k-Labelsets for Multimedia Classification with Global and Local Label Correlation

Yan Yan, Shining Li(✉), Xiao Zhang, Anyi Wang, Zhigang Li, and Jingyu Zhang

School of Computer Science, Northwestern Polytechnical University, Xi'an, China
yanyan.nwpu@gmail.com, {lishining,lizhigang}@nwpu.edu.cn, {zhang_xiao,hay,zhangjingyu}@mail.nwpu.edu.cn

Abstract. Multimedia data, e.g., text and images, can be associated with more than one label. Existing methods for multimedia data classification either consider label correlation globally by assuming that it is shared by all the instances; or consider label correlations locally by assuming that it is a pairwise label correlation and shared only in a local group of instances. In fact, both global and local correlations may occur in the real-world applications; and the label correlation cannot be confined to pairwise labels. In this paper, a novel and effective multi-label learning approach named *GLkEL* is proposed for multimedia data categorization. Briefly, a *H*igh-*O*rder *L*abel *C*orrelation *A*ssessment strategy named *HOLCA* is proposed by using approximated joint mutual information; and then *GLkEL*, which breaks the original label set into several of the most correlated and distinct combination of k labels (called k-lab*EL*sets) according to the *HOLCA* strategy, learns *G*lobal and *L*ocal label correlations simultaneously based on label correlation matrix. Comprehensive experiments across 8 data sets from different multimedia domains indicate that, it manifests competitive performance against other well-established multi-label learning methods.

Keywords: Multimedia data · Global and local label correlations
Multi-label classification

1 Introduction

With rich information presented in multimedia data, many classification tasks require to assign more than one label to each instance under real-world scenarios. For example, an image can be annotated with many tags [1], a magazine may be associated with multiple topics [2], a piece of music may belong to several genres [3]. Thus, multi-label classification problem, where each instance corresponds to a set of class labels, has received an increasing attention in recent years [4].

Current studies on multi-label learning try to incorporate label correlation of different orders to facilitate the learning process [4]. However, the existing approaches mostly concentrate on global label correlation by assuming that it is

K. Schoeffmann et al. (Eds.): MMM 2018, Part II, LNCS 10705, pp. 177–188, 2018.
https://doi.org/10.1007/978-3-319-73600-6_16

shared by all the instances [1,5]. For example, 'bird' and 'sky' are highly correlated labels, and so are 'Yao Ming' and 'basketball star'. On the other hand, some label correlations, which can be distinct in different data groups, may be only shared by a local data group [6,7]. For example, 'Jordan' is related to 'shoe' in life magazines, but 'basketball star' in basketball magazines. *ML-LOC* [6], which is the first work trying to model local label correlations, demonstrates that label correlations may be only shared by a local data group. It gets a satisfactory performance by exploiting label correlations locally, while may suffer from the fact that label correlations could not be confined locally. The previous research described above considers label correlation either from a global or local perspective, however, both of them may occur under real-world scenarios. So, taking both global and local label correlations into consideration is more desirable and practical.

Label correlations, which may change in different locales, are hard to mine manually. Several methods mine label correlations by label clustering based on the assumption of label hierarchies [8]; or construct a Bayesian network model based on the Bayesian assumption [9]. However, the hierarchical structure, which is not a universal architecture, only exists in few real-world applications. Others learn label dependency by the co-occurrence of labels in training data [10]. For example, *GCC* [7] estimates label correlations by the co-occurrence of each pair of labels. It is effective to some extent by considering pairwise label correlations locally, while may suffer from the fact that the label correlations may go beyond *second-order*. What's worse, co-occurrence is less meaningful, since there exists an intrinsic characteristic of multimedia data, i.e. the widely-existing class-imbalance among labels.

In this paper, we propose a novel and effective approach called '***k***-Lab**EL**sets for Multimedia Classification with **G**lobal and **L**ocal Label Correlation' (*GLkEL*). *GLkEL* tackles multi-label classification problem with four simple steps. Firstly, a *H*igh-*O*rder *L*abel *C*orrelation *A*ssessment strategy named *HOLCA* is proposed, which can estimate the correlation among multiple labels based on approximated joint mutual information. Secondly, *GLkEL* learns local label correlation for each group respectively on the base of clustering, and then combines global label correlation with it linearly. Thirdly, *GLkEL* exploits global and local label correlations simultaneously based on the integration of label correlation above, through breaking the original label set into several of the most correlated and distinct combination of k labels (called k-lab*EL*sets) according to the *HOLCA* strategy in different groups; and then, employs *Label Powerset* (*LP*) to train the linear classifier using specific *k-labelsets* for each group. Lastly, the classifier, which corresponds to the nearest group for a test instance, is used to predict.

