ICIC Express Letters
Volume **11**, Number **8**, August **2017**
 ISSN 1881-803X
pp. **1299–1305**

HISTOPATHOLOGICAL IMAGE CLASSIFICATION WITH GAUSSIAN PROCESS EXPERTS

Gang Zhang, Yiyu Lin, Huadong Lai and Dong Lin

School of Automation
Guangdong University of Technology
No. 100, Waihuan West Road, Guangzhou Higher Education Mega Center
Guangzhou 510006, P. R. China
ipx@gdut.edu.cn

Received March 2017; accepted May 2017

Abstract. *We propose a classification method for histopathological image based on Gaussian process (GP) experts. The main idea is that a set of GP experts is initialized and improved with a data point selection strategy, through which a data point is assigned to a GP expert by evaluating how close it is to the representation point of that expert. The model precisely divides the input space into regions and each expert is responsible for a certain region. The method is non-parametric and can be trained with less computational cost compared to traditional ones. We evaluate the proposed method on a histopathological image dataset of common skin diseases. The results show that the proposed method is superior to current state-of-the-art methods.*
Keywords: Histopathological image classification, Gaussian process expert, Bayesian classification

1. **Introduction.** Histopathological images are the imaging of slices of lesion tissues under a microscopic, which play an important role for disease diagnosis in many medical departments. Histopathological images can reflect the presence and/or severity of various serious diseases and in some cases they are essential hints for doctor's final decision. Altunbay et al. [1] proposed a classification method by detecting what the micro-structure units a histopathological image has. In their study, images from colon while in cancerous tissues regular structures would be destroyed significantly. Caicedo et al. [2] proposed a bag-of-feature (BOF) method based on color-pixel histogram for histopathological image classification. Molin et al. [3] proposed a method for automated image analysis in digital pathology incorporating some key issues of human-computer interaction (HCI). Current works mainly focused on methods of feature representation and classification models [4, 5], but seldom on the explanation of them, i.e., the models themselves cannot directly provide enough information of how they work [6]. On the contrary, probabilistic models are widely accepted for their ability of well explanation from input space to output variables [7, 8]. Though probabilistic models are well studied and many successful applications have been reported, they still met some problems from model complexity to difficulty of incorporating prior knowledge [9]. For medical image processing tasks, Bayesian inference, naive method and maximum a posterior (MAP) evaluators are widely studied and used. However, we should point out that these methods may suffer from the approximation of a single but very complex Gaussian distribution, especially in case of analyzing images with dense pixels.

In order to make a balance between the representation ability and complexity of the analysis model, we propose a novel method based on Gaussian process (GP) experts. A GP expert can be regarded as a small but well-trained model based on Gaussian process that performs well in a subspace of the original input space. Different from deriving a single mixture Gaussian distribution and approximating it at a high cost, we go another

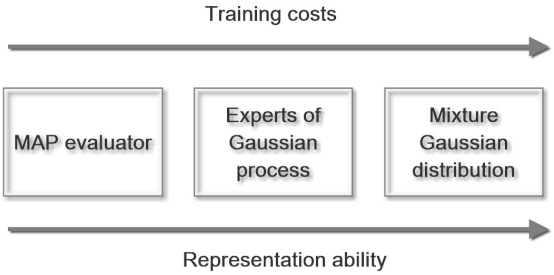

FIGURE 1. The relation between the proposed method and the mainstream Bayesian analysis methods

direction, i.e., combining several GP experts which are trained under GP principles. Figure 1 shows the relation between the proposed method and the mainstream Bayesian analysis methods.

Roughly speaking, our method trains several sub-models based on Gaussian process and makes each of them be responsible for certain region (maybe overlapping) of the decision surface and uses a voting strategy to get the final decision. It includes an interesting mechanism of Gaussian expert, i.e., an expert is weighed more if the query point is closer to its responsive region. Based on Nguyen and Bonilla's work [10], we adopt a training point allocation strategy to achieve a fast inference of the experts for determining their hyper-parameters.

The remainder of this paper is organized as follows. In Section 2 we formally present the main method and propose an effective algorithm for model training, as well as some discussion on the model parameters. In Section 3 we report the application on histopathological image analysis, including image preprocessing, evaluation methods and discussion of evaluation results. Finally, we conclude the paper in Section 4.

