

Research on Image Semantic Information Mining Based On Latent Dirichlet Allocation Model

Sa Yang
Guangdong University of Education, Guangdong, Guangzhou 510303, China
E-mail: yangsa@21cn.com

***Abstract*—Focusing on the issues of lacking of semantic description on image identification and methods of mapping from low-level semantics to high-level semantics, this paper describes the experiments of identification of image semantic information by using LDA model, which can achieve the mapping from image visual feature to high-level semantics, and experiments on the data sets of Corel 5k and Corel 30k. The experiments test and verify that LDA model performs a good stability on identification of image semantic information, and advantageous on dimension accuracy and recall rate, which provide a new solution and an embodiment of the identifying of image semantics intelligently and automatically.**

***Index Terms*—Image semantic information, Image identification, Visual word bag, LDA model**

I. INTRODUCTION

Image mining (IM) is an interdisciplinary research subject which achieves automatically the content and mode of implied image from a large number of images. It doesn't apply the technology of data mining to the field of image comprehension simply, but concentrates the technology of data mining; image management, computer visual, image retrieval, machine learning, pattern recognition database and artificial intelligence on realize the analysis and comprehension on image contents. The basic task of image mining is achieving the top image space objects and the correlations among them efficiently to collect implied information of image sequences[1][2].

Nowadays, the information technology develops rapidly. With the progress of digital imaging, internet and data storage, normal people are able to communicate and express themselves with others by sharing images, videos and other online multimedia. Digital images from all kinds of information source increase everyday. The issue on how to manage, visit, store and retrieve the image information becomes important in recent years. The image retrieval based on semantics proves to be an efficient method. It uses technologies of feature extraction and high-dimensional indexing to retrieve images and makes good progress[3]. However, because of the semantic gap between visual feature and human perception, this method of retrieval is not that ideal. For achieving semantically related retrieval results and avoiding the issues of large workload and strong subjectivity of manual marking, automatically marking and semantic learning technology becomes the key challenging issue[4-6].

The method in this article regards images and texts as peer-to-peer data, and attempts to figure out the combination by using the way of non-supervision. This method transforms the semantic learning issue to the statistical reasoning issue in probabilistic graphical model via simultaneous distribution between estimate features and key words. From the aspect of graphical representation, the existing automatic tagging model can be divided into discrete model and continuous model. The discrete model qualifies provincial features which have been divided to gain a visual words statistical histogram, which known as "bag of words". Then this histogram will be modeled or divided; the continuous model directly disposes these fields, which express as the standard feature collection (including colors, textures, shape and other information). The feature collection is also called "bag of features". The discrete model is more efficient, because it only disposes discrete magnitude. And it can choose and collect more features. In recent years, probability subject models which represented by probabilistic latent semantic analysis (PLA) and latent Dirichlet allocation (LDA) are widely used in computer vision fields.

Among these, LDA[7], which develops fast in recent years, is an important method of discrete data collection modeling. For the reason LDA is a total probability-generating model, it has an ample inner structure, and can calculate with an efficient probabilistic algorithm. Moreover, LDA model is more suitable to dispose the large-scale corpus for the reason the scale of parameter space of LDA is not related to the number of training documents. The LDA model has been widely applied in many fields of machine learning and information retrieval.

In the recent years, there are several research related to image semantic information mining as follows.

Liu et al. proposed a novel approach for image region-level semantic mining. In this method, images are segmented into several parts using an improved segmentation algorithm, each with homogeneous spectral and textural characteristics, and then a uniform region-based representation for each image is built. Once

doi:10.4304/jsw.8.9.2329-2335

the probabilistic relationship among image, region, and hidden semantic is constructed, the Expectation Maximization method can be applied to mine the hidden semantic[8].

Wang et al. proposed a method to extract semantic concepts from a large database of images effectively. The authors we tackle the problem by mining the decisive feature patterns. A decisive feature pattern is a combination of low-level feature values that are unique and significant for describing a semantic concept. Interesting algorithms are developed to mine the decisive feature patterns and construct a rule base to automatically recognize semantic concepts in images[9].

