2016 4th International Conference on Control, Instrumentation, and Automation (ICCIA)
27-28 January 2016, Qazvin Islamic Azad University, Qazvin, Iran

# *Feature Selection in Multi-label classification through MLQPFS*

Majid Soheili, Amir-Massoud Eftekhari Moghadam
Faculty of Computer and Information Technology Engineering,
Qazvin Branch, Islamic Azad University, Iran
Qazvin, Iran
{ma.soheili, eftekhari}@qiau.ac.ir

***Abstract*: In multi-label classification, each data instance is associated with a set of labels. Feature selection is one of the most significant challenges in multi-label classification. Irrelevant and dependent features can mislead the learning phase of multi-label classification. Therefore it is important to select the effective features. A large and growing body of literature has investigated multi-label feature selection problem. Most of these studies used incremental approach. In this paper a new algorithm has been proposed called MLQPFS which selects subset of features in such a way that redundancy among the selected features will be minimized meanwhile the relevancy between the selected features and class labels will be maximized. MLQPFS applies quadratic programming to optimize feature selection process. In order to evaluate the performance of proposed algorithm, MLQPFS and PMU have been compared in three multi-label data sets. The experimental results showed that MLQPFS has better performance than conventional incremental feature selection methods such as PMU.**

***Keywords: Feature selection; multi-label classification; Quadratic programming; mutual information;***

## I. INTRODUCTION

Multi-label classification is a specialization of classic classification problems, in which each training data instance is associated with a set of class labels, instead of single class label. Nowadays, the application of this problem is in variety of fields such as text categorization, gene function classification, and semantic annotation of images [1-5]. In such classification, because of the irrelevant or dependent features, learning algorithm to build a suitable model is difficult that leads to the performance decrease. Feature subset selection (FSS) is an answer to tackle this drawback [6, 7]. The major objective of FSS is to find a set of features considering discrimination among all features.

There are various approaches for the FSS that could be categorized into three methods including filter, wrapper and embedded. Filter method applies statistical approaches to remove features that have little chance to be useful in classification. This method has less computational cost compared with other methods which makes filter method to be generalized for wider problems[8]. In wrapper method, FSS utilizes final classification algorithm as a part of evaluation function and each set of the features are candidates to be evaluated with the estimation of quality of the final classification model. This way of evaluation results in increasing accuracy in the feature selection phase whereas computation cost will be increased[9]. Conversely, in embedded model, feature selection phase has been implicitly attached to the classification algorithm. Decision tree is a specimen of this model, in which selection of effective features has been embedded in the process of tree construction [10, 11].

In multi label feature selection (MLFSS), feature selection algorithm should scan all labels simultaneously which leads to more complexity of the problem. One approach to deal with this problem is to transform it to a single-label feature selection which can be solved with classic methods [12, 13]. Classic FSS method can be applied in single-label problem. Since the number of the classes in this transformation will be increased, this method leads to learning difficulty in learning phase [14].

In this paper, a new algorithm has been proposed (MLQPFS) based on mutual information between features and class labels and among features. The proposed algorithm uses *quadratic programming* for optimization and uses no transformation [15, 16]. MLQPFS attempts to select a subset of features in which the dependency among the selected features be minimized meanwhile the relevance between selected features and labels be maximized.

The rest of this paper is organized as being followed. Section 2 briefly reviews related work, section 3 introduces multi-label feature selection problem and describes the proposed algorithm MLQPFS in detail. Section 4 introduces evaluation measures and reports experimental results over three multi-label datasets. Section 5 concludes the paper.

## II. RELATED WORK

One approach to deal with multi-label feature selection problem is transforming this problem to single-label feature selection and then solve the resultant problem[17, 18]; in this way the most multi-label feature selection solutions include three steps. First step: transforming multi-label dataset to single-label dataset. Second step: independent evaluation of each features with some cost function such as mutual information, third step: selecting the most effective features[7]. Conversely, other approach uses no transformation and classic feature selection algorithms are adapted with multi-label feature selection, some of them are mentioned below.

