Multi-label Classification Using Hypergraph Orthonormalized Partial Least Squares

Gaofeng Luo
Hunan Provincial Key Laboratory of Information Service in Rural Area of Southwestern Hunan, Shaoyang University, Shaoyang 422000, China
E-mail: luogfeng@yeah.net

Tongcheng Huang and Zijuan Shi
Hunan Provincial Key Laboratory of Information Service in Rural Area of Southwestern Hunan, Shaoyang University, Shaoyang 422000, China
E-mail: luogfeng@yeah.net

Abstract — **In many real-world applications, human-generated data like images are often associated with several semantic topics simultaneously, called multi-label data, which poses a great challenge for classification in such scenarios. Since the topics are always not independent, it is very useful to respect the correlations among different topics for performing better classification on multi-label data. Hence, in this paper, we propose a novel method named Hypergraph Orthonormalized Partial Least Squares (HOPLS) for multi-label classification. It is fundamentally based on partial least squares with orthogonal constraints. Our approach takes into account the high-order relations among multiple labels through constructing a hypergraph, thus providing more discriminant information for training a promising multi-label classification model. Specifically, we consider such complex label relations via enforcing a regularization term on the objective function to control the model complexity and balance its contribution. Furthermore, we show that the optimal solution can be readily derived from solving a generalized eigenvalue problem. Experiments were carried out on several multi-label data sets to demonstrate the superiority of the proposed method.**

Index Terms—**Partial least squares; Orthogonal constraints; High-order relations; Regularization; Multi-label learning**

I. INTRODUCTION

Multi-label learning has gained increasing popularity from both the academia and the industry in recent years [1, 2, 3, 4]. It focuses on the data each of which is associated with more than one label. Such problems are omnipresent in many real-world applications, such as image annotation, video indexing and music style categorization [5, 6]. For example, an image might contain 'road' and 'house' at the same time since a road has a high probability to appear surrounding a house. For a video clip on Youtube, it might has several annotations, e.g., 'comedy', 'American' and 'humor'. In modern music information retrieval systems, the contents of the music refer to various styles, e.g., 'urban', 'country', 'jazz' and 'pop'. It can be easily observed from the above examples that there exist positive relations among the semantic topics, which can be explored to better address the multi-label problems.

In practice, multi-label data often reside in a high-dimensional space, where some noises or redundancy would affect the classification performance [23]. Therefore, it is of vital importance to learn a low-dimensional subspace which preserves the primary energy or components of the original data. In this work, we study extracting a latent subspace shared by all labels for multi-label classification. This is different from conventional methods, which construct the binary classifier for each label and neglect the label relations. In the low-dimensional data space, the performance of multi-label classification is expected to be significantly improved. Essentially, this problem can be called dimensionality reduction or feature extraction [26], which is a classical problem in machine learning and data mining. Thus, traditional unsupervised methods can be directly employed for multi-label problems, since none of the label information is required in this fashion, such as Principal Component Analysis (PCA) [7]. An obvious drawback is they fail to employ the label information, which plays a critical role in deriving a well-structured subspace. For supervised methods, we can achieve feature extraction on multi-label data from the perspective of maximizing the correlation or the covariance between the features and the labels, such as Canonical Correlation Analysis (CCA) [8] and Partial Least Squares (PLS) [9, 25] in this paper, we concentrate on a variant of PLS, named Orthonormalized Partial Least Squares (OPLS) [10], which imposes the orthogonal constraints onto the projected vectors. However, none of them have probed into the intrinsic relationships among the multiple labels, thus not capturing the decoupled effects from label relations.

To address this issue, we propose to encode the high-order label relations by a hyper-graph [11] to capture the correlated discriminant information for a better preserved structure in the projected data space. Hence, we present a novel method called Hyper-graph Orthonormalized

doi:10.4304/jcp.9.6.1364-1370

Partial Least Squares (HOPLS) for multi-label classification. In particular, each data point is treated as a vertex, and each label indicates a hyper-edge including all data points sharing a common label. With this method, the projection from the high-dimensional data space into a low-dimensional space is guided by the high-order relations among multiple labels, thus achieving improved multi-label classification performances. We formulate this model as a generalized eigenvalue problem, which can be readily solved. As a consequence, the extracted subspace is spanned by the eigenvectors corresponding to the leading eigenvalues from solving the eigenvalue decomposition problem. Since we take into account the complex label relations of multiple labels, as a result, the low-dimensional data space is able to characterize the discriminating power and approach the intrinsic data structure.

