Improved Dictionary Learning with Enriched Information for Biomedical Images

Shengda Luo and Alex Po Leung

Macau University of Science and Technology, Taipa, Macau, China
pleung@must.edu.mo

Abstract. With dictionary learning using k-means or k-means++, the optimal value of k is traditionally determined empirically using a validation set. The optimal k, which should depend on the particular problem, is chosen with previously determined values from prior work. We argue that there is rich information from clustering with a number of values of k. We propose a novel method to extract information from clustering with all reasonable values of k at the same time. It is shown that our method improves the performance of dictionary learning for the popular bag-of-features model in image classification with simple patterns like cells such as biomedical images. Our experiments demonstrate that, our proposed dictionary learning method outperforms popular methods, on two well-known datasets by 12.5% and 8.5% compared to k-means/k-means++ dictionary learning and by 8.9% and 6.1% compared to sparse coding.

1 Introduction

Many methods for the popular bag-of-features model (BoF) have been proposed in recent years [4], [7], [8]. On images with simple patterns, there is not enough information to be extracted by dictionary learning. There are much fewer visual words to be learned from the simple patterns of visual objects which severely limit the descriptive power of the visual words. The lack of sufficient resolution or descriptive power of object patterns makes it very hard for the classifier to classify the input images very accurately. Recently, a dictionary learning method in [11] is proposed to deal with this problem by extracting more information dictionary learning. The extra information is generated by the clustering results with different values of k using the k-means++ algorithm. However, the high-dimensional feature vectors used in the method [11] makes the curse of dimensionality a much bigger problem with performance degradation.

In this work, it is shown that our dictionary learning method significantly improves the accuracy of the BoF model for classifying the low-resolution images with simple patterns. The improvement is achieved by bringing in the extra information obtained through dictionary learning. In this work, there are three variants of our proposed method using the multi-channel kernel for the SVM, concatenation of features, and random sampling for k with clustering.

The variant of the proposed method using the multi-channel kernel is called Enriched Dictionary Learning with the Multiple-Channel Kernel (ED-M). The variant using concatenation of features is called as Enriched Dictionary Learning by Concatenation (ED-C). The computational efficiency of the ED-M method

can be improved using random sampling for different values of k for k-means++. We call this more efficient ED-M method Enriched Dictionary Learning with the Multiple-Channel Kernel and Random Sampling (ED-MRS). In the experiment section, the BoF model using our three proposed algorithms (ED-M, ED-C, and ED-MRS) are compared with the model using previous state-of-the-art dictionary learning methods (see Table 1). Experiments show that the best result can be obtained by the BoF model using our proposed dictionary learning method.

The differences between our three methods and the method [4] are as follows: (a) For concatenation, we combine the histograms of images in our ED-C method instead of combining clustering results in the method [4]. It has been shown in our experiments that this method for concatenation performs much better than the method in [4], and (b) In our methods ED-M and ED-MRS, extra information is added to the BoF model using the multi-channel kernel (again only concatenation is used in [4]).

Table 1. This table summarizes the dictionary learning methods that be compared in Section 4. The SVM kernels of the BoF model using them are shown in the rightmost column.

Dictionary learning method	SVM kernel
Proposed ED-M method	Multi-channel kernel
Proposed ED-C method	Chi-square kernel
Proposed ED-MRS method	Multi-channel kernel
K-means++	Chi-square kernel
K-means	Chi-square kernel
Method in [11]	Chi-square kernel
SIFT Sparse Coding [10]	Kernel in [10]

We divide the remainder of this paper into 4 sections. In Section 2, a brief introduction of the related work is given. Our dictionary learning methods are described in detail in Section 3. In Section 4, with the BoF model, our dictionary learning method and all its variants are evaluated on two datasets by comparing them with different dictionary learning methods. The last section concludes our work.

2   Related Work

The BoF model is one of the most popular methods for image classification. In this paper, we focus on the dictionary learning methods for this model.

