An Effective Multilabel Classification Using Feature Selection

S. S. Sane, Prajakta Chaudhari and V. S. Tidake

Abstract Recently, multilabel classification has received significant attention during the past years. A multilabel classification approach called coupled k-nearest neighbors algorithm for multilabel classification (called here as CK-STC) reported in the literature exploits coupled label similarities between the labels and provides improved performance [Liu and Cao in A Coupled k-Nearest Neighbor Algorithm for Multi-label Classification, pp. 176–187, 2015]. A multilabel feature selection is presented in Li et al. [Multi-label Feature Selection via Information Gain, pp. 346–355, 2014] and called as FSVIG here. FSVIG uses information gain that shows better performance when used with ML-NB, ML-kNN, and RandSvm when compared with existing multilabel feature selection algorithms.This paper investigates the performance of FSVIG when used with CK-STC and compares its performance with other multilabel feature selection algorithms available in MULAN using standard multilabel datasets. Experimental results show that FSVIG when used with CK-STC provides better performance in terms of average precision and one-error.

Keywords Algorithm adaptation · Coupled label similarity · Feature selection Information gain · Multilabel classification

S. S. Sane · P. Chaudhari (✉)
Department of Computer Engineering, K. K. Wagh Institute of Engineering Education and Research, SPPU, Nashik, India
e-mail: p.prajakta42@gmail.com

S. S. Sane
e-mail: sssane@kkwagh.edu.in

V. S. Tidake
Computer Engineering, Matoshri College of Engineering and Research Center, Savitribai Phule Pune University, Nashik, India
e-mail: vaishalitiake@yahoo.co.in

© Springer Nature Singapore Pte Ltd. 2018

S. Bhalla et al. (eds.), *Intelligent Computing and Information and Communication*, Advances in Intelligent Systems and Computing 673,
https://doi.org/10.1007/978-981-10-7245-1_14

1 Introduction

Previously in supervised learning, each example is associated with only one instance and single label is assigned to this instance. As single label is assigned to the instance, this classification is known to be single-label classification. In some situations, where real-world object cannot be handled by single label, this problem can be solved by using multilabel classification. The purpose of multilabel classification is to discover the set of labels for unseen instances. In multilabel classification, input data instances are related to multiple class labels. For example, particular news article can be associated with economical article, political article, and business article.

Multilabel classification is used in several areas such as bioinformatics, text categorization, tag recommendation, image classification, direct marketing, medical diagnosis, query categorization, and protein function prediction [3, 4].

2 Related Work

Multilabel classification techniques are grouped into two categories that are problem transformation and algorithm adaptation method [5].

Problem transformation approach transforms the multilabel problem into a set of single-label problem. Problem transformation approach is algorithm independent. Several problem transformation methods such as binary relevance (BR) [6], label powerset (LP) [5], and classifier chain (CC) [7] are used in multilabel classification.

The algorithm adaptation method extends the available machine learning algorithms so as to handle multilabel data. The several algorithm adaptation approaches have been proposed that are multilabel k-nearest neighbor (ML-kNN) [8], binary relevance k-nearest neighbor (BR-kNN) [9], and IBLR [10].

ML-kNN and BR-kNN do not consider label correlation whereas IBLR considers label correlation but not good in terms of hamming loss.

Liu and Cao [1] proposed coupled k-nearest neighbor algorithm for multilabel classification (CK-STC). CK-STC is based on coupled label similarity [11] which updates ML-kNN algorithm which can handle label correlation. An advantage of CK-STC is that it considers label correlations and provides better performance than that of ML-kNN, IBLR, BSVM, and BR-kNN but it is more complex.

Same as traditional classification, multilabel classification also suffers from curse of dimensionality which can be handled by feature selection. Feature selection selects the relevant or more efficient features from original set of features. Traditional feature selection methods handle multilabel data by transforming multilabel problem into single-label problem and then apply feature selection.

Lee et al. [12] proposed multivariate mutual information-based feature selection. Due to transformation of multilabel problem into single-label problem, there is damage to original label structures and which reduces the performance of classifier.

So new method called information gain feature selection for multilabel data [2] can perform feature selection on multilabel data directly.