2 Related Works

During the last decade, many multi-label learning methods have been witnessed [4]. The existing approaches, which mainly focus on exploiting label correlation, can be grouped into three categories based on the degree of label correlation used [9]:

First-order methods refer to addressing the problem of multi-label classification by breaking up it into several independent binary ones. For example, *BR* [1] trains a classifier for each label independently. Although simple and easy to implement, these methods are not so effective due to lack of label correlation.

Second-order methods deal with multi-label learning problem by taking relevance between pairwise labels into account. For example, *CLR* [11] deals with multi-label learning problem by the pairwise label ranking. Although be used to some extent and relatively effective, the correlation cannot be confined to pairwise labels under real-world scenarios.

High-order methods tackle multi-label learning problem by taking *high-order* relation among labels into account. For example, *RAkEL* [5] transforms multi-label learning problem into an ensemble of multi-class classification problems, where each multi-class classification problem is generated by adopting the *LP* [1] on a randomly selected *k-labelset* from label space. High-order methods could address more realistic label correlations, while it may have high model complexities.

All the studies mentioned above focus on learning label correlations globally. However, *ML-LOC* [6] presents that label correlations may be local and only shared by a local data group. Specifically, it exploits local label correlations by augmenting the original features with a *LO*cal *C*orrelation (*LOC*) code for each instance. It performs well with considering local label correlations, while may not get rid of the following drawbacks. Firstly, the *LOC* code may be less useful for high-dimensional data. Secondly, *ML-LOC* only learns label correlations locally, does not take global ones into account. *GCC* [7], which estimates pairwise label dependency by using co-occurrence pattern, enhances the feature representation of each instance by embedding the pairwise label correlations code into feature space; and then learns the label correlation graph locally in each group by clustering. It has a certain advantage in some cases by taking pairwise local label correlations into account, while may suffer from the following limitations. Firstly, co-occurrence is meaningless, since few positive instances exist in real-world scenes. Secondly, label correlations may be *high-order* and not be limited locally.

It is easily seen that considering both global and local label correlations simultaneously is more practical and beneficial.

3 The Proposed Approach

In this section, details of the proposed approach *GLkEL* will be presented. First, we propose a method named *HOLCA* to model *high-order* correlation among labels, and then give the details of the *GLkEL* framework.

3.1 Preliminaries

In multi-label learning, an instance can be associated with more than one label. Let $D = \mathbb{R}^n$ be the n-dimensional sample space and $L = \{l_1, l_2, \cdots, l_q\}$ be the finite set of q possible labels. $T = \{(x_i, Y_i)\big| i = 1, 2, \cdots, d\}$ denotes the multi-label training set with d samples, where $x_i \in D$ is a n-dimensional feature vector so that $x_i = [x_i^1, x_i^2, \cdots, x_i^n]$. We denote $Y = [Y_1, \cdots, Y_d]^T \in \{-1, 1\}^{d\times q}$ as the ground truth label matrix, where $Y_i = [l_{i1}, l_{i2}, \cdots, l_{iq}]$ is the set of label vector associated with x_i. Each element $l_{ij} = 1$ if the label l_{ij} is associated with x_i, otherwise $l_{ij} = 0$. k is the size of the labelsets for data set. g is the number of clusters for one data set.

3.2 HOLCA Strategy

In this subsection, a simple but effective *H*igh-*O*rder *L*abel *C*orrelation *A*ssessment strategy named *HOLCA* is proposed, which evaluates *high-order* label correlation based on approximated joint mutual information. The target of label correlation estimate strategy is to select the most correlated *k-labelset* S from label space L. The S with k labels $\{l_1, l_2, \cdots, l_k\}$ can be chosen from L through maximizing the Interaction Information shared by all k labels. The process is defined as:

$$\arg\max_{S\subseteq L} I(S) \equiv \arg\max_{S\subseteq L} I(l_1; l_2; \cdots; l_k) = \arg\max_{S\subseteq L} \sum_{\Gamma\subseteq S} -(-1)^{|S|-|\Gamma|} H(\Gamma) \quad (1)$$

It is an alternating (inclusion-exclusion) sum over all the subsets $\Gamma \subseteq S$, where $|S| = k$.

