College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach

Han-Jia Ye*, De-Chuan Zhan*‡, Xiaolin Li†‡, Zhen-Chuan Huang* and Yuan Jiang*
*National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210023, China
yehj@lamda.nju.edu.cn, zhandc@lamda.nju.edu.cn, corresponding author‡, {huangzc, jiangy}@lamda.nju.edu.cn
†School of Business, Nanjing University, Nanjing, 210023, China, lixl@nju.edu.cn, corresponding author‡

***Abstract*—Scholarships and financial aids in modern universities are the basic administrative plans to ensure and promote the completion of academic training and studies for students. Traditional grants allocation procedures are based on manual determination, which costs lots of human resources. In this paper, we investigate an assistance model for helping improve the scheme of granting. We first collect students information from multi-modal channels, including their behaviors of campus consumption, internet usage, daily trajectory together with their enrollment information. The approval status and amount of funds granted are converted as labels. We propose the College Student Scholarships and Subsidies Granting (CS^{3}G) approach to address the concrete problem. CS^{3}G approach overcomes 3 obstacles, i.e., complicated multi-label influences, private modal information protection and difficulties in label collection. In detail, based on the facts that scholarships mainly depend on academic achievements, subsidies granting is generally based on students financial hardships as well as credits, and there are implicit influences among scholarships and subsidies, the CS^{3}G approach handles types of interactions between multiple labels; it is notable that data from different modalities are collected by different divisions of a university, privacy protection is considered in CS^{3}G, i.e., no interaction between features from different modalities in the model training phase. Besides, due to the confidentiality of the concrete types/amounts of granting, only a portion of labels is collected in this application, CS^{3}G is trained in a semi-supervised style. Empirical investigations show good generalization ability of CS^{3}G on benchmark datasets, and a real assessment of a university also validates the power of our approach for tackling this type of problem well.**

***Keywords*-Student Scholarships and Subsidies Granting; Multi-Modal/Multi-Label Learning; Privacy-Preserving;**

I. Introduction

In modern universities, scholarships and financial aids ensure and promote the completion of academic training and studies for students. Students in colleges get scholarships or subsidies each year as rewards of their excellent behavior in academic study or financial aids for their tuition and fees. Scholarships, as rewards, will encourage students to make progress further; while subsidies, as types of assistance, will support the academic studies of students and release them from financial pressure. These grants are usually government funded or come from commercial companies, and the allocation of these financial funds is an important task to ensure its effectiveness and fairness.

Traditional grants allocation lies in two stages. First, students should apply for these grants and self-check whether they meet the minimum requirements. Then, college administrative authorities will rank the applications based on some grades of students: for scholarships, the grades are mainly decided by their daily performances, while the subsidy scores depend on more critical authentication of students' backgrounds. Both these two phases, i.e., application and score verification, have obvious drawbacks. First, due to reasons from amour-propre, those students, especially who are with financial hardships, are likely to avoid applying for subsidies. Similarly, due to the loss of self-confidence, some students may miss submitting their applications for scholarships. Thus, with voluntary application process, students who actually need financial aids or meet the requirements of academic performance may lose the opportunities of receiving awards. From the perspective of college administrative authorities, it is also hard to verify daily performance of a particular student. Consequently, the determination of grants might be subjective, which violates the original intention of fairness and effectiveness. Last but not least, both these two stages are resources consuming. Therefore, it is desired for universities and colleges to design a semi-automatic mechanism for funds granting.

As a matter of fact, there are some "easy" rules for facilitating the allocation, e.g., the less money of a student's expense for meals in a month, the more likely he should be granted with subsidies. However, impacts of daily expenses come from multiple factors and utilizing only one aspect can hardly yield accurate judgements. In this paper, legitimate multi-modal data about students behaviors from campus consumption, internet usage, daily trajectory together with their enrollment information, which are considered strongly related to student's performance and act as indicators of financial hardship, are collected for further analysis.

These pieces of information can be categorized into 2 aspects, namely personal information and equipment-collected information. Personal information is about students and their families, which is collected by questionnaires and entrance application forms. It is with high credibility and confidentiality; from another perspective, this type of information is half-baked, i.e., data from this modality is with missing values. As digital services in colleges developed, electronic campus cards with forgery-proof can provide a large amount of behavior description data with pre-authorizations. For example, campus card can provide the identification and time information of entering the studio, dormitory and library; can record the consumptions and utilities usage. These pieces of information from diverse channels serve as a reflection of students daily lives, and provide the possibility for funds allocation recommendations.

Although students information is collected with pre-authorizations, the private information should be protected in confidential, which leads to two difficulties for the semi-automatic funds granting problem. In detail, information collected by students campus card is partially gathered by different authorities of the university, e.g., Network Information Center, Office of Academic Affairs and Logistics Support Department. More important, there is no legal term for sharing the information among different authorities, which suggests there should be no raw information transfer among these departments during the multi-modal learning process. The 2nd issue lies in the fact that only a small portion of students is willing to share their grants information. This results in our problem a semi-supervised one, where effectively utilizing different channels of information from students with unknown grants amount is important in the mechanism design process. Besides, there will be more than one fund for a recipient at a time, which leads to our problem a multi-label one. Since there are complicated relationships among grants such as category correlations and restrictions, considering label relationship is necessary.

In this paper, we formalize the College Student Scholarships and Subsidies Granting problem as a privacy-preserving multi-modal multi-label learning approach (CS^{3}G), which can effectively make recommendations on grants portfolio. In our formulation, private data from different authorities are regarded as different modalities. To preserve the data privacy, raw data are not allowed to transmit in the training process among different modalities; to roughly control the total granting, the amount of funding is quantified into intervals, and CS^{3}G deals with multiple labels. It is notable that different from existing multi-label methods, CS^{3}G considers both positive and negative label affections for co-existence or exclusive scholarships and subsidies. Moreover, CS^{3}G can treat each modality unequally and has the ability of extracting the most useful modal features for final recommendation as well. The main contributions are:

- Real world College Student Scholarships and Subsidies Granting (CS^{3}G) is investigated and assessed.
- CS^{3}G approach utilizes multi-modal information in a privacy-preserving style to deal with multi-label tasks.

Section II gives the related work followed by the feature extraction descriptions. Section IV is the detailed CS^{3}G approach. Last are experiments and then conclusion.

II. Related Work

In this section, we briefly present state-of-the-art methods in multi-modal and multi-label fields as well as their differences between CS^{3}G.

Multi-modal learning aims to utilize the consistency and disagreement among feature groups from different channels [25]. Existing multi-modal methods can be categorized into four groups [28], namely pre-fusion [6] [7]; subspace learning [21]; disagreement based methods [1] [22] [24] and late fusion [27]. It is notable that in these multi-modal learning approaches, pre-fusion and subspace style approaches have to interact with features in different modalities, which will violate the privacy between information channels. Late fusion methods only build model on outputs of each modality, but make no effects on enhancing the learning ability of classifiers. Nevertheless, in this work, CS^{3}G can overcome these disadvantages of existing multi-modal models by optimizing modality classifier directly and using only predictions to keep modality consistency.

Different from taking each label independently in a binary relevance strategy [15], multi-label learning makes predictions on instances by taking label correlations into consideration [32]. For instance, [10] uses label prototypes to take local label relations into account; [4] and [13] propose to use sparse matrix norms such that related label classifiers have the same set of useful features, while [5] discovers exclusive label relations in image datasets and uses sub-gradient methods to solve it. Considering only a single type of relationship between labels is not sufficient for the scholarships and subsidies granting problem, while CS^{3}G solves this dilemma by a mixed structural regularizer, which can be solved using a reweighted method.

III. Feature and Label Generation in CS^{3}G

To achieve better performance on College Students Scholarships and Subsidies Granting (CS^{3}G), pre-authorized daily life information of students is collected from 4 authorities in a certain university. In this section, we will present the feature generation procedure for this problem briefly.