The k-means algorithm became one of the most widely used dictionary learning methods after it was proposed to generate a dictionary in [2]. There are three main issues in the standard k-means algorithm. The efficiency of the k-means algorithm is lower when the value of k or the number of feature descriptors is very big. The result of the k-means algorithm is easily influenced by the positions of its initial centers. Finally, the noise in the image can affect the result of the k-means algorithm because the algorithm treats each descriptor equally.

Some algorithms are proposed to improve the standard k-means algorithm and the BoF model. The Simple Random Sampling K-means (SRS-K) Algorithm [1] is proposed to improve the efficiency of the k-means algorithm by reducing the number of descriptors. In order to generate the initial centers of the k-means algorithm more reasonably, the k-means++ algorithm with a proven statistical guarantee is used to instead of the k-means algorithm. In addition,

the BISecting K-means (BIS-K) Algorithm [12] is proposed to address the same issue. The Spatial Pyramid Matching (SPM) [5] Method is one of the most commonly used variants of the BoF model. In order to improve the efficiency of the SPM Method, the ScSPM Method [10] is proposed. In the ScSPM Method, the sparse coding of the SIFT descriptors is instead of the k-means algorithm as the method for dictionary learning. Although these methods are meaningful, they neglect the information obtained from the clustering results of the k-means algorithms with different values of k. In [11], a dictionary learning method is proposed to improve the accuracy of the BoF model for image classification. It is a also method proposed to exploit extra information obtained from clustering. However, the method makes the curse of dimensionality a bigger problem when k for k-means++ is relatively large. We compare our proposed method with this previous method in our experiments.

3 The Proposed Dictionary Learning Method

3.1 The Motivation

What motivated us is the question: can a single empirically determined optimal k for k-means or k-means++ be improved for dictionary learning? The disadvantages to use an empirical k can be 1.) rich information might be lost if the empirical k is not carefully chosen by extensive experiments and 2.) sometimes, the optimal k, which should depend on the particular problem, is chosen with previously determined values from prior work or past experience. We, therefore, aim to extract rich information from various histograms using clustering with a large number of possible values of k.

In our experiments, it is demonstrated that our method significantly improves the performance of dictionary learning over state-of-the-art methods for the popular bag-of-features (BoF) model in image classification with biomedical images. However, the method is a general method to improve dictionary learning for images with simple patterns.

With enriched information from dictionary learning, we focus on datasets with low-resolution images with simple patterns for classification using BoF model. Traditionally, the performance of the BoF model with these images may not be ideal because there is not enough information to be obtained from such images using dictionary learning with simple patterns. It is our goal that extra information can be extracted to obtain an enriched dictionary for the BoF model.

We propose a novel dictionary learning method to learn the dictionary using extra information. Using our method, the accuracy of the BoF model can be improved significantly. It is unnecessary for the new method to obtain a very high-dimensional feature vector in the large ($\sum_{n=2}^{k} n$-word) dictionary which makes the curse of dimensionality a bigger problem. In our method, there are three variants of our dictionary learning method: the ED-M, the ED-C, and the ED-MRS (see Table 1).

3.2 Notations

To be more precise, notations used in the rest of this paper are defined in this subsection. N is defined as the number of the images used for training. S_j is the

set of SIFT descriptors extracted from the j-th image, and $\mathbb{S} = \{S_1 \cup ... \cup S_N\}$ is the set of all SIFT descriptors from all N training images. To extract enriched information from various histograms with multiple k's using k-means++, we have the set of all k's used, $k \in \mathcal{K} = \{2, 3, ..., K\}$. For clustering, c_k is the set of k cluster centers obtained with $k \in \mathcal{K}$ using k-means++. We have the set of all histograms obtained with k-means++ $h_{k \in \mathcal{K}, n \in \{1,2,...,N\}} \in \mathcal{H}$ where $h_{k,n}$ indicates the histogram generated for the n-th image by using the k-th clustering result.

3.3 Feature Extraction for Enriched Information

Our proposed enriched information step is acheived using Algorithm 1. With k-means++ algorithm for clustering, it has been shown that the performance is better due to its statistical guarantee. In Algorithm 1, for each $k \in \mathcal{K}$, we have one iteration with clustering and the associated histograms. The k-means++ is, therefore, performed $|\mathcal{K}|$ times with k values in $\mathcal{K}$. $|\mathcal{K}|$ sets of clustering centers are then used to obtain $|\mathcal{K}|$ sets of histograms for every single image.