However, direct calculation of Eq. (1) is impractical, and an exhaustive strategy has an unacceptable time complexity when k is large. To address the issue, a simple iterative greedy strategy is adopted: given a *k-labelset* S of k - 1 selected labels $\{l_1, l_2, \cdots, l_{k-1}\}$, the next label l_k is chosen by maximizing Eq. (2) in an incremental fashion, as follows:

$$\arg\max_{l_k \epsilon L\setminus S} I(l_k; S) \quad (2)$$

Simply, let us describe the approach for choosing the most correlated label l_i from L, so that it maximizes the mutual information (MI) between l_i and S, $I(l_i; S) = H(l_i) + H(S) - H(l_i; S)$, where $H(x) = -\sum_x p(x)\log(p(x))$ is the entropy of x. When $S = \{l_1, l_2\}$ has two labels, $I(l_i; S)$ can be computed according to Eq. (3),

$$I(l_i; S) = I(l_i; l_1, l_2) = H(l_i) + H(l_1, l_2) - H(l_i, l_1, l_2) \quad (3)$$

It is intractable to calculate $H(x)$ accurately when $x = (x_1, x_2, \cdots, x_n)$ is a multivariate discrete random variable. To estimate $H(x)$ effectively, we approximate it by Shearer's inequality [12], $H(l_i, l_1, l_2) \leq \frac{1}{2}(H(l_i, l_1) + H(l_i, l_2) + H(l_1, l_2))$; and then we can describe the lower bound of $I(f_i; S)$ as follows,

$$I(l_i; S) = I(l_i; l_1, l_2) \geq \frac{1}{2}(I(l_i; l_1) + I(l_i; l_2)) + \theta, \quad (4)$$

where $\theta = -\frac{1}{2}I(l_1; l_2)$ is a constant w.r.t l_i. It is easily seen that the lower bound of $I(l_i; S)$ has the same variation tendency with the correlation between the candidate label and the selected labels. The correlation between l_i and S can be approximated as Eq. (5) according to Eq. (4),

$$\hat{I}(l_i; S) = \sum_{l \in S, l_i \in L \setminus S} I(l_i; l) \tag{5}$$

After l_i with highest $\hat{I}(l_i; S)$ is chosen as the most correlated label with S, the next label l_{i+1} is chosen by maximizing *MI* when it is included. Using the approximation in Eq. (5) for the term, we can get the following estimate,

$$J(l_{i+1}) = \hat{I}(l_{i+1}; l_i, S) = \sum_{l \in S, l_{i+1} \in L \setminus (S, l_i)} I(l_{i+1}; l) + I(l_{i+1}; l_i), \tag{6}$$

The label l_{i+1} is chosen as subsequent labels by maximizing $J(l_{i+1})$. Suppose a labelset S with i labels $\{l_1, l_2, \cdots, l_i\}$ has been chosen, then we select the next label l_{i+1} in an incremental fashion by optimizing the following formula,

$$\arg\max_{l_{i+1} \in L \setminus S} \sum_{l \in S} I(l_{i+1}; l) \tag{7}$$

3.3 *GLkEL* Framework

Cluster the data into groups: To exploit label correlations locally, we cluster the data set into different groups. Suppose that multi-label training dataset T can be decomposed into g groups $\{T_1, \cdots, T_g\}$, where $T_m \in \mathbb{R}^{d_m \times n}$ has d_m instances. In our experiments, we simply choose *kmeans* as the clustering method, and the similarity is calculated by Euclidean distance. Since instances from distinct data groups may share different label correlations, so the local label correlations may vary from group T_1 to T_g. Moreover, both local and global label correlations, which may occur in real-world applications, will be learned simultaneously in the next step.