Tang et al. present an intelligent content-based image retrieval system called I-Browse, which integrates both iconic and semantic content for histological image analysis. The I-Browse system combines low-level image processing technology with high-level semantic analysis of medical image content through different processing modules in, the proposed system architecture[10].

In paper [11], Monay et al. present three alternatives to learn a Probabilistic Latent Semantic Analysis (PLSA) model for annotated images and evaluate their respective performance for automatic image indexing. Under the PLSA assumptions, an image is modeled as a mixture of latent aspects that generates both image features and text captions, and we investigate three ways to learn the mixture of aspects. The authors also proposed a more discriminative image representation than the traditional Blob histogram, concatenating quantized local color information and quantized local texture descriptors[11].

Luo et al. present a general-purpose knowledge integration framework that employs BN in integrating both low-level and semantic features. The efficacy of this framework is demonstrated via three applications involving semantic understanding of pictorial images. The first application aims at detecting main photographic subjects in an image, the second aims at selecting the most appealing image in an event, and the third aims at classifying images into indoor or outdoor scenes[12].

At present, the computer image recognition is not able to collect the semantic feature of images. There are two difficulties: one is the lack of description method of some conceptual semantic structure; the other is the "semantic gap" in the process of mapping from the low-level image visual feature to the high-level semantic features has not been solved completely, as is to say, there's no method to map from the low-level image visual feature to the high-level semantics.

This article achieve the mapping from the image visual feature to the high-level semantics via the research on the model description of image semantic, collection method and the use of image semantic mining, and provide a new solution and an embodiment of the identification of image semantics intelligently and automatically. The current image classification is based on three aspects: image low-level feature layer, middle semantic of middle layer and high-level image semantic. The images will be modeled. Using the image low-level feature to classify images can no longer be qualified for the database classification of a large number of images. Thus the research of this article mainly focuses on the middle layer of description and modeling and the discrimination of high layer scene semantics, which including the expression of visual bag of words, middle semantic modeling and semantic identification retrieval.

II. Explanation of the Visual Bag of Word Model

As the improvement of the performance of single low-level classification, the appearance such as human, sky, grass can solve the "semantic gap" between the image low-level features and high-level semantics. There are inconformity and huge diversity in the image objective visual information, which collected by the system, high-level semantic content of the image itself and the comprehension of the people who face the same picture in different environments. These inconformity and huge diversity make the so-called "semantic gap"

In the field of the computer vision, the feature of images is the important part which connects to its content tightly and has received widely attention by scholars. Image feature is formed by the obvious change of the grayscale of the local area of the image which effected by the physical and geometrical feature of the scene. The same kind of image scenes has similar distributed concepts. By establishing the areal distribution of the semantic concept, the scenes can be divided into specific semantic categories. These similar concepts are usually related to image features. How to maintain the invariance of features in the process of the expression of the middle-layer image to embody the spatial information of features and achieve the integration and compliment of local feature and global feature are the keys to the connection of the semantic gap between the low-level features and high-level semantics and the keys to the scene semantic modeling.

This paper achieves classification based on semantic distribution of words with bag of words expressing images. According to the different training data of M amplitude, the word of bags will be modeled as following steps.

1) Feature point detection: detect the local interested feature point by using image rasterized way and difference of Gaussian (DoG) operator method. For the former, the interested point will be confirmed by the grid position whose size is 4×3 pixels. For the latter, some remarkable features such as corners and spots will be detected.

2) Feature description: the descriptor which maintains the scale but changes the feature is insensitive to the change of factors like scale, zoom, rotation and illumination. Therefore in the article, the SIFT operator will be used to describe the feature point. The rasterized feature point will be expressed by Dense SIFT (DSIFT). And the DoG feature point will be expressed by DoG+SIFT.