A. Feature Extraction from Different Modalities

Four different modalities, namely campus consumption, network usage, routine tracking and enrollment information are used for characterizing the daily behavior of a voluntary college student. In the surveyed university, each student has a campus card equipped with RFID tags, and can be used as credentials of consumptions in canteens and supermarkets in the university. Consumption records of students basically serve as a reflection of their consumption level. In most cases, if a student has financial hardship, he will not select expensive meals or have other avoidable consumptions except three meals a day. What's more, Internet access for relaxing or academic searching, is also managed by NIC with campus cards. Network usage reflects students preferences and is also a daily consumption since most universities charge on time. Campus card equipped with RFID tags obviously can serve as an identification. Moreover, with the help of wireless hotspot signals and trilateration techniques, trajectory information including when a student leaving and coming back dormitory/library/working studios can be tracked. This type of information can be also helpful in CS^{3}G scenarios, e.g., hardworking students spend most of their time in the library, while those who have great interests on campus activities will stay a lot of time in college student center. For information from these three modalities, features are generated as in [8].

Students personal and family information will be recorded when they are enrolled and we have collected those information from voluntary students. This information is with high credibility and often useful for grants determination. We extracted 17 different types of information, such as gender, nationality, year of birth, department, family zip code and district. All features are treated as discrete ones and 1-of-K coding is used to do a further transformation. Missing attributes for some students are completed by a nearest neighbor strategy, i.e., neighbor candidates are selected by computing hamming distance first, then majority voting is used to approximate missing features.

B. Grants Quantification and Label Generation

Concrete amount of grants (both scholarships and subsidies) is quantified into 11 levels, namely [1, 160), [160,320), . . ., and more than 1600 in US dollars. These 11 levels of grants act as 11 different labels. We also introduce 2 dummy labels for indicating no scholarships and no subsidies granting, respectively. Due to the fact that students are rarely approved to get multiple grants from the same quantity level for avoiding cutthroat repetitions, we can describe the allocation results of grants by a label matrix, where value 1/-1 indicates the successful/unsuccessful applications.

C. Privacy-Preserving Feature Transformation

Features are mainly organized and transformed in a Label-Specific (LIFT) [30] manner, and different modalities construct features in the same way. In detail, information from a certain modality of one single student is treated as an instance. Instances are first partitioned based on their possession of a particular label, then each part is clustered. The transformed features for each instance are constructed by concatenating value of distances to all cluster centers based on all labels together, and different types of distances can be used to reflect diverse properties of the original features. In this type of feature transformation, label specific characteristics are considered; besides, another advantage of this transformation lies in the mono-directional property of transformation. Thus, it is hard to recover the original features, which facilitates data privacy preserving.

IV. Methodology

Given extracted features from different authorities and part of student grants allocation, a multi-modal multi-label approach can be built to make predictions for students scholarships and subsidies in a semi-supervised scenario. In this section, we first give an overview of CS^{3}G approach, then detailed mechanisms for modality and label relationship are described. Last are optimization and convergence analysis.

A. The CS^{3}G Approach

Features of totally N students provided and grouped by different authorities can be regarded as K different modalities $\{X_k \in \mathcal{R}^{N\times d_k}\}_{k=1}^K$ and d_k is the dimensionality of feature in modality k. $X = [X_1, X_2, \ldots, X_K] \in \mathcal{R}^{N\times(d_1+d_2+\cdots+d_K)}$ is a combination of all modalities. Given L levels of grants, w.l.o.g., we assume the first N_l students have grants allocation labels, i.e., $Y \in \{-1, 1\}^{N_l \times L}$. With non-negative weighting parameter λ_1 and λ_2, the College Students Scholarships and Subsidies Granting problem can be initially solved by:

$$\min_{f_1,\ldots,f_K} \sum_{k=1}^K \ell(Y, f_k(X_k)) + \lambda_1 \sum_{k=1}^K \Omega_1(f_k) + \lambda_2 \Omega_2(f_1, \ldots, f_K) \ . \tag{1}$$

In the learning process, K classifiers $\{f_k\}_{k=1}^K$ are learned on each modality respectively. $\ell(\cdot)$ is a convex loss function, which measures the performance of classifier using modal prediction value $f_k(X_k)$ and label matrix Y. The last two terms in Eq. 1 are regularizers on classifiers from the individual and the whole aspects, which can be used to inject prior knowledge such as label correlation or view consistency in the learning process. For efficiency and effectiveness [29], linear classifier and least square loss is used, i.e., $W_k \in \mathcal{R}^{d_k \times L}$ is used to rank labels in the k-th modality and each column of W_k corresponds to a particular label. Taking both labeled and unlabeled instances into consideration, CS^{3}G approach can be summarized as:

$$\min_{\mathcal{F}_K,\mathcal{W}_K} \sum_{k=1}^K \|X_k W_k + \mathbf{1}\mathbf{b}_k^\top - F_k\|_F^2 + \lambda_1 \sum_{k=1}^K \Omega_1(W_k) + \lambda_2 \Omega_2(\mathcal{F}_K),$$
$$s.t. \ \ F_k^{1,\ldots,N_l} = Y \ , -1 \le F_k \le 1, \ k = 1, \ldots, K \ . \tag{2}$$

In Eq. 2, $\{F_k \in \mathcal{R}^{N\times L}\}_{k=1}^K$ is the prediction values on each modality, and we have an equality constraints to force the first N_l predicted labels the same as the true label. On the one hand, this equality constraint assists learning with the given part of the label; on the other hand, in this semi-supervised procedure, the variable F_k will help predictions on both labeled and unlabeled part keep the same scale. Inequality constraint on F_k is used to avoid the trivial solution of Eq. 2 [4]. $\mathbf{b}_k$ is bias on the k-th modality and $\mathbf{1}$ is a vector the same size as $\mathbf{b}_k$ and all elements equal to 1. We use $\mathcal{F}_K$ to denote the prediction set $\{F_1, \ldots, F_K\}$, and $\mathcal{W}_K = \{(W_1, \mathbf{b}_1), \ldots, (W_K, \mathbf{b}_K)\}$ is the classifier set. After taking derivative w.r.t. $\mathbf{b}_k$, a closed form solution can be obtained as $\mathbf{b}_k = \frac{1}{N}(F_k^\top \mathbf{1} - W_k^\top X_k^\top \mathbf{1})$. Denoting $H = I - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$, the above objective can be simplified as:

$$\min_{\mathcal{F}_K,\mathcal{W}_K} \sum_{k=1}^K \|H X_k W_k - H F_k\|_F^2 + \lambda_1 \sum_{k=1}^K \Omega_1(W_k) + \lambda_2 \Omega_2(\mathcal{F}_K) \ . \tag{3}$$

Before detailed analysis, we first define some notations. M_i and M^j denotes the i-th row[1] and j-th column of matrix $M \in \mathcal{R}^{m \times n}$, respectively. $\|M\|_{2,1} = \sum_{i=1}^{m} \|M_i\|_2$ and $\|M\|_{1,2} = \| [\|M_1\|_1, \ldots, \|M_m\|_1] \|_2$. Threshold operator of a matrix M with rank r is based on its Singular Value Decomposition (SVD): $M = U \Sigma V^\top$, where $U \in \mathcal{R}^{m \times r}$, $V \in \mathcal{R}^{n \times r}$ are column orthogonal matrix, and $\Sigma \in \mathcal{R}^{r \times r}$ is a diagonal matrix whose elements are singular values $\sigma = [\sigma_1, \ldots, \sigma_r]$. We assume singular values are sorted in non-increasing order. Soft threshold operator truncates singular values which are lower than a threshold λ to zero: $\mathcal{S}_\lambda(M) = U \mathrm{diag}([\sigma - \lambda, 0]_+) V$. $\mathrm{diag}(\cdot)$ is an operator that transforms a vector to a diagonal matrix, and $[\cdot]_+$ is a threshold operator which only preserves the non-negative part of the input value. Nuclear norm $\|M\|_* = \sum_{i=1}^{r} \sigma_i$ is the sum of singular values of the input matrix M, usually it serves as a convex surrogate of matrix rank operator [2]. A generalized surrogate of rank operator is Truncated Nuclear Norm (TNN) [9], $\|M\|_z = \sum_{i=z+1}^{r} \sigma_i$, which is the sum of minimum $r - z$ singular values.

B. Mechanism for Multi-Modal and Multi-Label Problem

Formulation of the CS^{3}G approach in Eq. 2 has two types of regularizers, one for each classifier and one for concatenated predictions.