Algorithm 1 Feature Extraction for Enriched Information

Require: The SIFT descriptors of features in training images $\mathbb{S}$; The set of values $k \in \mathcal{K}$ for clustering with k-means++

Ensure: The set of all histograms of all training images $\mathcal{H}$.

1: **for** each value $k \in \mathcal{K}$ **do**
2:        Cluster all features $\mathbb{S}$ using k-means++ with k to obtain the set of cluster centers c_k
3:        **for** each image $j \in 1, 2, ..., N$ in the training set **do**
4:            Bin S_j for the histogram: $h_{i,j}$ with centers c_k
5:        **end for**
6: **end for**

The curse-of-dimensionality which leads to poor performance is avoided in our dictionary learning method with the multi-channel kernel considering that we can have a very-high dimensional feature vector from the $\sum_{n=2}^{k} n$-word dictionary. The kernel makes the feature vectors significantly shorter. To be more specific, in our Algorithm 1, for each image one histogram is generated using the cluster centers from k-means++ with one particular k, so the number of bins is exactly k in this histogram.

3.4 Enriched Dictionary Learning with the Multiple-Channel Kernel (ED-M)

For classification, we propose the ED-M algorithm after feature extraction (see Table 1) is shown in Algorithm 2. The ED-M utilizes the multi-channel kernel of the SVM to combine features obtained in Algorithm 1 for the classification in the BoF model.

With $\mathcal{H}$ from Algorithm 1, the multi-channel chi-square kernel is calcuated using histograms $\mathcal{H}$. For each channel k, we have histogram $h_{k,j} \in \mathcal{H}$ where j can be any image in the training set, i.e. $1 < j < N$. The chi-square kernel between two histograms from two images j_1 and j_2 in channel k is obtained by

$$\chi^2(h_{k,j_1}, h_{k,j_2}) = 1 - \sum_{z=1}^{k} \frac{\left[(h_{k,j_1})_z - (h_{k,j_2})_z\right]^2}{\frac{1}{2}\left[(h_{k,j_1})_z + (h_{k,j_2})_z\right]} \tag{1}$$

where z is a particular bin in histogram $(h_{k,j})$ and $(h_{k,j})_z$ is the value of bin z

in $h_{k,j}$.

After computing the multi-channel chi-square matrix, the SVM is trained with the matrix to obtain the support vectors. Each element in the matrix M_{j_1,j_2} of this matrix is computed using

$$M_{j_1,j_2} = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \chi^2(h_{k,j_1}, h_{k,j_2}) \tag{2}$$

where $\chi^2(.,.)$ is the chi-square kernel in Equation 1.

Algorithm 2 Enriched Dictionary Learning with the Multiple-Channel Kernel (ED-M).

Require: The SIFT descriptors of features in training images $\mathbb{S}$; The set of values $k \in \mathcal{K}$ for clustering with k-means++

Ensure: The Trained Model of the SVM, i.e. the support vectors.

1: Obtain all histograms for all training image $\mathcal{H}$ by invoking Algorithm 1

2: With $\mathcal{H}$ as input, the chi-square kernel matrix is computed for each channel (Equation 1)

$$\chi^2(h_{k,j_1}, h_{k,j_2}) = 1 - \sum_{z=1}^{k} \frac{[(h_{k,j_1})z - (h_{k,j_2})z]^2}{\frac{1}{2}[(h_{k,j_1})z + (h_{k,j_2})z]}$$

3: The matrix of the multi-channel kernel is computed (Equation 2)

$$M_{j_1,j_2} = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \chi^2(h_{k,j_1}, h_{k,j_2})$$

4: Train the SVM by using the matrix of the multi-channel kernel

---

3.5 Enriched Dictionary with the Multiple-Channel Kernel and Random Sampling (ED-MRS)

To improve computational efficiency, we propose a randomized version of the ED-M method using random sampling. The new algorithm is called Enriched Dictionary with the Multiple-Channel Kernel and Random Sampling (ED-MRS). Essentially, a random subset $\mathcal{K}_{MRS}$ from $\mathcal{K}$ is obtained before clustering. That means we have $\mathcal{K}_{MRS} \subset \mathcal{K}$. In our experiments, it is demonstrated that a significant speedup can be acheived with ED-MRS with some reasonable performance decrease.