3) Vocabulary generating: two kinds of SIFT operator of the training images will be gained in Step 2. Cluster them by using k-means clustering algorithm. The central cluster composes the vocabulary of the image, V. That

means at first the detecting feature points and describing features will be done, then express the image by using the $w=\{w_1,w_2,\ldots,z_N\}$ bag of words.

III. THE INTERMEDIATE SEMATIC MODEL

Among images in the same category of, there exist similar concepts that are diversified because of the diversity of the content. The way to capture the similar concept and abstract them as a subject is the key to provide a further intermediate sematic description for the image. Because of the combination way of the contents in a semantic scene is unknown, it is more reasonable to use the probability-generating model for derivation. Based on the probability-generating model, the universal subject of each image is established and the image visual similarity is captured. According to the internal semantic changes of each image, the category constraint mechanism is established; the related subject model is generated. Because the initial setting value of EM in the progress of the model derivation is too sensitive, the initial value will be tested firstly to speed up the convergence based on the priori testified information of the semantic content.

IV. IMAGE SEMATIC INFORMATION MINING BASED ON LDA MODEL

The probability generating models PLSA and LDA which are often used in the image semantic research belong to unsupervised learning style, and the classifier such as KNN, SVM are often use for help to complete the final classification task. The PLSA model can not provides a probability model for the potential subject to generate documentation and is unable to produce new document definition, thus the generalization capability is poor. Base on this shortcoming the LDA model; Blei promoted the LDA model. Based on the PLSA model and the probability distribution of the hidden variable Σ on the hyper-parameter layer, the LDA is formed. LDA model is a hierarchical bayesian model. It uses Dirichiet distribution to describe the subject distribution, i.e., using the Dirichict conjugate testified distribution to select the hybrid subject document sample. Firstly sample a group of subject in LDA model, and based on the distribution of which each document is generated, words are produced according to polynomial distribution of each related subject. So LDA overcome the shortcomings of PLSA model, it is a complete generation model and a hybrid model to view the probability of each document as probability of random occurrence of words in the potential subject.

This paper is based on the LDA model, and in the application, it also of get rid of the dependence on classifier. The maximum likelihood model comparing method is adopted to determine the semantic category of the image. The final determination of the established intermediate semantic subject of the image is made.

The LDA model bag contains K hidden subjects $z=\{z_1,z_2,\ldots,z_k\}$, z_i is the subject of i, each word od a image is produced by its corresponding theme. Each image is blended by K subjects, which are based on the certain probability θ. The formation process of image $w=\{w_1,w_2,\ldots,z_N\}$ is as follows

(1) Choose $\theta\sim \mathrm{Dirichlet}(\alpha)$,that is the formation of the subject is according with the probability of dirichlet's distribution, α is the prior of dirichlet's distribution;

(2) Choose words from the vocabulary which are accord with the $w_j\sim p(w_j|\theta,\beta)$, B is the polynomial distributed parameter of the subject - words layers, as shown in Figure 1.

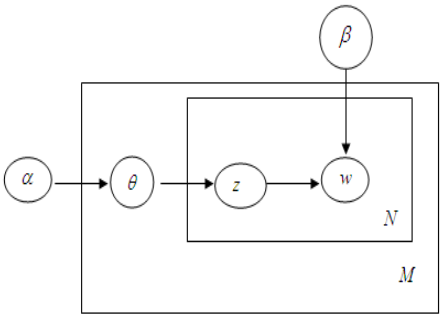

Figure 1. LDA model

In the LDA model, α,β are the hyper-parameter of the image collection layer which are presented by the assemblage $\Theta=\{\alpha,\beta\}$, w is the observable variable, $D=\{w_1,w_2\ldots w_M\}$ is the document collection that is formed by the M images. LDA model involves two processes of variation inference and parameter learning. Variation reasoning refers to the determination of the mixed probability θ and the probability of each word caused by the subject z when the hyper-parameter Θ and observation variable w is fixed, as shown in formula (1)

$$p(\theta,z|w,\Theta)=\frac{p(\theta,z,w|\Theta)}{p(w|\Theta)}=\frac{p(\theta|\alpha)\left(\prod_{i=1}^{N}p(z_i|\theta)\,p(w_i|z_i,\beta)\right)}{\int p(\theta|\alpha)\left(\prod_{i=1}^{N}p(z_i|\theta)\,p(w_i|z_i,\beta)\right)d\theta} \quad (1)$$

where $p(w|\Theta)$ is the Likelihood function of document w.