It is worthwhile to highlight the main contributions of this work as follows.

- A novel method is proposed for multi-label classification, i.e., Hypergraph Orthonormalized Partial Least Squares (HOPLS). This approach makes use of a hypergraph to encode the high-order label relations to guide the projection, so that the obtained low-dimensional data space characterizes a well preserved structure approaching the intrinsic one. Meanwhile, thanks to the supervised information, the derived data subspace has more discriminating power, leading to improved multi-label classification performances.
- We show this established model can be mathematically formulated as a generalized eigenvalue problem, which can be easily solved by eigenvalue decomposition technique. Thus, a set of eigenvectors corresponding to the leading eigenvalues span a subspace in the low-dimensional data space, where the multi-label classification are performed.
- To examine the performance of the proposed method, we conducted some interesting experiments on several real-world multi-label data collections. Results have demonstrated that our approach outperforms some competing alternatives.

The remainder of this work is structured as follows. Section II reviews some related works. We introduce the proposed Hypergraph guided Orthonormalized Partial Least Squares method in Section III. Experimental results are reported in Section IV with rigorous analysis. Finally, the concluding remarks are provided in Section 5.

II. Related Works

In this section, we give a brief review on recent works related to our method. Multi-label learning has gained increasing attention in the last decade, due to its widespread applications in many areas, e.g., image annotation, video retrieval and webpage categorization. A latest comprehensive review on multi-label learning algorithms can be referred to [12]. In these applications, the dimensionality of multi-label data is often very high, which is computationally expensive. Therefore, it is very meaningful to reduce the dimension of multi-label data prior to further processing them.

Heuristically, we can directly employ unsupervised methods in single-label learning, e.g., PCA[7]. Besides, we can use matrix factorization methods to obtain the low-dimensional representations, such as manifold kernel concept factorization [13] and discriminant orthogonal nonnegative matrix factorization [14]. But it is a common fact that the dimensionality reduction can be better performed while guided by supervised information, such as pairwise constraints or labels themselves. This poses a challenge for multi-label data since several labels might be associated with each data point. If we treat each label set as an individual, the number of label combinations is always too huge to handle and the label correlations are neglected as well. To this end, a number of methods have emerged to address this issue for regression and classification [8, 15, 16, 17, 18]. Among these methods, Partial Least Squares (PLS) [9] and Canonical Correlation Analysis (CCA) [8] are two representative ones, which are used for finding the relationships between two sets of variables.

Partial least squares maximizes the covariance along the maximum direction while canonical correlation analysis finds the directions of maximum correlation [10]. Specifically, PLS is shown to be useful when the number of observed variables is much larger than that of observations. Generally, PLS maximizes the covariance between different sets of variables to obtain orthogonal score vectors or components. Orthornormalized PLS (OPLS) is one of its variants to be studied in this work. Essentially, there exists a close connection between PLS and CCA in discrimination and the equivalence relation between OPLS and CCA has been proved [10]. Both of them can be naturally applied to multi-label data, in the sense that the label set with multiple dimensions caters to a set of multi-dimensional variables for CCA and PLS. They can be also performed in reproducing kernel Hilbert space by using kernel tricks [19]. Besides, some researchers attempt to extend the Linear Discriminant Analysis (LDA) to multi-label scenario by considering the label relations in the between-class and within-class scatter matrices [20]. In addition, some works try to maximize the dependence between the data features and the labels through using Hilbert-Schmidt Independence Criterion (HSIC) for multi-label dimensionality reduction [20]. Nevertheless, the label correlations still remain unclear thus requiring further explorations.

Hypergraph is employed to address this point in recent works for multi-label classification [21, 11]. In particular, a hypergraph is able to capture high-order relations among different categories in multi-label data. Each vertext represents an instance and each hyperedge includes all instances sharing the same label. Empirical studies have shown the effectiveness of hypergraph in revealing the intrinsic label relations, which inspires us to impose the hypergraph regularization onto orthonormalized partial least squares. Details are narrated in the following section.

III. Our Methods

In this section, we introduce the proposed Hypergraph Orthonormalized Partial Least Squares (HOPLS) algorithm for multi-label classification. We begin with the problem formulation.

