Direct Multi-label Linear Discriminant Analysis

Maria Oikonomou and Anastasios Tefas

Aristotle University of Thessaloniki, Department of Informatics,
Box 451, 54124 Thessaloniki, Greece
tefas@aiia.csd.auth.gr

Abstract. Multi-label problems arise in different domains such as digital media analysis and description, text categorization, multi-topic web page categorization, image and video annotation etc. Such a situation arises when the data are associated with multiple labels simultaneously. Similar to single label problems, multi label problems also suffer from high dimensionality as multi label data often happens to have large number of features. In this paper, the Direct Multi-label Linear Discriminant Analysis method is proposed for dimensionality reduction of multilabel data. In particular we extend Multi-label Discriminant Analysis (MLDA) and modify the between-class scatter matrix in order to improve classification accuracy. The problem that Direct MLDA overcomes is the limitation of the produced projections that in MLDA are defined as $K-1$ for a K class problem. Experimental results on video based human activity recognition for digital media analysis and description as well as on other challenging problems indicate the superiority of the proposed method.

Keywords: Direct Multi-label Discriminant Analysis, Dimensionality Reduction, Multi-label classification, Activity recognition, Video analysis.

1 Introduction

In recent years, with the development of the internet and the technologies used for media production, the amount of available information has been increased dramatically. The main problem of the huge amount of information is how the user can interpret the content of the information and how he can retrieve successfully the information he is interested in. Moreover, big data problems arise in several steps of the media production chain. Pictures, videos, texts, music can contain information on various topics. The user can not be sure that the correct information will be retrieved. This is the reason why it is necessary to classify the information into categories. The most popular way of classification is label classification. The label is attached to the item, indicating the category that the item belongs to. In label classification there are two categories, single-label classification and multi-label classification. The first refers to problems when each instance is associated with one label and the second when each instance is associated with multiple different categories. In most cases, items belong to the second category. For example in music information retrieval, a song could

L. Iliadis, H. Papadopoulos, and C. Jayne (Eds.): EANN 2013, Part I, CCIS 383, pp. 414–423, 2013.

belong to categories such as *piano*, *classical music* and *Mozart*. The interest of researchers on multi-label learning has increased due to the large number of applications, multi label data are associated with. Multi-label learning algorithms are examined in [1], [7], multilabel semantic image annotation methods are presented in [2], [3], [5]. In [22] music is categorized into emotions. Gene and protein function prediction are proposed in [4].

The *curse of dimensionality* often causes serious problems when high dimensional data are used for learning, and thus a lot of dimensionality reduction methods have been developed. Depending on whether the label information is used, those methods can be classified into two categories, *supervised* and *unsupervised*. In *unsupervised* learning, label information is not provided. There are two options available in order to reduce the dimensionality of multilabel data.

In the first approach, any *unsupervised* dimensionality reduction method built for single label problems can be used. A representative of *unsupervised* dimensionality reduction methods is *Principal Component Analysis (PCA)* that tries to find those projections that maximize the variance among data. *Random Projection* [17] is another method that projects the data on a random lower-dimensional orthogonal subspace that captures as much of the variation of the data as possible. *Latent semantic indexing (LSI)* [18] can also be used directly in multilabel data. *LSI* is widely applied to documents analysis and information retrieval. To apply *LSI*, documents are represented in a vector space model, and *Singular Value Decomposition (SVD)* is performed to find the sub-eigenspace with large eigenvalues. *Partial least squares (PLS)* [12] can also be applied directly to multi-label data by ignoring label correlation.

The second approach uses the provided labels in order to find a projection that enhances discriminality between labels. *Multi-Label Dimensionality Reduction via Dependence Maximization (MDDM)* [14] projects the original data into a lower-dimensional feature space maximizing the dependence between the original feature description and the associated class labels. *Multi-label latent semantic indexing (MLSI)* [15] is an extend of *LSI* so that it can properly manage multilabel data. A representative of supervised dimensionality reduction method is *Linear Discriminant Analysis (LDA)*, which aims at identifying a lower-dimensional space minimizing the inter-class similarity while maximizing the intra-class similarity. *LDA* cannot be directly applied to multilabel data, thus an extend of *LDA* is proposed in [13], named *Multi label Discriminant Analysis MLDA* that takes advantage of label correlation between label sets of multi label data.

