SPARSE NON-PARAMETRIC BAYESIAN MODEL FOR HEP-2 CELL IMAGE CLASSIFICATION

Shahab Ensafi† Star Agency for Science, Technology and Research (A*STAR/I2R), Email: stuse,slu@i2r.a-star.edu.sg
Ashraf A. Kassim† School of Computing National University of Singapore, Email: tancl@comp.nus.edu.sg

ABSTRACT

This paper studies automated classification of Human Epithelial Type-2 (HEp-2) cell images which is essential in diagnosing the Autoimmune Diseases (AD). The prevalent approach for this problem makes use of the Bag of Words (BoW) model and sparse coding scheme on over complete dictionaries, where the dictionary dimension is usually much larger than feature dimension. In addition, this approach usually requires manual selection of the dictionary dimension which is often troublesome and dependent highly on specific applications and datasets. We proposed a non-parametric Bayesian model that is capable of determining the dictionary dimension automatically by exploiting the Indian Buffet Process (IBP). This proposed model has been evaluated on two public HEp-2 benchmarking datasets, i.e., ICPR2012 and ICIP2013 where the SIFT and SURF features of the cell image are extracted in a grid manner and used as the input. Experiments show that the proposed model obtained state-of-the-art cell classification accuracy. More importantly, the dictionary dimension learned by the proposed model is around 29 and 8 times lower than the over complete dictionaries, respectively, for two benchmarking datasets. This low-dimensional dictionary helps to reduce the computational cost significantly for the cell classification task.

1. INTRODUCTION

Autoimmune Diseases (AD) are in the list of top mortality causes according to the American Autoimmune Related Diseases Association (AARD). The AD develops when the immune system of the body treat some healthy cells as foreigners and attack them. Diagnosis of these kinds of diseases plays an important role in AD treatment. Therefore, a preponderant Computer Aided Diagnosis (CAD) system is required to alleviate the costs and time of diagnosis and provide the repeatability of the diagnosis.

For diagnosing the AD, a high resolution images of the affected organ is needed. Indirect Immunofluorescence (IIF) is an imaging technique that is used to capture images of Human Epithelial Cells type 2 (HEp-2 cells). In this method, antibodies are first stained in a tissue and then bound to a fluorescent chemical compound. In case of antinuclear antibodies (ANAs), the antibodies bind to the nucleus and demonstrate different visual patterns that can be captured and visualized within microscope images. Classification of different patterns of the cell images gives the diagnosis which distinguishes the phase and severity of the disease [1].

Sparse representation of images using Bag of Word (BoW) model has become a promising image classification approach and attracted increasing interest in recent years. It has been successfully applied for the HEp-2 cell classification task [2, 3, 4, 5]. For sparse image representation, one critical parameter is the dictionary size and the optimal size depends heavily on the data as well as the visual features employed. A dictionary is called critically complete when the dimension of the features and dictionary are close to each other. On the other hand, biologically inspired over complete dictionary with dimension much larger than the feature dimension often gives better classification accuracy and is therefore widely adopted. Nevertheless, the high dimensional dictionaries use the high-dimensional features and codes which often reduce the classification speed significantly. The high-dimensional issue could be relieved by using the non-parametric Bayesian model that is capable of learning the optimal dimension with little human intervention.

We propose a Sparse Non-Parametric Bayesian (SNPB) model and implement it for the HEp-2 cell classification problem, targeting applications for computer-aided AD diagnosis. In the SNPB model, the dictionary learning exploits the non-parametric sparse factor analysis (NFSA) [6, 7, 8] that is capable of determining the dimension of the dictionary words automatically. In particular, the “non-parametric” here means that the dimension of the dictionary can be extended to infinity at the beginning of the learning process and lead to the correct and most efficient value at the end. The learned sparse representation of the codes is used as the final feature for the HEp-2 cell classification.

1.1. Related Works

A number of HEp-2 cell classification techniques have been reported recently [9, 10, 4]. A state-of-the-art method proposed by Ensafi et al.[2] uses sparse coding method on top of the SIFT and SURF features. The sparse codes are then scaled in three layers and the max-pooling scheme concatenates them to provide the final features. Finally these features are used to learn a SVM model for classification. A similar approach is used by Wiliem et al. [5]. They used the pyramids of intensity orders of the cells as their features to learn the dictionary and multiple kernel learning method for classifier.