A. Problem Formulation

Given a set of training multi-label data points $\{x_1,...,x_n\}$, each of which is stacked in the column of a data matrix $X \in \mathbb{R}^{m\times n}$. In other words, these data points reside on am-dimensional Euclidean space. Each data point might be associated with more than one labels and the maximum number of the labels is c. Thus, the corresponding label matrix can be denoted by $Y \in \mathbb{R}^{c\times n}$,which is an indicator matrix, where $Y_{ki}=1$ holds if x_i belongs to the $k-th$ class c_k, otherwise $Y_{ki}=-1$. In this work, both the data matrix and the label matrix are assumed to be centered, such that the cumulative column-wise sum in the data space is zero. Note that throughout this paper I is the identity matrix and $\mathrm{A}^{\dagger}$ denotes the pseudo-inverse of the matrix A.

Our goal is to learn a projection matrix $W \in \mathbb{R}^{m\times d}$ from the training data and the label matrix. By using this projection matrix, the original high-dimensional data points can be mapped onto a ddimensional data space, which suffices to $d \ll m$. In this way, the unseen data (e.g., out-of-sample problem) can be projected onto a much lower-dimensional data space via this transformation matrix W, such that the most prominent components are preserved for discrimination, thus benefitting multi-label classification.

B. Hypergraph

Hypergraph [6] is a generalization of traditional graph and yet it has some nice merits, namely the high-order relations among different objects can be captured for further analysis.

Mathematically, we define a hypergraph $G=(v,\varepsilon)$ where v is the vertex set containing the data points each of which acts as a vertex, and ε denotes the set of the hyperedges. Each hyperedge includes all samples sharing the same label, i.e., each label has an affiliated hyperedge e. Suppose that each hyperedge is assigned a weight $w(e)$ and the number of vertices in e is denoted by $\delta(e)$, i.e., the degree of a hyperedge. Thus, the degree in a conventional simple graph remain 2. The degree $d(v)$ of a vertex v is

$$d(v)=\sum_{v\in e, e\in\varepsilon} w(e) \quad (1)$$

The vertex-edge incidence matrix $J \in \mathbb{R}^{|V|\times|E|}$ is defined as

$$J(v,e)=\begin{cases}1, & if\ v\in e\\ 0 & othervise\end{cases} \quad (2)$$

As a result, we obtain

$$d(v)=\sum_{e\in\varepsilon} W(e)J(v,e) \quad (3)$$

$$\delta(e)=\sum_{e\in v} J(v,e) \quad (4)$$

The diagonal matrix forms for $d(v)$, $\delta(e)$ and $w(e)$ are respectively D_e, D_v and W_h. The Laplacian matrix from a traditional graph has been widely used to learn from graphs [11, 24]. It is clear that graph Laplacian is the discrete analog of the Laplace-Beltrami operator on compact Riemannian manifolds, which reflect the intrinsic structure of the data. In this work, we utilize a commonly used Clique Expansion algorithm to construct our hypergraph Laplacian.

Note that we add a subscript to the above notations for discrimination in clique expansion. Thus, the edge weight $w_c(u,v)$ of G_c is defined by

$$w_c(u,v)=\sum_{u,v\in e, e\in\varepsilon} w(e), \quad (5)$$

whose matrix form can be written as

$$W_c = HWH^T \quad (6)$$

We define

$$D_c(u,u)=\sum_{v} w_c(u,v), \quad (7)$$

then the combinatorial Laplacian L_c is shown as

$$L_c = D_c - W_c \quad (8)$$

Usually, we make use of its normalized version, i.e.,

$$L_n = D_c^{-1/2} L_c D_c^{-1/2} \quad (9)$$

In clique expansion, the similarity between two data points is positively proportional to the weights of their common labels, thus capturing the intrinsic relationship among different classes. This motivates us to enforce hypergraph as a regularizer to orthonormalized partial least squares in the following part.

C.Orthonormalized PLS with hypergraph regularization

Different from CCA that maximizes the correlation of two sets of variables in the transformed space, Orthonormalized PLS (OPLS) attempts to find the principal directions of maximum variance with orthogonality constraints. Formally, OPLS computes the orthogonal score vectors for X by solving

$$\max_{w} T_r(W^T XY^T YX^T W),$$
$$s.t. \quad W^T XX^T W = I, \quad (10)$$

Algorithm 1 Hypergraph Orthonormalized Partial Least Squares (HOPLS)
Input:

A collection of training data points $\left[x_{1,\cdots,}x_n\right] = X \in \mathbb{R}^{m\times n}$ the label matrix $Y \in \mathbb{R}^{c\times n}$ the constant parameters $a, \lambda > 0$

Output:

The projection matrix $W \in \mathbb{R}^{m\times d}$.