Learn global and local label correlations: It is well known that exploiting label correlations is an essential ingredient in multi-label learning. Let $M_0 = [M_{ij}] \in \mathbb{R}^{q \times q}$ be the global label correlation matrix of Y. *GLkEL* exploits *high-order* global label correlation using the matrix M_0, through partitioning L into $\alpha = \lceil q/k \rceil$ most correlated and disjoint *k-labelset* R_j according to *HOLCA* strategy, where $j = 1 \cdots \alpha$, $\bigcap_{j=1}^{\alpha} R_j = \emptyset$.

As mentioned in the cluster stage, label correlations may vary from group T_1 to T_g. Let Y_m be the label submatrix in Y corresponding to T_m, and $M_m \in \mathbb{R}^{q \times q}$ is the local label correlation matrix of group m. Similar to global label correlation, we learn *high-order* local label correlations according to the same strategy by using the corresponding matrix M_m.

By combining global and local label correlation matrix, we have the Eq. (8) as follows,

$$\hat{M}_m = \lambda_1 M_0 + \lambda_2 M_m, \tag{8}$$

where $m = 1 \cdots g$, and λ_1, λ_2 are trade-off parameters for controlling the weight between global and local label correlations.

Then, we can learn global and local label correlation structures simultaneously, through choosing α *k-labelsets* according to *HOLCA* strategy using the corresponding matrix $\{\hat{M}_m\}_{m=1}^{g}$ which is a linear combination of label correlation matrices using Eq. (8) for each group. Namely, g label correlation structures $(\{G_i\}_{i=1}^{g})$ can be obtained.

Build the multi-label classifier: In this stage, we build g multi-label classification models $(H_i(x)_{i=1}^{g})$ based on the corresponding learned label correlation structure G_i and the training subset T_i.

$$H_i(x) = \{h_{i1}(x), h_{i2}(x), \cdots, h_{i\alpha}(x)\} \tag{9}$$

Each binary classifier $h_{i\hat{\alpha}}(x)$, which confronts a single-label classification job having the label values (namely, all the *k-labelset* of R_j found in T_i), is learned by using *LP*. The training set for $h_{i\hat{\alpha}}(x)$, denoted as t_i, only contains a group of the training set associated with the intersection of their original annotations and R_j: $t_i = \{(x_i, Y_i \bigcap R_j) \big| i = 1, 2, \cdots, d_i\}$.

It is noted that the label correlation structure $\{G_i\}_{i=1}^{g}$ is constructed by α distinct *k-labelsets* according to *HOLCA* strategy, which means that multi-label classification model $\{H_i(x)\}_{i=1}^{g}$ for all the labels need to be trained only once.

Predict: Intuitively, similar instances may have similar labels. So, we find the nearest group T_i of the test instance x_t by computing Euclidean distance. Suppose that x_t shares similar label dependency with the instances in group T_i. Classification model $H_i(x)$ is learned by the label correlation structure G_i which is composed of the local label correlations on group T_i and global label correlation. $H_i(x)$ is used to predict x_t, and we expect that $H_i(x)$ works better than other classification models which are learned on the group-specific local label correlations and global label correlation simultaneously.

4 Experiments

4.1 Data Sets

In order to evaluate the effectiveness of the proposed multi-label learning approach, 8 real-world multimedia data sets have been adopted in this paper. For each data set $D = \{(X_i, Y_i) \big| 1 \leqslant i \leqslant p\}$, we use $\big|D\big|$, *dim(D), L(D)* and *F(D)* to denote the *number of examples, number of features, number of possible class labels, and feature type* for D respectively. Moreover, some other multi-label properities [4] are denoted as follows:

- $LCard(D) = \frac{1}{p}\sum_{i=1}^{t}\left|Y_i\right|$: *label cardinality* measures the average number of labels for each instance;
- $LDen(D) = \frac{LCard(D)}{L(D)}$: *label density* normalizes *LCard(D)* via the size of label space;
- $DL(D) = \left|\{Y \middle| (X,Y) \in D\}\right|$: *distinct label sets* counts the size of distinct label combinations in D;
- $PDL(D) = \frac{DL(D)}{|D|}$: *proportion of distinct label sets* which normalizes *DL(D)* by the size of instances.