Multilabel learning considers both *multi-label classification (MLC)* and *label ranking (LR)*. *MLC* is concerned with learning a model that outputs a bipartition of the set of labels into relevant and irrelevant with respect to a query instance. *LR* extends conventional multiclass classification in the sense that it gives an ordering of all class labels. We can group the existing methods for multi-label learning problems into two main categories a) *transformation methods* and b) *adaption methods* [7]. The first approach transforms the multilabel classification problem into one or more single label classification, regression or ranking task.

The second approach extends specific learning algorithms in order to handle multilabel data directly. The most popular approach as a transformation method is *Binary Relevance (BR)* [7]. *BR* creates k datasets each for one class label and trains a binary classifier, one for each different dataset. In [8] *Calibrated Label Ranking (CLR)* is proposed. A label ranking method is able to predict a ranking of all topics in decreasing order of relevance to a specific instance but it is not able to distinguish between the sets of relevant and non-relevant topics. To overcome this problem *CLR* adds an extra label to the original label set which is interpreted as "neutral element". ML-kNN [9] adapts lazy learning techniques to solve multi-label problems. To identify the label set for a given instance it uses *maximum a posteriori (MAP)*, based on prior and posterior probabilities for each k nearest neighbor label. *BP-MLL* [19] is an adaptation of *back-propagation* algorithm for multi-label learning. The algorithm introduces a new error function that takes multiple labels into account. Multi-label perceptron based algorithms have also been extended for multi-label learning. *Multi-class Multi-layer Perceptron (MMP)* is proposed in [10] where the perceptron algorithms weight update is performed in such a way that it leads to correct label ranking. *Support Vector Machine (SVM)* is used in [11] where *RankSVM* is proposed. *RankSVM* defines a specific cost function and the corresponding margin in order to solve multilabel problems. In this paper we extend MLDA and propose the Direct Multi-label Discriminant Analysis method for dimensionality reduction of multi-label data. MLDA cannot find a space with a larger dimensionality than the number of the labels. The proposed method overcomes this limitation and projects data onto a subspace that gives more than $K-1$ dimensions. This way we can find a reduced space that improves classification accuracy.

2 Direct Multi-label Discriminant Analysis

The proposed Direct MLDA extends MLDA by modifying the between class scatter matrix in order to improve classification accuracy. Given a multi label data set with n samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$ and K classes where $\mathbf{x}_i \in \mathcal{R}^p$ and $\mathbf{y}_i \in \{0,1\}^K$, $y_i(k) = 1$ if $\mathbf{x}_i$ belongs to the k-th class, and 0 otherwise. Let the input data be partitioned into K groups as $\{\pi_k\}_{k=1}^K$ where π_k denotes the sample set of the k-th class with n_k data points. We write $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]^T = [\mathbf{y}_{(1)}, \ldots, \mathbf{y}_{(K)}]$ where $\mathbf{y}_i \in \{0,1\}^n$ is the class-wise label indication vector for the k-th class.

In order to improve classification accuracy multi-label learning takes into account label correlation that takes advantages of label interactions. The label correlation between two classes is formulated as following [16]:

$$C_{kl} = cos(\mathbf{y}_{(k)}, \mathbf{y}_{(l)}) = \frac{\langle \mathbf{y}_{(k)}, \mathbf{y}_{(l)} \rangle}{\|\mathbf{y}_{(k)}\| \|\mathbf{y}_{(l)}\|}. \tag{1}$$

In this Section, we discuss *MLDA* in Section 2.1 and in Section 2.2 we propose *Direct MLDA* for dimensionality reduction of multilabel data.