1: Initialize the weights $w(e)$ for the hyperedges.

2: Compute the normalized graph Laplacian matrix L_n for the hypergraph using Clique Expansion as shown in (9).

3: Compute the matrix $S = I - L_n$.

4: Optimize the objective function in (11) by solving the generalized eigenvalue problem in (13).

5: Construct the projection matrix W by the eigenvectors corresponding to the top d eigenvalues.

where W is a projection matrix for learning a low-dimensional data representation. Inspired by the success of hypergraph in multi-label learning [11], we impose it onto the objective function in (10) as a regularization term, thus additionally capturing the high-order relations among different labels. Now, we can readily formulate the proposed HOPLS algorithm as

$$\max_w \ T_r[W^T X(Y^T Y + \alpha S)X^T W],$$
$$s.t. \quad W^T XX^T W = I, \qquad (11)$$

Where $S = I - L_n$ captures the high-order label relationships and $a > 0$ is a tradeoff parameter for balancing the contribution of the hypergraph regularizer to the objective function. Thus, the proposed method is able to maximize the relation between the data points and the corresponding labels as well as to respect the high-order relations among different class labels. Note that the matrix $\hat{S} = Y^T Y + aS$ is symmetric and semi-definite positive. If we assume $\hat{S}$ is full-rank, then it is a well-defined matrix having the inverse.

Observing the objective function in (11), we find it is a generalized eigenvalue problem given by

$$X\left(Y^T Y + aS\right)X^T w = \eta XX^T w \qquad (12)$$

Where η is the eigenvalue variable and w is the corresponding eigenvector. Therefore, the optimal projection matrix W can be derived from solving this eigen-decomposition problem. The eigenvectors corresponding to the topdeigenvectors spans the row space of W.

Furthermore, it is commonly believed that regularization is a popular technique to penalize the complexity of a learning model and regularized CCA is shown to have natural statistical interpretations [10]. Hence, we can directly show the regularized HOPLS by adding a regularization term to XX^T, leading to the following formulation:

$$X\left(Y^T Y + aS\right)X^T w = \eta\left(XX^T + \lambda I\right)w, \qquad (13)$$

Where λ is positive constant to avoid overfitting and also control the model complexity.

In summary, the complete procedures of our approach is structured clearly in Algorithm 1.

Moreover, we provide the whole framework of our proposed HOPLS method using MLKNN as the classifier in Fig.2.

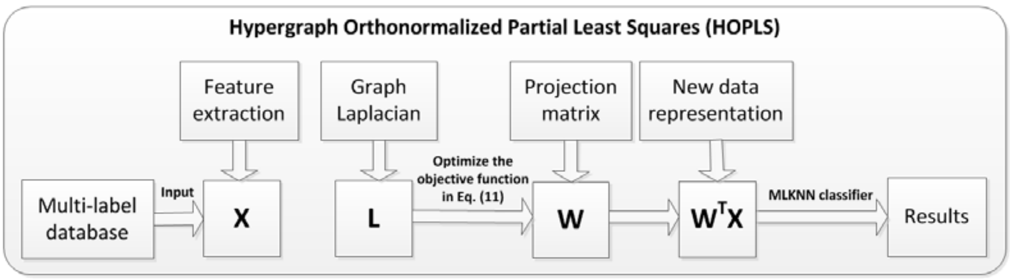

Fig.2 Framework of HOPLS using MLKNN as the classifier

IV. Experiments

We have conducted a number of experiments on several multi-label data sets to justify the proposed method as compared to the competitive alternatives. For all algorithms, we use MLKNN[4] as the base classifier to predict the possible label sets in a lower-dimensional data space. For each data set, they are randomly divided into two parts, leading to 70% as the training data and the rest as the test data. Parameter settings are as follows. We set the number of nearest neighbor for MLKNN to 10 as default. The number of projected components is set to the number of classes in the data set. The regularization parameters are searched from the grid $\{10^i \mid i = -3:3\}$ and the best parameters are derived from cross-validation on the training data. The hyperedge weights of HOPLS are set as in [6]. Note that we tuned the tradeoff parameters in smaller bins, i.e., $[0.1:0.1:1]$, for better performance. To eliminate the bias, we repeat the experiments ten times, and report the averaged values.

