Image Classification via Semi-Supervised pLSA

Liansheng ZHUANG, Lanbo SHE, Yuning JIANG, Ketan TANG,Nenghai YU

MOE-MS Key Laboratory of Multimedia Computing and Communication,
University of Science and Technology of China, Hefei, 230027, P.R.China

Email:{lszhuang@ustc.edu.cn}

Abstract

In this paper, we propose Semi-Supervised pLSA(SS-pLSA) for image classification. Compared with the classic non-supervised pLSA, our method overcomes the shortcoming of poor classification performance when the features of two categories are quite similar. By introducing category label information into EM algorithm, the iteration process can be directed carefully to the desired result. SS-pLSA greatly prevents the inter-impact between different categories. The experiment results show that the proposed SS-pLSA significantly improves the performance of image classification, especially when two categories' features are similar and difficult to distinguish by classic pLSA. In contrast to these totally supervised algorithm, SS-pLSA almost has no loss in detection rate while sharply reduces the difficulty of collecting training samples. With highly flexibility, SS-pLSA enables users to explore the trade-off between labeled number and accuracy.

1 Introduction

Object classification is to predict the class label of unknown data with classifier obtained from experiential data , which is a basic problem in pattern recognition, machine learning and statistics, as well as in data mining. Learning the classifier is the key problem of object classification. Probabilistic latent semantic analysis (pLSA) is a well-known algorithm to solve the problem [4–6, 10], which is first applied in the field of text retrieval and document classification. In recent years, Sivic and some other researchers introduced pLSA algorithm into the field of computer vision and applied it to the image classification [6, 11]. They treated images as documents, and represented each image as a single histogram of visual words. A visual word is an equivalent notion of a text word, and defined by clustering various local descriptors, such as SIFT [9].

The classic pLSA adopts non-supervised learning style, which doesn't need any prior information. The lack of useful information leads to some serious problems, such as slow convergence and decreasing of classification accuracy. A lot of researchers try to solve this problem, such as [8]. In [8], the authors combine attention model and classic pLSA to create AM-pLSA. They use attention model to alleviate the background's impact to object. However, this kinds of algorithm just add a pre-processing to pLSA. They didn't change the essence of pLSA.

One thing is for sure, that what we really care about is just the dominant topic in an image and the less important ones are regarded as background. Also it is true that the context of images is much more complex than texts. For example, faces always coexist with indoor background, so that it's difficult to distinguish indoor face images and office images. In order to get the distribution on all possible topics of one image, including the less important ones, the classic pLSA takes quite a lot of iterations. These redundant iterations caused by the lack of directionality not only waste plenty of time, but also make the classifier sensitive to less important topics, especially when the images contain closely related topics.

According to the analysis above, an advanced pLSA algorithm in semi-supervised mode is proposed. By adding the label information of training sample into iteration process, more accurate parameters are achieved to calibrate the classifier. SS-pLSA inherits the advantages of both non-supervised pLSA and totally supervised pLSA algorithm (S-pLSA), since more effective classifier is obtained while we only need to label a few samples (in fact, the labeled number can be set discretionarily). The experiment results show that SS-pLSA significantly reduces the redundant iterations and increases the detection rate on these images which are usually difficult to classic pLSA.

The rest of this paper is organized as follows: Section 2 briefly describes the classic pLSA and makes a detail explanation of our SS-pLSA. Experiments and result analysis are presented in Section 3 and a conclusion is made in Section 4.

2 Image classification via Semi-Supervised pLSA

In this section we will detailedly introduce our Semi-Supervised pLSA algorithm. Before that we shall make a brief explanation of classic pLSA algorithm.

2.1 Classic pLSA algorithm

Classic pLSA is a semantic model. The model assumes that, under the texts or images we observed, there is another implicit level: the topic level. An article(document) has a certain probability distribution on a variety of topics, and similarly, topics also have different distribution on a set of items(words).

$$P(d,w) = P(d)P(w|d), \tag{1}$$

$$P(w|d) = \sum_{z \in Z} P(w|z)P(z|d), \tag{2}$$

$$P(d,w) = \sum_{z \in Z} P(z)P(d|z)P(w|z), \tag{3}$$

where d, w, z represents document, word and topic respectively.

Model parameters $P(z)$, $P(d|z)$, $P(w|z)$ can be calculated by the classic EM algorithm:

E step(Expectation):

$$P(z|d,w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z'} P(z')P(d|z')P(w|z')}. \tag{4}$$

M step(Maximization):

$$P(w|z) = \frac{\sum_d n(d,w)P(z|d,w)}{\sum_{d,w'} n(d,w')P(z|d,w')}, \tag{5}$$

$$P(d|z) = \frac{\sum_w n(d,w)P(z|d,w)}{\sum_{d',w} n(d',w)P(z|d',w)}, \tag{6}$$

$$P(z) = \frac{1}{R} \sum_{d,w} n(d,w)P(z|d,w), \tag{7}$$

$$R = \sum_{d,w} n(d,w). \tag{8}$$

Iteratively perform E step and M step until the probability values are stable.