2.1 Multi-label Linear Discriminant Analysis

Multi label Linear Discriminant Analysis (MLDA) is a multi-label version of *Linear Disriminant Analysis (LDA)* that has been adapted for multilabel data. The proposed scatter matrices are calculated class-wise as:

$$\mathbf{S}_b = \sum_{k=1}^{K} \mathbf{S}_b^{(k)}, \mathbf{S}_b^{(k)} = \left(\sum_{i=1}^{n} Y_{ik} \right) (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^T. \tag{2}$$

$$\mathbf{S}_w = \sum_{k=1}^{K} \mathbf{S}_w^{(k)}, \quad \mathbf{S}_w^{(k)} = \sum_{i=1}^{n} Y_{ik} (\mathbf{x}_i - \mathbf{m}_k)(\mathbf{x}_i - \mathbf{m}_k)^T. \tag{3}$$

where $\mathbf{m}_k$ is the mean vector of class k and $\mathbf{m}$ is the multi-label global mean vector. *MLDA* takes advantage of label correlation (1), and constructs the correlation matrix $\mathbf{C} \in \mathcal{R}^{K \times K}$ to define the correlations between labels. Moreover solves the over-counting problem, when a datapoint $\mathbf{x}_i$ is used more than once on the calculation of the scatter matrices, by the following normalized matrix $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]^T \in \mathcal{R}^{n \times K}$:

$$\mathbf{z}_i = \frac{\mathbf{C}\mathbf{y}_i}{\|\mathbf{y}_i\|_{\ell 1}}. \tag{4}$$

where $\|.\|_{\ell 1}$ is the $\ell 1$-norm of a vector. In equations (2) and (3) $\mathbf{Y}$ is replaced by $\mathbf{Z}$.

MLDA projects the original data $\mathbf{X}$ onto a lower q-dimensional feature space by taking into account both the within-class scatter matrix and the between-class scatter matrix. *MLDA* tries to minimize $\mathbf{S}_w$ in order to keep each class compact and maximize $\mathbf{S}_b$ in order to separate classes as much as possible. Thus the following criterion is maximized:

$$J = \frac{tr\{\mathbf{G}^T \mathbf{S}_b \mathbf{G}\}}{tr\{\mathbf{G}^T \mathbf{S}_w \mathbf{G}\}}. \tag{5}$$

where $\mathbf{G}$ the transformation matrix that consists of q eigenvectors $\mathbf{g}$ that correspond to the q largest eigenvalues of the eigenanalysis problem, $\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{g} = \lambda \mathbf{g}$.

2.2 Direct Multi-label Discriminant Analysis

A very strong limitation that *MLDA* suffers from, is the number of the produced projections that are defined as $K-1$ for a K class problem. Indeed, the between scatter matrix as defined in 2 is comprised of K rank 1 matrices that consider only the mean vectors for each label and the global mean vector. This can be interpreted as information less in the definition of $\mathbf{S}_b$ since the data samples are represented by their mean vectors. This representation results to a low rank matrix $\mathbf{S}_b$ that can produce at most $K-1$ projections for dimensionality reduction. *Direct MLDA* overcomes this problem and searches for a subspace that gives more than $K-1$ dimensions and better classification accuracy. In *Direct MLDA* the definition of the between-class scatter matrix $\mathbf{S}_b$ changes in order to distinguish data that do not have a specific label from the mean vector of the data belong to this label. $\mathbf{S}_b$ is defined as:

$$\mathbf{S}_b = \sum_{k=1}^{K} \mathbf{S}_b^{(k)}, \mathbf{S}_b^{(k)} = \sum_{i=1}^{n} (1 - Y_{ik})(\mathbf{x}_i - \mathbf{m}_k)(\mathbf{x}_i - \mathbf{m}_k)^T. \quad (6)$$

The definition of the within-class scatter matrix remains the same as we try to minimize the covariance of data that belong to the same class.

$$\mathbf{S}_w = \sum_{k=1}^{K} \mathbf{S}_w^{(k)}, \mathbf{S}_w^{(k)} = \sum_{i=1}^{n} Y_{ik}(\mathbf{x}_i - \mathbf{m}_k)(\mathbf{x}_i - \mathbf{m}_k)^T. \quad (7)$$

where $\mathbf{m}_k$ is the mean vector of class k which is defined as:

$$\mathbf{m}_k = \frac{\sum_{i=1}^{n} Y_{ik}\mathbf{x}_i}{\sum_{i=1}^{n} Y_{ik}}. \quad (8)$$

