A Hierarchical Model with Pseudoinverse Learning Algorithm Optimazation for Pulsar Candidate Selection

Shijia Li, Sibo Feng, Ping Guo*, Qian Yin
Image Processing and Pattern Recognition Laboratory
Beijing Normal University
Beijing 100875, China
{sjli,sibofeng}@mail.bnu.edu.cn
pguo@ieee.org
yinqian@bnu.edu.cn

***Abstract*—Pulsars search has always been one of the most concerned problem in the field of astronomy. Nowadays, with the development of astronomical instruments and observation technology, the amount of data is getting bigger and bigger. Radio pulsar surveys have generated and will generate vast amounts of data. To handle big data, developing new technologies and frameworks to efficiently and accurately analyze these data become increasing urgent. The number of positive and negative samples in pulsar candidate data set is very unbalanced, if we only use these a few positive samples to train a deep neural network (DNN), the trained DNN is prone because of the problem of overfitting and will affect the generalization ability. Motivated by the mixtures of experts network architecture, we proposed a hierarchical model for pulsar candidate selection which assembles a set of trained base classifiers. Moreover, training a neural network always takes a lot of time because of using gradient descent (GD) based algorithm. In this work, we utilize the pseudoinverse learning algorithm instead of GD based algorithm to train proposed model. With the designed network architecture and adopted training algorithm, our model has the advantages not only with high steady-state precision but also good generalization performance.**

***Index Terms*—Pulsar candidate selection, Pseudoinverse learning algorithm, Hierarchical mixtures of experts, Big data, Stacked generalization**

I. Introduction

In big data era, how to effectively analyze the massive data is a difficult task on the limited hardware resources. In recent years, deep learning techniques have been widely used in various scenes and have achieved remarkable performance. However, it cannot be ignored that training a model of deep learning requires a lot of computing resources and a long time. Therefore, how to train an effective model within tolerable elapsed time is one of the focuses of current research. The main directions of research for handling big data include developing new frameworks and algorithms and increasing computing power with new hardware. In this work, we will focus on frameworks and algorithms to deal with big data, and take the data of pulsar candidate as the example to examine the validity of the proposed model.

Searching for real pulsars is an important task in astronomical research. Pulsars are highly magnetized rotating neuron stars which can always emit a beam of electromagnetic radiation. It can be used as probes for a wide range of physics and astrophysics researches, such as the equation of state of dense matter, the properties of the interstellar medium, dark matter and dark energy, stellar evolution, the formation and evolution of binary and multiple-star systems. The discovery of a new pulsar is therefore a very important and significant astronomical task. From the existing radio telescope projection such as the Pulsar Arecibo L-band Feed Array [1], High Time Resolution Universe (HTRU) [2], and the Green Bank North Celestial Cap [3]. We could obtain massive data onto pulsar candidates. However most of these pulsar candidate data are contaminated by radio frequency interference (RFI) signals, that makes it difficult to recognize the real pulsars from the candidates produced using simple metrics such as the signal-to-noise ratio (S/N). Human experts could observe diagnostic plots which are made by folding the radiation signals from radio pulsar surveys and identify if the corresponding candidate is from real pulsars or non-pulsar noises with experience, and then perform further verifications on prospective candidates. But it is impractical to inspect millions of candidates that way. Therefore, it is necessary to develop a machine learning model to automatically filter out pulsar candidates.