Multilabel classification tools MEKA [13] and MULAN [14] are available for handling multilabel data. MEKA is a multilabel extension to WEKA data mining tool. MULAN provides framework for implementations of many multilabel learning algorithms. In MULAN feature selection methods, multilabel data is transformed into single-label data and then traditional single-label dimensionality reduction technique is applied on these single-label data.

3 Implementation Details

3.1 FSVIG Multilabel Feature Selection Algorithm Details

FSVIG multilabel feature selection algorithm is implemented using information gain measure. Feature selection method removes the redundant and irrelevant features from database. Feature selection technique used in this work is reported in [2]. In this feature, selection filter approach is used. Information gain is used as information metric for measuring correlation degree between features and labels that are present in database. Information gain (IG) between features and entire label set is used to quantify the importance of features. IG is calculated with the help of label entropy H(L) and feature entropy H(fi). A bigger value of IG represents greater correlation between feature and labels. But in some situation, information gain of each label set may not be in same range of measurement. So for comparison, the normalized processing was performed on information gain (SU). Figure 1 shows pseudo-algorithm for feature selection FSVIG which is as follows.

3.2 CK-STC Multilabel Classification Algorithm Details

CK-STC algorithm for multilabel classification is reported in [1, 15]. In CK-STC algorithm, coupled label similarity is estimated using intra-coupling similarity between label value w_i^x and w_j^y and inter-coupling label similarity between label value w_i^x and w_j^y w.r.t feature value w_p^z. Inter-coupling label similarity is calculated with the help of co-occurrence frequency (CF) and intra-coupling similarity is calculated with the help of occurrence frequency (F). For every instance in training data, k-nearest neighbor is calculated. Coupled label similarity is considered while calculating prior probabilities and frequency arrays. Estimation of k-nearest neighbor for unseen instance is done. For each label, statistical information is calculated with the help of k-nearest neighbors of unseen instances. Unseen instance labels are evaluated via MAP (Maximum a posterior) rule. MAP rule is based on Bayes theorem.

Input:
F { f1,f2,f3,……fn}
L {l1,l2,l3……lq}
Output:
Relevant set of features
1.begin
2.Initialize IGS = null
3.for(i=1 to m)
4.IG(fi,L)= H(fi)+H(L)-H(fi,L)
5.SU(fi,L) = $\frac{2*IG(fi,L)}{H(fi)+H(L)}$
6.IGSi = SU(f i,L)
7.IGS = IGS ∪IGSi
8.End for
9.threshold = $\frac{1}{m}\sum_{i=1}^{m} IGS_i$
10.Relevant feature = F
11.for(i=1 to m)Do
12.if IGSi < threshold then
13.Relevant feature = Relevant feature - {fi}
14.end for
15.end

Fig. 1 Pseudo-algorithm for feature selection FSVIG

Figure 2 shows pseudo-algorithm for multilabel classification CK-STC which is stated as follows:

4 Experimental Setup

Table 1 shows benchmark multilabel datasets information including number of features, total number of labels, and number of instances. Experiments were performed on Genbase, Medical, and Enron dataset.

Experiments were performed on i5 processor and on Windows 7 operating system. Implementation of multilabel classification algorithm CK-STC and multilabel feature selection algorithm FSVIG was done in MS VISUAL STUDIO.NET and development tool is visual studio 10. In experiments, a proposed (CK-STC with feature selection) method was compared with existing CML-kNN method. Euclidean distance was used to calculate k-nearest neighbors. The experiments were performed on k = 5, 7, 9 and then the average is calculated. Tenfold cross-validation is performed on the above dataset.