Direct MLDA projects the original p-dimensional feature vectors into a new reduced q-dimensional feature space while keeping the discrimination information between classes. We wish to determine a transform $\mathbf{w}_i = \mathbf{G}^T\mathbf{x}_i$, where $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_q]$ the projection matrix, such that the projected classes are well separated. After the projection onto $\mathbf{g}_t$ the between-class scatter matrix is defined as $\mathbf{S}_b^{g_t} = \mathbf{g}_t^T\mathbf{S}_b\mathbf{g}_t$ and the within-class scatter matrix is defined as $\mathbf{S}_w^{g_t} = \mathbf{g}_t^T\mathbf{S}_w\mathbf{g}_t$. To enhance the separability of the classes we wish to maximize $\mathbf{S}_b^{g_t}$ and keep each cluster compact by minimizing $\mathbf{S}_w^{g_t}$, thus the following criterion function should be maximized:

$$J = \frac{\mathbf{g}_t^T\mathbf{S}_b\mathbf{g}_t}{\mathbf{g}_t^T\mathbf{S}_w\mathbf{g}_t}. \quad (9)$$

We wish to maximize criterion J for each $\mathbf{g}_t$ of matrix $\mathbf{G}$, thus we define the following eigenanalysis problem:

$$\mathbf{S}_w^{-1}\mathbf{S}_b\mathbf{g} = \lambda\mathbf{g}. \quad (10)$$

The projections vectors of *Direct MLDA* are the q eigenvectors $\mathbf{g}_t$ corresponding to the q largest eigenvalues of $\mathbf{S}_w^{-1}\mathbf{S}_b$. Label correlation was also examined in *Direct MLDA* according to (1) and the over-counting problem was solved similar to *MLDA* by the following normalized matrix $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]^T \in \mathcal{R}^{n\times K}$:

$$\mathbf{z}_i = \frac{\mathbf{C}\mathbf{y}_i}{\|\mathbf{y}_i\|_{\ell 1}}. \quad (11)$$

where $\|.\|_{\ell 1}$ is the $\ell 1$-norm of a vector. In equations (6) and (7) $\mathbf{Y}$ is replaced by $\mathbf{Z}$.

3 Experimental Results on Video Analysis

The proposed approach has been applied to several challenging problems that arise in digital media analysis and description. That is, in digital video analysis one objective is to annotate the video according to the appearances of the actors and to automatically recognize their faces [25], their activities [26] or even their facial expression [27]. All these recognition problems refer to the same multi-label video data and thus, it would be useful to reduce the video dimensionality keeping the discriminality for each label.

Performance evaluation in multi-label learning is more complicated than traditional single label classification as multi-label data are associated with more than one label simultaneously. The multi-label evaluation metrics that are used in this paper consider the performance of label set prediction as well as the performance of label ranking. For the evaluation of label prediction *macro* (Precision, Recall, F-measure, Accuracy) and *micro* (Recall, F-measure, Accuracy) measures were used that average difference of the actual and the predicted sets of labels over all labels (*micro* measures) or for each label and subsequently average over all labels (*macro* measures). Hamming Loss was also used to evaluate the fraction of misclassified instance-label pairs. For the evaluation of label ranking One Error, Coverage, Ranking Loss and Average Precision were used. One error evaluates the fraction of examples whose top-ranked label is not in the relevant label set. Coverage evaluates how far on average a learning algorithm needs to move down in the ranked label list in order to cover all the relevant labels of the example. Ranking Loss evaluates the fraction of reversely ordered label pairs and Average Precision computes for each relevant label the fraction of relevant labels ranked higher than it, and finally averages over all relevant labels.