In this work, we employ a well know "divide and conquer" strategy to dealt with big data problem. Especially, we propose to utilize a hierarchical mixtures of experts model [4] to implement the task of pulsar candidate selection. In the literatures, most researchers had to face the problem of imbalance between positive and negative samples, while most of them use sampling methods to obtain a balanced data set, which can improve the prejudice of the model to a certain extent. But at the same time, it has the risk of overfitting when these samples are used to train a deep neural network model. Therefore, we use the method of cross validation partition samples in stacked generalization [5] to obtain multiple experts

978-1-5090-6017-7/18/$31.00 ©2018 IEEE

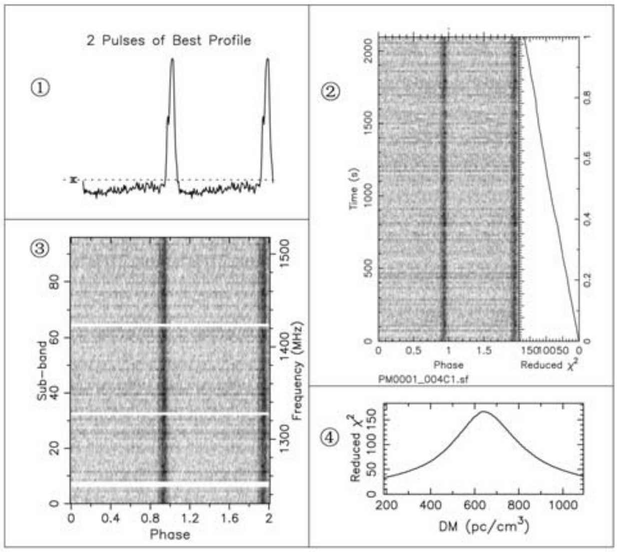

Fig. 1. An example of the four key feature plots. No.1.for summed pulse profile, No.2. for time vs. phase plot, No.3. for frequency vs. phase plot, and No.4. for DM curve

classifiers as the base learners and take a gating network as the meta learner to integrate all outputs of the base learners, and expect to reduce over-fitting to obtain the better generalization ability of the model.

From the astronomy project we can obtain massive data onto pulsar candidate. These data could be projected into some important feature plots against pipeline software. Human experts usually use patterns to identity pulsar candidate from these feature plots. Emulating the human experts. We will use the four most important standard diagnostic feature plots to select pulsar candidate: Summed profile, time versus phase plot, frequency versus phase plot and dispersion measure (DM) curve. Fig. 1 gives an example of the feature plots. These plots consist of one-dimensional (1D) data arrays (summed profile or DM curve) and two-dimensional (2D) data (time versus phase and frequency versus phase plots). Naturally, we consider using four sub-modules to deal with four kinds of feature plots respectively, and then summarize the outputs to obtain the final result.

For 1D data, we use ordinary feedforward neural network as expert classifiers which are trained by pseudoinverse learning (PIL) algorithm [6], [7]. PIL algorithm has been proved to be a very efficient method of training neural networks and it is especial for 1D data because of s no spatial structure. But for 2D data, flattened into 1D data will have adverse effect on the performance of recognition. So we first extract features of 2D data to make it into 1D data, and then use the PIL algorithm to train the base learner. Therefore, for 2D data, we use the method which is proposed by Feng et al. [8]. And the method is the combination of histogram of oriented gradient (HOG) [9] and pseudoinverse learning autoencoders (PILAE) [10], [11], [12], [13]. Therefore we have two kinds of experts: the PIL expert and the HOG+PILAE expert. Moreover, deep learning such as convolutional neural network (CNN) has developed rapidly and has defeated humankind in many difficult tasks in recent years. For comparison, we can build a baseline model that uses CNN and feedforward neural network trained via error back propagation algorithm (BPNN) as base classifiers.

The structure that we designed is shown in Fig. 2. In the first layer of our model, we use four diagnostic features plots as inputs to train a set of expert classifiers. Then we use the gating network to combine classification results from these base expert classifiers. At the second layer, we would synthesize the forecast outputs obtained from the first layer and then put into the second layer's classifier to get the final classification result. To demonstrate the practicality of our model on real world application, we examine it we examine it on two pulsar candidate datasets.

Experiments on the HTRU dataset and the PMPS-26k dataset shows that our proposed pulsar candidate selection model offers a significant improvement to other method.