$$p(w|\Theta)=\int p(\theta|\alpha)\left(\prod_{i=1}^{N}p(z_i|\theta)\,p(w_i|z_i,\beta)\right)d\theta \quad (2)$$

Owing to the coupled relation between θ and β, $p(w|\Theta)$ can not be calculated directly. According to Figure 1, this coupled relation is caused by the conditional relation of θ, z, and w. By deleting the line between θ and z in Figure 1,and the point w, the

approximate distribution of $p(\theta, z|w, \Theta)$ is $q(\theta, z|\gamma, \phi)$, as shown in formula (3), the simplified model is shown in Figure 2.

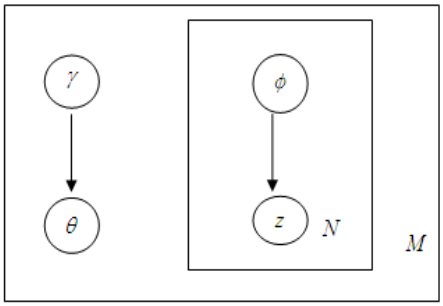

Figure2. The approximate solution figure of LDA model

Parameter γ is the dirichlet distribution parameter of θ, ϕ is the polynomial distributed parameter of z. Using variaty method to solve the parameters. As for the logarithmic likelihood function $\log p(w|\Theta)$ of the document, using Jensen inequality to obtain the lower bound $L(\gamma, \phi; \Theta)$ of the likelihood function.

$$q(\theta, z|\gamma, \phi) = q(\theta|\gamma)\prod_{n=1}^{N} q(z_n|\phi_n) \quad (3)$$

$$\log p(w|\Theta) = L(\gamma, \phi; \Theta) + D\left(q(\theta, z|\gamma, \phi)\| p(\theta, z|w, \Theta)\right) \quad (4)$$

where $D\left(q(\theta, z|\gamma, \phi)\| p(\theta, z|w, \Theta)\right)$ refers to the Kull-back — Leibler divergence between the q of approximate distribution model and p of the LDA model. The smaller of KL divergence, the closer of q and p. The minimum KL divergence can be achieved through the low bound $L(\gamma, \phi; \Theta)$ of the maximum likelihood function, thus model parameters γ and ϕ. Based on γ and ϕ, θ s and z can be obtained through sampling.

Parameter learning process is the determination process of the hyper-parameter $\Theta = \{\alpha, \beta\}$, when the observation variable $D = \{w_1, w_2, \cdots, w_M\}$ is set, the EM – iteration can be achieved. It uses the variation inference method to calculate the variation parameters γ and ϕ of each document. Collecting the variation parameters of all documents, calculating the derivatives of the hyper-parameters, and based on the low bonder $L(\gamma, \phi, \Theta)$ of the maximum likelihood function, realizing the estimation of hyper-parameter. When applying the LDA model into image semantic information identification, the subjects mixing probability of each image is θ, by the usage of variational inference method. Then based on maximum component θ_i of θ the sematic information can be determined.

V. Experimental Results and Analysis

For testing the effectiveness and accuracy of the LDA model in the process of image information mining, we had experiments on the data set of Core15k and Core130k. Core15k includes 5000 images. Each has 1~5 original mark of key words. 4500 images are used to training, and 500 are used for testing. There are 371 key words in the data set. The key words that mark more than 8 images will be chosen into the vocabulary, which includes 260 key words in total. Most semantic learning models inform the result of this data set. So using this data set can be compared with other typical method of automatic image remark. Core130k is a larger scale data set. It can solve the issues of image and word shortage in Core15k. The use of Core130k is for proving the extendibility and robustness of LDA. Core150k includes 31695 images and 5587 words, in which 90% of the images (28525 images) are used for training and 10% (3170 images) are used for testing. The vocabulary will choose the key words that mark more than 10 images. There are 950 key words in total.

