Technology and Health Care 24 (2016) S665–S674
DOI 10.3233/THC-161194
IOS Press

The research on medical image classification algorithm based on PLSA-BOW model

C.H. Cao[a,b,c,*] and H.L. Cao[a,b,c]
[a] *College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning, China*
[b] *Key Laboratory of Medical Image Computing of Ministry of Education, Northeastern University, Shenyang, Liaoning, China*
[c] *State Key Lab. for Novel Software Technology, Nanjing University, Nanjing, Jiangsu, China*

Abstract.
BACKGROUND: With the rapid development of modern medical imaging technology, medical image classification has become more important for medical diagnosis and treatment.
OBJECTIVE: To solve the existence of polysemous words and synonyms problem, this study combines the word bag model with PLSA (Probabilistic Latent Semantic Analysis) and proposes the PLSA-BOW (Probabilistic Latent Semantic Analysis-Bag of Words) model.
METHODS: In this paper we introduce the bag of words model in text field to image field, and build the model of visual bag of words model.
RESULTS: The method enables the word bag model-based classification method to be further improved in accuracy.
CONCLUSIONS: The experimental results show that the PLSA-BOW model for medical image classification can lead to a more accurate classification.

Keywords: Medical image classification, bag of words model, PLSA

1. Introduction

When doctors visit medical imaging library, they often need to get a certain type of image for narrowing the scope of the search, which requires medical image classification. Medical image classification in early stage used manual methods, which means annotate the images to distinguish the categories of them, but with the massive growth of medical imaging, artificial text annotation is no longer feasible. Therefore, the automatic annotation of medical imaging has become an urgent demand. Traditional medical image classification is mostly based on the basic characteristics of the image, such as color features, texture features, shape features [1,2], which failed to solve the "Semantic gap" problem [3–5]. Because the basic characteristics of the image cannot reflect the underlying information in the images, for example the image may imply information of the specific organizational structure which cannot be obtained from the basic characteristics of the image and be a kind of potential information which only can be obtained by the doctors' experience.

*Corresponding author: C.H. Cao, College of Information Science and Engineering, Northeastern University, Shenyang 110819, Liaoning, China. Tel.: +86 15502406519; E-mail: caochunhong@ise.neu.edu.cn.

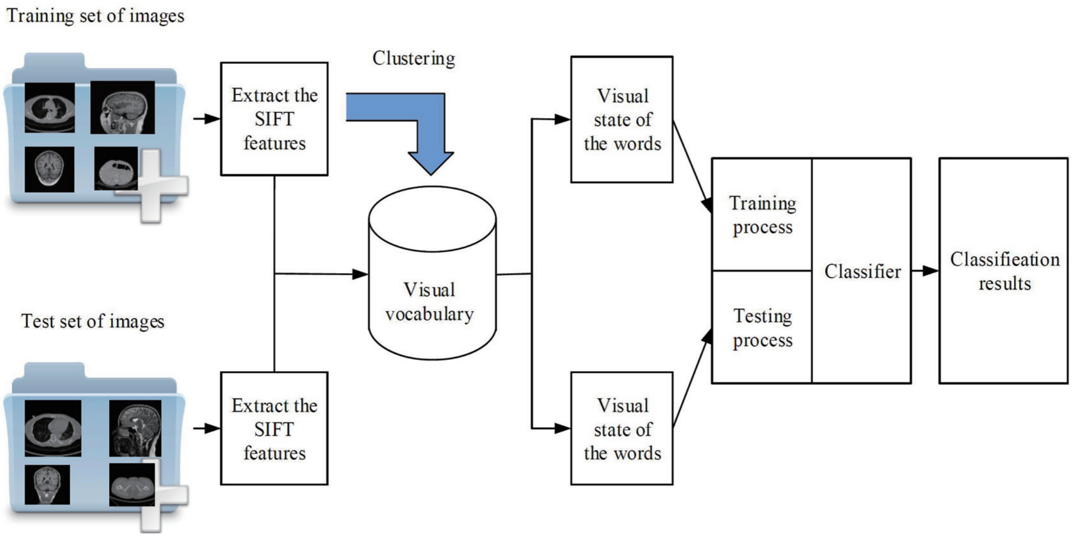

Fig. 1. Flow chart of the medical image classification based on the word bag model.