II. Related work

a) Pulsar Candidate Selection: Some successful algorithms have been applied to pulsar candidate indentation. For example, Lee et al. [14] using a combination of six carefully designed heuristic scores. Morello et al. [15] developed six hand-crafted features and trained a single-hidden-layer neural network of binary classification on candidates. Zhu et al. [16] proposed a two-layer artificial intelligence system to select pulsar candidate, which take diagnostic plots as inputs, and train single-hidden-layer artificial neural networks (ANN) and support vector machines (SVMs) on histogram plots and convolutional neural networks (CNNs) on two-dimensional image plots. The predictions of all classifiers are then assembled together with a logistic regression classier. In order to solve the problem of limited training samples, Guo et al [17] utilize DCGAN to generate samples and learn features. And SVM are adopted as the classifier for predicting candidate's labels in their work.

b) Mixtures of Experts: Mixtures of Experts (MoE) is originally proposed by Jacobs et al. [18] in 1991. It is an adaptive model that could global optimize the ensemble networks. And Jordan et al. [4] developed MoE model and proposed hierarchical mixtures of experts (HME). The HME model consists of several expert networks and gating network. If we treat it as a tree network, the leaf nodes are expert networks that handle the regression or classification of data, and the root nodes is the gating networks that are responsible for regulating the output of the expert network. It is a powerful supervised learning framework.

A standard MoE model has a set of expert networks f_i and a gating network g. Each expert network f_i receive the vector x as input and produces an output vector μ_i, while gating network g maps the input x to a distribution over experts $i = 1,, N$ which sums to 1. The calculation process of final