Input:
Unlabeled instance (u_t), multilabel data.
Output:
Label set of unlabeled instance.
1. begin
2. Calculate intra coupling similarity $\delta^{ia}\left(w_i^x, w_j^y\right) = \frac{F\left(w_i^x\right).F\left(w_j^y\right)}{F\left(w_i^x\right)+F\left(w_j^y\right)+F\left(w_i^x\right).F\left(w_j^y\right)}$
3. Calculate inter coupling similarity $\delta^{ier}\left(w_i^x, w_j^y | w_p^z\right) = \frac{min\left(CF\left(w_p^{zx}\right), CF\left(w_p^{zy}\right)\right)}{max\left(F\left(w_i^x\right), F\left(w_j^y\right)\right)}$
4. Calculate coupled label similarity (CLS) $CLS\left(w_i^x, w_j^y\right) = \delta^{ia}\left(w_i^x, w_j^y\right).\sum_{k=1}^{m} \delta^{\mathrm{ier}}\left(\mathrm{w_i^x}, \mathrm{w_j^y} | \mathrm{w_k}\right)$
5. for i=1 to m do
6. Estimate k nearest neighbor k(u_i) for each u_i
7. end for
8. for j=1 to q do
9. Estimate prior probability P(Hj) and P(-Hj)
10. Maintain frequency array
11. End for
12. Estimate k nearest neighbor k(u_t) for unseen instance u_t
13. for j = 1 to q
14 .Estimate unseen instance label statics
15. End for
16. Return label set of unlabeled instance u_t according to MAP rule
17. end

Fig. 2 Pseudo-algorithm for multilabel classification CK-STC

Table 1 Dataset Description

Datasets	Number of instances	Number of features	Number of total labels	Label cardinality	Label density	Feature type
Genbase	662	1185	27	3.378	0.064	Categorical
Medical	978	1449	45	1.252	0.046	Categorical
Enron	1702	1001	53	1.245	0.028	Categorical

5 Experimental Results

Performance of multilabel classification is measured in terms of hamming loss, one-error, and average precision. For each evaluation metric, "↓" indicates "smaller value has better results" and "↑" indicates "bigger value has better results". Bold value indicates winner of the classifier. Experiments were carried out in four sets as follows:

(1) **Multilabel classification using CK-STC without feature selection**
Table 2 shows results of CK-STC algorithm.

Table 2 Results of CK-STC

Dataset	One-error ↓	Hamming loss ↓	Average precision ↑
Enron	0.303	0.087	0.597
Genbase	0.008	0.003	0.991
Medical	0.158	0.013	0.874
Average	**0.156**	**0.034**	**0.821**

(2) **Multilabel feature selection using FSVIG algorithm**
Table 3 indicates results of feature selection FSVIG.

(3) **Multilabel classification using MULAN feature selection technique binary relevance attribute evaluator and its comparison with CK-STC with feature selection FSVIG**

The results of multilabel feature selection algorithm FSVIG and MULAN attribute selection algorithms such as binary relevance attribute evaluator algorithm represented in Tables 4, 5 and 6.

Table 4 indicates that CC algorithm with binary relevance attribute evaluator performs better than other algorithms w.r.t. hamming loss. Table 5 indicates that CK-STC algorithm with FSVIG feature selection technique performs better than other algorithms w.r.t. one-error. Table 6 shows that CK-STC algorithm with FSVIG feature selection technique performs better than other algorithms w.r.t average precision.

(4) **Multilabel classification using MULAN feature selection technique multiclass attribute evaluator and its comparison with CK-STC with feature selection FSVIG**

The results of multilabel feature selection algorithm FSVIG and MULAN attribute selection algorithms such as multiclass attribute evaluator algorithm represented in Tables 7, 8, and 9.

Table 7 indicates that CK-STC algorithm with FSVIG feature selection technique performs better than other algorithms w.r.t. hamming loss. Table 8 indicates that CK-STC algorithm with FSVIG feature selection technique performs better than other algorithms w.r.t. one-error. Table 9 indicates that CK-STC algorithm with FSVIG feature selection technique performs better than other algorithms w.r.t average precision.

Table 3 Results of feature selection as per FSVIG

Dataset	Number of features before feature selection	Number of features after feature selection
Enron	1001	596
Genbase	1185	316
Medical	1449	400

Table 4 Comparisons of CK-STC, feature selection FSVIG with MULAN binary relevance attribute evaluator w.r.t. hamming loss ↓

Dataset	Number of attribute used	CK-STC	FSVIG	MULAN (binary relevance attribute evaluator)							
			CK-STC	CK-STC	BR	CC	RAKEL	ML-kNN	LP	BR-kNN	IBLR
Enron	596	0.087	0.054	0.052	0.049	0.052	0.047	0.052	0.069	0.056	0.055
Genbase	316	0.003	0.002	0.004	0.001	0.001	0.001	0.005	0.001	0.003	0.002
Medical	400	0.013	0.012	0.012	0.011	0.001	0.011	0.012	0.013	0.014	0.013
Average	–	0.034	0.022	0.023	0.020	**0.018**	0.020	0.023	0.027	0.025	0.024