The bag model is introduced into the image field to build the model of visual word bag [6], so that the algorithm of text field is introduced into the image field. We use the model of visual word bag to construct the visual word expression, and then carry out the classification of medical images. In order to solve the existence of polysemous words and synonyms problem, we combine the word bag model with PLSA [7] and propose the PLSA-BOW model to solve the traditional word bag model semantic problem. It makes a semantic-based image classification achieved and the method further improved in accuracy.

This article is divided into four parts. The first part is introduction, the second part is the Classify the Medical Image with the Bag in the Model, the third part is the Medical Image Classification Algorithm – based on PLSA-BOW Model, the fourth part is the Experimental Results and Analysis of Medical Image Classification Algorithm – based on PLSA-BOW Model.

2. Classify the medical image with the bag model

The step of the medical image classification based on the bag model is approximately divided into: extracting local feature, constructing the visual vocabulary, training and testing the classifier. Process of classification is shown in Fig. 1.

Firstly, use SIFT to extract the local features of each image. Then extract all the SIFT features for clustering to form a visual vocabulary, which is equivalent to the dictionary in the text field. Secondly, use the vocabulary to represent each image as a vector, which means SIFT features in the image with comparing similarity of features in the dictionary and find the most similar features; the vector obtained is equivalent to a word frequency histogram. Finally, input the visual words of each image to the SVM for training the classification model. If it is necessary to test the accuracy of the classification model, transforming test image into the visual word said in the same way, and then inputting it to the classifier to test the effect.

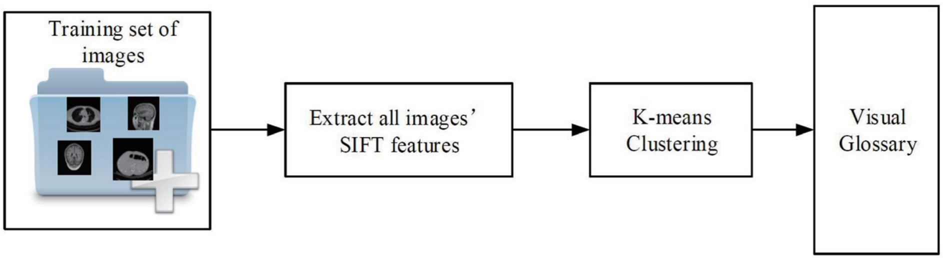

Fig. 2. The generation process of the visual vocabulary.

2.1. *Construct the visual vocabulary*

Construct the visual vocabulary (hereinafter dictionary), which is a set of significance words and can be imagined as more representative of local feature when corresponding to the image classification. In order to get the dictionary, firstly, extract the SIFT features of all images, and then cluster all the features to select the cluster centers as the word in the dictionary, the set of all words is the dictionary. The generation process of visual vocabulary shown in Fig. 2.

Specifically, the generation process of the dictionary as follows:

(1) Normalize all the training images of the medical images to 320*320 pixel size.
(2) Extract the SIFT descriptor from each training images.
(3) Cluster all the SIFT descriptors whose center constitute the dictionary with the k-means algorithm.

2.2. *Construct the visual word said*

Use the dictionary to represent each image as an n-dimensional vector, where n is the number of dictionary words, after the dictionary constructed. The more traditional method is a brute-force method, which means compare each feature in the image with all the words in the dictionary in distance (e.g. Euclidean distance) to find the closest visual word. Count the frequency of all the words appearing to construct images' the visual word said. Use the kd-tree algorithm to build the visual vocabulary's high-dimensional indexing, which used to construct images' the visual word said when each image calculates visual word said. The kd-tree algorithm [8] is a typical high-dimensional indexing technology. kd-tree use the divide and conquer thinking to divide the whole space into several smaller parts for efficiently finding the nearest neighbor. It is necessary to build the kd-tree and search data for using kd-tree to match the data.