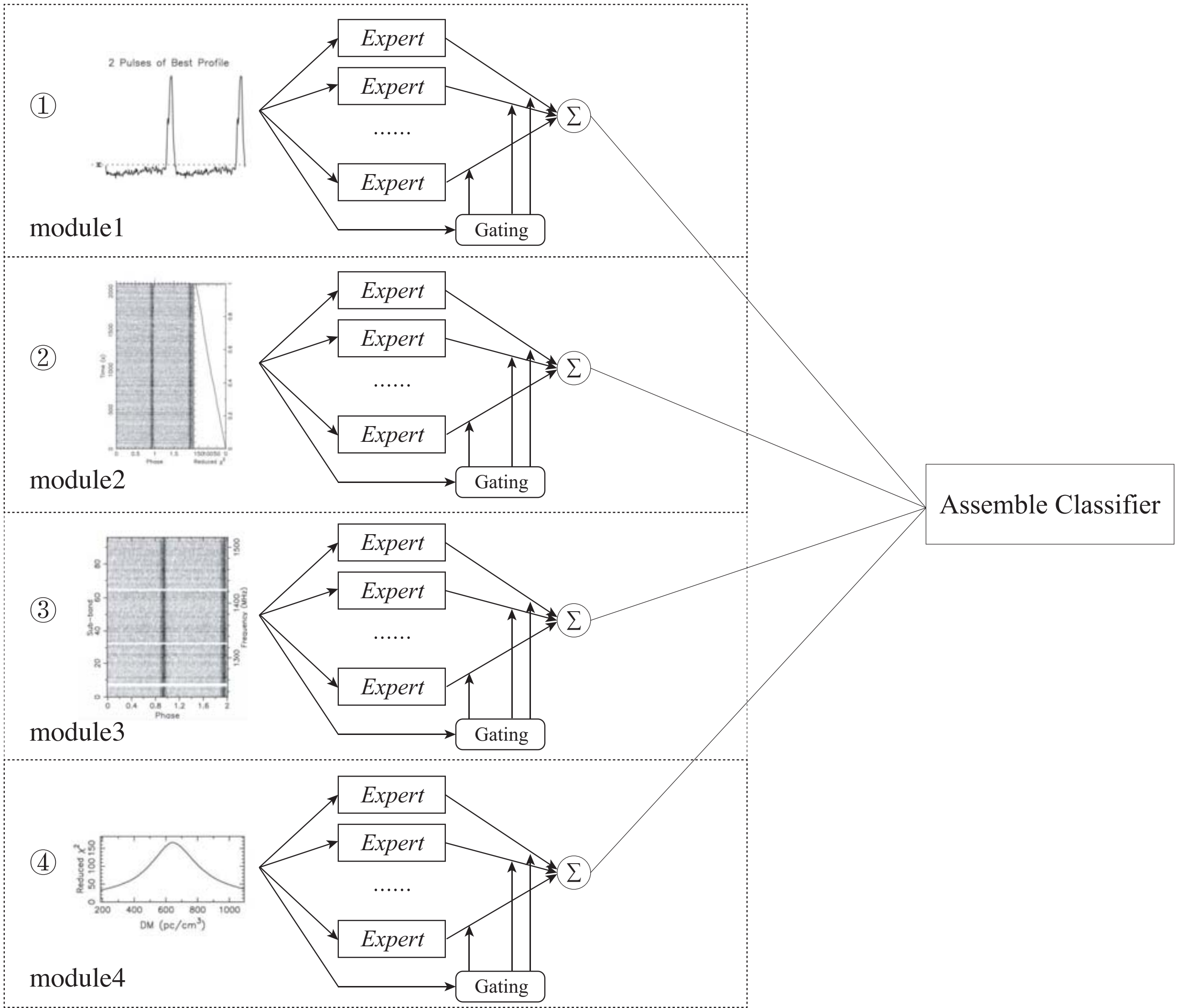

Fig. 2. The structure of our proposed model: it has two layers, and in the first layer, it consists four sub-modules into four sub-modules that correspond to four kinds of feature plots. Each module has a group of expert classifiers and a gating network to rate how pulsar-like a particular feature plot. And the second layer as an assemble classifier takes the four outputs into account and form a final consensus on how pulsar-like a candidate is.

output is shown as follow:

$$F_{MoE}(x) = \sum_{i=1}^{N} g_i(x) softmax(f_i(x)) \tag{1}$$

c) Histogram of Oriented Gradient with Pesudoinverse Learning Algorithm for training Auto-Encoders: Histogram of Oriented Gradient(HOG) [9] was proposed by Navneet et al., it is a hand-craft feature descriptor with many applications in image processing. The main idea of HOG is to describe the local information of the image in a gradient and edge direction density. The paper [8] combining the HOG and the pseudoinverse learning algorithm for training autoencoders(PILAE) to further learn feature for classification.

PILAE set hidden unit smaller than the dimension of the input vector. The number of hidden unit determined by the rank of the matrix and the dimension of the input vector. The rank obtained by decomposing the input matrix with singular value decomposition(SVD) method. Furthermore, the low rank approximation used in calculating encoder weight matrix, the weight tied used in encoder and decoder for reduce the degree of freedom of parameter.

Assume that the input matrix $\mathbf{X} \in R^{N\times r}$ is a row feature matrix. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$, where vector $\mathbf{x}_i = [x^1, x^2, ..., x^r]$ denote the i-th training data.

1) Singular value decomposition. We compute the pseudoinverse matrix of the input matrix with the thin singular value decomposition method,

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{2}$$

where $\mathbf{U} \in R^{N\times r}$, $\mathbf{\Sigma} \in R^{r\times r}$ and $\mathbf{V}^T \in R^{r\times r}$. We calculate the pseudoinverse matrix of the input matrix $\mathbf{X}^+$through the SVD result.

$$\mathbf{X}^+ = \mathbf{V}\mathbf{\Sigma}'\mathbf{U}^T, \tag{3}$$

where $\mathbf{\Sigma}'$ is the transposed diagonal matrix composed of the reciprocal of nonzero elements in matrix $\mathbf{\Sigma}$.

The number of the hidden layer units p determined by the rank of the matrix k and the dimension of the input vector r.

$$p = \beta Dim(\mathbf{x}), \beta \in [0, 1], \tag{4}$$

where β is a empirical parameter which depend on the what kind degree of dimension should be reduced.