The winner of the IPR 2012 cell classification contest used a modified version of the Local Binary Patterns (LBP) and SVM classifier [11]. Xiangfei and Kuan [12] represented a cell image by using a frequency histogram of textons and classified it by using kNN with \( \chi^2 \) distance metric. Fisher tensors on the Riemannian manifold are also exploited by Faraki et al. [13]. They used the BoW model and k-means algorithm for dictionary learning procedure. Additionally, Shen et al. [4] pooled the gradient features based on the intensity orders of local grid points for their BoW method [14][15].

All the methods described above assigned manual values for the dictionary learning procedures and nearest neighbors. However, there is a trade off between performance of the system and the di-
mension of the dictionary [3], where the large values for dictionary dimension results in better accuracies but affect the computational cost. To the best of our knowledge, there is no study of finding the best low-dimensional dictionary for the HEp-2 cell classification problem, which is one of the novelties of this paper.

In the rest of the paper, we describe the proposed CAD system in section 2 and explain the experiments and results in section 3 on two benchmarking datasets introduced in section 3.1. Finally, we have the discussion and conclusion sections in 3.4 and 4 respectively.

2. METHOD

The proposed SNPB method is depicted in Fig. 1. First, the SIFT and SURF features are extracted from the masked images in a grid manner. Then a dictionary is learned by using the non-parametric Bayesian method, which can estimate the dimension of the dictionary automatically. By transferring the input features to sparse codes by means of the learned dictionary, we scaled them to three layers and then the max pooling approach makes the output feature vectors.

In other words, three layers of codes are used. The first layer is the all sparse codes, second one is the divided image to four regions and the last layer is divided to 16 regions. Totally there are 21 regions where in each region the maximum bin of the histograms of codes are calculated. Finally, the concatenated sparse codes of each region is used as the final feature vector for each image. By assuming the dimension of each feature vector is \( D \), the final feature vector dimension is \( 21 \times D \). In training stage, these features are then fed to the multi class (One-Versus-All) linear Support Vector Machine (SVM) to classify the input training images with their ground truth labels. The same procedure is applied on the test images by using the pre-learned dictionary and SVM classifier.

2.1. Dictionary Learning

The dictionary learning method which is used is based on non-parametric Bayesian method [16] that makes use of the sparse prior knowledge on coefficients of dictionary words based on Indian Buffet Process (IBP) [17]. The graphical model of this method is shown in Fig. 1 in the dictionary part [6].

Let \( F \) be a set of features in a \( D \)-dimensional space, \( F = [f_1, f_2, \ldots, f_N] \in \mathbb{R}^{(D \times N)} \), and \( D = [d_1, d_2, \ldots, d_K] \in \mathbb{R}^{(D \times K)} \) are \( K \) words of our dictionary to be estimate. We can write:

\[
f_n = D x_n + e_n, \quad D = \gamma \odot Z
\]

where \( \gamma_k \) is the precision (inverse variance) of the \( k^{th} \) word in dictionary and \( Z \) is a binary matrix. The indicator function \( X = [x_1, x_2, \ldots, x_N]^T \) contains the weights of the words in the dictionary and we need it to be as sparse as possible and \( e \) is the noise vectors for each dictionary words, usually assumed to be Gaussian with diagonal covariance matrix \( \Sigma_e \) for each dimension. Here we assume both indicator function \( F \) and dictionary \( D \) are hidden variables of our non-parametric model and we want to infer the posterior distribution given the input feature vectors. Now we can model our dictionary by “spike and slab” distribution as

\[
P(D_{dk}|Z_{dk}, \gamma_k) = Z_{dk}N(D_{dk}; 0, \gamma_k^{-1}) + (1 - Z_{dk})\delta_0(D_{dk}) = \Gamma(D_{dk}; \gamma_k, 0)
\]