Classifier Structured Regularizer: The importance of each modality varies when predicting different labels. For example, the weight of student family information is more important in subsidies prediction than for scholarships. Diversity of weights among modalities can be depicted by a group sparse characteristic on classifiers [23]. For the k-th modality, we have regularizer $\Omega_{1,1}(W_k) = \|W_k^\top\|_{2,1} = \sum_{l=1}^{L} \|W_k^l\|_2$. Minimizing the $\ell_{2,1}$-norm makes the column of matrix W_k sparse, which corresponds to feature weights on diverse levels of grants. Thus, for each label, only important classifiers with discriminative information among modalities will be selected.

Label relationship is also important to make predictions well, especially in such a multi-label case [32]. Some labels often appear simultaneously. For example, if a student does not have high academic grades but takes part in a lot of campus activities, he may try to (and should) apply for some special scholarships, which often have similar amount of funding. In addition, a student granted with subsidies may be hardworking, and are likely to get high scholarships as well. One characteristic of related labels lies in the shared features of their classifiers [13]. For instance, students applying for special scholarships which focus on organization activities/services may have similar daily trajectories in campus. Label co-existence relation can be represented in a group set $\mathcal{G}^1$ such that $\mathcal{G}^1 = \{g = (l, m) \mid \text{label } l \text{ and } m \text{ are related}\}$. We use pairwise label relationship for illustration, but it can be generalized into high order relationship easily. Given $\mathcal{G}^1$, the regularizer used for correlated label feature discovering can be formed as $\Omega_{1,2} = \sum_{g \in \mathcal{G}^1} \|W I_{\mathcal{G}_g^1}\|_{2,1}$. $I_{\mathcal{G}_g^1} = \{0,1\}^{L \times L}$ is a diagonal matrix with only elements indicated by the group g is set to 1, i.e., if $g = (l, m)$, then the (l, l) and (m, m) elements of $I_{\mathcal{G}_g^1}$ is set to 1. Thus, $W I_{\mathcal{G}_g^1}$ selects the classifiers in group g and combine them together. The $\ell_{2,1}$-norm used makes all classifiers in the group have zero values along the same dimension (corresponds to some modal attributes), so they share some features with non-zero values in the classifier.

However, in most multi-label learning cases, label relationship is not confined to the co-occurrence type that "If one instance has label A it is also likely to possess label B". Some labels are exclusive [5], i.e., "an instance cannot own label A and label B at the same time". For instance, if a student successfully applies for the highest scholarship, it is unlikely for him to have other types of financial funds, because the total amount of funding to be granted may have an upper bound. From the feature perspective, irrelevant labels may share no common features [12], which can be formalized by a $\ell_{1,2}$-norm: $\Omega_{1,3} = \sum_{g \in \mathcal{G}^2} \|W I_{\mathcal{G}_g^2}\|_{1,2}^2$, given exclusive relationship set $\mathcal{G}^2$. For the selected classifier $W I_{\mathcal{G}_g^2}$, ℓ_1 and ℓ_2 norm are used successively to produce intra-group exclusive feature sparsity and a total measure.

Modality Consistency Regularizer: In a multi-modal scenario, prediction results from different modalities have disagreements on the unlabeled part of data due to their diversity. In the ideal case, an instance will have the same decision over modalities results from the consistency among channels. Traditional multi-modal methods use some diversity measure to make predictions among modalities as consistent as possible. For example, square loss is often used to measure pairwise prediction diversity [22], which performs poorly when scales of instances in different modalities vary a lot and has a heavy computational burden when the number of modalities increases. Inspired by [28], we use matrix rank operator to measure the consistency among modalities, which is scale invariant and can deal with multiple modalities easily. So $\Omega_2 = \mathrm{rank}([F_1, \ldots, F_K])$. When all modalities have the same predictions (or only a scale difference), the rank of the combination prediction matrix will be less than the number of labels (L). If some labels are highly related, the rank will be even smaller. Since directly optimizing the rank operator is NP-hard, we use Truncated Nuclear Norm (TNN) [9] $\Omega_2 = \|[F_1, \ldots, F_K]\|_z$ as a surrogate. It is notable that when $z = 0$, TNN can be transformed into nuclear norm which is a convex regularizer. The proposed CS^{3}G process can be summarized in Fig. 1, and our approach can be formulated as:

$$
\begin{aligned}
&\min_{\mathcal{F}_K, \mathcal{W}_K} \sum_{k=1}^{K} \|H X_k W_k - H F_k\|_F^2 + \lambda_2 \|[F_1, \ldots, F_K]\|_z + \\
&\lambda_1 \sum_{k=1}^{K} (\|W_k^\top\|_{2,1} + \sum_{g \in \mathcal{G}^1} \|W_k I_{\mathcal{G}_g^1}\|_{2,1} + \sum_{g \in \mathcal{G}^2} \|W_k I_{\mathcal{G}_g^2}\|_{1,2}^2) \\
&s.t.\ \ F_k^{1\ldots,N_l} = Y\ , -1 \leq F_k \leq 1,\ k = 1, \ldots, K\ . \qquad (4)
\end{aligned}
$$

[1] Subscript of a matrix can also be used to denote a certain modality, e.g., W_k is the classifier from the k-th modality. We will neglect the detailed description when they are distinguishable.

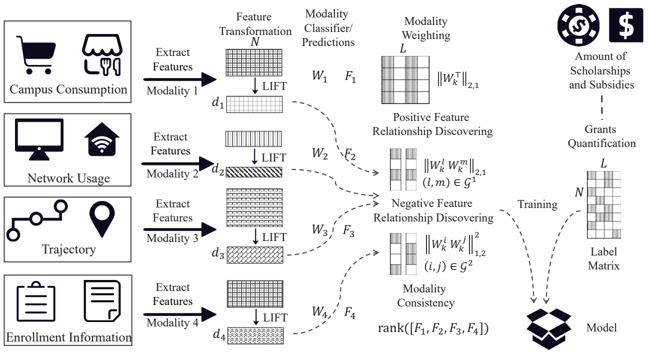

Figure 1: A brief illustration of CS^{3}G approach. Features are extracted from multiple different modalities and then transformed with LIFT. Scholarship and subsidy allocations are quantified into multiple levels as labels. Classifiers $\{W_k\}$ are learned on each modality with regularizers to weigh different modalities and discover the positive/negative affections between features in multi-label scenarios. Prediction results $\{F_k\}$ from different modalities are regularized by the "rank" operator to ensure modal consistency.

C. Optimization and Convergence Guarantees

The formulation of CS^{3}G approach in Eq. 4 contains two parts, i.e., prediction results $\mathcal{F}_K$ and classifier $\mathcal{W}_K$, which can be optimized in an alternative style.

Optimizing $\mathcal{W}_K$ when $\mathcal{F}_K$ is fixed: optimization on each modality can be conducted separately. Considering the k-th modality, the optimization problem over W_k contains smooth square loss function and non-smooth regularizers:

$$\min_{F_K, W_K} \|HX_kW_k - HF_k\|_F^2 + \lambda_1(\|W_k^\top\|_{2,1} + \sum_{g\in\mathcal{G}_1} \|W_k I_{\mathcal{G}_g}\|_{2,1} + \sum_{g\in\mathcal{G}_2} \|W_k I_{\mathcal{G}_g}\|_{1,2}^2) \ . \tag{5}$$

Due to multiple non-smooth regularizers in Eq. 5, proximal gradient cannot be directly used due to the complexity of proximal sub-problem. Hence we propose a re-weighted method to solve Eq. 5, which can deal with all three non-smooth terms simultaneously. By taking derivatives of $\ell_{2,1}$-norm regularizers in Eq. 5 w.r.t W_k^l $(l = 1, \dots, L)$, the classifier in modality k for label l, we can get

$$2D_1W_k^l + \sum_{g\in\mathcal{G}^1, l\in g} 2D_2W_k^l \ . \tag{6}$$

$D_1 \in \mathcal{R}^{d_k\times d_k}$ is a diagonal matrix whose diagonal elements are all equal to $\frac{1}{2\|W_k^l\|_2}$. The second part is a summation over a set of groups which contains label l and in the co-existed label group set $\mathcal{G}^1$, i.e., classifiers corresponding to labels that often co-exist with label l are considered together. $D_2 \in \mathcal{R}^{d_k\times d_k}$ is also a diagonal matrix with (d,d)-element equals to $\frac{1}{2\|(W_kI_{\mathcal{G}_g^1})_d\|_2}$. Re-weighted term D_1 only concerns the norm of classifier for label l in current modality, which is applied to select the modality information based on the discriminative power on labels; while D_2 takes group information into consideration, which is used to select features for classifiers in the same group.