2. Gaussian Process Experts.

We begin with a formal definition of the problem to be solved. Suppose there is a training set $D = \{(x_i, y_i)|x \in R^d, y \in R, i = 1, 2, \ldots, N\}$. The goal is to learn a function $f : R^d \to R$, such that $\forall(x, y) \in D$ the loss between $f(x)$ and y is minimized. Meanwhile, for an unseen data point x_t, $f(x_t)$ is as close as possible to the ground truth value y_t. And a good regression model can be balanced between these two issues. In order to obtain good explanation ability, a probabilistic model is preferred, either as discriminative or generative one. In a probabilistic perspective, the function f represents a predictive distribution $p(y|x_t, D)$. We denote the value of y that makes $p(y|x_t, D)$ achieve its maximum value as y^* and it is the model output for x_t. And the value $p(y^*|x_t, D)$ provides the confidence of this output of the model.

However, as mentioned in many Gaussian process and Bayesian learning literature [7, 9], the cost of training f can be extremely expensive especially for large datasets. We propose an ensemble-style method to solve this problem effectively. The method is originated from the Gaussian process experts algorithm. Let us initialize K independent experts, each of which models a function f_i according to certain Gaussian process, i.e., $f_i(x) \sim GP(0, k(x, x'; \theta_i))$, in which GP stands for a standard Gaussian distribution with zero mean and a co-variance matrix determined by a kernel function k on the training points $x \in D$. θ_i is the hyper-parameter of the kernel function associated with f_i. For each f_i, there is a set of inducing points composing of m elements, denoted as U_i, which means that the behaviors of the experts are completed determined by this set of points. For each inducing point $u_{ij} \in U_i$, we have $y_i = f_i(u_{ij}) + \epsilon$, where $\epsilon \sim N(0, \delta^2)$ is a natural Gaussian noise.

There are two subproblems to be solved. The first is that we must derive an inducing point selection strategy given the number of experts K. The second is that an ensemble mechanism is required to generate a unified model with K trained experts. For the first

subproblem, we perform the allocation procedure by finding their centroids and variances. A new point x is allocated to all experts with weights determined by the Euclidean distance to the centroids. Let $D_E(x)$ be the set of distances between the new point and all centroids, and the weights are given by:

$$i^* = \arg\max_i = N(d_i|0, E), \quad d_i \in D_E(x) \tag{1}$$

$$w_i = \frac{1}{1 + e^{-d_{i^*}}} \tag{2}$$

where d_i stands for the Euclidean distance between the test point and the ith centroid. Note that w_is should be normalized before applying to the output of experts, i.e., multiply a factor to make their sum 1. Finally, we present an algorithm for inducing point selection, as shown in Algorithm 1.

Algorithm 1 Inducing point selection for Gaussian process experts

Require:
training data set $D = \{(x_i, y_i), i = 1, \ldots, N\}$
the number of experts K

Ensure:
K point sets for training experts, combination weights w

1: Cluster D into K clusters measured by Euclidean distance using KMeans
2: Compute the centroid of each cluster
3: Train K experts with the K clusters
4: **while** true **do**
5: **for** $t = 1, 2, \ldots, N$ **do**
6: Read a training sample x_t from D
7: Solve the optimization problem described in Equation (1), denoted the result as t^*
8: **for** $j = 1, 2, \ldots, K$ **do**
9: Compute w_j according to Equation (2)
10: **end for**
11: Update the K clusters according to the weights
12: **end for**
13: **if** no update performs **then**
14: break WHILE
15: **end if**
16: Train K experts with the K clusters
17: **end while**
18: **return** K clusters and w

Algorithm 1 performs a batch update with iterative point allocation and model training. In lines 1 and 2 it initially clusters all the training points and finds the centroids of all clusters. From line 4 to line 17 it iteratively re-allocates each point between clusters and in each round trained the K experts so as to adjust them fitting the current training data subsets well. Note that for a fast convergence of Algorithm 1, in line 4 the condition of **while** statement can be replaced by a threshold on the number of re-allocated points.

The K experts are Gaussian process models with kernel function $k(\cdot, \cdot)$. We use RBF kernel function with random parameter θ to initialize the experts, as shown in Equation (3):

$$k(x, y) = \exp\left(-\frac{d^2(x, y)}{\theta^2}\right) \tag{3}$$

Now we come to the second subproblem, i.e., an ensemble mechanism for generating a union model with K trained experts. It is observed that though the K experts are

identically trained through the method indicated by Algorithm 1, they have different weights on each test sample. The explanation of this observation is as follows.