A. Data Sets

In this test, we collect four multi-label data sets to investigate the performance of our approach from various domains, including music (emotions), image (scene), gene expression (genbase) and text document (medical). They are available at http://mlkd.csd.auth.gr/multilabel.html.

The music data emotions contains 593 songs, each of which has 72 features, and can be affiliated with 6 different styles. The scene data consists of 2407 images and refers to six scenarios, where each image is represented as a 294 dimensional vector. The biology data genbase is composed of 662 instances that are associated with at most 27 labels. The medical data includes 978 documents with 1449 features, covering 45 different topics. We summarize the statistics of these data in Table 1.

TABLE I.
SUMMARY OF STATISTICS OF THE DATA SETS.

Data set	domain	instances	features	labels	cardinality	density	distinct
emotions	music	593	72	6	1.869	0.311	27
scene	image	2407	294	6	1.074	0.179	15
genbase	biology	662	1185	27	1.252	0.046	32
medical	text	978	1449	45	1.245	0.028	94

B. Performance Evaluation

We make use of macro F1-score, precision and recall[3] as the evaluation criteria to examine the performance of the proposed HOPLS method and all the compared algorithms in the following.

- Principal Component Analysis (PCA) [7].
- Locality Preserving Projection (LPP) [22].
- Canonical Correlation Analysis (CCA) [8].
- Multi-label informed Latent Semantic Indexing (MLSI) [16].
- Multi-label Dimensionality reduction via Dependence Maximization (MDDM) [20].
- Orthonormalized Partial Least Squares (OPLS) [10].

TABLE II.
THE F1-SCOREON DIFFERENT COMPARISON METHODS

Data set	PCA	LPP	CCA	MLSI	MDDM	OPLS	HOPLS
emotions	0.4282	0.4831	0.6158	0.4208	0.6421	0.6479	**0.6615**
scene	0.6277	0.6331	0.6313	0.6423	0.6358	0.6445	**0.6538**
genbase	0.9230	0.9287	0.9065	0.9520	0.9488	0.9509	**0.9569**
medical	0.6338	0.6317	0.6776	0.7178	0.6964	0.7205	**0.7352**

TABLE III.
THE PRECISIONON DIFFERENT COMPARISON METHODS

Data set	PCA	LPP	CCA	MLSI	MDDM	OPLS	HOPLS
emotions	0.6162	0.6045	0.6834	0.6211	0.6888	0.6971	**0.7127**
scene	0.7006	0.7077	0.7023	0.7073	0.6972	0.7087	**0.7290**
genbase	0.9896	0.9905	0.9888	0.9916	0.9958	0.9947	**0.9972**
medical	0.7624	0.7559	0.7128	0.7688	0.7426	0.7635	**0.7748**

TABLE IV
THE RECALLON DIFFERENT COMPARISON METHODS

Data set	PCA	LPP	CCA	MLSI	MDDM	OPLS	HOPLS
emotions	0.3305	0.4022	0.5613	0.3225	0.6020	0.6074	**0.6369**
scene	0.5678	0.5668	0.5740	0.5886	0.5847	0.5914	**0.6114**
genbase	0.9008	0.9103	0.8915	0.9109	0.9062	0.9109	**0.9147**
medical	0.5437	0.5479	0.6468	0.6738	0.6561	0.6826	**0.6984**

Among the above methods, PCA and LPP are unsupervised methods while the rest are all supervised methods. Multi-label classification were conducted in the lower-dimensional data space derived from these approaches.

C. Results and Analysis

Experimental results are reported in Table 2 to Table 4. The best performances on each data set are highlighted in boldface. To examine the classification performances with the varied size of training data, we depict the curves of the three evaluation metrics onemotionswith increasing ratios of training samples in Fig. 1.

A number of interesting points can be observed from these results.

- The proposed HOPLS algorithm systematically and consistently outperforms other competing methods. We attribute this to the fact that a hypergraph is utilized as a regularizer in HOPLS to encode the high-order relations among different labels, thus capturing the intrinsic label structure for more discriminating power of the model. This justifies our theoretical analysis.