In EM algorithm the parameters $P(z)$, $P(d|z)$, $P(w|z)$ are randomly initialized. As a result, the iteration process will increase the volatility of the parameters and impact the final classification accuracy.

2.2 Semi-Supervised pLSA

To solve the problems described above, we proposed Semi-Supervised pLSA. Firstly, we extract the feature points and feature description from the training images. In this article, we use SIFT descriptor. And then we calculate the model parameters by our modified EM algorithm, in which category labels are added as additional information to direct the iteration process.

The biggest difference between SS-pLSA and pLSA is that the category labels of some training image are introduced into the EM iteration process. At the beginning of training, we assume that part of the images' dominant topics are provided, so that the corresponding terms of label vector lv are set to a specific value called topic label. For example, if the first 5 images of 100 training images are labeled topic No.1, then

$$lv(1..5) = 1, lv(6..100) = 0, \tag{9}$$

where 1 means sample labeled topic No.1 and 0 means unlabeled sample. In the iteration process, a threshold will be adopted to determine the image's dominant categorization. When the probability of some images on one dominant topic is greater than the threshold, the image will be marked with the category label and the probability will be set a certain large value. New affirmed labels will continually be added in as the iteration process goes on. Category labels guarantee SS-pLSA to get more accurate model parameters quickly. Therefore the classification accuracy of SS-pLSA is much higher than pLSA. The detail of iteration process is described as follows.

Firstly, perform the E step and M step of EM algorithm as described above.

Secondly, Renew the label information. Calculate the $P(z|d)$ from $P(d|z)$ and $P(z)$:

$$P(z|d) = \frac{P(d|z)P(z)}{\sum_z P(d|z)P(z)}, \tag{10}$$

Traverse all the pictures, when find an image's distribution on a particular topic reaches a threshold (0.8, for example) for the first time, we can say that we find the image's topic. The corresponding label vector term will be changed from zero to the specific category label number:

$$lv(i) = a, \text{ if } P(z_a|d_i) > 0.8 \text{ and } lv(i) = 0. \tag{11}$$

Once the label vector term being set, it won't change any more.

Thirdly, update $P(d|z)$. Traverse the label vector and record the already labeled images' indexes and category they belong to. Then, set the corresponding part of $P(z|d)$ a certain larger number, such as 0.9:

$$P(z_a|d_i) = 0.9, \text{ if } lv(i) = a. \tag{12}$$

At last, take the formulation below to renew $P(d|z)$:

$$P(d|z) = \frac{P(z|d) \sum_z P(d|z)P(z)}{P(z)}. \tag{13}$$

Go back to the E step and go on with the iteration until all the parameters are learned.

From the above steps we can see that the adding in of label information weakens the inter-impact of different topics in an image so that the iteration process is more directional. Moreover, the number of input category labels is not fixed, that is, supervision amount can be adjusted easily by adjusting the input label vector, so that the SS-pLSA is very flexible. When needed, user can set the entire label vector to corresponding category labels at the training stage and SS-pLSA algorithm becomes totally supervised pLSA, named Supervised pLSA (S-pLSA). SS-pLSA inherits the advantages of both supervised and non-supervised algorithm, so that it keeps a balance between the difficulty of sample collection and classifier performance according to the actual situation.

2.3 Image classification

For image classification, we need to extract visual words from images, and then all the procedures are the same as text classification. The following is our algorithm framework.

Table 1. Algorithm framework of image classification via SS-pLSA

Input: a. training images
   b. category numbers
   c. label information of part images

Output: image classifier

Procedure:
   1. Find interest points of images
   2. Extract visual words from interest points
   3. Label part of training images as label vector
   4. Run modified SS-pLSA algorithm to learn model parameters

3 Experiments

In this section, we conduct Supervised pLSA (S-pLSA), SS-pLSA and classic pLSA on Caltech-101 dataset. Six classes (airplane, face, leopard, motorbike[1], chelonian, bonsai) are selected as positive samples and background_caltech [1] as negative sample. We choose these categories because they have relatively complex background

---

[1]We select the motorbike pictures with complex background because pure motorbike pictures are too easy to classified. The classification accuracy is always 100%

Figure 1. Sample images of our dataset.

and thus the strength of our model can be better demonstrated. Some typical examples are shown in Fig.1.

To validate our method, we compare the binary classification performance of pLSA, SS-pLSA and S-pLSA. In our experiments, we extract SIFT feature as local feature, and use k-means clustering algorithm to quantize the SIFT descriptor as visual words. The codebook size is 300. To keep the authenticity, all experiments are repeated 10 times. The final result is the average of 10 runs. For each run, we randomly select 50% of the images from each category as our training set, and the rest as testing dataset. For SS-pLSA, 10% of training images were marked category label. And for S-pLSA, all training images were marked category label. Table 1 shows classification accuracy of six object classifiers. And Fig.2 is ROC curve of face & background_caltech binary classifier.