In 2009, Zhang et al. published a paper, in which they proposed a new algorithm called MLNB[19]. In MLNB, traditional naïve

978-1-4673-8704-0/16/$31.00 ©2016 IEEE

Bayes classifier has been adapted to multi-label called MLNB_BASIC. Then to select effective features, combination of filter and wrapper method is applied. In the first step (filter) irrelative features is removed by PCA (Principal component analysis), in the second step, the genetic algorithm is used to select effective features. Therefore, candidate features is sent to MLNB_BASIC as a chromosome, then average of hamming loss and ranking loss is calculated as fitness of a typical chromosome.

In 2010, Zhang et al. published a paper in which they proposed a multi-label dimensionality reduction method called MDDM [20]. In MDDM, two kinds of projection strategies are considered. MDDM attempts to determine a lower dimensionality space maximizing the dependency between original features and class labels associated with all data instances. MDDM uses Hilbert-Schmidt Independence Criterion [21] to measure dependency.

In 2013, Lee et al. proposed a new algorithm called PMU[7]. PMU attempts to determine subset of features in such a way that mutual information between selected features and labels be maximized. Since the calculation cost of mutual information between all features and all labels is too expensive, PMU uses an approximation to calculate mutual information. PMU applies incremental approach to find effective features.

## III. PROPOSED ALGORITHM

### A. notation of Multi-label feature selection

Before presenting the proposed algorithm, introduction of multi-label feature selection's notations is necessary.

**Input**: matrix $X = R^{n\times d}$ which each row denotes a d-dimensional learning instance and $L = \{L_1, L_2, \dots L_q\}$ denotes the label-class set with q different class labels. Fig.1 is a schema of the input for the problem.

**Multi-label learning**: multi-label learning algorithm aims to build a function or classifier $h: X \to 2^l$ in such a way each data instance x assigns to a set of relevant labels $h(x) \subseteq L$.

**Output:** multi-label feature subset selection (MLFSS) attempts to find a subset of features which keeps discrimination original feature space.

| | $F_1$ | $F_2$ | .... | $F_d$ | $L_1$ | $L_2$ | | $L_q$ |
|---|---|---|---|---|---|---|---|---|
| $X_1$ | 0.1 | 1 | 1 | 2 | 0 | 1 | 0 | 0 |
| $X_2$ | 0.3 | 0.4 | 2.5 | 8 | 1 | 1 | 0 | 0 |
| ⋮ | 3 | 2 | 0.6 | 7 | 0 | 0 | 0 | 1 |
| $X_n$ | 2 | 0.9 | 0.4 | 0.2 | 0 | 0 | 1 | 0 |

Fig. 1. *a schema of the input*

### B. Redundancy and relevancy

There are many studies that apply the redundancy and relevancy in MLFSS [7, 16, 22, 23]. The redundancy refers to the dependency among all features and relevancy refers to the dependency between the features and the class labels. In MLFSS, a subset of features is effective if it has minimal dependency among all the selected features and maximal relevancy between the selected features and labels. The concept of dependency could be represented through variety of measures, such as correlation and mutual information (MI). In this paper mutual information has been applied. The relationship between MI and entropy for two arbitrary random variables x and y has been seen in equations (1) and (2)

$$H(x) = -\sum p(x)\log_2\big(p(x)\big) \tag{1}$$

(2)

$$MI(X;Y) = H(X) - H(X|Y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p(x,y)\log_2\frac{P(x;y)}{p(x)p(y)}$$

MRMR is one of the most famous single-label feature subset selection algorithms that uses the redundancy and relevancy based on MI [24]. The proposed algorithm in this paper has been inspired by MRMR and it has been adapted to MLFSS.