2) Encoder weight. Here, we get the low rank approximation matrix $\hat{\mathbf{X}}^{+}$of $\mathbf{X}^{+}$ by truncating the left singular matrix $\mathbf{U}$ in the singular value decomposition. The approximation matrix of $\mathbf{X}^{+}$:

$$\hat{\mathbf{X}}^{+} = \hat{\mathbf{V}}\mathbf{\Sigma}'\mathbf{U}^{T}, \tag{5}$$

where $\hat{\mathbf{X}}^{+}$ is composed of $r \times p$ consist of first p column of matrix $\mathbf{U}^{T}$. According the idea of PIL, we take matrix $\hat{\mathbf{X}}^{+}$ as the encoder matrix $\mathbf{W}_e = \hat{\mathbf{X}}^{+}$. Then the hidden layer output:

$$\mathbf{H} = \sigma(\mathbf{X}\hat{\mathbf{X}}^{+}), \tag{6}$$

3) Decoder weight. The output of an autoencoder is approximately equal to the input, that is $\mathbf{X} = \mathbf{O}$. According the optimization target of the pseudoinverse learning algorithm, the loss function is:

$$E = \|\mathbf{H}\mathbf{W}_d - \mathbf{X}\|^2, \tag{7}$$

To avoid over fitting, we add a weight decay regularization, the loss function shown as follow:

$$E = \frac{1}{2}\|\mathbf{H}\mathbf{W}_d - \mathbf{X}\|^2 + \frac{k}{2}\|\mathbf{W}\|^2, \tag{8}$$

Where k is a regularization parameter setting by user.

$$\mathbf{W}_d = (\mathbf{H}^T\mathbf{H} + k\mathbf{I})^{-1}\mathbf{H}^T\mathbf{X}, \tag{9}$$

where $(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the psuedoinverse matrix of $\mathbf{H}$.

After the decoder weight is calculated, we tie decoder weight and encoder weight together which takes transposed decoder weight matrix as the encoder weight matrix:

$$\mathbf{W}_e = (\mathbf{W}_d)^T. \tag{10}$$

Weight tied reduce the parameters of the network and increase the regularization to avoid over fitting.

III. Method

In this section, the structure of our model will be described in detail. Besides, we will briefly introduce the data preparation and the specifics of training in this work.

A. Structure

The four diagnostic feature plots are an important basis of human experts on identity a pulsar. Therefore, emulating human experts, we consider to build a model that could efficiently combine all the information about these feature plots. The structure of our model is described in Fig. 2. As show in the figure, the model has two layers, and in the first layer, it could be divided into four sub-modules that correspond to four kinds of feature plots. A sub-module which has the similar structure of MoE, consists of a group of expert classifiers and one gating network. These expert classifiers are generated using bootstrap sampling methods. Each expert classifier rates how "pulsar-like" a particular feature plot of the candidate is with a number between 0 (not a pulsar) and 1 (a pulsar). And the gating network is responsible for combining the results of all classifiers in its sub-module. Four sub-modules would give four prediction matrices that are the outputs of the first layer. Then these predict results are fed into the second layer's classifier that evaluates which sub-modules prediction is more reliable and form the final prediction that determines the candidate's type. We know that introducing randomness into the model through the bagging method can make the model avoid the problem of overfitting and improve its generalization ability.

It should be noted that the classifiers and gating networks in our model could use many kinds of machine learning algorithms. Considering efficiency and accuracy, in this work we adopted single-hidden layer feedforward neural network training via PIL as expert classifiers in module 1 and module 4, and HOG+PILAE as expert classifiers in module 2 and module 3. The gating networks and the second layer classifier are also single-hidden layer feedforward neural networks training via PIL.