Table 1 summarizes some detailed characteristics of multimedia data sets adapted in our experiments. The 8 data sets, which are ordered by $\left|D\right|$, are chosen from four different application areas, such as text, images, audio and music. Therefore, comprehensive data sets with a broad range of multimedia properties are used in our experiments. For each dataset, we randomly select 90% of the instances for training, and the rest for testing.

Table 1. Characteristics of the experimental data sets

Data set	$\|D\|$	$dim(D)$	$L(D)$	$F(D)$	$LCard(D)$	$LDen(D)$	$DL(D)$	$PDL(D)$	Domain	URL*
Flags	194	10	7	Numeric	3.392	0.485	54	0.278	Images	URL 1
Emotion	593	72	19	Numeric	1.869	0.311	27	0.046	Music	URL 1
Birds	645	258	6	Numeric	1.014	0.053	133	0.206	Audio	URL 1
Enron	1702	1001	53	Nominal	3.378	0.064	753	0.442	Text	URL 2
Image	2000	294	5	Numeric	1.236	0.247	20	0.010	Images	URL 2
Scene	2407	294	6	Numeric	1.074	0.179	15	0.006	Images	URL 1
Rev1 (subset 1)	6000	944	101	Numeric	2.880	0.029	1028	0.171	Text	URL 1
Bibtex	7395	1836	159	Nominal	2.402	0.015	2856	0.386	Text	URL 1

*URL 1: http://mulan.sourceforge.net/datasets.html
URL 2: http://cse.seu.edu.cn/people/zhangml/Resources.htm#data

4.2 Performance Evaluation

To evaluate the performance of multi-label learning algorithms effectively, four popular evaluation metrics in multi-label learning are adopted in this paper. Let $\tau = \{(x_i, Y_i) \middle| 1 \leqslant k \leqslant t\}$ be a multi-label testing set, where $Y_i \in \{0,1\}^q$ is the ground truth labels of the ith instance, and $\hat{Y}_i$ is the predicted labels. Let $\{f_1, f_2, \cdots, f_q\}$ be a group of q learned functions, and P_j^+, N_j^- be the sets of positive and negative instances belonging to the jth label.

- *Average AUC:*
  $$avgauc = \frac{1}{q}\sum_{j=1}^{q}\frac{\left|\hat{P}_j\right|}{\left|P_j^+\right|\left|P_j^-\right|}$$

where $\hat{P}_j = \{(i^{'}, i^{''})\middle| f_j(x_{i^{'}}) \geq f_j(x_{i^{''}}), (i^{'}, i^{''}) \in P_j^+ \times N_j^-\}$. Average AUC evaluates the quality of the prediction for each class label pairs and returns the averaged value across all the class labels.

– *Macro F1*:
$macroF1 = \frac{1}{q} \sum_{i=1}^{q} \frac{2p_i r_i}{p_i + r_i}$

where p_i and r_i are the precision and recall for the ith label. $MacroF1$ is defined as the harmonic mean of precision and recall.

– *Coverage:*
$coverage = \frac{1}{q} \left(\frac{1}{t} \sum_{i=1}^{t} \max_{L_k \in Y} rank(x_i, L_k) - 1 \right)$

where $rank(x_i, L_k) = \sum_{j=1}^{q} \tau\left(f_j(x_i) \geq f_k(x_i)\right)$ returns the sequence of L_k when all class labels in Y are ranked by descending order according to $\{f_1(x_i), f_2(x_i), \cdots, f_q(x_i)\}$. Coverage evaluates the average depth to move down the ranked label list so as to cover all the possible labels of instance.

– *Accuracy*:
$accuracy = \frac{1}{t} \sum_{i=1}^{t} \frac{\left|Y_i \wedge \hat{Y}_i\right|}{\left|Y_i \vee \hat{Y}_i\right|}$

Accuracy evaluates Jaccard similarity between the ground truth labels Y_i and the predicted labels $\hat{Y}_i$.

It is noted that *Coverage* metric is normalized by the number of possible class labels (i.e. q) in this paper; and the values of the four multi-label metrics vary in [0, 1]. For *Average AUC, Macro F1, Accuracy*, the larger value the better performance; While for *Coverage*, the smaller value the better performance.