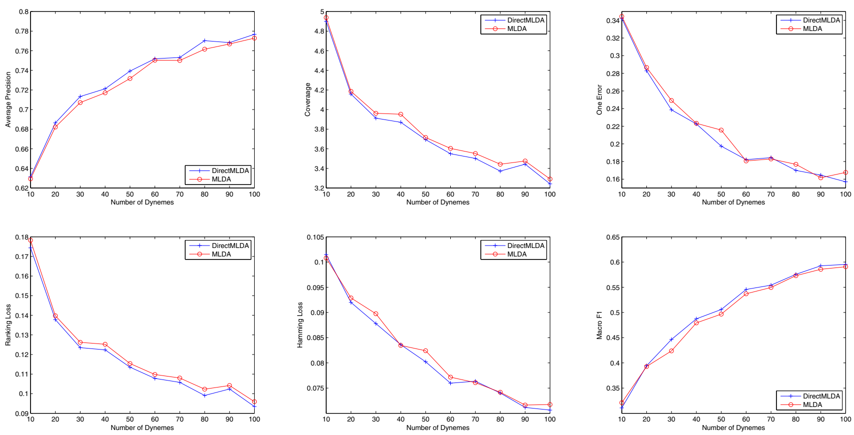

Fig. 1. Performance of *MLDA* and *Direct MLDA* for different evaluation metrics and number of dynemes for the database *i3DPost mask*

To examine *Direct MLDA* performance we used the following databases that refer mostly to multi-label activity recognition for video analysis:

i3DPost [20] is an image sequence database that contains 64 high-resolution image sequence of eight persons performing eight actions and two person interaction. Eight cameras having a wide 45° viewing angle difference was used to provide 360° coverage of the capture volume. *Mobiserv*[24] is an image sequence database, that depicts twelve persons performing three actions activities "eat", "drink" and "apraxia". The total number od activity video is 954.

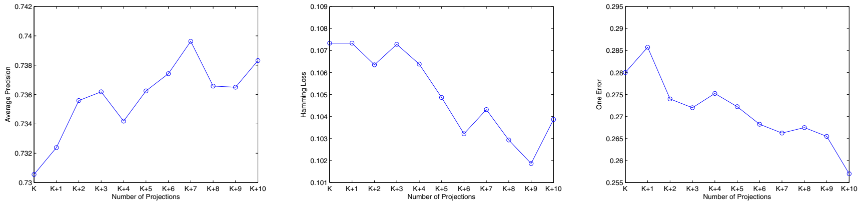

Fig. 2. Performance of *Direct MLDA* for different number of projections and metrics

IXMAS[21] database contains 330 low resolution image sequences of ten persons performing eleven activities. Each sequence has been captured by five cameras. The persons freely change position and orientation.

The activity representation of those databases was formed with the use of dynemes that refer to sequences of movement primitives. By calculating the similarity from multi-view dynemes the final representation of human activity was formed. We used 5-fold cross validation to evaluate the performance of the proposed approach to those databases for different number of dynemes [24]. In all these activity video datasets there are multiple labels per sample (i.e., video) that correspond to the identity of the person performing the activity, the different activity type and the different camera view. That is, the proposed approach is tested on recognizing simultaneously the correct person id, the correct activity and the correct camera view using the same dyneme based representation and the direct MLDA for dimensionality reduction.

Emotion [22] categorizes 593 songs into 6 emotions. A 5-fold cross validation was also used on *Emotion* database.

Reference [23] database, from Yahoo dataset refers to multi-label web pages. In *Reference's* label set we removed topics with less than 100 web pages. The high dimensionality was reduced using *Random Projections*.

Multi-label K-Nearest Neighbor (ML-KNN) was used for classification after dimensionality reduction by *MLDA* and *Direct MLDA*. The selected number of K-Nearest Neighbors was defined to 5.

Evaluation of *MLDA* and the proposed *Direct MLDA* are depicted in Tables 1 and 2. The results show that *Direct MLDA* is superior to *MLDA* in terms of both label prediction and label ranking. Although *Direct MLDA* can project data to more than $K-1$ projections to improve classification accuracy, we prove that *Direct MLDA* is superior to *MLDA* also for $K-1$ projections. In Figure 1 we project the original database onto a $K-1$-dimensional space in both cases (*MLDA*, *Direct MLDA*) and show that *Direct MLDA* gives better results for the database *i3DPost mask* for different number of dynemes for different evaluation metrics. In Figure 2 performance of *Direct MLDA* for *Reference* is presented for different number of projections for the evaluation metrics Average Precision, OneError and Hamming Loss.