B. Data Preparation and Training

For the task of pulsar candidate selection, the data of pulsar candidate usually be folded into a group of diagnostic plots against a pipeline software. Pulsar candidate diagnostic representations mainly include summed profile histogram, time-vs-phase plot, frequency-vs-phase plot and dispersion-measure curve. These are also the four most important features that human experts look at when classifying candidates and similarly they are the inputs to our model. For our model to work on these plots, the features should have the same size and scale, so we have to down-sample or interpolate the data onto a uniform size. For 1D data, the dimension was compressed into 100 numbers using piecewise linear interpolation and for 2D data we use cubic spline Interpolation to compress it into the size 48×48. We also normalize the data to zero median and unit variance to remove the absolute scale of the plots. For the 2D image arrays, normalization is performed line by line along the phase axis, which removes instrumental variations over the course of the observation and across the observing band but maintains the variance in signal across the phase this should be dominated by the pulsar signal.

Once the features are well prepared, we can use them to train our model. The training process of the model is implemented in a layer-by-layer way. In the first stage, the first layer's classifiers are trained on four features of each sample in the training set. Since we have the trained classifiers, then we can train the gating network to learn how to assemble the outputs of classifiers. After the first layer is trained, then we combine the four outputs from the first layer and feed into the second layer classifier to getting the final result. The inference is a forward propagation process after we get a trained model.

TABLE I
NUMBER OF EXAMPLES IN TWO PULSAR CANDIDATE DATASETS

Datasets	#Positive	#Negative	#Total Examples
HTRU medlat	1,196	89,996	91,192
PMPS-26k	2,000	22,000	26,000

TABLE II
BINARY CLASSIFICATION CONFUSION MATRIX

Outcomes	Prediction -	Prediction +
Groundtruth -	True Negative (TN)	False Positive (FP)
Groundtruth +	False Negative (FN)	True Positive (TP)

IV. Experiment

To evaluate the performance of our model, we examine two pulsar candidate datasets. In this section, we will introduce the details of experiments and discussing the results.

A. Datasets and Evaluation Metrics

The experiment is trained and tested on two pulsar candidate datasets: HTRU medlat and PMPS-26k. HTRU medlat dataset is the first public pulsar candidate dataset form the intermediate Galactic latitude area of the HTRU survey which was introduced by Morello et al. [15]. This dataset consists of 1196 positive samples which are observed from 521 distinct pulsars, and 89996 negative samples which are non-pulsar candidates. PMPS-26k is another pulsar candidate dataset which was built on the PMPS observation data by Manchester et al. [19]. It comprises 2000 positive samples that are obtained from real pulsars, 2000 negative samples which comes from non-pulsar signals, 20000 radio frequency interference signals which are also denoted as negative examples, and 2000 samples that were labeled as unknown. The number of examples in each dataset is listed in Table.I.

In consideration of the imbalance between positive samples and negative ones, we could use F1 score, that is defined as the harmonic average of precision and recall rate, to evaluate the performance of our model. The binary classification confusion matrix is commonly used to calculate the F1 score which is presented in Table. II. And it is calculated as follows:

$$Precesion = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F1score = \frac{2 * Precesion * Recall}{Precesion + Recall} \tag{13}$$

Because of random fluctuations, the F1 score from validation tests may vary from test to test. Hence we repeat our experiment to find a reliable estimate.

B. Implementation Details

Observing the two pulsar candidate datasets, we found that the positive samples are much less than the negative ones. Therefore, we should randomly sample the same number of negative samples as positive ones first. Then the all positive samples and selected negative samples are randomly split into three folds for training, validation and testing. The proportion of the parts are 40%, 30%, 30%. And all the unselected negative samples are collected into the testing set.

In the first layer of our model, we setup n classifiers at each module. Specially, n is equal to 5 in the experiment. So there are 20 classifiers in all. These classifiers should be biased so that each classifier could learn different distribution of datasets. In order to achieve this goal, we randomly split the training set into n non-overlapping parts. When we train the i-th classifier, we could use all the training set except the i-th part.