In the ED-MRS algorithm, $\mathcal{K}_{MRS} \subset \mathcal{K}$ is obtained using random sampling. $\mathcal{K}_{MRS}$ is then used as the set of the k values for the k-means++ instead of $\mathcal{K}$. It means that the only difference between the ED-MRS algorithm and the ED-M algorithm, i.e. Algorithm 2 is the use of randomly sampled values of $k \in \mathcal{K}_{MRS}$ for ED-MRS. More specifically, $|\mathcal{K}_{MRS}|$ is used in ED-MRS, instead of $|\mathcal{K}|$ in the Equation 2.

3.6 Enriched Dictionary Learning with Concatenation (ED-C)

Instead of using the multi-channel kernel for different values of k, it is natural to also consider a simple concatenation for the proposed method with the chi-square kernel. Therefore, we have enriched dictionary learning with concatenation (ED-C) which combines enriched information from clustering to learn the dictionary (see Algorithm 3 and Table 1).

Like ED-M, in Algorithm 3, $\mathbb{S}$, and $\mathcal{K}$ are needed as the input to Algorithm 1 which computes $\mathcal{H}$. With histograms $\mathcal{H}$, for the same image j, $\{h_{k,j} : k \in \mathcal{K}\}$ can be concatenated to form one feature vector. With only one feature vector for one image, the SVM is then trained using the chi-square kernel.

Algorithm 3 Enriched Dictionary Learning with Concatenation (ED-C)

Require: The SIFT descriptors of features in training images $\mathbb{S}$; The set of values $k \in \mathcal{K}$ for clustering with k-means++

Ensure: The Trained Model of the SVM, i.e. the support vectors.

1: Obtain all histograms for all training image H by invoking Algorithm 1

2: Concatenate the histograms with different values of k to obtain one feature vector for image j, i.e. $H_j = [h_{2,j}, h_{3,j}, ..., h_{k,j}]$

3: Computing the chi-square kernel matrix (Equation 1)

$$\chi^2(H_{j_1}, H_{j_2}) = 1 - \sum_{z=1}^{i} \frac{[(H_{j_1})_z - (H_{j_2})_z]^2}{\frac{1}{2}[(H_{j_1})_z + (H_{j_2})_z]}$$

4: Training the SVM by using the chi-square kernel matrix

---

4   Experiments

4.1   Datasets and Performance Metrics

Our experiments [1] are conducted using two medical datasets: human epithelial type 2 cells dataset (SNPHEp-2 Cell Dataset) [2] [9] and pap-smear cells dataset [3] [3]. One motivation for us to use medical datasets is that medical image classification is a very important problem with challenging domain specific sub-problems to tackle. With fast-growing volumes of medical data/images collected by various systems or projects, medical data analysis has become a big data problem. One way to solve this problem is to use automated medical image classification. Cell image classification is an important type of medical image classification. Recognizing cells is a challenge in image classification because the resolutions of the cell images are usually very low. In addition, the image patterns on the cells are simple so not a lot of features can be extracted from these patterns. We show that our methods performs well in these medical images. The performance of our method in the experiment is evaluated by accuracy which is computed using true negative (tn), false negative (fn), true positive (tp), false positive (fp) and the number of classes (l): $accuracy = \left(\sum_{i=1}^{l} (tp_i + tn_i) / (tp_i + tn_i + fp_i + fn_i)\right) / l$.