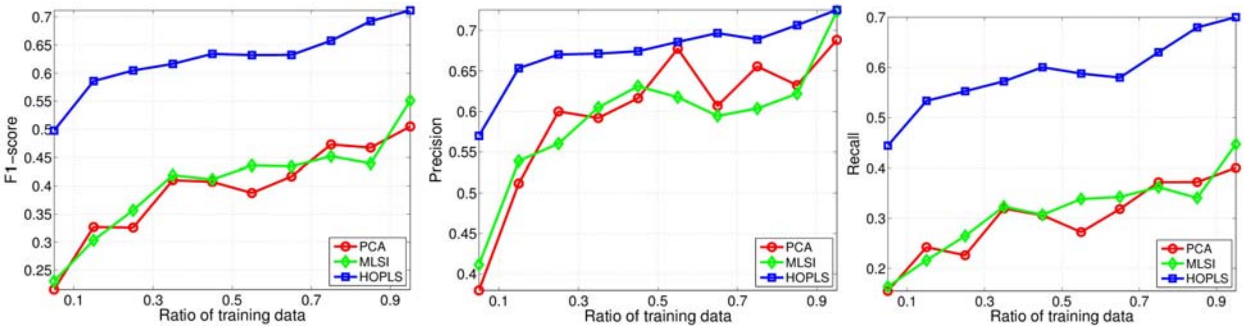

Fig.1: Multi-label classification performance with increasing ratios of training data onemotions. Results are averaged over five test runs.

- Mostly supervised methods perform better than unsupervised methods (i.e., PCA and LPP), which confirm that projection under the guidance of label information will lead to a more discriminant low-dimensional data space, thus improving the multi-label classification performance.
- With the increase of the training data points, the classification performance would be boosted gradually regardless of supervised or unsupervised methods. The reason for this is that more data points lead to a more robust and discriminating model for learning a betterstructured data space.

V. CONCLUSION

This paper presents a novel method for multi-label classification, i.e., Hypergraph Orthonormalized Partial Least Squares (HOPLS). Essentially, it is strongly motivated by the success of hypergraph to encode the high-order relations among different labels. In this work, we incorporate the intrinsic label information into the orthonormalized partial least squares, which has been shown to have satisfying performance on multi-label problems. To consider the complex label relations, we impose the hypergraph regularizer onto the objective function of OPLS, leading to a generalized eigenvalue problem. Empirical studies on some multi-label data sets have shown more promising performances by the proposed method in comparison with others.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful advices to improve this paper. This work was supported in part by a grant from the Construction Program of the Key Discipline in Hunan Province of China and the Key Scientific Research Program of Department of Education of Human Province (NO.13A091).

REFERENCES

[1] X. Feng, H. Zhang, H. Ji, S. Zhang, "TMLLC: Transductive multi-label learning based on local correlation", *Journal of Information & Computational Science*, 10(1), 2013, 297-311

[2] H. Li, Y. Guo, M. Wu, P. Li, Y. Xiang, "Combine multi-valued attribute decomposition with multi-label learning", *Expert Systems With Applications*, 37(12), 2010, 8721-8728

[3] G. Tsoumakas, I. Katakis, I. Vlahavas, "Mining multi-label data", Data mining and knowledge discovery handbook, Springer, 2010, 667-685

[4] M. Zhang, Z. Zhou, Ml-knn: "A lazy learning approach to multi-label learning", *Pattern Recognition*, 40(7), 2007, 2038-2048

[5] G. Qi, X. Hua, Y. Rui, J. Tang, T. Mei, H. Zhang, "Correlative multi-label video annotation", in *MM'07: Proc. of the 15th International Conference on Multimedia*, ACM, 2007, 17-26

[6] F. Wang, X. Wang, B. Shao, T. Li, M. Ogihara, "Tag integrated multi-label music style classification with hypergraph", in *ICMIR'09: Proc. of the 10th International Conference of Music Information Retrieval*, 2009, 363-368

[7] H. Abdi, L. Williams, "Principal component analysis", Wiley Interdisciplinary Reviews: *Computational Statistics*, 2(4), 2011, 433-459

[8] L. Sun, S. Ji, J. Ye, Canonical correlation analysis for multilabel classification: "A least-squares formulation, extensions and analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1), 2011, 194-200

[9] C. M. Bishop, N. M. Nasrabadi, "Pattern recognition and machine learning", *Springer*, New York, 2006