**Relevancy:** The relevancy between feature $F_i$ and class label L set is defined as follows.

$$Rel(F_i) = MI(F_i; L) \tag{3}$$

Because class label set L has q different labels, equation (3) can be rewritten to equation (4). The large numbers of the class label in set L can result in high computational cost in equation (4), so it has approximated by the equation (5)

(4)

$$Rel(F_i) = MI\big(F_i; L_1, L_2, \dots, L_q\big) = \sum_{j=1}^{q} MI\big(F_i; L_j \big| L_{j-1}, L_{j-2}, \dots, L_1\big)$$

$$Rel(F_i) = \frac{1}{|L|}\sum_{j=1}^{q} MI(F_i; L_j) \tag{5}$$

**Redundancy:** The redundancy between features $F_i$ and other features is expressed as equation (6) in which variable S refers to F-$F_i$.

$$Red(F_i|S) = \frac{1}{|S|}\sum_{F_j\epsilon S} MI(F_i; F_j) \tag{6}$$

The objective in MLFSS is to determine S such that equation (7) be maximized.

$$\max_{F_i\epsilon F}\{Rel(F_i) - Red(F_i|S)\} \tag{7}$$

The most approaches to optimize equation (7) have incremental in nature. As such these methods lead to suboptimal decision as selected features in earlier step cannot deselected in later steps. In this paper, Quadratic programming (QP) is applied to make global optimal decision. In this method, MI among all features will be considered simultaneously. The way that QP should be used is expressed as follows.

### C. Optimization by quadratic programming

Quadratic programming is used to minimize multi variable function with linear constrains. Previously, QP has been

applied to variety of problems successfully [15] as well as feature selection problem in 2010 [16]. MLQPFS uses QP to optimize multi-label feature selection. In [15], standard form of QP has been introduced as follows:

$$\min_{w}\left\{\frac{1}{2}w^{T}Hw - F^{T}W\right\} \tag{8}$$

Equation (7) can be transformed to equation (8) by some reforms [16]. In equation (8), w is a d-dimensional vector and $H\epsilon R^{d\times d}$ is a symmetric-semi positive matrix and F is a non-negative d-dimensional vector. In order to apply equation (8) in multi-label feature selection, H is assumed as similarity or redundancy among features and F is assumed as relevant or relation vector between feature and class labels. After solving the equation (8), every entry in w represents the weight of each feature. Features with higher weight, will be more useful in multi-label learning phase. Because $w_i$ represents weight of i-th feature, therefore, following constrains should be considered in equations (9) and (10)

$$w_i \geq 0 \; for \; all \; i = 1 \ldots q \tag{9}$$

$$\sum_{i=1}^{q} w_i = 1 \tag{10}$$

Quadratic and linear terms in equation (8) can have different relative importance in the objective function therefore a parameter (α) as balancing factor is considered, so equation (8) will be rewritten to equation (11).

$$\min_{w}\left\{\frac{1}{2}(1-\alpha)w^{T}Hw - \alpha F^{T}W\right\} \tag{11}$$

In equation (11), the parameter α is in range [0, 1]. When α =1, quadratic term will be removed and equation (11) is just minimized based on relevancy between features and class labels, whereas, if α =0, the linear term will be removed and equation (11) is just minimized based on the similarity among all features. Thereby, a reasonable choice of α should balance the linear and quadratic terms of equation (11). The value of the parameter α will be evaluated according to the equation (12). Assuming that $\hat{h}$ represents mean of Matrix H and $\hat{f}$ represents mean of vector F.

$$(1-\alpha) * \hat{h} = \alpha * \hat{f} \rightarrow \alpha = \frac{\hat{h}}{\hat{h}+\hat{f}} \tag{12}$$

MLQPFS applies equation (5) to construct vector F and uses mutual information among all features to construct matrix H.

## IV. EXPERIMENTAL RESULTS

In this section, efficiency and effectiveness of the proposed algorithm will be evaluated through comparing with another MLFSS algorithm (PMU). In the following section, evaluation metrics, examined data sets and compared algorithm will be described, respectively.

### A. Evaluation metrics

In order to assess the effectiveness of the proposed algorithm, following steps will be done. Firstly, the selected features will be determined by the execution of MLQPFS. Secondly, original dataset will be filtered based on the selected features. Thirdly, a multi label model classifier will be constructed based on filtered dataset and finally, evaluation metrics of the multi label model classifier will be calculated. Fig.2 illustrates this process.