4.2   The SNPHEp-2 Dataset

The anti-nuclear antibody test is a useful diagnostic method for autoimmune diseases and the Indirect Immunofluorescence protocol using human epithelial type 2 (HEp-2) cells is the important standard for the test [6]. The SNPHEp-2 Cell Dataset contains 1,884 cell images with the image size roughly 80-by-80 pixels and the images are divided into 5 classes: Homogeneous, Coarse speckled, Fine speckled, Nucleolar and Centromere. In the public SNPHEp-2 dataset, all the 1,884 cell images are extracted from 40 specimen images and these 40 specimens

---

[1] Our code is made available on the web at https://sites.google.com/site/apleung/

[2] Available from: http://staff.itee.uq.edu.au/lovell/snphep2/

[3] Available from: http://mde-lab.aegean.gr/downloads

are equally split into 20 for training and 20 for testing. The numbers of the cell images extracted from different specimens are different. We use the cell images. Hence, the numbers of the cell images for training (905 cell images) or testing (979 cell images) are different. As shown in Table 2, when the value of the k is large, it takes almost 10 days to run a single experiment. For a fair comparison, for each k and each method, we run the program 20 times. However, it takes weeks to finish computation with a 5-fold cross validation. In the SNPHEp-2 dataset, a five-fold validation for training and testing were created by randomly selecting the training (905 cell images) and test (979 cell images) images. The "Split-50-1" means that the first fold of these five-fold validation. Because of limited computational resources, only the first split (Split-50-1) which contain 450 images for training and 493 images for testing, is used for our experiments.

4.3   The PAP-Smear Dataset

The term pap-smear refers to human cell samples stained with the Papanicolau method for the ease of observation under a microscope [3]. The classification of pap-smear cells helps to detect cancer cells. The pap-smear cell dataset collected by Jan Jantzen et al. in 2005 [3]. The dataset consists of 917 cell images which are divided into 7 classes: superficial squamous epithelial, intermediate squamous epithelial, columnar epithelial, mild squamous non-keratinizing dysplasia, moderate squamous non-keratinizing dysplasia, severe squamous non-keratinizing dysplasia, and squamous cell carcinoma in situ intermediate. The whole pap-smear cell dataset [3] is used in our experiments. The dataset is randomly divided into the training set (50% images) and the test set (50% images) for the experiments. The results of the methods are compared with the reported accuracy of image classification and the computational efficiency.

4.4   Comparison with State-of-the-Art Algorithms

In the experiments, the BoF model using different dictionary learning methods (see Table 1) are compared using the two datasets. All experiments are conducted on an Intel Xeon E5-2690 CPU. We use MATLAB to implement six BoF model variants: three variants using our methods (see Table 1) and the three variants using the k-means, the kmeans++ and the dictionary method in [11]. The remaining one variant is from [10] which is a publicly available version of ScSPM based on sparse coding. To have a fair comparison, all the variants use the SIFT descriptor as the feature descriptor. In order to obtain reliable results, we repeat our experiments for each variant 20 times using the same datasets. The averages over 20 experiments are taken as our results for the accuracy.

To clearly show the comparison for the accuracy at different values of k, Figure 1 is plotted to show the comparison among six variants (three variants using our methods and three variants based on k-means, k-means++ or the dictionary method in [11]). In Figure 1, with the value of k increased between 20 and 350, the accuracy rate of the BoF model using our ED-M are larger than that of the model using other methods. On the SNPHEp-2 Cell Dataset and the pap-smear dataset, The BoF model using our ED-M method outperforms that using the k-means++ algorithm by nearly 12.4% and 8.6%.

In Table 2, the highest accuracies from the six BoF model variants in Figure 1 are compared with the accuracy of ScSPM. With the use of our ED-M method for dictionary learning, the BoF model obtains a much higher accuracy in the comparison. The BoF Model using our ED-MRS method speeding up our ED-M method outperforms the BoF model using other methods except for the ED-M and the ED-C.
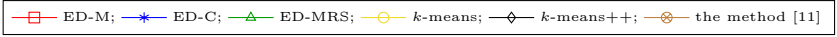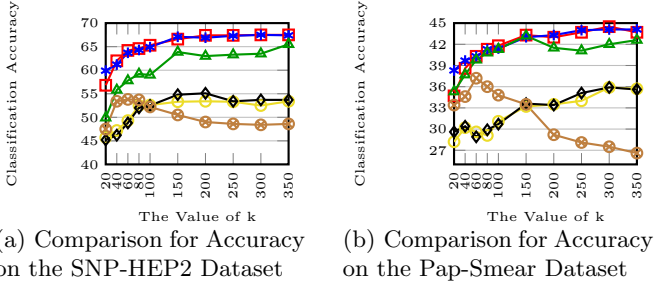