[10] L. Sun, S. Ji, S. Yu, J. Ye, "On the equivalence between canonical correlation analysis and orthonormalized partial least squares", in IJCAI'09: *Proc. of the 21st International Joint Conference on Artificial Intelligence*, 2009, 1230-1235

[11] L. Sun, S. Ji, J. Ye, "Hypergraph spectral learning for multi-label classification", in KDD'08: *Proc. Of the 14th*

International Conference on Knowledge Discovery and Data mining, ACM, 2008, 668-676

[12] M. Zhang, Z. Zhou, "A review on multi-label learning algorithms", *IEEE Transactions on Knowledge and Data Engineering, in press*, 2013, doi : http://dx.doi.org/10.1109/TKDE.2013.39

[13] P. Li, C. Chen, J. Bu, "Clustering analysis using manifold kernel concept factorization", *Neurocomputing*, 87, 2012, 120-131

[14] P. Li, J. Bu, Y. Yang, R. Ji, C. Chen, D. Cai, "Discriminative orthogonal nonnegative matrix factorization with flexibility for data representation", *Expert Systems With Applications*, in press, 2013, doi: http://dx.doi.org/10.1016/j.eswa.2013.08.026

[15] G. Tsoumakas, I. Katakis, "Multi-label classification: An overview". *International Journal of Data Warehousing and Mining*, 3(3), 2007, 1-13

[16] K. Yu, S. Yu, V. Tresp, "Multi-label informed latent semantic indexing", in SIGIR'05: *Proc. of the 28th ACM SIGIR Conference on Research and Development in Information Retrieval, ACM*, 2005, 258-265

[17] H. Li, P. Li, Y. Guo, M. Wu, "Multi-label dimensionality reduction based on semi-supervised discriminant analysis", *Journal of Central South University of Technology*, 17, 2010, 1310-1319

[18] P. Li, H. Li, M. Wu, "Multi-label ensemble based on variable pairwise constraint projection", *Information Sciences*, 222, 2013, 269-281

[19] J. Arenas-Garc´\u305Xa, G. Camps-Valls, "Efficient kernel orthonormalized pls for remote sensing applications", *IEEE Transactions on Geoscience and Remote Sensing*, 46(10), 2008, 2872-2881

[20] Y. Zhang, Z. Zhou, "Multilabel dimensionality reduction via dependence maximization", *ACM Transactions on Knowledge Discovery from Data*, 4(3), 2010, issue 14

[21] G. Chen, J. Zhang, F. Wang, C. Zhang, Y, Gao, "Efficient multi-label classification with hypergraph regularization", in CVPR'09: *Proc. of the International Conference on Computer Vision and Pattern Recognition, IEEE*, 2009, 1658-1665

[22] X. He, P. Niyogi, "Locality preserving projections", *Advances in Neural Information Processing Systems*, 16, 2004, 153-160

[23] Wenbin Zheng, Yuntao Qian, Minchao Ye. "A Grouped Structure-based Regularized Regression Model for Text Categorization". *Journal of Software*, Vol 7, No 9 (2012), 2119-2124.

[24] Wencheng Wang, Faliang Chang. "A Multi-focus Image Fusion Method Based on Laplacian Pyramid". *Journal of Computers*, Vol 6, No 12 (2011), 2559-2566.

[25] Zhengguang Xu, Jinxia Wu, Shoude Qu. "Prediction Model based on Moving Pattern". *Journal of Computers*, Vol 7, No 11 (2012), 2695-2701.

[26] Shi Qu, Ronghuan Yu, Yingmei Wei, Lingda Wu. "Gaussian Process Latent Variable Models for Inverse Kinematics". *Journal of Multimedia*, Vol 6, No 1 (2011), 48-55

Gaofeng Luo received the M.S. degree in Information and Communication Engineering from Central South University, China, in 2010. He is currently a lecturer in Shaoyang University, China. Also, he is the Office Director of the Key Laboratory of Southwest Hunan Rural Information Service, Hunan Province. His research interests include Internet of things and artificial intelligence.

Tongcheng Huang received the PhD degree in Information and Communication Engineering from Shanghai University, China, in 2008. Before that, he received the M.S. degree from National University of Defense Technology, China, in 2000. He is currently a Professor and Master supervisor in Shaoyang University, China. Also, he is the Leading Director of the Key Laboratory of Southwest Hunan Rural Information Service, Hunan Province. His research interests include Intelligent Information Processing and artificial intelligence.