LDA chooses 5 key words with the highest posterior probability as the mark result of each image, and compares them with calculating the accuracy and recall rate of every key word. The result of estimating the semantic retrieval can take MAP (mean average precision) as standard. First, every average precision (AP) can be defined as the sum of accuracy that is calculated by inquiring the location of the related images correctly. The sum will be divided by the related image number which is inquired. Then MAP can be defined as the average value of the AP which is gained by the retrieval system's inquiring for many times. So results of the MAP can estimate the performance of the whole system.

TABLE 1.
PERFORMANCE COMPARISON OF ALL KINDS OF AUTOMATIC IMAGE MARK MODEL

Model	Words with recall>0	Mean Per-word Recall		Mean Per-word Precision	
		Best 49 key words	All 260 key words	Best 49 key words	All 260 key words
TM	47	0.37	0.03	0.24	0.08
CRM	95	0.70	0.25	0.74	0.24
MBRM	115	0.78	0.22	0.56	0.13
PLSA	110	0.75	0.20	0.66	0.21
LDA	125	0.80	0.25	0.75	0.26

In order to evaluate the performance of the proposed approach, we utilize Precision, Recall and F1 as metric. The precision and recall for a given dataset is defined as follows.

$$P(i) = \frac{NF(i) \cap NG(i)}{NF(i)} \quad (5)$$

$$R(i) = \frac{NF(i) \cap NG(i)}{NG(i)} \quad (6)$$

Where $NF(i)$ and $NG(i)$ denote the number of elements found by the proposed algorithm and ground truth ones respectively.

Table 1 shows the mark performance of several of semantic learning models, and has counted the average recall rate and AP which are the result of the calculation made under the set of the two key words: one is the result under the set of 49 key words with the best performance; the other one is the result under the set of all 260 keys words.

The data in the table shows that the performance of LDA is obviously higher than other common model. It indicates that the modeling method of LDA and learning algorithm is practicable and efficient.

In the semantic retrieval of images, the inquiry of each key word can achieve the relevant AP value through calculating, so the mAP value of every key word set can be calculated. Table 2 shows several mAP value which achieved by the semantic retrieval of the common mark model. The data in the table indicates that the performance of LDA is also higher than other methods.

TABLE 2.
PERFORMANCE COMPARISON OF ALL KINDS OF IMAGE MODEL RETRIEVALS

Model	All 260 words	Words with recall ≥ 0
TM	0.19	0.20
CRM	0.26	0.27
MBRM	0.23	0.25
PLSA	0.30	0.32
LDA	0.33	0.37

Compared with Core15k, Core130k provides a large-scale image and semantic key words data base. In this data set, the PLSA, which performed well, and LDA will be compared. In Table 3, from each mAP value we can see that LDA still has a higher semantic retrieval performance than PLSA. From the above all, the experiments show that the LDA model has a well stability in image semantic identification and has an upper advantage in labeling accuracy and recall rate, which decide directly the good effect of its semantic retrieval. Besides, it has an upper robustness and extendibility.

TABLE 3:
RESULT COMPARISON OF SEMANTIC IDENTIFICATION ON CORE130K

Model	All 950 Words	Words with recall ≥ 0
PLSA	0.16	0.18
LDA	0.22	0.25

To make performance comparision, we compare the proposed algorithm with the methods proposed in Paper [10], paper [11] and paper [12] respectively.