4.3 Compared Methods

To validate the effectiveness of the proposed approach, we compare *GLkEL* against the following state-of-art multi-label learning algorithms:

1. *BR* [1], which is a first-order method, trains a binary classifier for each label independently;
2. *RAkEL* [5], which is a high-order method, transforms multi-label learning problem into several multi-class learning problems by exploiting *high-order* global label correlation;
3. *ML-LOC* [6], which is a high-order method, exploits label correlations locally by embedding them into instances' feature space;
4. *GCC* [7], which is a second-order method, exploits pairwise label correlations locally by learning a local label dependency graph.

Note that *BR* does not consider label correlation. *RAkEL* only takes *high-order* label correlation globally into account. Both *ML-LOC* and *GCC* only consider local label correlations.

For simplicity, we set $\lambda_1 = 1$ and $k = 3$ in *GLkEL*. The other parameters, as well as the compared methods, are chosen by *ten-fold cross validation (10-CV)* on the training set.

In order to ensure the fair comparison, *LIBSVM* [13] is adopted as the base learner for classifier induction to instantiate *GLkEL, BR, RAkEL, ML-LOC, GCC*. 3 is set to the size of labelsets k and distinct fashion is used in *RAkEL*. All the methods are performed using Matlab.

Table 2. Experimental result of each comparison algorithm (mean±std. deviation) on the eight multimedia data sets

Evaluation criterion	Algorithm	flags	emotion	birds	enron	image	scene	rev1-s1	bibtex
Average AUC ↑	GL*k*EL	**0.707±0.016**	**0.762±0.009**	**0.755±0.012**	**0.734±0.002**	0.711±0.007	**0.835±0.007**	0.688±0.002	0.673±0.002
	BR	0.581±0.006	0.699±0.008	0.581±0.006	0.732±0.003	0.608±0.007	0.785±0.003	0.603±0.002	0.656±0.001
	RAKEL	0.678±0.010	0.754±0.012	0.677±0.010	0.733±0.004	0.700±0.005	0.797±0.003	0.683±0.002	0.683±0.002
	ML-LOC	0.644±0.004	0.755±0.086	0.707±0.036	**0.734±0.002**	**0.804±0.045**	0.787±0.082	0.632±0.003	0.696±0.005
	GCC	0.706±0.004	0.735±0.031	0.752±0.000	**0.734±0.031**	0.701±0.061	0.792±0.061	**0.871±0.005**	**0.888±0.007**
Macro F1 ↑	GL*k*EL	**0.530±0.020**	**0.651±0.016**	**0.407±0.018**	0.193±0.007	0.550±0.007	**0.742±0.011**	0.237±0.007	0.309±0.006
	BR	0.117±0.009	0.526±0.124	0.117±0.009	0.177±0.003	0.442±0.011	0.667±0.005	0.236±0.006	0.305±0.000
	RAKEL	0.288±0.016	0.643±0.018	0.288±0.016	0.190±0.005	0.538±0.007	0.686±0.007	0.237±0.006	0.309±0.006
	ML-LOC	0.288±0.038	0.634±0.049	0.388±0.034	**0.505±0.017**	0.545±0.055	0.712±0.030	**0.298±0.005**	**0.366±0.008**
	GCC	0.016±0.010	0.621±0.031	0.250±0.000	0.137±0.013	**0.551±0.048**	0.691±0.048	0.236±0.004	0.247±0.010
Coverage ↓	GL*k*EL	**0.029±0.001**	**0.387±0.009**	**0.161±0.012**	0.570±0.008	**0.054±0.002**	**0.028±0.001**	0.429±0.005	0.457±0.003
	BR	0.276±0.014	0.435±0.012	0.276±0.161	0.573±0.006	0.282±0.013	0.170±0.005	0.442±0.004	0.473±0.005
	RAKEL	0.228±0.013	0.393±0.120	0.228±0.132	0.577±0.007	0.285±0.012	0.158±0.004	0.442±0.003	0.458±0.003
	ML-LOC	0.257±0.023	0.390±2.959	0.204±0.034	0.571±0.006	0.271±0.012	0.068±0.006	0.430±0.010	0.464±0.003
	GCC	0.185±0.012	0.394±2.959	0.173±0.026	**0.226±0.020**	0.106±0.013	0.088±0.009	**0.190±0.005**	**0.138±0.008**
Accuracy ↑	GL*k*EL	**0.540±0.025**	**0.553±0.014**	**0.222±0.011**	**0.426±0.006**	0.520±0.011	**0.725±0.013**	**0.303±0.005**	**0.329±0.005**
	BR	0.073±0.005	0.422±0.016	0.073±0.005	0.382±0.006	0.498±0.013	0.575±0.006	0.285±0.004	0.323±0.002
	RAKEL	0.171±0.008	0.540±0.020	0.171±0.008	0.387±0.006	0.512±0.009	0.632±0.007	0.291±0.003	0.322±0.002
	ML-LOC	0.204±0.045	0.533±0.043	0.188±0.049	0.135±0.015	0.519±0.047	0.647±0.005	0.294±0.003	0.320±0.004
	GCC	0.157±0.033	0.481±0.041	0.158±0.033	0.161±0.026	**0.524±0.054**	0.594±0.043	0.297±0.030	0.323±0.006