3. Medical image classification algorithm based on PLSA-BOW model

3.1. *Basic principles of PLSA and the solution of synonyms and antonyms problem in image classification*

PLSA is a probabilistic latent semantic analysis model, whose variables except to the observable variable word $w \in W\{w_1, w_2, \ldots, w_m\}$, and document $d \in D\{d_1, d_2, \ldots, d_n\}$ is a implicit theme variable. PLSA regard a document as a set of theme and each theme in the document have their own weigh, so that a document can be represented as a vector whose dimensions are the weight of each theme in the document. If choose a document with the probability as $p(d)$, there are

$$P\left(d\right) = \sum_{i}^{k} P(z_i|d) \tag{1}$$

Where k is the theme number.

On the other hand, if the chosen theme is z, each word in the theme has form of probability which is

$$P\left(z\right) = \sum_{i}^{m} P(w_i|z) \tag{2}$$

Where m is the total number of words.

PLSA assume that the generation of each observation point (d, w) is dependent of each other, and they correlate through potential variable of a class theme; the generation of word w is independent on the document, and is dependent on potential variable of a class theme z.

Base on these two assumptions and the above two formulas, the generation of each observation point (d, w) can be completed through potential variable of a class theme z.

Choose document d with the probability as $P(d)$; the probability of each theme of the document can be sure since the document is chosen; the probability of word w is $P(w|z)$ in the condition of the theme is chosen, so that the generation of each observation point (d, w) can be completed.

N is the number of text in the document set. W_d is the number of word in each document. The above process can be represented by formula as follow:

$$P(d_i, w_j) = P(d_i)P(w_j|d_i) = P(d_i)\sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{3}$$

The thinking of the maximum likelihood estimation can be used to estimate $P(z|d)$ and $P(w|z)$.

The logarithmic form of likelihood function is as follows:

$$\begin{aligned} L &= \sum_{i=1}^{N}\sum_{j=1}^{M} n\left(d_i, w_j\right) logP\left(d_i, w_j\right) \\ &= \sum_{i=1}^{N} n(d_i)\left[\log P(d_i) + \sum_{j=1}^{M}\frac{n(d_i, w_j)}{n(d_i)}\log\sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i)\right] \end{aligned} \tag{4}$$

$n(d_i)$ is the sum of word frequency of all the word in the document d_i. $n(d)$ is not the model parameter, so that which have no effect on the model,

$$L \propto \sum_{i=1}^{N}\sum_{j=1}^{M} n(d_i, w_j) log \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{5}$$

Attention only should be paid on the effect $P(z|d)$ and $P(w|z)$ on the likelihood value. The maximum is difficult to seek when summation operations are integral, so that the EM algorithm can solve the problem [9].