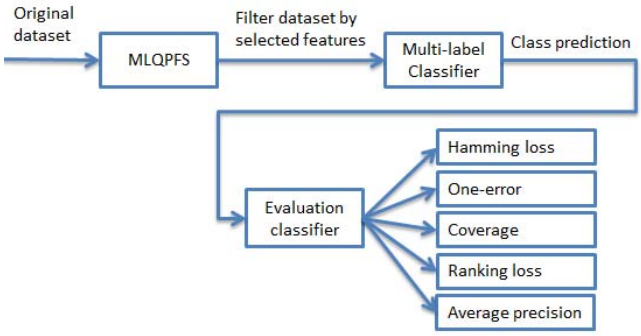


Fig. 2. *The process of experiments*

In multi-label classifier, evaluation metrics differ from traditional single-label classifier, therefore, in this paper, Hamming loss, One-error, Coverage, Ranking loss and Average precision has been used as evaluation metrics instead of accuracy, precision and recall.

Let $= \{(x_i, Y_j)|1 \leq i \leq t\}$ denotes a given test set where $x_i$ represents unseen data instance and $Y_i$ is a predicted label set through multi-label classifier. $h(x_i)$ is also a correct label set. Five evaluation metrics is used in multi-label classification.

- Hamming Loss: this measure evaluates the fraction of instance-label pairs which is misclassified.

$$hloss = \frac{1}{t}\sum_{i=1}^{t}|h(x_i) \vartriangle Y_i| \tag{13}$$

- One-error: this measure evaluates the fraction of top-ranked predicted label is not in relevant label set.

$$oneError = \frac{1}{t}\sum_{i=1}^{t}\left\|\left[argmax_{y\epsilon Y} f(x_i, y)\right] \in Y_i\right\| \tag{14}$$

- Coverage: this measure evaluates how many steps are needed, on average, to move down the label list of an example in order to cover all its the relevant labels

$$coverage = \frac{1}{t}\sum_{i=1}^{t}\max_{y\in Y_i} rankf(x_i, y) - 1 \tag{15}$$

- Ranking Loss: this measure evaluates the average fraction of reversely ordered label pairs for the instance.

$$rLoss = \frac{1}{t}\sum_{i=1}^{t}\frac{1}{|Y_i||\overline{Y}_i|}|\{(l_k, l_j)|\, f(x_i, l_k) \leq f(x_i, l_j), (x_i, l_k) \in (Y_i \times Y_j)\}| \tag{16}$$

- Average precision: this measure evaluates relevant labels ranked above a particular label $l_k \epsilon Y_i$

$$avgprec = \frac{1}{t}\sum_{i=1}^{t}\frac{1}{|Y_i|}\sum_{l_k \in Y_i}\frac{|R(x_i,l_k)|}{rank(x_i,l_k)}, \qquad where \qquad (17)$$
$$R(x_i,l_k) = \{l_j | rank(x_i,l_j) \le rank(x_i,l_k), l_j \in Y_i\}$$

In [25], equations 13-17 have been elaborated. Note that for Hamming-loss, One-error, Coverage, Ranking-loss, the performance of the classification will be adequate if the results of the multi-label classification be low. Conversely, for the average precision, the performance of the classification will be proper if the results of multi-label classification be high.

### B. Data Sets

In the experiments, three datasets has been used from different application areas such as: bioinformatics, semantic scene analysis, and text categorization, Table I, summarizes characteristics of the datasets used in the experiments.

TABLE I. CHARACTERISTICS OF MULTI-LABEL DATASETS

| Name | Domain | patterns | Features | labels |
|---|---|---|---|---|
| Scene | Image | 2407 | 294 | 6 |
| Enron | Text | 1702 | 1001 | 53 |
| Yeast | Biology | 2417 | 103 | 14 |

### C. Experimental results

In order to evaluate the effectiveness of the MLQPFS algorithm, two experiments has been organized and results of MLQPFS and another multi-label feature selection (PMU) have been compared. Firstly p% (p = 10 , 15) of the top selected features (PPTF) by MLQPFS and PMU is fetched. Secondly, filtered dataset by PPTF is sent to the LIFT and finally, the evaluation metrics are computed. Note that in the LIFT, parameter ratio is set to 0.1

Tables II-IV present the experimental results with 10% of PPTF in all datasets. The less number of the selected features means more emphasis on the effectiveness of the selected features in the experiments. Evaluation metrics are placed on rows of the tables and the best result is shown as bold.