(a) Comparison for Accuracy on the SNP-HEP2 Dataset

(b) Comparison for Accuracy on the Pap-Smear Dataset

ED-M;   ED-C;   ED-MRS;   k-means;   k-means++;   the method [11]

Fig. 1. This figure is plotted to show the comparison among six k-means based variants. The dictionary methods for the BoF model are used to denote the BoF model variants.

Table 2. This table shows the comparison on the two datasets. The highest accuracies for the six BoF model variants in Figure 1 obtained with their respective optimal values of k are compared with the accuracy of ScSPM.

Dictionary learning method used in the BoF model	For the SNPHEp2 dataset			For the pap-smear dataset		
	Accuracy	optimal k	Execution time	Accuracy	optimal k	Execution time
proposed our ED-M	**67.5**	200	44853 s	**44.5**	300	91755 s
proposed ED-C	**67.5**	300	46569 s	44.1	300	96796 s
proposed ED-MRS	65.5	350	1695 s	41.9	150	1921 s
Sparse Coding	58.6	NA	452 s	38.4	NA	644 s
k-means++	55.1	200	416 s	35.9	300	813 s
k-means	53.4	200	353 s	35.9	300	1215 s
the method in [11]	53.8	60	28731 s	37.2	60	32022 s

The improvements of the proposed ED-M and ED-C over sparse coding and k-means come at a computational cost which makes our methods slower than traditional k-means dictionary learning. However, the proposed methods can be made more efficient using random sampling techniques like our proposed ED-MRS. With the speed up using ED-MRS to improve the performance of dictionary learning, the computational cost of our improved methods is lowered as can be seen in Table 2.

5   Conclusions

In this paper, a novel dictionary learning method is proposed to improve the accuracy of the BoF model for the image classification with simple patterns such as biomedical images. The improvement is achieved by the adding of the extra information from clustering with all reasonable values of k at the same time. Our experiments demonstrate that, with medical imaging datasets, the proposed dictionary learning method outperforms state-of-the-art methods, such as k-means/k-means++ clustering and sparse coding.

6   Acknowledgement

References

1. Ayech, Walid, M., Ziou, D.: Segmentation of terahertz imaging using k-means clustering based on ranked set sampling. Expert Systems with Applications **42**(6), 2959–2974 (2015)
2. Huang, et al., J.Z.: Automated variable weighting in k-means type clustering. IEEE T-PAMI **27**(5), 657–668 (2005)
3. Jantzen, Jan, e.a.: Pap-smear benchmark data for pattern classification. In: Nature inspired Smart Information Systems (NiSIS) (2005)
4. Jegou, Herve, et al.: Accurate image search using the contextual dissimilarity measure. IEEE T-PAMI **32**(1), 2–11 (2010)
5. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2006)
6. Meroni, P.L., Schur, P.H.: Ana screening: an old test with new recommendations. Annals of the rheumatic diseases **69**(8), 1420–1422 (2010)
7. Philbin, James, e.a.: Object retrieval with large vocabularies and fast spatial matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2007)
8. Philbin, James, e.a.: Improving particular object retrieval in large scale image databases. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2008)
9. Wiliem, Arnold, e.a.: Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors. In: IEEE Workshop on Applications of Computer Vision (WACV) (2013)
10. Yang, Jianchao, et al.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2009)
11. Zhang, S., Leung, A.P.: A novel approach to dictionary learning for the bag-of-features model. In: Wavelet Analysis and Pattern Recognition (ICWAPR) (2017)
12. Zhao, Chunhui, Li, X., , Cang, Y.: Bisecting k-means clustering based face recognition using block-based bag of words model. Optik-International Journal for Light and Electron Optics **126**(19), 1761–1766 (2015)