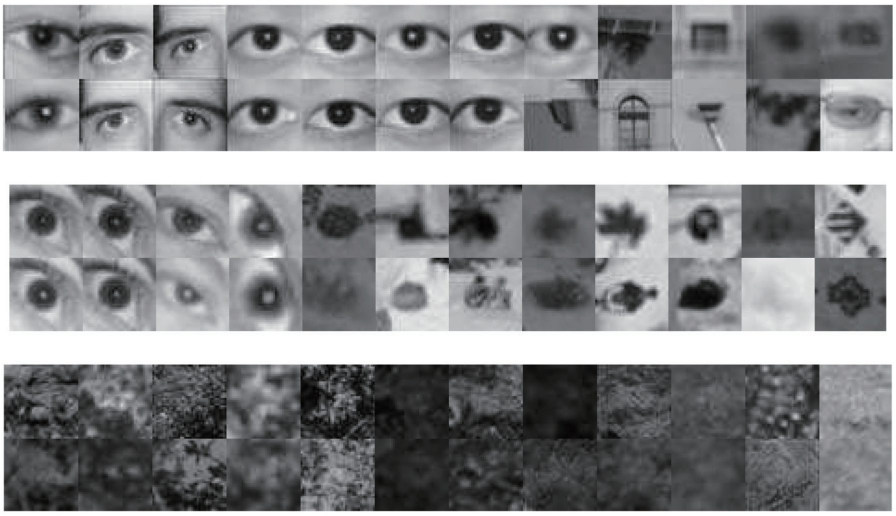

Fig. 3. The problem of synonyms and polysemy in the bag model.

PLSA aims to extract the theme of the text, after bring in the bag model in classification field, use PLSA to extract the theme of the image for removing the noise in the image and eliminating the problem of synonyms and polysemy, so that the accuracy of the classification can be improved.

In the image description, the problem of synonyms is that a number of visual words corresponding to the local area contain the same or similar information, and the problem of polysemy is that a visual word corresponding to the local area contains a variety of information. Pedro extracted three groups of visual words from the visual vocabulary to explain this phenomenon, shown in Fig. 3.

The first and second line is the visual word of eyes, which corresponds to the information on the local area. It is clear that most of the visual word is correct but some of them are actually windows, this is the problem of polysemy. The third and fourth line is also the visual word of eyes, which is the synonym of the first line.

The using of PLSA and the visual bag model can solve the problem of polysemy and synonym in a certain extent. Use PLSA to extract the theme of the image, and represent each image as the distribution of visual themes to classify. Images of the same category, whose visual word is different, but must share some common words, theme PLSA extracted can be regarded as the visual word co-occurrence information representation. Use PLSA to build the visual word co-occurrence information can ease the problem of synonyms and polysemy in the word bag model and reduce the redundant information of the image in the word bag model. Use PLSA to model "words – image" co-occurrence matrix of word bag model, abstract semantic distribution of the image, and map the underlying characteristics of low-level image to the middle-level semantic features, so use low-dimensional semantic to represent the image.

3.2. Classify the medical image with the PLSA-BOW model

Combining the bag model and the PLSA algorithm can classify the medical image, through extracting the visual themes of the medical image.

Firstly, all the training images of the medical images should be normalized to 320*320 pixel sizes. Secondly, use PLSA model to extract the visual semantic said of each image for inputting to the classifier to classify, basing on constructing the visual word said of the each image. Finally, the images which have not classified show, use PLSA model to extract the visual semantic said of them to input to the classifier to classify.

Images contain themes, constructing the visual word said is equivalent to establishing the mapping between the semantic features of the image and the underlying characteristics of such images, which means simulating the manner which human describe things to complete the classification of the images.

The process of classifying the medical image with the PLSA-Bow model can be divided into three steps, the first step, construct the visual word said of the each image; the second step, use PLSA model to extract the visual semantic said of each image; the third step, input to the classifier to classify.

3.2.1. Construct the visual word said of the image

Firstly, extract the visual semantic said of the training image. After the constructing of the visual word said of the training image, use the PLSA theme model to construct the visual theme said of them. In the process of above, it is necessary to set the number of topics T manually, after the generation process of the matrix $P(z|d)$ and the matrix $P(w|z)$ completed. $P(z|d)$ is the distribution of each document theme, which is the visual theme said of each image in the image classification, when the number of theme is decided to be T, the i-th row in the matrix $P(z|d)$ is $(p(z_1|d_i), p(z_2|d_i), \ldots, p(z_T|d_i))$, which is the theme said of the *i-th* image. For keeping the Probability of the normalization, each row of the matrix should be normalized.

Another matrix which trained by the PLSA model is a word theme matrix $P(w|z)$. $P(w|z)$ is each word's distribution in the different themes, which means the probability of each word generated by each theme and in the image field is the probability of each word generated by each visual theme. The aim of PLSA training is to obtain the matrix $P(w|z)$, which is equivalent to the knowledge of machine learning. Machine can judge once owned the knowledge.

