1678-1697.
[17] Belohlavek R, Vychodil V(2010). Discovery of optimal factors in binary data via a novel method of matrix decomposition. Journal of Computer and System Sciences, 76(1): 3-20.
[18] Koedinger K R, McLaughlin E A, Stamper J C(2012). Automated Student Model Improvement. International Educational Data Mining Society.
[19] Zhang, S., DeCarlo, L., Ying, Z(2013). Non-identifiability, equivalence classes, and attribute-specific classification in Q-matrix based cognitive diagnosis models. Tech. rep., Columbia University .
[20] Monson S D, Pullman N J, Rees R(1995). A survey of clique and biclique coverings and factorizations of (0, 1)matrices. Bull. Inst. Combin. Appl, 14: 17-86.
[21] Schmidt G. Relational mathematics(2011). Cambridge University Press.
[22] Tatsuoka K K. Cognitive assessment(2009). An introduction to the rule space method[M]. Routledge.
[23] Streich A P, Frank M, Basin D, et al(2009). Multi-assignment clustering for Boolean data[C]//Proceedings of the 26th annual international conference on machine learning. ACM, 969-976.
[24] Lu H, Vaidya J, Atluri V, et al(2012). Constraint-Aware Role Mining via Extended Boolean Matrix Decomposition. IEEE Transactions on Dependable and Secure Computing, 9(5).
[25] Drineas P, Kannan R, Mahoney M W(2006). Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix. SIAM Journal on computing, 36(1): 158-183.
[26] Drineas P, Mahoney M W, Muthukrishnan S(2008). Relative-error CUR matrix decompositions. SIAM Journal on Matrix Analysis and Applications, 30(2): 844-881.
[27] Wicker J, Pfahringer B, Kramer S(2012). Multi-label classification using boolean matrix decomposition[C]//Proceedings of the 27th Annual ACM Symposium on Applied Computing. ACM, 179-186.
[28] Sun Y, Ye S, Sun Y, et al(2015). Improved algorithms for exact and approximate boolean matrix decomposition[C]//Data Science and Advanced Analytics (DSAA). 36678 2015. IEEE International Conference on. IEEE, 1-10.
[29] Miettinen P(2008). The boolean column and column-row matrix decompositions. Data Mining and Knowledge Discovery, 17(1): 39-56.
[30] Cerri R, Barros R C, de Carvalho A C(2012). A genetic algorithm for hierarchical multi-label classification[C]//Proceedings of the 27th annual ACM symposium on applied computing. ACM, 250-255.
[31] Zhang M L, Zhou Z H(2007). ML-KNN: A lazy learning approach to multi-label learning. Pattern recognition, 40(7): 2038-2048.