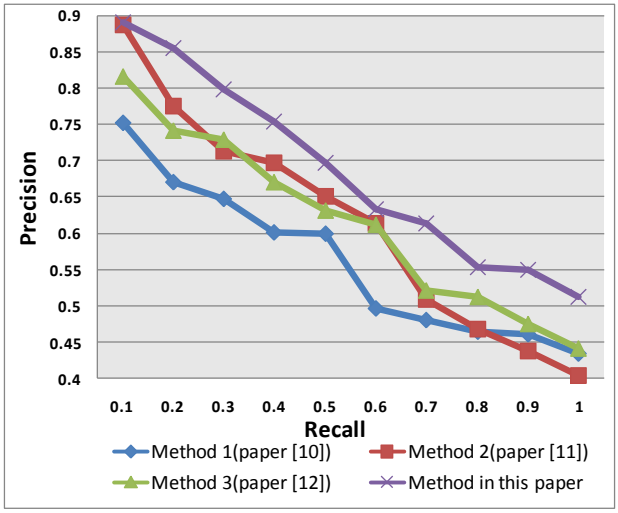

Figure 3. Precision-Recall curve for different methods under the Corel5K dataset

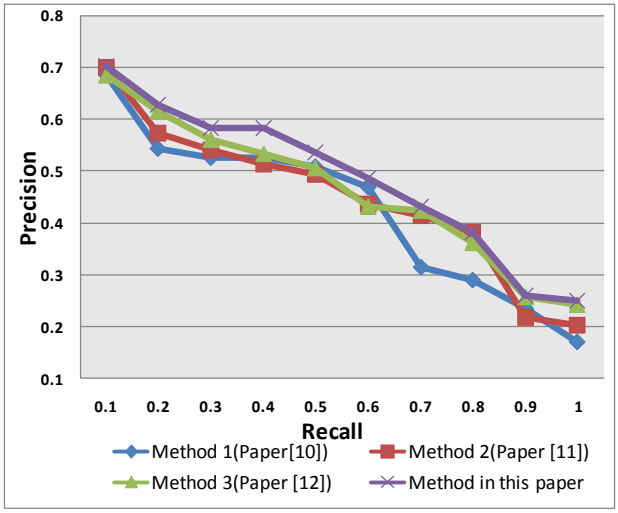

Figure 4. Precision-Recall curve for different methods under the Corel30K dataset

Image					
Semantic information	Mountain Snow Tree	Sky Mountain Road	Aircraft Sky	Sky Cloud Mountain	Mountain Sky Sea
Image					
Semantic information	Tiger Water	Fox Grass Sky	Fox	Train	Car Road Tree

Figure 5. Examples of image semantic mining for Corel 5K dataset.

From the experimental results shown in Fig.3 and Fig.4, we can see that compared with other existing methods, our algorithm can effectively semantic information from images. The reasons lie in that the LDA model can effectively the problem of "semantic gap". To illustrate image semantic information mining results of the proposed, we show some examples of image semantic mining for Corel 5K dataset in Fig.5. It can be seen that the proposed can accurately extract semantic information from images.

VI. Conclusion

The research on image semantic information is becoming the hotspot in recent years. Image information handling based on semantics plays an more important role in image comprehension and computer visual field. It has a significant practical meaning of the development of the image and video retrieval based on semantics. In recent years, the method of image semantic information mining is under research in how to jump over the semantic gap between the low-level image feature and the high-level semantics. In the process of watching from people's vision, the image semantic content can be analyzed through spatial arrangement without perceiving the target in the scene. Therefore, modeling learning scene semantics becomes the research hotspot and key point of the image comprehension and classification. The research in the article processes in the way of hierarchical images. The bonding point between the visual words and advanced semantics becomes the start of the research. With the idea of" visual word of bags modeling-semantic subject modeling-semantic classification", it takes the description of the middle-layer middle image semantic as the main research content. The image semantic information mining will lead the direction of the development for a long time. The article achieves the mapping from the image visual feature to the high-level semantics through the description of the image semantics, methods of collection and its application in image semantic mining with the image semantic information identified by the LDA. Therefore, it provides a new solution and an embodiment of the identifying image semantics intelligently and automatically.

Reference

[1] Lam Tony, Singh Rahul, "Semantically Relevant Image Retrieval by Combining Image and Linguistic Analysis"[C], Proceedings of ISVC(06) II, pp.770-779, 2006.