Table 5 Comparisons of CK-STC, feature selection FSVIG with MULAN binary relevance attribute evaluator w.r.t. one-error ↓

Dataset	Number of attribute used	CK-STC	FSVIG	MULAN (binary relevance attribute evaluator)							
			CK-STC	CK-STC	BR	CC	RAKEL	ML-kNN	LP	BR-kNN	IBLR
Enron	596	0.303	0.225	0.285	0.369	0.397	0.268	0.282	0.832	0.441	0.332
Genbase	316	0.008	0.007	0.013	0.003	0.003	0.010	0.006	0.006	0.016	0.013
Medical	400	0.158	0.146	0.192	0.190	0.197	0.238	0.200	0.262	0.283	0.231
Average	–	0.156	**0.126**	0.163	0.187	0.200	0.172	0.163	0.366	0.247	0.192

Table 6 Comparisons of CK-STC, feature selection FSVIG with MULAN binary relevance attribute evaluator w.r.t. average precision ↑

Dataset	Number of attribute used	CK-STC	FSVIG	MULAN (binary relevance attribute evaluator)							
			CK-STC	CK-STC	BR	CC	RAKEL	ML-kNN	LP	BR-kNN	IBLR
Enron	596	0.597	0.617	0.643	0.613	0.595	0.622	0.643	0.211	0.578	0.626
Genbase	316	0.991	0.996	0.986	0.992	0.992	0.989	0.989	0.990	0.982	0.986
Medical	400	0.874	0.913	0.855	0.839	0.820	0.772	0.850	0.727	0.803	0.829
Average	–	0.821	**0.842**	0.828	0.815	0.802	0.794	0.827	0.642	0.788	0.813

Table 7 Comparisons of CK-STC, feature selection FSVIG with MULAN binary relevance attribute evaluator w.r.t. hamming loss ↓

Dataset	Hamming loss ↓										
	Number of attribute used	CK-STC	FSVIG	MULAN (multiclass attribute evaluator)							
			CK-STC	CK-STC	BR	CC	RAKEL	ML-kNN	LP	BR-kNN	IBLR
Enron	596	0.087	0.054	0.056	0.056	0.060	0.055	0.056	0.070	0.061	0.060
Genbase	316	0.003	0.002	0.004	0.001	0.001	0.004	0.004	0.001	0.038	0.002
Medical	400	0.013	0.012	0.012	0.010	0.010	0.011	0.011	0.011	0.013	0.013
Average	–	.034	**0.022**	0.024	0.022	0.023	0.024	0.024	0.027	0.037	0.025

Table 8 Comparisons of CK-STC, feature selection FSVIG with MULAN multiclass attribute evaluator w.r.t one-error ↓

Dataset	One-error ↓										
	Number of attribute used	CK-STC	FSVIG	MULAN (multiclass attribute evaluator)							
			CK-STC	CK-STC	BR	CC	RAKEL	ML-kNN	LP	BR-kNN	IBLR
Enron	596	0.303	0.225	0.366	0.442	0.473	0.383	0.367	0.954	0.565	0.431
Genbase	316	0.008	0.007	0.013	0.003	0.003	0.010	0.006	0.004	0.016	0.013
Medical	400	0.158	0.146	0.117	0.162	0.167	0.184	0.185	0.215	0.233	0.217
Average	–	0.156	**0.126**	0.185	0.202	0.214	0.192	0.186	0.391	0.271	0.220

Table 9 Comparisons of CK-STC, feature selection FSVIG with MULAN multiclass attribute evaluator w.r.t average precision ↑

Dataset	Number of attribute used	CK-STC	FSVIG	MULAN (multiclass attribute evaluator)							
			CK-STC	CK-STC	BR	CC	RAKEL	ML-kNN	LP	BR-kNN	IBLR
Enron	596	0.597	0.617	0.587	0.568	0.549	0.515	0.590	0.108	0.555	0.569
Genbase	316	0.991	0.996	0.986	0.992	0.992	0.989	0.989	0.991	0.982	0.986
Medical	400	0.874	0.913	0.862	0.858	0.838	0.818	0.855	0.768	0.825	0.829
Average	–	0.821	**0.842**	0.812	0.806	0.793	0.774	0.812	0.622	0.787	0.795