The matrix $P(z|d)$ and the matrix $P(w|z)$ are obtained in the process of PLSA training. In the process of training, firstly construct of the visual word said of the training image, and then use the PLSA theme model to construct the visual theme said of them. Extracting the semantic said of training image is not simply using the PLSA algorithm through the EM iteratively to obtain the matrix $P(w|z)$. In the process of testing, combining the matrix $P(w|z)$ which is the knowledge obtained in the process of training, then each EM iteration keeping the matrix $P(w|z)$ unchanged and only updating the matrix $P(z|d, w)$ and the matrix $P(z|d)$. When meet the convergence conditions, the matrix $P(z|d)$ of the test image is obtained.

3.2.2. Classify the medical image with the visual semantic said

Obtaining the training image and the matrix $P(z|d)$ of the training image is equivalent to obtaining the theme said of the image and another abstract with the bag model. If the abstract of the image with the bag model is equivalent to a simple statistics frequency of each feature appears in the image and regard the image as a collection of many visual features, then abstract the visual word said with the PLSA is equivalent extracting some abstract semantic information, which means some specific information in the image, from the words.

4. Experimental results and analysis of medical image classification algorithm based on PLSA-BOW model

4.1. Experimental results

The Neusoft NSR medical imaging library is adopted as the experimental data, which include 12 categories, which are CT cross-sectional lung, CT cross-sectional abdominal, CT cross-sectional femur,

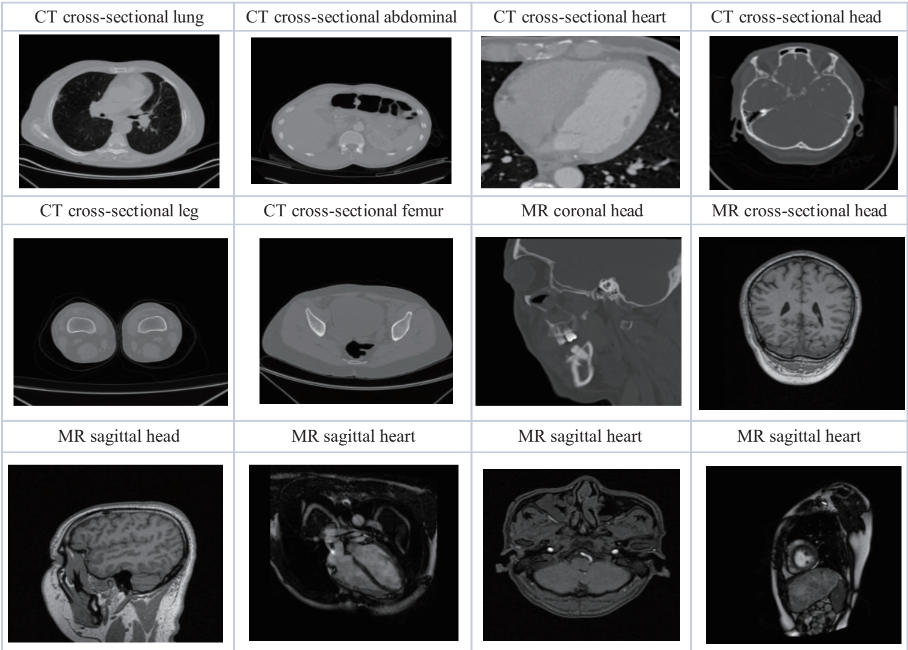

Fig. 4. NSR medical imaging set.

CT cross-sectional head, CT cross-sectional leg, CT cross-sectional heart, MR coronal head, MR cross-sectional head, MR cross-sectional heart, MR sagittal head, MR sagittal heart, shown in Fig. 4. Due to the less of CT cross-sectional femur and CT sagittal head images, excluding these two types of image.