Table 1. Performance evaluations of *MLDA* and *Direct* MLDA for *i3DPost mask*, *i3DPost stips* and *mobiserv* databases

	i3DPost mask (180)		i3DPost STIPs (30)		mobiserv (50)	
	MLDA	DMLDA(K-1)	MLDA	DMLDA(K+5)	MLDA	DMLDA(K+1)
F1-macro ↑	0.4940	**0.6215**	0.4324	**0.4510**	0.4183	**0.4237**
Macro Precision ↑	0.6812	**0.7985**	0.6477	**0.6648**	0.5881	**0.5895**
Macro Recall ↑	0.4328	**0.5498**	0.3756	**0.3887**	0.3747	**0.3833**
F1-micro ↑	0.5697	**0.6788**	0.5164	**0.5292**	0.6619	**0.6661**
Micro Recall ↑	0.4659	**0.5778**	0.3783	**0.3887**	0.5368	**0.5484**
Hamming Loss ↓	0.0810	**0.0684**	0.0883	**0.0863**	**0.0715**	0.0717
Accuracy ↑	0.9190	**0.9316**	0.9116	**0.9137**	**0.9285**	0.9283
Average Precision ↑	0.6822	**0.6864**	0.6822	**0.6848**	0.7874	**0.7879**
Coverage ↓	4.0871	**3.3483**	8.5161	**8.3545**	6.0526	**5.9573**
One Error ↓	0.2433	**0.1517**	0.1093	**0.1054**	**0.1094**	0.1126
Ranking Loss ↓	0.1387	**0.0982**	**0.1312**	0.1325	0.0907	**0.0898**

Table 2. Performance evaluations of *MLDA* and *Direct MLDA* for *IXMAS mask*, *Emotion* and *Reference* databases

	IXMAS mask (10)		Emotion		Reference	
	MLDA	DMLDA(K-1)	MLDA	DMLDA(K+2)	MLDA	DMLDA(K+16)
F1-macro ↑	0.0505	**0.0624**	0.5794	**0.6009**	**0.3221**	0.3105
Macro Precision ↑	0.1628	**0.1841**	**0.6752**	0.6519	0.3014	**0.3303**
Macro Recall ↑	0.0335	**0.0439**	0.5334	**0.5802**	**0.3644**	0.3177
F1-micro ↑	0.0628	**0.0794**	0.6093	**0.6248**	0.5374	**0.6100**
Micro Recall ↑	0.0333	**0.0431**	0.5509	**0.5964**	0.5543	**0.5775**
Hamming Loss ↓	**0.0826**	0.0827	**0.2198**	0.2230	0.1331	**0.1026**
Accuracy ↑	**0.9174**	0.9173	**0.7802**	0.7761	0.8669	**0.8964**
Average Precision ↑	0.3518	**0.3579**	0.7695	**0.7758**	0.6808	**0.7479**
Coverage ↓	11.4778	**11.3227**	1.9360	**1.9044**	4.0580	**3.8367**
One Error ↓	0.7069	**0.7009**	**0.3170**	0.3256	0.3765	**0.2485**
Ranking Loss ↓	0.3164	**0.3102**	0.1948	**0.1903**	0.1410	**0.1187**

4 Conclusions

In this paper we have proposed the *Direct Multi-label Linear Discriminant Analysis* method, as an extend of *Multi-label Linear Discriminant Analysis.* We reformulated the between-class scatter matrix in order to distinguish data that do not have a specific label. Direct MLDA gives more than $K - 1$ projections and search this way for the better reduced space that improves classification accuracy to the maximum. The theoretical advantages of our method are confirmed in experimental evaluations on multi-label video analysis that show that *Direct MLDA* performs better than*MLDA* in terms of both label prediction and label ranking.

Acknowledgment. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART). This publication reflects only the authors views. The European Union is not liable for any use that may be made of the information contained therein.