In our model, we use 3 kernels extract HOG feature maps and three layers for PILAE as our HOG+PILAE experts, meanwhile we use single hidden layer neural networks as our PIL experts. Besides, we built a baseline model that used CNN and BPNN as expert classifiers. CNN experts are used on 2D data (time versus phase and frequency versus phase plots) and BPNN experts are used on 1D data arrays (summed profile or DM curve). And we also use BPNN as its gating networks and second layer classifier. The structure of the CNN is similar to LeNet-5. The BPNN experts have one hidden layer. All the best hyper-parameters of neural networks are determined through cross-validation tests. We conduct all the experiments on the same computer with 6 Xeon 2.00GHz processors.

C. Performance of Models

The main results on the testing sets of two datasets are listed in the Table. III. In the first layer of our model, we have four sub-modules corresponding four kinds of feature plots. Each sub-module would give an output of how pulsar like the candidate is. The experts classifiers in sub-modules have different F1 score. For the BPNN & CNN model, the range of F1 score of the individual first layers expert classifiers are from 88% to 94% on HTRU medlat datasets. While for the PIL & HOG+PILAE model, the range of F1 score of the individual first layer expert classifiers are from 85% to 92%. And the two models final F1 score achieved 95.44% and 94.65%. It shows that our structure could improve the performance of signal classifiers. Compare the performance on two datasets, the BPNN & CNN model performed a little better, however it took much longer time even more than five times longer than the PIL & HOG+PILAE model on training.

V. Conclusion

On consideration of big data and the imbalance class data for the task of data mining in the astronomy which most researchers have to face, we proposed a hierarchical model which is trained with the PIL algorithm to solve the problems and improve model's generalization performance. The architecture of the model is designed by taking into account the characteristics of the Pulsar candidate samples, and adopts the strategy of divide and conquer to deal with big data problem. And for speed up the training process, we utilize PIL algorithm which can avoid the problems of local minima, hyper parameters selection, gradient vanish, and

TABLE III

MAIN RESULTS ON THE TWO DATASETS

Datasets	Models	Precision(%)	Recall(%)	F1-score(%)	Training time (s)
HTRU medlat	BPNN & CNN	95.33	95.56	95.44	1,697.05
	PIL & HOG+PILAE	93.59	95.74	94.65	319.41
PMPS-26k	BPNN & CNN	89.10	90.14	89.61	2,234.35
	PIL & HOG+PILAE	87.83	87.50	87.66	406.87

so on. Experimental investigation on two pulsar candidate datasets demonstrates that our model has an advantage in learning speed, meanwhile could achieve a high steady-state precision.

Acknowledgment

This work is fully supported by National Natural Science Foundation of China (NSFC) (61472043) and Joint Research Fund in Astronomy under cooperative agreement between the National Natural Science Foundation of China (NSFC) and Chinese Academy of Sciences (CAS) (U1531242). Prof. Ping Guo and Qian Yin are the authors to whom all the correspondence should be addressed.

References

[1] J. M. Cordes, P. Freire, D. R. Lorimer, F. Camilo, D. J. Champion, D. J. Nice et al., “Arecibo pulsar survey using alfa. i. survey strategy and first discoveries,” *The Astrophysical Journal*, vol. 637, no. 1, pp. 446, 2006.

[2] M. J. Keith, A. Jameson, W. Van Straten, M. Bailes, S. Johnston, M. Kramer et al., “The high time resolution universe pulsar survey–i. system configuration and initial discoveries,” *Monthly Notices of the Royal Astronomical Society*, vol. 409, no. 2, pp. 619-627, 2010.

[3] J. Boyles, R. S. Lynch, S. M. Ransom, I. H. Stairs, D. R. Lorimer, M. A. McLaughlin et al., “The Green Bank Telescope 350 MHz drift-scan survey. I. Survey observations and the discovery of 13 pulsars,” *The Astrophysical Journal*, vol. 763, no. 2, pp. 80, 2013.

[4] M. I. Jordan, “Hierarchical mixtures of experts and the EM algorithm,” *Neural computation*, vol. 6, no. 2, pp. 257-290, 1993.