- Each Gaussian expert is a distribution function of the whole sample space.
- The classification decision of an expert to a certain test point is more important if the point is closer to the corresponding inducing point of the expert.

We propose a joint weighting mechanism considering two criteria. On the one hand, to evaluate the impact of an expert to a test sample, the distance between the corresponding inducing point and the sample should be considered, which is determined by the definition of Gaussian distribution. On the other hand, the quality of each expert should be considered, which is evaluated by its performance on a preset validation set. Algorithm 2 summarizes the whole ensemble procedure for the prediction of a feature vector of an unseen histopathological image.

Algorithm 2 Weighted ensemble for Gaussian process experts

Require:
a test data sample x
the set of trained Gaussian experts $H = \{h_1, h_2, \ldots, h_K\}$
the validation set V

Ensure:
the classification label of x

1: $acc = \emptyset$
2: **for** $i = 1$ to K **do**
3: add $val(h_i, V)$ to acc
4: **end for**
5: $res = \emptyset$
6: **for** $i = 1$ to K **do**
7: $d = \dfrac{1}{1 + \exp(-dist(x, h_i.ind))}$
8: $res(i) = \Delta(d, acc(i)) * h_i(x)$
9: **end for**
10: **return** the mean of res

In Algorithm 2, lines 1 to 4 evaluate the accuracy of the K experts on a validation set V. If the validation set is identical to all test samples, this procedure can be run only once. Lines 5 to 9 describe the weighted ensemble mechanism by incorporating the distance between x and each inducing point of an expert, and the performance of each expert on V. Following the idea of Zhang et al.'s work [11], we implement a function $\Delta(\cdot, \cdot)$ for combination of the two criteria which balances their importance, as shown in Equation (4):

$$\Delta(x, y) = \frac{s(x) \cdot s(y)}{s(x) + s(y)} \tag{4}$$

where $s(x) = \frac{1}{1+\exp(x)}$ is the sigmoid function.

3. **Histopathological Image Analysis with Gaussian Process Experts.** We evaluate the proposed method on a clinical image dataset from the biopsy laboratory of a big local hospital. The dataset has about 16000 histopathological images of lesion skin tissues, attached with diagnostic information. The size and channel of each image are 600*800 and 24bit-RGB. We implement our evaluation programs using WEKA [12], a Java-based machine learning algorithms library. The JDK version is Java SE 8 Update 121 and the IDE is Eclipse Neon 64 bit. Figure 2 shows some samples of the dataset.

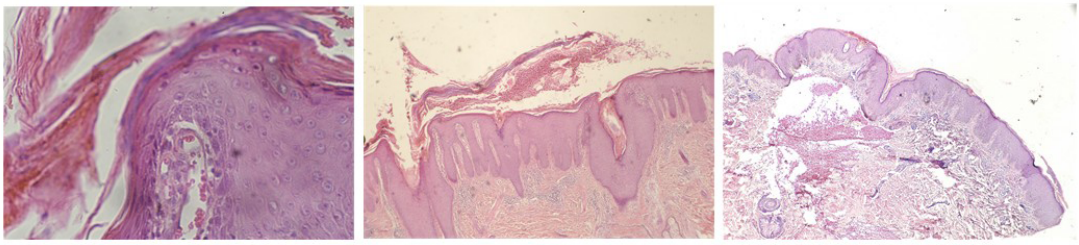

FIGURE 2. Sample images of the evaluation dataset

TABLE 1. Five histopathological labels

No.	Name	Ratio
F1	inflammatory granuloma	26.2%
F2	abscess center giant neutrophils	18.1%
F3	intercellular edema	21.7%
F4	papillomatous hyperplasia	25.9%
F5	absence of granular layer	31.8%

The dataset has totally 33935 images with multiple-labels which are generated according to their diagnostic information. We consider 5 classification labels which are the most significant among all labels, as shown in Table 1.

We evaluate the proposed model on the aforementioned dataset in comparison to two current state-of-the-art methods which were proposed by Murthy et al. [13] and Zhang et al. [14]. For brevity, we denote these two methods as M1 and M2. Generally speaking, M1 is a deep learning based method that automatically represents each histopathological image as a feature vector through a convolutional neural network (CNN) and performs classification with a standard classifier. M2 regards each image as a multiple-instance sample and solves the problem with a multiple-instance multiple label classifier. These two methods reflect different ideas in solving the same problem and both achieve relatively good results.