4.4 Experiment Results and Analysis

In this paper, various multimedia data sets shown in Table 1, such as images, audio, music and text, are adopted to evaluate the effectiveness of *GLkEL*. Performance on the multimedia data sets is shown in Table 2, which demonstrates the detailed experimental results about the comparison of the proposed approach with other four multi-label learning approaches on the 8 data sets. For each evaluation measure, "↓" denotes "the smaller, the better" while "↑" denotes "the larger, the better". In addition, the best performance among the five compared algorithms is highlighted in boldface.

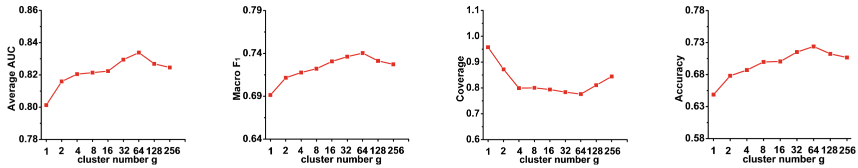
Across all the 32 configurations (i.e. 8 data sets × 4 criteria as shown in Table 2), *GLkEL* ranks in *1st* place among the five comparison algorithms at 65.6% cases, in *2nd* place at 31.3% cases, and only 3.1% cases rank after *2nd*.

By comparing with the most popular multi-label learning methods (*BR*, *RAkEL*, *ML-LOC*, and *GCC*), we can see that *BR* is the worst of all, since it does take label correlations into account. Compared with *RAkEL* which is a degenerated version of *GLkEL* without considering label correlations effectively, *GLkEL* is always superior to it for all the four evaluation measures under each dataset. We can see that *GLkEL* gets the satisfactory predictive results on most of the multimedia data sets, which signifies that the performance of classifier can be improved by exploiting label correlations; and the methods (*GLkEL*, *ML-LOC* and *GCC*) which consider label correlations locally get the better performance to some extent than the method (*RAkEL*) which exploits label correlation globally. It indicates that label correlations may be only shared in a local group, and exploiting the group-specific label correlations can perform better. It is worth to mention that *GLkEL* performs better than *ML-LOC* and *GCC* in most cases, as it models *high-order* label correlation and takes both global and local label correlations into account.

4.5 Sensitivity to Parameters

To evaluate the influence of parameters, we study the number of clusters g and the weight of local label correlation parameter λ_2 in our experiment. We vary one parameter, while maintaining the other fixed at its best set. Due to the page limit, we only report the results on the *scene* dataset in Figs. 1 and 2, whereas others get similar results.

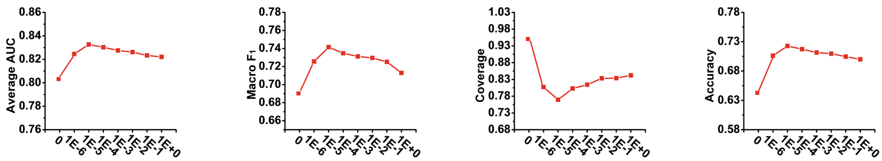
Influence of Parameter g (The Number of Clusters)
Figure 1 shows the influence on the *scene* dataset. When there is only one cluster, global label correlation is considered. With the number of clusters g becomes larger, the performance of *GLkEL* improves as more local label correlations are considered. When overmuch clusters are adopted, local label correlations, which cannot be estimated reliably by few instances contained in each group, may be less meaningful. Thus, the performance of *GLkEL* begins to decrease.