To optimize the $\ell_{1,2}$-norm term, we first conduct a vectorization transform. We use $\mathbf{w}_k = vec(W_k) \in \mathcal{R}^{Ld_k}$ as the vector version of classifier W_k, which vertically combines classifier for all labels together. Group index in the exclusive group set $\mathcal{G}^2$ can be decomposed. For instance, if labels pair $(i,j) \in \mathcal{G}^2$, then the square of $\ell_{1,2}$-norm will be imposed on the selected part $[W_k^i\ W_k^j]$, where each row of the combined classifier stays in one group. So groups induced by (i,j) can be transformed to d_k groups on $\mathbf{w}_k$:

$$\{(\mathbf{w}_{k,(i-1)d_k+1}, \mathbf{w}_{k,(j-1)d_k+1}), \dots, (\mathbf{w}_{k,id_k}, \mathbf{w}_{k,jd_k})\} \ .$$

If we denote the transformed group on $\mathbf{w}_k$ as $\hat{\mathcal{G}}^2$, the $\ell_{1,2}$-norm in above equation can be transformed as:

$$\|W_kI_{\mathcal{G}_g^2}\|_{1,2}^2 = \sum_{g\in\hat{\mathcal{G}}^2} \|\mathbf{w}_{k,\hat{\mathcal{G}}_g^2}\|_1^2 = \mathbf{w}_k^\top D_3 \mathbf{w}_k \ . \tag{7}$$

$\mathbf{w}_{k,\hat{\mathcal{G}}_g^2}$ denotes the elements in $\mathbf{w}_k$ selected by group $g \in \hat{\mathcal{G}}^2$ and $D_3 \in \mathcal{R}^{d_kL\times d_kL}$ is a diagonal matrix with (d,d)-element equals to:

$$D_{3,dd} = \sum_{g\in\hat{\mathcal{G}}^2} \frac{(I_{\hat{\mathcal{G}}_g^2})_d \|\mathbf{w}_{k,\hat{\mathcal{G}}_g^2}\|_1}{|\mathbf{w}_{kd}|} \ .$$

$I_{\hat{\mathcal{G}}_g^2} \in \{0,1\}^{d_kL}$ is a vectorized group indicator. D_3 can be decomposed to L blocks with equal size and the l-th part is D_3^l. Combine Eq. 6-7 and take derivative w.r.t. W_k^l:

$$X_k^\top HXW_k^l - X_k^\top HF_k^l + D_1W_k^l + \sum_{g\in\mathcal{G}^1, l\in g} D_2W_k^l + D_3^lW_k^l = 0 \ .$$

Hence we can get a closed form solution of W_k^l:

$$W_k^l = [X_k^\top HX_k + D_1 + \sum_{g\in\mathcal{G}_1, l\in g} D_2 + D_3^l]^{-1} X_k^\top HF_k^l \ . \tag{8}$$

It is noteworthy that the update of W_k^l in Eq. 8 needs to compute an inverse. However, $X_k^\top HX_k$ can be pre-computed before training, and other terms are summation over *diagonal* matrices, so the inverse can be calculated easily using matrix Woodbury Identity. The update for W_k can be computed in parallel for each label l. Since D_1, D_2 and D_3 all depend on W_k, the reweighted process works in an alternative style with all updates in closed form. It can be proved that the update of reweighted form truly optimizes the non-smooth $\ell_{2,1}$-norm and $\ell_{1,2}$-norm terms.

Theorem 1: The alternative procedure to optimize W_k decreases the objective value of Eq. 5 and will be converged.

Proof: Update method on each modality is same, so we only consider the process over one modality k. We neglect the modality subscript and tuning parameter λ_1 for simplicity. In the t-th iteration, given current solution W^t, the loss function in Eq. 5 for current modality is:

$$J_1^t = l^t + \|W^{t\top}\|_{2,1} + \sum_{g\in\mathcal{G}^1} \|W^tI_{\mathcal{G}_g^1}\|_{2,1} + \sum_{g\in\mathcal{G}^2} \|W^tI_{\mathcal{G}_g^2}\|_{1,2}^2 \ ,$$

where $l^t = \|HXW^t - HF\|_F^2$ is the square loss function. The reweighted objective with matrices D_1, D_2 and D_3 in the t-th iteration can be denoted as J_2^t:

$$J_2^t = l^t + W^tD_1^tW^{t\top} + \sum_{g\in\mathcal{G}^1} (W^tI_{\mathcal{G}_g^1})^\top D_2^t(W^tI_{\mathcal{G}_g^1}) + \mathbf{w}^{t\top} D_3^t\mathbf{w}^t \ .$$

By update strategy, the value of smooth objective J_2 decreases once an update, thus $J_2^{t+1} \leq J_2^t$. So the goal equals to use the variation of J_2 to bound the difference of J_1.

$$
\begin{aligned}
& J_1^{t+1} - J_1^t - (J_2^{t+1} - J_2^t) \\
=& \|{W^{t+1}}^\top\|_{2,1} + \sum_{g \in \mathcal{G}^1} \|W^{t+1} I_{\mathcal{G}_g^1}\|_{2,1} + \sum_{g \in \mathcal{G}^2} \|W^{t+1} I_{\mathcal{G}_g^2}\|_{1,2}^2 \\
& - \|{W^t}^\top\|_{2,1} - \sum_{g \in \mathcal{G}^1} \|W^t I_{\mathcal{G}_g^1}\|_{2,1} - W^t D_1^t {W^{t+1}}^\top \\
& - \sum_{g \in \mathcal{G}^1} (W^{t+1} I_{\mathcal{G}_g^1})^\top D_2^t (W^{t+1} I_{\mathcal{G}_g^1}) + {\mathbf{w}^{t+1}}^\top D_3^t \mathbf{w}^{t+1} \\
& + W^t D_1^t {W^t}^\top + \sum_{g \in \mathcal{G}^1} (W^t I_{\mathcal{G}_g^1})^\top D_2^t (W^t I_{\mathcal{G}_g^1}) \\
\leq & \sum_{g \in \mathcal{G}^2} \|W^{t+1} I_{\mathcal{G}_g^2}\|_{1,2}^2 - {\mathbf{w}^{t+1}}^\top D_3^t \mathbf{w}^{t+1} \leq 0 \, .
\end{aligned}
$$

The first equality comes from the property of the reweighted matrix that ${\mathbf{w}^t}^\top D_3^t \mathbf{w}^t = \sum_{g \in \mathcal{G}^2} \|W^t I_{\mathcal{G}_g^2}\|_{1,2}^2$. The second inequality is the result of Lemma 1 that:

$$
\begin{aligned}
J_3 =& \|{W^{t+1}}^\top\|_{2,1} + \sum_{g \in \mathcal{G}^1} \|W^{t+1} I_{\mathcal{G}_g^1}\|_{2,1} - \|{W^t}^\top\|_{2,1} \\
& - \sum_{g \in \mathcal{G}^1} (W^{t+1} I_{\mathcal{G}_g^1})^\top D_2^t (W^{t+1} I_{\mathcal{G}_g^1}) + \sum_{g \in \mathcal{G}^1} (W^t I_{\mathcal{G}_g^1})^\top D_2^t (W^t I_{\mathcal{G}_g^1}) \\
& - W^{t+1} D_1^t {W^{t+1}}^\top + W^t D_1^t {W^t}^\top - \sum_{g \in \mathcal{G}^1} \|W^t I_{\mathcal{G}_g^1}\|_{2,1} \leq 0 \, ,
\end{aligned}
$$

and the third inequality is the result of Lemma 2. Therefore, $J_1^{t+1} - J_1^t \leq (J_2^{t+1} - J_2^t) \leq 0$, i.e., $J_1^{t+1} \leq J_1^t$, which indicates the reweighted update will converge at last. □

Lemma 1: $J_3 \leq 0$.