[5] D.H. Wolpert, “Stacked generalization”, *Neural Networks* Vol. 5, pp. 241–259, 1992.

[6] P. Guo, C.L.P. Chen and Y.G. Sun, “An Exact Supervised Learning for a Three-Layer Supervised Neural Network,” in *Proc. Proceedings of the International Conference on neural Information Processing* (ICONIP 1995), Beijing, China, pp.1041-1044, 1995.

[7] P. Guo, M. R. Lyu, and N. E. Mastorakis, “Pseudoinverse learning algorithm for feedforward neural networks,” *Advances in Neural Networks and Applications*, no. 321-326, 2001.

[8] S. Feng, S. Li, P. Guo, and Q. Yin, “Image Recognition with Histogram of Oriented Gradient Feature and Pseudoinverse Learning AutoEncoders,” in *Proc. The 24th International Conference on Neural Information Processing* (ICONIP 2017), Guangzhou, China, Nov. 2017, pp. 740-749.

[9] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*(CVPR 2005), San Diego, CA, USA, June. 2005, pp. 886-893.

[10] K. Wang, P. Guo, Q. Yin, A. Luo, and X. Xin, “A pseudoinverse incremental algorithm for fast training deep neural networks with application to spectra pattern recognition,” in *Proc. International Joint Conference on Neural Networks, Vancouver* (IJCNN 2016), Canada, July. 2016, pp. 3453-3460.

[11] K. Wang, P. Guo, and A. Luo, “A new automated spectral feature extraction method and its application in spectral classification and defective spectra recovery,” *Monthly Notices of the Royal Astronomical Society*, vol. 465, no. 4, pp. 4311-4324, 2016.

[12] K. Wang, P. Guo, A. Luo, X. Xin, and F. Duan, “Deep neural networks with local connectivity and its application to astronomical spectral data,” in *Proc. IEEE International Conference on Systems, Man, and Cybernetics* (SMC 2016), Budapest, Hungary, Oct. 2016, pp. 2687-2692.

[13] P. Guo, K. Wang, and X. Xin, “Autoencoder, Low Rank Approximation and Pseudoinverse Learning Algorithm,” in *Proc. IEEE International Conference on Systems, Man, and Cybernetics* (SMC 2017), Banff, Canada, Oct. 2017, pp. 948-953.

[14] K. J. Lee, K. Stovall, F. A. Jenet, J. Martinez, L. P. Dartez, A. Mata et al., ”PEACE: pulsar evaluation algorithm for candidate extraction–a software package for post-analysis processing of pulsar survey candidates,” *Monthly Notices of the Royal Astronomical Society*, vol. 433, no. 1, pp. 688-694, 2013.

[15] V. Morello, E. D. Barr, M. Bailes, C. M. Flynn, E. F. Keane, and W. van Straten, “SPINN: a straightforward machine learning solution to the pulsar candidate selection problem,” *Monthly Notices of the Royal Astronomical Society*, vol. 443, no. 2, pp. 1651-1662, 2014.

[16] W. W. Zhu, A. Berndsen, E. C. Madsen, M. Tan, I. H. Stairs, A. Brazier et al., “Searching for pulsars using image pattern recognition,” *The Astrophysical Journal*, vol. 781, no. 2, pp. 117, 2014.

[17] P. Guo, F. Duan, P. Wang, Y. Yao, and X. Xin, “Pulsar Candidate Identification with Artificial Intelligence Techniques,” arXiv preprint arXiv:1711.10339, 2017.

[18] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, “Adaptive mixtures of local experts,” *Neural computation*, vol. 3, no. 1, pp. 79-87, 1991.

[19] R. N. Manchester, A. G. Lyne, F. Camilo, et al, “The Parkes multibeam pulsar survey–I. Observing and data analysis systems, discovery and timing of 100 pulsars,” *Monthly Notices of the Royal Astronomical Society*, vol. 328, no. 1, pp. 17-35, 2001.