TABLE II. EXPERIMENTAL RESULTS ON SCENE DATASET (P=10%)

| Metrics | PMU | MLQPFS |
|---|---|---|
| HammingLoss↓ | 0.1470 | **0.1112** |
| RankingLoss↓ | 0.1951 | **0.1093** |
| OneError↓ | 0.4581 | **0.3043** |
| Coverage↓ | 1.0953 | **0.6471** |
| AveragePrecision↑ | 0.7063 | **0.8161** |

TABLE III. EXPERIMENTAL RESULTS ON YEAST DATASET (P=10%)

| Metrics | PMU | MLQPFS |
|---|---|---|
| HammingLoss↓ | 0.2101 | **0.2041** |
| RankingLoss↓ | 0.1931 | **0.1810** |
| OneError↓ | 0.2606 | **0.2453** |
| Coverage↓ | 6.7764 | **6.5921** |
| AveragePrecision↑ | 0.7308 | **0.7480** |

TABLE IV. EXPERIMENTAL RESULTS ON ENRON DATASET (P=10%)

| Metrics | PMU | MLQPFS |
|---|---|---|
| HammingLoss↓ | 0.0491 | **0.04715** |
| RankingLoss↓ | 0.0915 | **0.0873** |
| OneError↓ | 0.2538 | **0.2487** |
| Coverage↓ | 13.4732 | **13.038** |
| AveragePrecision↑ | 0.6622 | **0.6741** |

Tables II-IV show that MLQPFS on all evaluation metrics with 10% of top selected features have better performance than PMU. Tables V-VII present the experimental results with top 15% of the selected features on all datasets.

TABLE V. EXPERIMENTAL RESULTS ON SCENE DATASET (P=15%)

| Metrics | PMU | MLQPFS |
|---|---|---|
| HammingLoss↓ | 0.1315 | 0.1035 |
| RankingLoss↓ | 0.1318 | 0.0953 |
| OneError↓ | 0.3812 | 0.2734 |
| Coverage↓ | 0.7968 | 0.5744 |
| AveragePrecision↑ | 0.7689 | 0.8353 |

TABLE VI. EXPERIMENTAL RESULTS ON YEAST DATASET (P=15%)

| Metrics | PMU | MLQPFS |
|---|---|---|
| HammingLoss↓ | 0.2089 | **0.2014** |
| RankingLoss↓ | 0.1895 | **0.1780** |
| OneError↓ | 0.2508 | **0.2420** |
| Coverage↓ | 6.7480 | **6.5332** |
| AveragePrecision↑ | 0.7378 | **0.7498** |

TABLE VII. EXPERIMENTAL RESULTS ON ENRON DATASET (P=15%)

| Metrics | PMU | MLQPFS |
|---|---|---|
| HammingLoss↓ | 0.0482 | **0.0471** |
| RankingLoss↓ | 0.0892 | **0.0849** |
| OneError↓ | **0.2504** | 0.2538 |
| Coverage↓ | 13.022 | **12.816** |
| AveragePrecision↑ | 0.6652 | **0.6806** |

It can be seen from the data in Tables V-VII that performance of the MLQPFS is better than PMU on all evaluation metrics except one-error on Enron dataset.

### D. Conclusion

In this paper, an algorithm called MLQPFS is presented for Multi-label feature selection (MLFSS). MLFSS is derived from classic feature subset selection. One of the most significant challenges in MLFSS is to associate each data instance to multi labels simultaneously. The MLFSS in many applications, such as text categorization, gene function classification and semantic annotation of images, is common. More recent studies have confirmed that the relevancy and the redundancy concepts are two important factors to measure features. The relevancy refers to dependency between features and class labels and the redundancy refers to dependency among all features. In order to measure the dependency, mutual information has been used. Then, the MLFSS has been introduced based on the relevancy and the redundancy as optimization equation. For solving the optimization equation, the MLQPFS algorithm has been proposed. The MLQPFS uses quadratic programming therefore, it can find the global optimum. The results show that MLQPFS has better performance than PMU. Overall, this study strengthens the idea that combination of a proper

estimation for relevancy and redundancy and using Quadratic programming can lead to better results than conventional incremental feature selection methods such as PMU.