Proof: From [17], for non-zero vectors $\mathbf{w}$ and $\mathbf{w}^t$, the ℓ_2-norm has the following property:

$$
\|\mathbf{w}\|_2 - \frac{\|\mathbf{w}\|_2^2}{2\|\mathbf{w}^t\|_2} \leq \|\mathbf{w}^t\|_2 - \frac{\|\mathbf{w}^t\|_2^2}{2\|\mathbf{w}^t\|_2} \, .
$$

From the definition of $\ell_{2,1}$-norm, we have:

$$
\begin{aligned}
J_3 =& \sum_l \|W_l^{t+1}\|_2 - \|W_l^t\|_2 - \frac{\|W_l^{t+1}\|_2^2}{2\|W_l^t\|_2} + \frac{\|W_l^t\|_2^2}{2\|W_l^t\|_2} \\
& + \sum_{g \in \mathcal{G}^1} \sum_l \|Q_{g,l}^{t+1}\|_2 - \|Q_{g,l}^t\|_2 - \frac{\|Q_{g,l}^{t+1}\|_2^2}{2\|Q_{g,l}^t\|_2} + \frac{\|Q_{g,l}^t\|_2^2}{2\|Q_{g,l}^t\|_2} \leq 0 \, .
\end{aligned}
$$

We use $Q_{g,l}^t$ to denote the l-th column of selected group classifier $W^t I_{\mathcal{G}_g^1}$ in t-th iteration. □

Lemma 2: $\sum_{g \in \mathcal{G}^2} \|W^{t+1} I_{\mathcal{G}_g^2}\|_{1,2}^2 - {\mathbf{w}^{t+1}}^\top D_3^t \mathbf{w}^{t+1} \leq 0$.

Proof: Using the definition of $\ell_{1,2}$-norm, we can get

$$
\begin{aligned}
& \sum_{g \in \hat{\mathcal{G}}^2} \|\mathbf{w}_{\hat{\mathcal{G}}_g^2}^{t+1}\|_1^2 - {\mathbf{w}^{t+1}}^\top D_3^t \mathbf{w}^{t+1} \\
=& \sum_{g \in \hat{\mathcal{G}}^2} \|\mathbf{w}_{\hat{\mathcal{G}}_g^2}^{t+1}\|_1^2 - \sum_d \frac{(I_{\hat{\mathcal{G}}_g^2})_d \|\mathbf{w}_{\mathcal{G}_g^2}^t\|_1}{|\mathbf{w}_d^t|} (\mathbf{w}_d^{t+1})^2 \\
=& \sum_{g \in \hat{\mathcal{G}}^2} ((\sum_{d \in \hat{\mathcal{G}}_g^2} |\mathbf{w}_d^{t+1}|)^2 - (\sum_{d \in \hat{\mathcal{G}}_g^2} |\mathbf{w}_d^t|)(\sum_{d \in \hat{\mathcal{G}}_g^2} \frac{|\mathbf{w}_d^{t+1}|^2}{|\mathbf{w}_d^t|})) \leq 0 \, .
\end{aligned}
$$

Last inequality can be proved by Cauchy inequality [12]. □

Optimizing $\mathcal{F}_K$ when $\mathcal{W}_K$ is fixed: Given updated $\mathcal{W}_K$, optimization on $\mathcal{F}_K$ becomes:

$$
\begin{aligned}
& \min_{\mathcal{F}_K} \sum_{k=1}^K \|F_k - T_k\|_F^2 + \lambda_2 \|F\|_z \\
& s.t. \;\; F_k^{1\ldots,N_l} = Y \;, -1 \leq F_k \leq 1, \; k = 1, \ldots, K \, . \qquad (9)
\end{aligned}
$$

$T_k = X_k W_k + \mathbf{1}\mathbf{b}_k^\top$ and $F = [F_1, \ldots, F_K] \in \mathcal{R}^{N \times LK}$, which is the combination of the prediction values. Considering Eq. 9 is a combination of a smooth square loss and a non-smooth norm, it can be solved by proximal algorithms [18]. The TNN regularizer degenerates to the nuclear norm if $z = 0$. Given $T = [T_1, \ldots, T_K] \in \mathcal{R}^{N \times LK}$, the objective of Eq. 9 is:

$$
\min_F \|F - T\|_F^2 + \lambda_2 \|F\|_* \; ,
$$

which has a closed form solution $F = \mathcal{S}_{\lambda_2/2}(T)$. After each update of F, it should be projected to the feasible domain, i.e., making the first N_l prediction values equal to the training label Y and other predictions in the range $[-1, 1]$. The objective function in Eq. 4 can be proved to be convex [4] given nuclear norm regularizer, hence by alternatively updating $\mathcal{F}_K$ and $\mathcal{W}_K$, the objective in Eq. 4 will decrease and finally find a global optimal solution.

When $z > 0$, general TNN is used. By the following lemma from [9], sub-problem in Eq. 9 can be solved in an alternative manner as well:

Lemma 3: Given a matrix $F \in \mathcal{R}^{N \times KL}$ and any non-negative integer z ($z \leq \min(N, KL)$), for any column orthogonal matrix $A \in \mathcal{R}^{r \times N}$ and $B \in \mathcal{R}^{r \times KL}$, truncated nuclear norm can be reformulated as:

$$
\|F\|_z = \|F\|_* - \max_{A,B} \mathrm{Tr}(AFB^\top) \, .
$$

It can be solved by optimizing on F and (A, B) alternatively. When (A, B) is fixed, F can be updated as:

$$
F = \mathcal{S}_{\lambda_2/2}(T + \frac{\lambda_2}{2} B^\top A) \, . \qquad (10)
$$

When F is fixed, if its SVD is $F = U \Sigma V^\top$ and singular values in Σ are sorted in non-increasing order, A and B can be obtained by assigning the first z column(s) of U and V. The whole procedure of the CS^{3}G approach ($z > 0$) can be summarized in Alg. 1. The objective value decreases in both two sub-problems over $\mathcal{W}_K$ and $\mathcal{F}_K$, so the algorithm converges at last.

Remark: It is notable that in the two-stage updates in Alg. 1, the update of W_k is independent for each modality, while only the renewal of F_k is based on modality prediction results. Thus, only $F_k \in \mathcal{R}^{N \times L}$ is shared among modalities. It is hardly to recover raw features X_k from F_k for other modalities. Moreover, the feature transformation [30] before the training process is also privacy-keeping. Therefore, features of each modality are protected from other modalities and cannot be accessed by learners on other modalities, so privacy is preserved in CS^{3}G.

Algorithm 1 The whole procedure of CS^{3}G Approach

```
Require: $\lambda_1$, $\lambda_2$, initialize $\{F_k, W_k\}_{k=1}^{K}$ with random matrix, project $F_k$ to the feasible domain in Eq. 4
while Outer stop criterion doesn't meet do
    Update $W_k$, $k = 1 \ldots, K$ in each modality separately
    while Reweighted stop criterion doesn't meet do
        Using $W_k$ to update $D_1$, $D_2$ and $D_3$
        Calculate $W_k$ as in Eq. 8
    end while
    Collecting $T_k$ and $F_k$ from modalities to form $T, F$
    while TNN update stop criterion doesn't meet do
        Get SVD of $F$ as $F = U\Sigma V^\top$
        Set $A$ and $B$ as the first $L$ columns of $U$ and $V$
        Update $F$ as in Eq. 10
    end while
end while
return $\mathcal{F}_K$ and $\mathcal{W}_k$
```

V. Experiments

In this section, we validate the effectiveness of our proposed CS^{3}G approach. It is notable that although CS^{3}G approach is designed for scholarships and subsidies allocation, it is also a general multi-modal multi-label learning approach. Therefore, in this section, we first compare CS^{3}G with both multi-label and multi-modal learning methods on real multi-modal multi-label datasets as benchmarks, and then present the assessment of scholarships and subsidies allocation in a university as well.