where \( \delta_0 \) is the delta function. In this model we want to estimate the number of dictionary words \( K \). Therefore, the \( Z \) matrix should have infinite columns in initial step. To do so, we make use of IBP, which provides a sparse matrix of intuitively infinite dimension. In this regard, we can assume that we have finite \( K \) model and then take the limit to \( K \to \infty \). To provide the \( Z \) matrix by IBP, we assume that the rows are generated separately and a probability of source contributing to any dimension is \( \pi_k \). Then we can write

\[
P(Z | \pi) = \prod_{k=1}^{K} \prod_{d=1}^{D} P(Z_{dk}|\pi_k) = \prod_{k=1}^{K} \pi_k^{m_k} (1 - \pi_k)^{D - m_k} = \Gamma(D; \sum_{k=1}^{K} m_k)
\]

where \( m_k \) is the number of nonzero elements of column \( k \) in \( Z \). Because the product is the binomial distribution, we can use the conjugate Beta distribution for \( \pi_k \). We can assume \( \alpha \) as the strength parameter in \( r = \frac{\alpha}{K} \) and \( s = 1 \). Then we can define the model as

\[
Z_{dk}|\pi_k \sim Bernoulli(\pi_k)
\]

\[
\pi_k|\alpha \sim Beta\left(\frac{\alpha}{K}, 1\right)
\]

By integrating out \( \pi \) we have

\[
P(Z) = \prod_{k=1}^{K} \frac{\Gamma(m_k + \frac{\alpha}{K})\Gamma(D - m_k + 1)}{\Gamma(D + 1 + \frac{\alpha}{K})} \prod_{k=1}^{K} \frac{\Gamma(m_k + \frac{\alpha}{K})\Gamma(D - m_k + 1)}{\Gamma(D + 1 + \frac{\alpha}{K})}
\]

where \( \Gamma(.) \) is the Gamma function. By defining the method proposed by Griffiths and Ghahramani [17] we can have the infinite limit of equation (6) as

\[
P(Z) = \frac{\alpha^{K+}K_+}{\prod_{h>0}^{K_+}}\exp\left(-\alpha H_D\right) \prod_{k=1}^{K} \frac{(D - m_k)!}{N!} \prod_{h>0}^{K_+} \frac{\Gamma(m_k + \frac{\alpha}{K})}{\Gamma(D - m_k + 1)}
\]

where \( K_+ \) is the number of non-zero column of \( Z \), \( H_D = \sum_{j=1}^{D} \frac{1}{j} \) is the \( D^{th} \) harmonic number and \( K_+ \) is the number of rows whose entries correspond t the binary number \( h \).

To provide a sparse matrix with the distribution in (7), the Indian Buffet Process starts from the first row and samples Poisson\((\alpha)\) columns. To generate the \( ith \) row, IBP samples from the columns which have been sampled in previous rows with the probability of \( \frac{m_k}{\alpha} \) and samples Poisson\((\frac{\alpha}{K})\) from the new columns. Here \( m_k \) is the number of nonzero element of column \( k \) in \( Z \). The large values of \( \alpha \) produce the Matrix with relatively large number of columns.

For inference, the Markov Chain Monte Carlo (MCMC) method is used, which defines a Markov chain on the hidden variables \( (Z, X) \) and maximizes the posteriors. In other words, in each iteration the \( Z, X \) matrices are sampled using Gibbs sampling strategy (which is a simple form of MCMC) and the posterior probability is maximized.

Additionally, we can sample the hyper-parameter of IBP \( \alpha \) as well using conjugate Gamma\((a_1, a_2)\) prior by the likelihood term of equation (7),

\[
P(\alpha|Z) \propto P(Z|\alpha)P(\alpha) = \Gamma(\alpha + a_1, H_D + a_2).
\]

where \( a_1 \) and \( a_2 \) are constant values [6].

Fig. 1. The SNPB framework of the HEp-2 cell classification.
2.2. Sparse Coding

Next stage after calculating the optimal dictionary is to code the input images sparsely [18]. We use the efficient sparse coding scheme [19] by fixing the dictionary words, which results in (9),

\[
\min_{\mathbf{x}} \sum_{n=1}^{N} \| f_n - \mathbf{D}\mathbf{x}_n \|^2 + \lambda |\mathbf{x}_n| \tag{9}
\]

However, if we know the sign of each code in \( \mathbf{x}_n \) then we can replace it with either \( \mathbf{x}_n \) or \( -\mathbf{x}_n \), then the resulting formulation will change to a simple Quadratic optimization Problem (QP), by gussing the initial values, we can refine it by solving this QP using least squares.