## REFERENCES


[1] R. Schapire and Y. Singer, "BoosTexter: A Boosting-based System for Text Categorization," *Machine Learning,* vol. 39, pp. 135-168, 2000/05/01 2000.

[2] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.,* vol. 34, pp. 1-47, 2002.

[3] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, "RCV1: A New Benchmark Collection for Text Categorization Research," *J. Mach. Learn. Res.,* vol. 5, pp. 361-397, 2004.

[4] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern recognition,* vol. 37, pp. 1757-1771, 2004.

[5] S. Diplaris, G. Tsoumakas, P. Mitkas, and I. Vlahavas, "Protein Classification with Multiple Algorithms," in *Advances in Informatics*. vol. 3746, P. Bozanis and E. Houstis, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 448-456.

[6] Y. Yang and J. O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," presented at the Proceedings of the Fourteenth International Conference on Machine Learning, 1997.

[7] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognition Letters,* vol. 34, pp. 349-357, 2/1/ 2013.

[8] L. Talavera, "An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering," in *Advances in Intelligent Data Analysis VI*. vol. 3646, A. F. Famili, J. Kok, J. Peña, A. Siebes, and A. Feelders, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 440-451.

[9] P. Langley, "Selection of Relevant Features in Machine Learning," in *In Proceedings of the AAAI Fall symposium on relevance*, ed: AAAI Press, 1994, pp. 140--144.

[10] J. R. Quinlan, "Induction of decision trees," *Machine Learning,* vol. 1, pp. 81-106, 1986/03/01 1986.

[11] S. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Machine Learning,* vol. 16, pp. 235-240, 1994/09/01 1994.

[12] W. Chen, J. Yan, B. Zhang, Z. Chen, and Q. Yang, "Document Transformation for Multi-label Feature Selection in Text Categorization," presented at the Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, 2007.

[13] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multi-label classification of music by emotion," *EURASIP Journal on Audio, Speech, and Music Processing,* vol. 2011, pp. 1-9, 2011/09/18 2011.

[14] J. Read, "A pruned problem transformation method for multi-label classification," in *In: Proc. 2008 New Zealand Computer Science Research Student Conference (NZCSRS)*, ed, 2008, pp. 143--150.

[15] D. P. Bertsekas, *Nonlinear Programming*, 1999.

[16] I. Rodriguez-Lujan, R. Huerta, C. Elkan, and C. S. Cruz, "Quadratic Programming Feature Selection," *J. Mach. Learn. Res.,* vol. 11, pp. 1491-1516, 2010.

[17] G. T. a. I. Katakis, "Multi-label classification: An overview," *Int J Data Warehousing and Mining,* vol. 2007, pp. 1--13, 2007.

[18] G. Doquire and M. Verleysen, "Feature Selection for Multi-label Classification Problems," in *Advances in Computational Intelligence*. vol. 6691, J. Cabestany, I. Rojas, and G. Joya, Eds., ed: Springer Berlin Heidelberg, 2011, pp. 9-16.

[19] M.-L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," *Information Sciences,* vol. 179, pp. 3218-3229, 9/9/ 2009.

[20] Y. Zhang and Z.-H. Zhou, "Multilabel dimensionality reduction via dependence maximization," *ACM Trans. Knowl. Discov. Data,* vol. 4, pp. 1-21, 2010.

[21] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring Statistical Dependence with Hilbert-Schmidt Norms," in *Algorithmic Learning Theory*. vol. 3734, S. Jain, H. Simon, and E. Tomita, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 63-77.

[22] G. Doquire and M. Verleysen, "Mutual information-based feature selection for multilabel classification," *Neurocomputing,* vol. 122, pp. 148-155, 12/25/ 2013.

[23] J. Lee and D.-W. Kim, "Mutual Information-based multi-label feature selection using interaction information," *Expert Systems with Applications,* vol. 42, pp. 2013-2025, 3// 2015.

[24] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 27, pp. 1226-1238, 2005.

[25] Z. Min-Ling and Z. Zhi-Hua, "A Review on Multi-Label Learning Algorithms," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 26, pp. 1819-1837, 2014.