Table 2. Classification accuracy of pLSA, SS-pLSA, S-pLSA

categories	pLSA	SS-pLSA	S-pLSA
airplane	0.814	0.944	0.980
face	0.792	0.878	0.915
motorbike	0.780	0.866	0.920
bonsai	0.906	0.922	0.980
chelonian	0.934	0.938	0.980
leopard	0.918	0.918	0.940

As can be seen from Table 1, S-pLSA and SS-pLSA outperform classic pLSA in all six cases. This proved that the iteration process with additional label information can effectively help the classifier to find suitable probability parameters and improve the detection rate. On the other hand, compared with S-pLSA, SS-pLSA almost had no loss in detection rate while significantly reduced the amount of la-
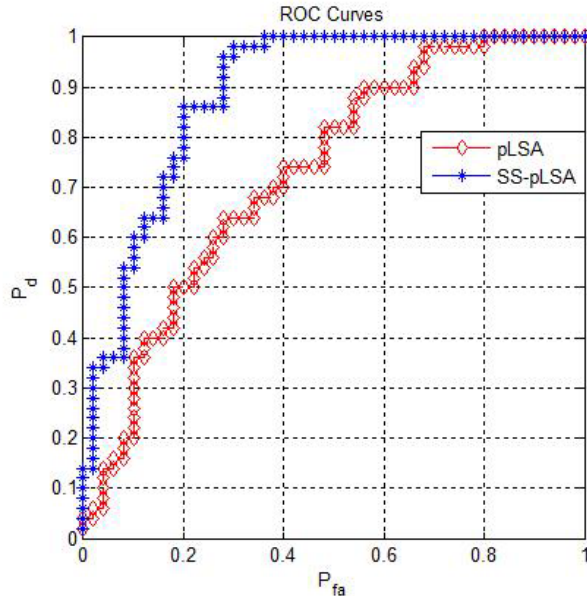
Figure 2. ROC curve of face & background_caltech classifier.

beled images. In particular, SS-pLSA algorithm can greatly improve the classifier's performance in category "airplane" and "face", compared with classic pLSA. It is because in these two cases the features of the two categories are quite similar. In this situation, pLSA cannot separate the two categories successfully. However, because SS-pLSA algorithm has additional labeling information, it can distinguish the subtle difference of the two categories' features.

We also compared S-pLSA with other supervised classification algorithms, such as SVM [2, 3, 7]. Our algorithm reaches almost the same performance as SVM on most cases. Furthermore, our algorithm is much easier to understand and implement. This proves that our algorithm is not only flexible, but also powerful, which ensures that our SS-pLSA can be used in practical situations.

4 Conclusion

In this paper, we propose Semi-Supervised pLSA(SS-pLSA) on image classification aimed to improve the performance of classic pLSA. By introducing category information into EM iteration, the process can be carefully directed, which greatly prevents the inter-impact between different categories. The experiment results show us that, SS-pLSA not only reduces the times of iteration, but also significantly improves the performance of image classification, especially when two categories' features are similar.

Furthermore, compared with these rigid supervised algorithm, our proposed SS-pLSA has great flexibility. Users

can set all the training images corresponding category labels and make SS-pLSA become totally supervised, S-pLSA. The amount of supervision can be easily adjusted depending on the situation and users' demand.

References

[1] http://people.csail.mit.edu/fergus/iccv2005/bag_words_demo.tgz.

[2] R. F. E. Osuna and F. Girosi. Support vector machines: Training and applications. *A.I. Memo, MIT A.I.Lab.*, 1996.

[3] R. F. E. Osuna and F. Girosi. Training support vector machines: An application to face detection. *CVPR*, 1997.

[4] T. Hofmann. Probabilistic latent semantic indexing. *In SIGIR*, 1999.

[5] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 43:177–196, 2001.

[6] A. E. A. Z. W. F. J. Sivic, B.C. Russell. Discovering objects and their location in images. *IEEE International Conference on Computer Vision (ICCV2005),*, Vol.1:370–377, 2005.

[7] T. Joachims. Text categorization with support vector machines. *European Conference on Machine Learning (ECML)*, 1998.

[8] N. Y. W. Z. Liansheng Zhuang, Ketan Tang. Unsupervised object learning with am-plsa. *CSIE*, 2009.

[9] D. Lowe. Object recognition from local scale-invariant features. *IEEE International Conference on Computer Vision (ICCV1999)*, 2:1150–1157, 1999.

[10] P. P. A. Z. R. Fergus, L. Fei-Fei. Learning object categories from googles image search. *IEEE International Conference on Computer Vision (ICCV2005)*, Vol.2:1816–1823, 2005.

[11] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. *IEEE International Conference on Computer Vision (ICCV2003),*, Vol.2:1470–1477, 2003.