(a) Average AUC (b) Macro F1 (c) Coverage (d) Accuracy

Fig. 1. Varying the number of clusters g on the scene dataset

Influence of Label Correlation Parameter $\boldsymbol{\lambda_2}$
A larger λ_2 means higher weight of local label correlation. Figure 2 shows the influence on the *scene* dataset. When $\lambda_2 = 0$ which means that only global label correlation is considered, the performance is poor. With the increase of λ_2, the



(a) Average AUC (b) Macro F1 (c) Coverage (d) Accuracy

Fig. 2. Varying the local label correlation parameter λ_2 on the scene dataset

performance improves as the local label correlation joins and works gradually. However, when λ_2 is too large, performance deteriorates as the local label correlations dominate.

5 Conclusion

In multimedia classification tasks, both global and local label correlations may occur under real-world scenarios. Contrasting to the existing methods which either consider label correlation globally or locally, the paper proposes a novel and effective multi-label learning approach *GLkEL* for multimedia data categorization by exploiting both global and local label correlations simultaneously. Compared with the previous studies, it is effective to exploit both global and local label correlations in multimedia classification, which learns *high-order* label correlations by using a simple but effective *H*igh-*O*rder *L*abel *C*orrelation *A*ssessment strategy *HOLCA*. Experimental results show that *GLkEL* performs better than the state of art multi-label learning methods which do not consider label correlation and only exploit label correlation globally or locally. In our work, we handle the case that label correlation is positive without considering negative correlation. In many situations, the labels can be mutually exclusive with each other, e.g., label "sea" and "train", label "ship" and "car". So it is desirable to take negative label correlation into account in our future work.

Acknowledgments. This work was supported by R&D program of Shannxi Province Grant No. 2017ZDXM-GY-018, National Key Technologies R&D Program of China Grant No. 2014BAH14F01.

References

1. Boutell, M.R., Luo, J., Shen, X.P., Brown, C.M.: Learning multi-label scene classification. Pattern Recogn. **37**(9), 1757–1771 (2004)
2. Ueda, N., Saito, K.: Parametric mixture models for multi-label text. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, vol. 15, pp. 721–728. MIT Press, Cambridge (2003)
3. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.: Semantic annotation and retrieval of music and sound effects. TASLP **16**(2), 467–476 (2008)
4. Zhang, M.L., Zhou, Z.H.: A review on multi-label learning algorithms. IEEE Trans. Knowl. Data Eng. **26**(8), 1819–1837 (2014)
5. Tsoumakas, G., Katakis, I., Vlahavas, I.: Random k-labelsets for multi-label classification. IEEE Trans. Knowl. Discov. Data Eng. **23**(7), 1079–1089 (2010b)
6. Huang, S.J., Zhou, Z.H., Zhou, Z.H.: Multi-label learning by exploiting label correlations locally. In: AAAI (2012)
7. Huang, J., et al.: Group sensitive classifier chains for multi-label classification. In: 2015 IEEE International Conference on Multimedia and Expo, pp. 1–6 (2015)
8. Punera, K., Rajan, S., Ghosh, J.: Automatically learning document taxonomies for hierarchical classification. In: Proceeding WWW 2005, Special Interest Tracks and Posters of the 14th International Conference on World Wide Web, pp. 1010–1011 (2005)

9. Zhang, M.L., Zhang, K.: Multi-label learning by exploiting label dependency. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 999–1008 (2010)
10. Gibaja, E., Ventura, S.: Multi-label learning: a review of the state of the art and ongoing research. Wiley Interdisc. Rev. Data Min. Knowl. Discov. **4**(6), 411–444 (2014)
11. Fürnkranz, J., Hüllermeier, E., Loza Mencía, E., Brinker, K.: Multilabel classification via calibrated label ranking. Mach. Learn. **73**(2), 133–153 (2008)
12. Chung, F.R.K., Frankl, P., Graham, R.L., Shearer, J.B.: Some intersection theorems for ordered sets and graphs. J. Comb. Theory Ser. A **43**, 23–37 (1986)
13. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) **2**, 27:1–27:27 (2011)