In our experiments, both multi-label and multi-modal methods are compared. For the multi-label learners, all modalities of a dataset are concatenated together as a single modal input. In detail, Binary Relevance (BR) [15] treats labels independently and uses linear SVM as base classifier; Label specIfic FeaTures (LIFT) [30] transforms features first then BR is invoked. Besides, Multi-Label Local Output Coding (MLLOC) [10], L2G21 [3], SubJ21 [16], students financial Hardship Discovery model (DisH) [8] and Multi-Label k Nearest Neighbor (MLKNN) [31] are supervised methods; Convex model for Semi-Supervised multi-label feature selection (CSFS) [4] and Label Correlation model with relaxed visual Graph Embedding (LGME) [26] can be trained with labeled together with unlabeled data.

For multi-modal methods, we treat each label independently, i.e., for each label, a method trains classifiers using different modalities. Five methods are compared, namely Alignment-Based Multiple Kernel Learning (ABMKL) [19] and Localized Multiple Kernel Learning (LMKL) [6], Generalized Multiview Analysis (GMA) [21] using LDA as base classifier, Robust Late Fusion (RLF) [27] using linear SVM as pre-training classifier and semi-supervised Rank Consistency multi-view learning method (RANC) [28].

Table I: Datasets description. #N, #L and #V denote the number of instances, labels and modalities in each dataset, respectively. #D shows the dimensionality of each modality.

Name	#N	#L	#V	#D
FCVID	4388	28	5	400,400,400,400,400
ML2000	2000	5	3	500,1040,576
MSRA	15000	50	7	256,225,64,144,75,128,7
MSRC	591	24	3	500,1040,576
Taobao	2079	30	4	500,48,81,24

It is remarkable that in the training process of all multi-label methods and some multi-modal methods, raw features from each modality are concentrated and combined, which violates the privacy among modalities. However, CS^{3}G learns classifiers for each modality without interactions among raw features. All methods are tested by 30 trials on computational servers with 24 cores (2.53GHz) and 48G RAM. In each trial, 70% of the data are split for training and the remains are used for test. While in the training set, only 30% of instances are labeled. All compared methods keep the default/recommended configurations as in literatures. Six criteria are used to evaluate performance, namely Coverage, Ranking Loss, Average Precision, Macro AUC, Example AUC and Micro AUC [32]. Coverage and Ranking Loss are *the smaller the better*, and the other four are *the larger the better*, respectively.

In the feature transformation stage of CS^{3}G approach, instances in one modality are clustered into multiple positive and negative centers by K-Means, where the number of clusters is set to 10% of the minimum number of instances from two classes. Both Euclidean distance and nonlinear RBF similarity are used to measure the difference between an instance and each cluster center, and the transformed features are obtained by concatenating these distance values. Label relationship matrices $\mathcal{G}_1$ and $\mathcal{G}_2$ are built on pairwise label similarities [13]. Pairs with highest/smallest similarity values are selected to indicate the co-existence/exclusive label relationship. Parameters of CS^{3}G are fixed as $\lambda_1 = 0.1$ and $\lambda_2 = 1000$. The large value for λ_2 takes full advantage of complementary information among modalities, and is expected to improve predictions; while small value for λ_1 makes CS^{3}G tend to pick up fine classifier structures. TNN is used in CS^{3}G and we set z the same as the number of labels. Given a test instance, CS^{3}G makes prediction by majority voting among modalities.

A. Benchmark Comparisons

CS^{3}G is compared with other methods on 5 benchmark datasets to demonstrate its ability as a general multi-modal multi-label method. Five benchmark datasets are collected or generated as follows. FCVID is the Fudan-Columbia Video Dataset [11], a subset of 4388 videos with most frequent category names are tested. Each video may come from more than one category and features can be extracted in

Table II: Comparison results (mean ± std.) of CS^{3}G with both multi-label and multi-modal methods on 5 benchmark datasets. 6 commonly used criteria are evaluated. The best performance for each criterion is bolded. ↑/↓ indicate *the larger/smaller the better* of a criterion.

Comparison	Coverage ↓					Comparison	Macro AUC ↑				
Algorithm	FCVID	ML2000	MSRA	MSRC	Taobao	Algorithm	FCVID	ML2000	MSRA	MSRC	Taobao
BR	1.710±.114	.715±.041	14.782±.255	7.576±.343	6.216±.386	BR	.944±.004	.902±.007	.683±.005	.849±.023	.708±.022
MLLOC	1.698±.110	.726±.041	18.839±.511	7.782±.410	6.819±.355	MLLOC	.946±.004	.904±.008	.668±.005	.848±.019	.718±.015
LIFT	1.941±.147	.828±.041	14.299±.230	8.368±.427	7.324±.271	LIFT	.932±.010	.876±.009	.655±.007	.815±.017	.553±.028
CSFS	1.741±.114	.786±.045	15.804±.246	8.368±.468	7.736±.486	CSFS	.936±.004	.887±.008	.677±.007	.833±.017	.641±.021
LGME	3.434±.146	1.290±.064	14.167±.235	9.037±.431	7.136±.419	LGME	.887±.005	.749±.015	.684±.006	.807±.020	.639±.024
L2G21	2.803±.123	.845±.045	14.638±.282	8.738±.524	10.150±.454	L2G21	.900±.005	.868±.009	**.707±.007**	.819±.018	.605±.018
SubJ21	5.153±.185	.851±.047	18.001±.285	8.390±.556	10.789±.482	SubJ21	.811±.008	.867±.009	.660±.007	.834±.018	.591±.021
DisH	2.057±.157	.851±.047	N\A	8.390±.556	10.789±.482	DisH	.927±.005	.867±.009	N\A	.834±.018	.591±.021
MLKNN	2.872±.163	1.058±.051	14.110±.196	8.793±.532	6.939±.266	MLKNN	.877±.006	.809±.014	.613±.007	.753±.024	.565±.020
LRF	1.649±.096	.806±.032	N\A	13.971±.439	9.550±.797	LRF	.959±.002	.886±.007	N\A	.846±.013	.641±.020
RANC	1.117±.075	.726±.034	21.577±.287	13.934±.385	5.846±.265	RANC	.966±.002	.906±.006	.720±.005	.841±.017	**.746±.016**
ABMKL	2.598±.174	.719±.042	31.581±1.038	8.719±.591	18.061±.807	ABMKL	.909±.007	.898±.008	.525±.010	.854±.017	.552±.019
LMKL	1.566±.116	.741±.039	23.926±.733	8.566±.601	12.991±1.104	LMKL	.949±.004	.892±.008	.598±.011	.846±.020	.598±.025
GMA	22.799±.333	3.465±.035	37.418±.376	21.961±.545	15.982±.454	GMA	.175±.011	.160±.009	.352±.005	.219±.024	.476±.018
CS^{3}G	**1.117±.103**	**.680±.035**	**13.144±.210**	**7.390±.377**	**5.405±.276**	CS^{3}G	**.970±.003**	**.916±.007**	.644±.007	**.865±.017**	.730±.020

Comparison	Ranking Loss ↓					Comparison	Example AUC ↑				
Algorithm	FCVID	ML2000	MSRA	MSRC	Taobao	Algorithm	FCVID	ML2000	MSRA	MSRC	Taobao
BR	.061±.004	.114±.009	.171±.003	.081±.006	.209±.013	BR	.939±.004	.886±.009	.829±.003	.919±.006	.791±.013
MLLOC	.061±.004	.116±.010	.221±.008	.085±.009	.229±.012	MLLOC	.939±.004	.884±.010	.779±.008	.915±.009	.771±.012
LIFT	.070±.005	.143±.009	.158±.004	.097±.008	.244±.010	LIFT	.930±.005	.857±.009	.842±.004	.903±.008	.756±.010
CSFS	.062±.004	.131±.010	.177±.003	.096±.009	.260±.016	CSFS	.938±.004	.869±.010	.823±.003	.904±.009	.740±.016
LGME	.104±.005	.235±.013	.156±.004	.107±.009	.245±.014	LGME	.875±.005	.744±.015	.845±.003	.887±.010	.760±.014
L2G21	.101±.005	.146±.011	.162±.003	.100±.009	.334±.018	L2G21	.899±.005	.854±.011	.838±.003	.900±.009	.657±.015
SubJ21	.187±.007	.147±.011	.206±.003	.097±.011	.365±.016	SubJ21	.813±.007	.853±.011	.794±.003	.903±.011	.635±.016
DisH	.074±.006	.147±.011	N\A	.097±.011	.365±.016	DisH	.926±.006	.853±.011	N\A	.903±.011	.635±.016
MLKNN	.104±.006	.200±.012	.155±.002	.109±.010	.233±.009	MLKNN	.896±.006	.800±.012	.845±.002	.892±.010	.767±.009
LRF	.059±.004	.137±.007	N\A	.248±.012	.321±.026	LRF	.941±.004	.863±.007	N\A	.752±.012	.679±.026
RANC	**.040±.003**	.117±.007	.278±.005	.240±.010	.195±.009	RANC	**.960±.003**	.883±.007	.723±.005	.760±.010	.805±.009
ABMKL	.094±.007	.115±.010	.415±.013	.103±.016	**.028±.012**	ABMKL	.906±.007	.885±.010	.585±.013	.897±.016	.386±.028
LMKL	.056±.004	.121±.009	.300±.011	.103±.016	.078±.028	LMKL	.944±.004	.879±.009	.700±.011	.897±.016	.558±.038
GMA	.842±.012	.834±.010	.662±.009	.774±.086	.544±.014	GMA	.158±.012	.166±.010	.338±.009	.226±.086	.456±.014
CS^{3}G	**.040±.004**	**.107±.008**	**.145±.003**	**.078±.009**	.181±.010	CS^{3}G	**.960±.004**	**.893±.008**	**.855±.003**	**.922±.009**	**.819±.010**