Histogram and SIFT descriptor are used for feature representation which is the same as that in Caicedo et al.'s work [15]. An image is represented as a feature vector of 64 histogram- and 128 SIFT-features. We use the WEKA project [12] to implement the training and test programs of the proposed method. All methods are evaluated through a ten-fold cross validation procedure. Note that the method M1 and our method need validation dataset. We use the following procedure to achieve this. In each round, 10% data is reserved for test and among the rest data 10% is for validation and the rest is for training. M1 requires a CNN to extract feature and we implement it with a Matlab based deep learning framework, MatConvNet [16]. Three methods are evaluated by accuracy in a one-vs-rest binary classification schema. Table 2 reports the performance of all methods.

TABLE 2. The overall performance of the three methods

	M1	M2	Gaussian Experts ($K = 5$)
F1	$69.2\% \pm 2.4\%$	$67.8\% \pm 3.1\%$	$74.3\% \pm 1.5\%$
F2	$76.1\% \pm 1.9\%$	$72.5\% \pm 2.3\%$	$76.8\% \pm 2.0\%$
F3	$70.5\% \pm 3.7\%$	$71.9\% \pm 2.4\%$	$77.7\% \pm 2.2\%$
F4	$80.4\% \pm 2.0\%$	$83.2\% \pm 2.5\%$	$85.0\% \pm 2.3\%$
F5	$77.8\% \pm 2.6\%$	$78.9\% \pm 3.0\%$	$80.7\% \pm 2.9\%$

In Table 2, the column **Gaussian Experts** stands for the proposed method. F1 to F5 stand for the five concerning histopathological characteristics. Since we use a ten-fold cross validation procedure for evaluation, the means and variances of each concerning histopathological characteristic are recorded in the table. From Table 2 we can see that the proposed method outperforms M1 and M2 in five histopathological characteristics.

We also present an evaluation of how the number of experts K affects the performance of the proposed model. This is achieved by varying the number of experts K from 1 to 15 with step 1 and we record the proposed model performance on the test dataset, also with a ten-fold-cross validation setting, which is the same as the previous evaluation case. Figure 3 shows the results (F1 and F2).

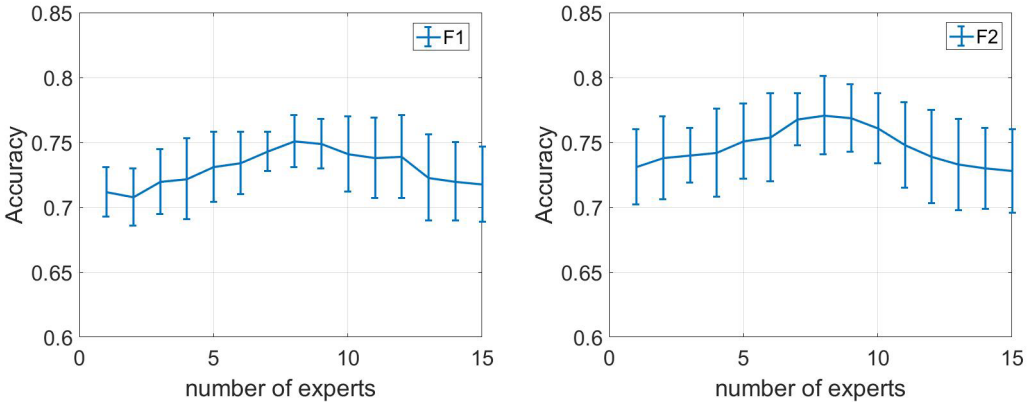

FIGURE 3. Performance of the proposed model with different K (F1 and F2)

From Figure 3 it can be concluded that the number of experts K affects the model performance and there may exist an optimal value for a certain task. Too small or large K may lead to performance downgrade. We can see in the figure that for F1 and F2 the optimal value is different. This is because the feature spaces defined by classification tasks (F1 or F2) are different leading to overlapping of subspaces induced by the trained GP experts. However, even we have domain knowledge it is not easy to determine the optimal number of GP experts. From Figure 3, it can be empirically concluded that the model performance is a convex function of the number of GP experts.