3. EXPERIMENTS AND RESULTS

3.1. Datasets

Two publicly available datasets namely MIVIA HEp-2 (ICPR2012) and ICIP2013 are used in this experiment. The former dataset has training and test sets in six classes but the latter one has huge number of cells in its training set and the test set remained as an evaluation set for the organizers, which is not published so far. Both datasets contain IIF images with several cells in them. The mask of the cells are provided in order to classify the cells without considering other neighboring cells named cell level classification, as shown in Fig. 2. Additionally, it is assumed that the cells in each image belong to one class, which defines the image level classification problem.

3.2. Optimizing the Dimension of Dictionary

The proposed non-parametric Bayesian method is used to estimate the optimum dictionary dimension in both datasets. As can be seen in Fig. 3, the dimension of dictionaries (\( K \)) increases in the starting iterations and finally converge to their optimums. By this method, the dimension of the positive and intermediate intensity level dictionaries in ICPR2012 are calculated 28 and 18, respectively. For ICIP2013 datasets, these values are estimated 139 and 123 respectively. whereas, these values are manually selected to 1024 as the state-of-the-art results in [2][3]. The increasing slope of the charts in Fig. 3 proves that the size of the dictionary matrix is intuitively infinite and by optimizing the model, it decreases to a minimum value in its steady state. Additionally, by having low dimensional dictionaries, the dimension of final sparse codes and complexity of calculating them are decreased as well.

3.3. Evaluation

To evaluate the method, the Mean Class Accuracy (MCA) is used as the ICIP2013 dataset suggests \( MCA = \frac{1}{C} \sum_{c=1}^{C} CCR_c \), where \( CCR_c \) is the correct classification rate for class \( c \) and \( C \) is equals to the number of classes. Additionally, in all the evaluation procedures, the dictionaries are learned on the training set only.

For the ICPR2012 dataset, the test set is available for evaluation. Table 1 shows the accuracies of ICPR2012 dataset versus the other methods. As can be seen in this table, although we have learned a low dimension dictionary, a better accuracy is achieved in cell level and positive intensity level in comparison with other methods.

For the ICIP2013 dataset, in order to compare our results with the method of Han et al. [20], the accuracies are achieved using 600 randomly selected images for training and the rest for testing. This evaluation is performed on positive and intensity level images as well and the results are stated in TABLE 2. Additionally, the method of Ensafi et al. [3], which we call it Sparse Coding (SC) method is evaluated using the same randomly selected images. In this method the dictionary dimension is manually defined to 1024 as the authors suggest. As we can see in the TABLE 2, the SNPB results are better than the SC method by having low dimensional dictionaries as well.

3.4. Discussion

The SNBP model obtained state-of-the-art result in Cell Level (75.2%) and positive intensity cells (82.6%) respectively in comparison with other methods for ICPR2012 dataset. Additionally, the dimension of the learned dictionary for positive and intermediate intensity levels are 28 and 18 respectively, which are more than 36 times smaller than the other dictionary based models as in [2], which is manually selected to 1024. This dominant reduction of dictionary size is a great beneficial for calculating the sparse codes and classifying the test images.

For the ICIP2013 dataset, the state-of-the-art accuracies are obtained with comparing to Han et al. [20] and Ensafi et al. [3] by
4. CONCLUSION

A Sparse Non-Parametric Bayesian (SNPB) model is proposed for automatic classification of the HEp-2 cell images. The prevalent approach uses sparse coding and Bag of Words models which depends highly on the dictionary size that is usually selected in a manual manner. The Indian Buffet Process provides prior knowledge of sparse codes and take advantage of intuitively infinite matrix dimension which is exploited to produce an optimal dictionary size automatically. Experiments show that the dimension of the proposed model is 28 and 8 times smaller than the similar BoW methods in ICPR2012 and ICIP2013 datasets respectively. Additionally, the lower dimension of learned dictionary leads to lower computational time in the test procedure.

5. REFERENCES