Comparison	Average Precision ↑					Comparison	Micro AUC ↑				
Algorithm	FCVID	ML2000	MSRA	MSRC	Taobao	Algorithm	FCVID	ML2000	MSRA	MSRC	Taobao
BR	.767±.010	.854±.011	.470±.004	**.833±.012**	.462±.021	BR	.942±.004	.904±.007	.835±.003	.919±.006	.782±.014
MLLOC	.768±.008	.855±.012	.438±.008	.827±.015	.465±.014	MLLOC	.944±.004	.905±.008	.784±.007	.915±.010	.770±.011
LIFT	.723±.012	.822±.010	.506±.005	.801±.012	.327±.024	LIFT	.933±.005	.881±.008	.850±.002	.900±.009	.758±.010
CSFS	.791±.009	.840±.011	.496±.004	.813±.014	.408±.017	CSFS	.938±.004	.888±.009	.831±.003	.903±.010	.733±.017
LGME	.653±.015	.720±.013	.511±.005	.786±.015	.416±.015	LGME	.884±.005	.751±.015	**.856±.003**	.886±.012	.763±.013
L2G21	.701±.010	.823±.011	**.513±.005**	.802±.015	.298±.056	L2G21	.900±.004	.869±.009	.847±.003	.899±.008	.655±.015
SubJ21	.494±.013	.822±.011	.479±.005	.813±.015	.292±.017	SubJ21	.807±.007	.868±.009	.800±.003	.900±.012	.634±.017
DisH	.764±.010	.822±.011	N\A	.813±.015	.292±.017	DisH	.928±.006	.868±.009	N\A	.900±.012	.634±.017
MLKNN	.650±.011	.761±.010	.490±.004	.775±.016	.342±.013	MLKNN	.906±.006	.828±.012	.853±.002	.890±.010	.771±.009
LRF	.722±.010	.827±.009	N\A	.566±.020	.267±.039	LRF	.946±.003	.888±.006	N\A	.750±.011	.676±.026
RANC	**.791±.010**	.850±.010	.455±.005	.589±.015	**.507±.014**	RANC	.962±.002	.906±.005	.727±.004	.759±.010	.808±.009
ABMKL	.658±.015	.854±.010	.307±.013	.789±.040	.112±.023	ABMKL	.908±.007	.901±.008	.586±.013	.894±.015	.390±.025
LMKL	.773±.009	.848±.009	.362±.017	.787±.043	.149±.025	LMKL	.952±.004	.893±.008	.696±.011	.894±.016	.559±.037
GMA	.048±.001	.290±.005	.071±.003	.190±.094	.101±.008	GMA	.180±.013	.155±.009	.341±.010	.237±.082	.459±.013
CS^{3}G	.790±.016	**.860±.011**	.485±.005	.828±.014	.463±.016	CS^{3}G	**.964±.003**	**.917±.007**	.849±.003	**.922±.008**	**.829±.009**

diverse ways. Five types of features, namely HOF, HOG, CNN, Trajectory and SIFT are extracted for each video, then PCA is conducted to reduce the dimension of each view to 400. Given different modalities of description for a video, the task is to predict its possible categories. ML2000 is an image dataset from [31], where 2000 images from 5 categories (desert, mountains, sea, sunset and trees) are collected. Each image may be affiliated to two or more classes. We extract BoW, FV and HOG features for each image, which constitut 3 modalities of data. MSRC is used for object class recognition [20]. Same types features with ML2000 are extracted. MSRA subset is a salient object recognition database which contains 15000 instances from 50 categories [14]. Seven groups of features are extracted for each images, including 256 RGB color histogram features, 225 dimension block-wise color moments, 64 HSV color histogram, 144 color correlogram, 75 distribution histogram, 128 wavelet features and 7 face features. Taobao dataset

Table III: Comparison results (mean ± std.) of CS^{3}G with both multi-label and multi-modal methods on college students scholarships and subsidies granting dataset. 6 commonly used criteria are evaluated. The best performance for each criterion is bolded. ↑/↓ indicate *the larger/smaller the better* of a criterion.

Compar. Alg.	Coverage ↓	Ranking Loss ↓	Average Precision ↑	Macro AUC ↑	Example AUC ↑	Micro AUC ↑
BR	6.861±2.666	.208±.095	.557±.120	.537±.021	.792±.095	.756±.084
MLLOC	3.171±.123	.063±.004	.830±.012	.496±.023	.937±.004	.921±.003
LIFT	2.782±.037	.053±.001	.852±.003	.511±.021	.947±.001	.929±.002
CSFS	2.851±.063	.054±.002	.857±.004	.630±.032	.946±.002	.938±.004
LGME	11.241±.086	.492±.009	.337±.007	.499±.021	.644±.003	.625±.003
L2G21	3.752±.113	.154±.004	.800±.003	.679±.004	.919±.004	.919±.003
SubJ21	4.430±.376	.118±.019	.828±.012	.789±.004	.882±.019	.882±.016
DisH	2.944±.100	.062±.004	.845±.009	.516±.017	.938±.007	.921±.006
MLKNN	2.823±.037	.054±.001	.849±.003	.377±.014	.946±.001	.928±.001
LRF	11.126±.110	.371±.004	.490±.010	.694±.016	.629±.004	.649±.005
RANC	8.301±.165	.233±.007	.686±.008	**.840±.011**	.767±.007	.787±.007
ABSMKL	3.906±.186	.078±.006	.850±.010	.657±.025	.922±.006	.918±.007
LMKL	3.612±.252	.068±.008	.860±.011	.682±.034	.932±.008	.928±.009
GMA	12.571±.699	.553±.041	.220±.037	.433±.025	.447±.041	.446±.042
CS^{3}G	**2.701±.381**	**.047±.013**	**.876±.022**	.732±.024	**.953±.013**	**.956±.013**

is used for shopping items classification, which has 2079 instances and 30 labels. Description images of items are crawled from a shopping website, and four types of features, i.e., BoW, Gabor, HOG, HSVHist, are extracted to construct 4 modalities of data. Corresponding tags/categories path of an item provides the label sets. Concrete information of each dataset can be found in Tab. I. Results of compared methods and CS^{3}G are listed in Tab. II, where notation "N/A" means a method cannot give a result in 48 hours. From the results, it is obvious that our CS^{3}G approach can achieve the best performance on most datasets, which clearly reveals that the CS^{3}G approach is a high-competitive multi-modal multi-label learning method. Moreover, compared with LIFT method, CS^{3}G approach can get better performance as well. This validates that CS^{3}G approach can make use of the information provided by unlabeled data as well as exploit the multi-modal/multi-label information.