4. **Conclusions.** We propose a histopathological image classification method based on weighted ensemble Gaussian process experts. The method initializes a set of GP experts and improves them with a data point selection strategy. Then we use a weighted ensemble method to combine all trained GP experts into a final classifier. The proposed method is evaluated on a histopathological image dataset and the results show that it outperforms two current state-of-the-art methods. Our future work includes designing a customized method to initialize the GP experts for ensemble. Specifically, we are interested in developing an initialization method which makes use of prior knowledge or distribution obtained from historic data. Meanwhile, for the ensemble mechanism of the trained GP experts, we will study an optimal algorithm which can balance the concerning criteria for ensemble.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (Nos. 81373883, 81573827), the Science and Technology Planning Project of Guangdong Province (No. 2016A030310340), the Science and Technology Program of Guangzhou, China (No. 201604016086), the Special Fund of Cultivation of Technology Innovation for University Students (No. pdjh2016b0150), the College Student Career and Innovation Training Plan Project of Guangdong Province (yj201611845593, yj201611845074, yj201611845075, yj201611845366), the Higher Education Research Funding of Guangdong

University of Technology (No. 2016GJ12) and the 2015 Research Project of Guangdong Education Evaluation Association (No. G-11).

REFERENCES

[1] D. Altunbay, C. Cigir, C. Sokmensuer and C. Gunduz-Demir, Color graphs for automated cancer diagnosis and grading, *IEEE Trans. Biomedical Engineering*, vol.57, no.3, pp.665-674, 2010.

[2] J. C. Caicedo, A. Cruz-Roa and F. A. González, Histopathology image classification using bag of features and kernel functions, *Proc. of AIME, ser. Lecture Notes in Computer Science*, C. Combi, Y. Shahar and A. Abu-Hanna (eds.), vol.5651, pp.126-135, 2009.

[3] J. Molin, P. W. Woźniak, C. Lundström, D. Treanor and M. Fjeld, Understanding design for automated image analysis in digital pathology, *Proc. of the 9th Nordic Conference on Human-Computer Interaction, ser. NordiCHI'16*, New York, NY, USA, pp.58:1-58:10, 2016.

[4] R. Ali, C. Gunduz-Demir, T. Szilagyi, B. Durkee and E. E. Graves, Semi-automatic segmentation of subcutaneous tumours from micro-computed tomography images, *Physics in Medicine and Biology*, vol.58, no.22, p.8007, 2013.

[5] T. Chankong, N. Theera-Umpon and S. Auephanwiriyakul, Automatic cervical cell segmentation and classification in pap smears, *Comput. Methods Prog. Biomed.*, vol.113, no.2, pp.539-556, 2014.

[6] C. Stoean, R. Stoean, A. Sandita, C. Mesina, C. L. Gruia and D. Ciobanu, Evolutionary search for an accurate contour segmentation in histopathological images, *Proc. of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation, ser. GECCO Companion'15*, New York, NY, USA, pp.1491-1492, 2015.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[8] L. Sesma-Sanchez, Y. Zhang, A. Bulling and H. Gellersen, Gaussian processes as an alternative to polynomial gaze estimation functions, *Proc. of the 9th Biennial ACM Symposium on Eye Tracking Research & Applications, ser. ETRA'16*, New York, NY, USA, pp.229-232, 2016.

[9] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[10] N. Trung and B. Edwin, Fast allocation of gaussian process expert, *Proc. of the 31st International Conference on Machine Learning*, pp.145-153, 2014.

[11] D. Zhang, M. M. Islam and G. Lu, A review on automatic image annotation techniques, *Pattern Recogn.*, vol.45, no.1, pp.346-362, 2012.

[12] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, Weka – Experiences with a Java open-source project, *J. Mach. Learn. Res.*, vol.11, pp.2533-2541, 2010.

[13] V. N. Murthy, S. Maji and R. Manmatha, Automatic image annotation using deep learning representations, *Proc. of the 5th ACM on International Conference on Multimedia Retrieval, ser. ICMR'15*, New York, NY, USA, pp.603-606, 2015.

[14] G. Zhang, J. Yin, X.-Y. Su, Y.-J. Huang, Y.-R. Lao, Z.-H. Liang, S.-X. Ou and H.-L. Zhang, Augmenting multi-instance multilabel learning with sparse Bayesian models for skin biopsy image analysis, *BioMed Research International*, vol.2014, 2014.

[15] J. C. Caicedo, F. A. González and E. Romero, Content-based histopathology image retrieval using a kernel-based semantic annotation framework, *J. of Biomedical Informatics*, vol.44, no.4, pp.519-528, 2011.

[16] A. Vedaldi and K. Lenc, Matconvnet – Convolutional neural networks for MATLAB, *Proc. of the ACM Int. Conf. on Multimedia*, 2015.