Fix the experimental data in the experiment in order to ease the evaluation of the experiment. Choose 100 images from each type for training and another 30 images from each type for testing.

The main parameter of the experiment is the size of the visual vocabulary, which can be adjusted to test the accuracy in different circumstances. Use the vocabulary expressed each image as the visual word said, which should be inputted to the SVM for classification and prediction after the visual vocabulary constructed.

In the experiment of medical image classification algorithm based on PLSA-BOW model, the adjusting the size of the visual vocabulary can be used to test the accuracy under different circumstances. Tables 1, 2 and 3 have given the accuracy of the word bag model in the different visual vocabulary size.

It can be seen from Tables 1–3 that the accuracy of the classification can reach the highest 90% when the number of words is around 1000, at the same time it can be seen that the classification accuracy decrease when the number of words deviate from 1000. This is caused by the problem of synonyms and polysemy. When the number of words is large, some words represent the same information but are divided into multiple categories. When the number of words is small, the problem of polysemy may be caused, which means different category of words may be divided in the same category and make different words to represent the same information to low classification accuracy.

4.2. Compare with other classification methods

Compare the classification based on PLSA-BOW model with the classification based on gray level histogram method and the classification based on texture features, the results shown in Table 4.

Table 1
The classification accuracy comparison of the number of words is 200, 500 and 1000

Image type	200	500	1000
CT cross-sectional lung	0.6333	0.7333	0.9
CT cross-sectional abdominal	0.7333	0.8	0.93
CT cross-sectional head	0.7	0.8667	0.97
CT cross-sectional leg	0.8667	0.9	0.9
CT cross-sectional heart	0.8	0.8667	0.9667
MR coronal head	0.7667	0.8333	0.9333
MR cross-sectional head	0.4667	0.5333	0.5333
MR cross-sectional heart	0.8333	0.9	0.9333
MR sagittal head	0.7333	0.8667	0.9667
MR sagittal heart	0.8	0.9333	0.9667
Total classification accuracy	73.33%	82%	90%

Table 2
The classification accuracy comparison of the number of words is 1500, 2000 and 2500

Image type	1500	2000	2500
CT cross-sectional lung	0.93	0.8667	0.8667
CT cross-sectional abdominal	0.8	0.9	0.8333
CT cross-sectional head	0.96	0.9	0.8667
CT cross-sectional leg	0.8667	0.8	0.8667
CT cross-sectional heart	0.9333	0.9333	0.9
MR coronal head	0.9	0.9667	0.8333
MR cross-sectional head	0.5333	0.5333	0.4667
MR cross-sectional heart	0.8667	0.9667	0.8333
MR sagittal head	0.8	0.9	0.8
MR sagittal heart	0.9333	0.9	0.9333
Total classification accuracy	85.33%	86.67%	82.00%

Table 3
The classification accuracy comparison of the number of words is 3000, 3500 and 4000

Image type	3000	3500	4000
CT cross-sectional lung	0.9	0.8667	0.7667
CT cross-sectional abdominal	0.8	0.3333	0.33
CT cross-sectional head	0.8	0.7667	0.7
CT cross-sectional leg	0.9	0.9	0.8
CT cross-sectional heart	0.9	0.8667	0.8667
MR coronal head	0.8333	0.9333	0.9
MR cross-sectional head	0.3	0	0
MR cross-sectional heart	0.8333	0.9	0.8333
MR sagittal head	0.8667	0.9	0.8667
MR sagittal heart	0.9	0.8667	0.8
Total classification accuracy	80.33%	73.33%	68.33%

From the Fig. 4 we can see that the optimal classification accuracy of the classification based on gray level histogram method is 67.9%, the optimal classification accuracy of the classification based on texture features is much higher than the optimal classification accuracy of the classification based on gray level histogram method, which is 83.67%, and the optimal classification accuracy of the classification based on PLSA-BOW model is higher than the optimal classification accuracy of the optimal classification accuracy of the classification based on texture features, which is 90%. This shows that the classification accuracy of the classification model based on the word bag model is far higher than the classification accuracy of two methods based on global features.