B. Students Grants Study

In this subsection, CS^{3}G approach has been tested on the college student scholarship and subsidies granting problem. There are totally 13,570 students in collection, and 4 modalities from different authorities are used to help their grants allocation. Specifically, there are 114 behavior features of campus card usage, 53 features for internet usage, 83 extracted from daily trajectories together with 173 for enrollment information. In each trial, only about 4,100 students provide grants labels. Comparison results against both multi-label and multi-modal methods are listed in Tab. III. Similarly, 6 measurement criteria are used as in previous subsection, i.e., Coverage, Ranking Loss, Average Precision, Macro AUC, example AUC and Micro AUC.

It can be found that our CS^{3}G approach can get the best results over 5/6 criteria, which validates the effectiveness of our method solving the grants allocation problem. In addition, this approach has assisted successfully for Dept. Undergraduates Student Affairs of the university which takes participate in these assessments.

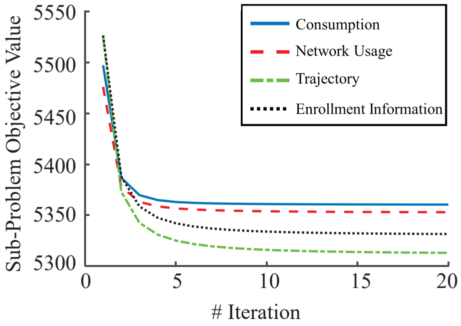

Figure 2: Empirical analysis of convergence when sub-problems in Eq. 5 solved with reweighted method.

To better analyze the property of CS^{3}G, we explore the changes of objective value when solving the reweighted sub-problem of Eq. 5. Due to the alternative optimization manner, changes of objective value in 1st outer iteration are recorded in Fig. 2. It can be found that objective values of 4 sub-problems decrease as the number of iterations increases, which validates Theorem 1. It is noteworthy that although three non-smooth regularizers are incorporated in Eq. 4, the sub-problem often converges in about 5 - 10 iterations, which suggests the efficiency of our CS^{3}G approach.

VI. Conclusion

Students scholarships and subsidies granting in universities is an important problem which consumes lots of human resources. Benefited from modern sensors integrated in campus cards and the collected enrollment information, we can build the CS^{3}G model and provide granting allocation recommendations based on multi-modal features from different authorities in the university. In CS^{3}G, we formulate the problem as a semi-supervised multi-model multi-label learning one. By utilizing both positive and negative affections between different types of grants, the model can impose different weights on different labels. Due to the privacy of students information, the proposed model is designed to be able to avoid the direct feature access in the learning process. In experiments, we first test the CS^{3}G model on real multi-modal multi-label benchmark data, and validate its ability of handling multi-modal and multi-label information. The final assessment on student grants allocation also points out that our model can achieve better recommendation for granting of scholarships and subsidies. More real assessments of universities possessing different cultural backgrounds should be a future work. In addition, how to predict and recommend with partial modality features for further preserving students privacy can also be a valuable direction.

Acknowledgment

The authors would like to thank reviewers for helpful comments. This work was supported by NSFC (61272217, 61273301) and JiangsuSF BK20131278.

References

[1] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *COLT*, Madison, WI., 1998, pp. 92–100.

[2] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Jour. Opt.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[3] X. Cai, F. Nie, W. Cai, and H. Huang, "New graph structured sparsity model for multi-label image annotations," in *ICCV*, Sydney, Australia, 2013, pp. 801–808.

[4] X. Chang, F. Nie, Y. Yang, and H. Huang, "A convex formulation for semi-supervised multi-label feature selection," in *AAAI*, Quebec, Canada, 2014, pp. 1171–1177.

[5] X. Chen, X.-T. Yuan, Q. Chen, S. Yan, and T.-S. Chua, "Multi-label visual classification with label exclusive context," in *ICCV*, Barcelona, Spain, 2011, pp. 834–841.

[6] M. Gönen and E. Alpaydin, "Localized multiple kernel learning," in *ICML*, Helsinki, Finland, 2008, pp. 352–359.

[7] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *JMLR*, vol. 12, pp. 2211–2268, 2011.

[8] C. Guan, X. Lu, X. Li, E. Chen, W. Zhou, and H. Xiong, "Discovery of college students in financial hardship," in *ICDM*, Atlantic, NJ., 2015, pp. 141–150.

[9] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE TPAMI*, vol. 35, no. 9, pp. 2117–2130, 2013.

[10] S.-J. Huang and Z.-H. Zhou, "Multi-label learning by exploiting label correlations locally," in *AAAI*, Toronto, Canada, 2012, pp. 949–955.

[11] Y.-G. Jiang, Z. Wu, J. Wang, X. Xue, and S.-F. Chang, "Exploiting feature and class relationships in video categorization with regularized deep neural networks," *arXiv:1502.07209*, 2015.

[12] D. Kong, R. Fujimaki, J. Liu, F. Nie, and C. Ding, "Exclusive feature learning on arbitrary structures via l1,2-norm," in *NIPS*. Cambridge, MA.: MIT Press, 2014, pp. 1655–1663.

[13] Y.-F. Li, J.-H. Hu, Y. Jiang, and Z.-H. Zhou, "Towards discovering what patterns trigger what labels," in *AAAI*, Toronto, Canada., 2012, pp. 1012–1018.

[14] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE TPAMI*, vol. 33, no. 2, pp. 353–367, 2011.

[15] O. Luaces, J. Díez, J. Barranquero, J. J. del Coz, and A. Bahamonde, "Binary relevance efficacy for multilabel classification," *Prog. AI*, vol. 1, no. 4, pp. 303–313, 2012.

[16] Z. Ma, F. Nie, Y. Yang, J. R. Uijlings, and N. Sebe, "Web image annotation via subspace-sparsity collaborated feature selection," *IEEE TMM*, vol. 14, no. 4, pp. 1021–1030, 2012.

[17] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint l2,1-norms minimization," in *NIPS*. Cambridge, MA.: MIT Press, 2010, pp. 1813–1821.

[18] N. Parikh and S. P. Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[19] S. Qiu and T. Lane, "A framework for multiple kernel support vector regression and its applications to sirna efficacy prediction," *ACM TCBB*, vol. 6, no. 2, pp. 190–199, 2009.

[20] F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting image databases from the web," *IEEE TPAMI*, vol. 33, no. 4, pp. 754–766, 2011.

[21] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *CVPR*, Providence, RI., 2012, pp. 2160–2167.

[22] V. Sindhwani, P. Niyogi, and M. Belkin, "A co-regularization approach to semi-supervised learning with multiple views," in *Proc. ICML Workshop on Learning Multi-Views*, Bonn, Germany, 2005, pp. 74–79.

[23] H. Wang, F. Nie, and H. Huang, "Multi-view clustering and feature learning via structured sparsity," in *ICML*, Atlanta, GA., 2013, pp. 352–360.

[24] W. Wang and Z.-H. Zhou, "A new analysis of co-training," in *ICML*, Haifa, Israel, 2010, pp. 1135–1142.

[25] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv:1304.5634*, 2013.

[26] Y. Yang, F. Wu, F. Nie, H. T. Shen, Y. Zhuang, and A. G. Hauptmann, "Web and personal image annotation by mining label correlation with relaxed visual graph embedding," *IEEE TIP*, vol. 21, no. 3, pp. 1339–1351, 2012.

[27] G. Ye, D. Liu, I.-H. Jhuo, and S.-F. Chang, "Robust late fusion with rank minimization," in *CVPR*, Providence, RI., 2012, pp. 3021–3028.

[28] H.-J. Ye, D.-C. Zhan, Y. Miao, Y. Jiang, and Z.-H. Zhou, "Rank consistency based multi-view learning: A privacy-preserving approach," in *CIKM*, Melbourne, Australia, 2015, pp. 991–1000.

[29] J. Ye and T. Xiong, "Svm versus least squares svm," in *AISTATS*, San Juan, Puerto Rico, 2007, pp. 644–651.

[30] M.-L. Zhang and L. Wu, "Lift: Multi-label learning with label-specific features," *IEEE TPAMI*, vol. 37, no. 1, pp. 107–120, 2015.

[31] M.-L. Zhang and Z.-H. Zhou, "Ml-knn: A lazy learning approach to multi-label learning," *Patt. Recog.*, vol. 40, no. 7, pp. 2038–2048, 2007.

[32] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE TKDE*, vol. 26, no. 8, pp. 1819–1837, 2014.