[2] MURPHY-CHUTORIAN E, TRIVEDI M M, "Head pose estimation computer vision: a survey"[J], IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31,no. 4, pp. 607-626,2009.

[3] WU Junwen, TRIVEDI M M, "A two-stage head pose estimation framework and evaluation"[J], Pattern Recognition, vol. 41, no. 3, pp.1138-1158, 2008.

[4] RAYTCHEV B, YODA I, SAKAUE K, "Head pose estimation by nonlinear manifold learning"[C], Proceedings of International Conference on Pattern Recognition, Cambridge, UK: IEEE Computer Society Press, pp.462-466, 2004.

[5] WANG Lianxi, JIANG Shengyi, "Unsupervised feature selection method for categorical features"[J], Journal of Chinese Computer Systems, vol. 32, no. 1, pp.47-50, 2011.

[6] ARANDELA R, S NCHEZ J S, GARC A V, "Strategies for learning in class imbalance problems"[J], Pattern Recognition, vol. 36, no. 3, pp.849-851, 2003.

[7] BLEI D, NG A, JORDAN M, "Latent dirichlet allocation" [J], Journal of Machine Learning Research, vol. 3, pp.:993-1022, 2003.

[8] Liu Tingting, Zhang Liangpei, Li Pingxiang. "Remotely sensed image retrieval based on region-level semantic mining"[J], Eurasip Journal On Image And Video Processing, article no. 4 2012.

[9] Wang, W; Zhang, AD, "Extracting semantic concepts from images: a decisive feature pattern mining approach"[J], Multimedia Systems, vol. 11, no. 4, pp. 352-366 2006.

[10] Tang HL, Hanka R, Ip HHS, "Histological image retrieval based on semantic content analysis"[J], IEEE Transactions

on Information Technology in Biomedicine, vol. 7, no. 1, pp. 26-36 2003.
[11] Monay Florent, Gatica-Perez Daniel, "Modeling semantic aspects for cross-media image indexing"[J], IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 10, pp. 1802-1817 2007.
[12] Luo JB, Savakis AE, Singhal A. "A Bayesian network-based framework for semantic image understanding"[J], Pattern Recognition, vol. 38 no. 6, pp. 919-934, 2005.
[13] NaPhade M, Huang T, "A Probabilistic framework for semantic video indexing", filtering and Retrieval", IEEE Transaction on Multimedia, vol. 3, no. 1, pp.141-151, 2001.
[14] Bosch A, Zisserman A, Munoz X, "Scene Classification Using a Hybrid Generativem/ Discriminative Approach" [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 4, pp.712-728, 2008.
[15] Yang Jianchao, Yu Kai, Gong Yihong, et al, Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification" [C], Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA: IEEE Computer Society, pp.1794-1801, 2009.
[16] ELAZMEH W, JAPKOWICZ , MATWIN S, "Evaluating misclassification in imbalanced data"[J], LNCS , vol. 4212:, pp.126-137, 2006.
[17] BARANDELA R, S NCHEZ J S, GARCA V, "Strategies for learning in class imbalance problems "[J], Pattern Recognition, vol.36, no. 3, pp.849-851, 2003.
[18] LIU X Y, WU J, ZHOU Z H, "Exploratory under-sampling for class-imbalance learning"[J], IEEE Transactions on Systems, Man and Cybernetics-part B, vol.39, no. 2, pp.539-550, 2009.
[19] YOU Mingyu, CHEN Yan, LI Guozheng, "Im-IG: a novel feature selection method for imbalanced problems"[J], Journal of Shandong University: Engineering Science, vol. 40, no .5, pp.123-128, 2010.
[20] WANG Lianxi, JIANG Shengyi, "Unsupervised feature selection method for categorical features"[J], Journal of Chinese Computer Systems, vol.2, no. 1, pp.47-50, 2011.
[21] HOFMANN T, "Unsupervised learning by probabilistic latent semantic analysis"[J], Journal of Machine Learning Research, vol.42, no.(1-2), pp.177-196, 2011.