Table 4
The optimal classification accuracy comparison of histogram, texture features, and the word bag model

Image type	Based on gray level histogram method	Based on texture features	Based on PLSA-BOW
CT cross-sectional lung	0.43	0.5667	0.9
CT cross-sectional abdominal	0.77	0.3333	0.93
CT cross-sectional head	0.93	0.9667	0.97
CT cross-sectional leg	1	1	0.9
CT cross-sectional heart	0.5	1	0.9667
MR coronal head	0.9	0.9333	0.9333
MR cross-sectional head	0.1	0.8333	0.5333
MR cross-sectional heart	1	1	0.9333
MR sagittal head	0.3	0.7333	0.9667
MR sagittal heart	1	1	0.9667
Total classification accuracy	69.70%	83.67%	90%

In this paper, we use the word bag model classifying in neusoft NSR image set, and local characteristics use the SIFT descriptor. Experiment from images pulled out 10 classes, in each kind of 100 images to the training of the classifier is put forward, in addition to take out the 30 to be tested. The experimental results show that the model based on word bag much higher classification accuracy of classification method based on gray histogram and texture characteristics of precision. After many experiments, the two kinds of algorithm to select the optimal parameters, based on global feature classification accuracy of 69.7%, based on texture feature classification accuracy of 83.67%, and the word bag model accuracy can reach 90%.

Experimental results show that the word bag model, on the basis of combination of PLSA algorithm constructs the image semantic representation, medical image classification, application effect than single word bag model. Corresponding to each kind of vocabulary in the experiment, testing the different theme values influence on classification accuracy, the results show that the theme of each case there is a optimal number, and at the time of other topics with little deviation, classification accuracy is still higher than the single word bag used the classification precision of the model. PLSA algorithm in the case of the optimal number of words, can solve the problem of bag model polysemy and synonyms, which can improve the classification accuracy.

5. Conclusions

The bag model is introduced into image classification field, to classify the medical image with the SIFT features, and then introduce the PLSA model into the experiment, which means to construct the visual semantic said based on the visual word said. The experimental results show some improvement of the medical image classification based on the PLSA-BOW model as compared with the traditional classification methods.

The LDA algorithm can be considered instead of the PLSA algorithm in the future, which can be combined with the word bag model to classify. LDA is a more perfect model of the themes, which can solve the problem of over-fitting of the PLSA model.

Acknowledgments

This work is supported by the National Nature Science Foundation of China (No. 61300096); this work is supported by the Fundarnental Research Funds for the Central Universities (N130404013).

References

[1] M. Szummer and R.W. Picard, Indoor-outdoor image classification, *Content-Based Access of Image and Video Database* **42** (1998).

[2] S.L. Wang, Information-based color feature representation for image classification, *Image Processing* **6** (2007), 353.

[3] J. Yang, Narrowing semantic gap in content-based image retrieval, *Computer Distributed Control and Intelligent Environmental Monitoring* **433** (2012).

[4] S.Z. Deng, M. Islam and G.J. Lu, A review on automatic image annotation techniques, *Pattern Recognition* **45** (2012), 346.

[5] Z.J. Zha, D.C. Tao and T.-S. Chua, Semantic-gap-oriented active learning for multipliable image annotation, *IEEE Transaction on Image Processing* **21** (2012), 2354.

[6] F.F. Li and P. Perona, A Bayesian hierarchical model for learning natural scene categories, *Computer Vision and Pattern Recognition* **2** (2005), 524.

[7] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, *Machine Learning* **42** (2001), 177.

[8] J.L. Bentley, Multidimensional binary search trees used for associative searching, *Communication of the ACM* **18** (1975), 509.

[9] C.M. Bishop, Pattern recognition and machine learning, *Corr* (2007), 424.