Tables 6, 7, 8, and 9 indicate that feature selection technique FSVIG with CK-STC gives better performance in terms of one-error and average precision than feature selection algorithms reported in MULAN. Multilabel classification with FSVIG feature selection does not provide best performance in terms of hamming loss because of weak connection between labels in the datasets. Feature selection algorithms in MULAN library transform multilabel problem into single-label problem and on single-label dataset, feature selection methods are applied; so due to transformation there is damage to original label structures which reduces the performance of classifier.

For evaluating the statistical significance of the results, the Friedman test for paired data is used. The level of significance of the Friedman test was determined at $\alpha = 0.05$.

6 Conclusions

A lazy learning approach CK-STC has been reported in the literature for multilabel classification using coupled similarity between labels that improves the prediction performance, but irrelevant features present in database affect the prediction accuracy of classifier.

A multilabel feature selection FSVIG is incorporated in CK-STC algorithm. FSVIG uses information gain that shows better performance than existing multilabel feature selection algorithms when used with ML-NB, ML-kNN, and RandSvm classifier. This paper investigates a performance of FSVIG when used with CK-STC and compares its performance with other multilabel feature selection algorithms available in MULAN using standard multilabel datasets.

Experimental results show that FSVIG when used with CK-STC provides better performance in terms of average precision and one-error.

References

1. Chunming Liu and Longbing Cao, "A Coupled k-Nearest Neighbor Algorithm for Multi-label Classification", Springer PAKDD, Part I, LNAI 9077, pp. 176–187, (2015).
2. Ling Li, Huawen Liu, Zongjie Ma, Yuchang Mo, Zhengjie Duan, Jiaqing Zhou, Jianmin Zhao, "Multi-label Feature Selection 6a Information Gain", 345–355, (2014).
3. Tsoumakas, G., Katakis, I. "Multi-label classification: An over6ew" International Journal of Data Warehousing and Mining (IJDWM) 3(3), 1–13 (2007).
4. Gjorgji Madjarov, DragiKocev, Dejan Gjorgje6kj, Saso Dzeroski, "An extensive experimental comparison of methods for multi-label learning", Pattern Recognition 45 (2012) 3084–3104.
5. Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas, "Mining Multi-label Data", Data Mining and Knowledge Discovery Handbook, Springer, 2010, pp. 667–685.
6. Boutell, M.R., Luo, J., Shen, X., Brown, C.M. "Learning multi-label scene classification", Pattern recognition 37(9), 1757–1771 (2004).

7. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. Machine Learning 85(3), 333–359 (2011).
8. Zhang, M.L., Zhou, Z.H., "ML-kNN: A lazy learning approach to multi-label learning", Pattern recognition 40(7), 2038–2048 pp. 401–406. Springer (2007).
9. Spyromitros, E., Tsoumakas, G., Vlahavas, I., "An empirical study of lazy multilabel classification algorithms", Artificial Intelligence: Theories, Models and Applications, pp. 401–406. Springer (2008).
10. Cheng, W., HÂ´lullermeier, E., "Combining instance-based learning and logistic regression for multi-label classification", Machine Learning 76(2–3), 211–225 (2009).
11. Cao, L., "Coupling learning of complex interactions", Information Processing and Management (2014).
12. Lee, J., Kim, D.W. "Feature selection for multi-label classification using multivariate mutual information", Pattern Recognition Letters 34(3), 349 – 357 (2013).
13. Tidake Vaishali S. And Shirish S. Sane, "Multilabel Learning with MEKA", CSI communications (2016).
14. G. Tsoumakas, E. Spyromitros-Xioufis, J. 6lcek, and I. Vlahavas, "MULAN: A Java library for multi-label learning" Machine Learning. Res., vol. 12, pp. 2411–2414, Jul.(2011).
15. Prajakta Chaudhari and S. S. Sane, "Multilabel Classification Exploiting Coupled Label Similarity with Feature Selection", IJCA ETC (2016).