References

1. Zhang, M., Zhou, Z.: A Review on Multi-Label Learning Algorithms. IEEE Transactions on Knowledge and Data Engineering (2013)
2. Yang, S., Kim, S.-K., Ro, Y.-M.: Semantic Home Photo Categorization. IEEE Transactions on Circuits and Systems for Video Technology, 324–335 (2007)
3. Tang, J., Hua, X.-S., Wang, M., Gu, Z., Qi, G.-J., Wu, X.: Correlative Linear Neighborhood Propagation for Video Annotation. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 409–416 (2008)
4. Borges, H.B., Nievola, J.C.: Multi-Label Hierarchical Classification using a Competitive Neural Network for protein function prediction. In: The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2012)
5. Wang, H., Hu, J.: Multi-label image annotation via Maximum Consistency. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 2337–2340 (2010)
6. Huang, S., Jin, L.: A PLSA-Based Semantic Bag Generator with Application to Natural Scene Classification under Multi-instance Multi-label Learning Framework. In: Fifth International Conference on Image and Graphics, pp. 331–335 (2009)
7. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining multi-label data-instance Multi-label Learning Framework. In: Data Mining and Knowledge Discovery Handbook, pp. 667–685 (2010)
8. Fürnkranz, J., Hüllermeier, E., Loza Mencía, E., Brinker, K.: Multilabel classification via calibrated label ranking, pp. 133–153 (2008)
9. Zhang, M.-L. and Zhou, Z.-H.: A k-nearest neighbor based algorithm for multi-label classification. In: 2005 IEEE International Conference on Granular Computing, pp. 718–721 (2005)
10. Crammer, K., Singer, Y., Jaz, K., Hofmann, T., Poggio, T., Shawe-taylor, J.: A Family of Additive Online Algorithms for Category Ranking. Journal of Machine Learning Research (2003)

11. Elisseeff, A., Weston, J.: A kernel method for multi-labelled classification. In: Advances in Neural Information Processing Systems, pp. 681–687 (2001)
12. Arenas-garcia, J., Petersen, K.B., Hansen, L.K.: Sparse Kernel Orthonormalized PLS for feature extraction in large data set. In: Advances in Neural Information Processing Systems (2007)
13. Wang, H., Ding, C., Huang, H.: Multi-label linear discriminant analysis. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 126–139. Springer, Heidelberg (2010)
14. Zhang, Y., Zhou, Z.-H.: Multilabel dimensionality reduction via dependence maximization. ACM Transactions on Knowledge Discovery from Data (TKDD), 1–14 (2010)
15. Yu, K., Yu, S., Tresp, V.: Multi-label informed latent semantic indexing. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 258–265 (2005)
16. Wang, H., Huang, H., Ding, C.: Image annotation using multi-label correlated Green's function. In: IEEE 12th International Conference on Computer Vision, pp. 2029–2034 (2009)
17. Bingham, E., Mannila, H.: Random projection in dimensionality reduction: applications to image and text data. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 245–250 (2001)
18. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science, 391–407 (1990)
19. Zhang, M.L., Zhou, Z.H.: Multi-label neural networks with applications to functional genomics and text categorization. IEEE Transactions on Knowledge and Data Engineering, 1338–1351 (2006)
20. Gkalelis, N., Kim, H., Hilton, A., Nikolaidis, N., Pitas, I.: The i3DPost multiview and 3D human action/interaction database. In: 6th Conference on Visual Media Production, pp. 159–168 (2009)
21. Weinland, D., Ronfard, R., Boyer, E.: Free viewpoint action recognition using motion history volumes. Computer Vision and Image Understanding, 249–257 (2006)
22. Trohidis, K., Tsoumakas, G., Kalliris, G., Vlahavas, I.: Multilabel classification of music into emotions. In: Proc. of ISMIR (2008)
23. Ueda, N., Saito, K.: Single-shot detection of multiple categories of text using parametric mixture models. In: Proc. of SIGKDD, pp. 626–631 (2002)
24. Iosifidis, A., Tefas, A., Pitas, I.: Activity-Based Person Identification Using Fuzzy Representation and Discriminant Learning. IEEE Trans. on Information Forensics and Security, 530–542 (2012)
25. Kyperountas, M., Tefas, A., Pitas, I.: Dynamic training using multistage clustering for face recognition. Pattern Recognition, 894–905 (2008)
26. Gkalelis, N., Tefas, A., Pitas, I.: Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition. IEEE Transactions on Circuits and Systems for Video Technology, 1511–1521 (2008)
27. Kyperountas, M., Tefas, A., Pitas, I.: Salient feature and reliable classifier selection for facial expression classification. Pattern Recognition, 972–986 (2010)