Shuffled Frog Leaping Algorithm Based Neural Network and Its Application in Big Data Set

Wei Shan, Shaohua Cui, Jiangtao Wang
School of Physics and Electronic Information, Huaibei Normal University, Huaibei 235000, China
shanwei@chnu.edu.cn

Abstract: This paper studies the random optimization algorithm to reorganize the leapfrog algorithm. The big data challenge requires an effective optimization algorithm to explore potential data structures using deep neural networks. At first, we introduce the neural network classifier and compare it with the support vector machine. Neural networks are suitable for large data sets and have the complex ability to extract high-level abstract data. And then we have to introduce a large dataset covering cancer data and voice data. Both datasets have large numbers of samples with complex low-level variance. At last we have to use the reorganized leapfrog algorithm to optimize neural network parameters. The random leapfrog algorithm is efficient and robust to a local minimum. The experimental results show that the algorithm has extensive application prospects and is suitable for the classification of big dataset. The neural network parameters can effectively optimized by the improved shuffled frog leaping algorithm.

1. Introduction
The neural network method has attracted more and more research attention and has been widely used in classification of many real-scenario data problems [1-5]. Classification of large dataset is very different from small dataset. Traditional algorithms for classification such as Support Vector Machines (SVM) etc are only effective on small datasets [6,7].

SVM are known for training and modeling with small samples, but on large datasets, their generalization ability is limited. In addition, deep neural networks can be able to build meaningful advanced functions for large datasets. It needs a lot of data for training. So it’s important to search the optimization algorithms that can efficiently deal with random searches on low-level original data by optimizing the neural network structure at the same time.

The use of large datasets will change the performance of the support vector machine classifier because the training process must deal with more parameters so the speed of convergence become slow. Nowadays the Shuffled Frog Leaping Algorithm (SFLA) [8-13] is receiving more and more attention in project optimization. SFLA has a powerful global optimal search capability and is more efficient than existing optimization algorithms. In our article, we suggest using the improved SFLA to configure neural networks parameters, and use the BP algorithm to compare with support vector machines and basic neural networks. We also recommend using the improved shuffled frog leaping algorithm to find out the global optimal value to initialize the neural networks classifier. The whole system flow chart is shown as in Fig. 1.

The rest of our article is arranged below:
Part 2 gives a basic and brief introduction to deep neural networks. Part 3 describes the reorganized leapfrog algorithm. Part 4 is a concise introduction to the dataset used in the experiment, the experimental results are demonstrated in Part 5. Finally, we provide conclusions in Part 6.

Figure 1. Flowchart of the training recognition system.

2. Experimental method
The flow chart of the training recognition system is shown in Fig. 1. It shows that classification algorithm based on basic NN trains parameters from a small database by using shallow structure, and Deep Neural Network (DNN) learns from a big data by using deep network structure [14-17]. However, it often has the problem of parameter convergence. In speech recognition and target recognition, it outperforms many of the most advanced machine learning algorithms. An ideal characteristic of NN is that it can express high-level features and improve modeling accuracy. It’s much more robust to the variability of data.

2.1 neural network
DNN comes from multilayer neural network. It consists of many hidden layers, as shown in Figure 2.

Figure 2. Deep neural network structures for classification of big database
DNN is a kind of supervised learning algorithm, which helps in modeling the posterior probability:

$$ p_{y|x}(y = s | x) $$ (1)

Note that $s$ represents the category, $x$ represents the observation vector and $y$ represents the outputs of the neural networks.

We let the first one $L$ level as a vector $v^l$, where $l$ is the node index. The hidden binary vector is written as $h^l$. Set $h^l_m$ as the hidden unit and $m$ is the unit index. The hidden units total number is $N^l$. The formula of posterior probability calculation is [18]:

$$ p^l(h^l | v^l) = \prod_{m=1}^{N^l} \frac{e^{z^l_m(v^l)h^l_m}}{e^{z^l_m(v^l)} + 1} \text{ (2)} $$

where $z^l(v^l) = (W^l)^T v^l + a^l$. $W$ is the weight and $a$ is the deviation vector. The output layer calculates the posterior probability as:

$$ p_{y|x}(y = s | x) = \sum_{y} e^{z^l_s(v^l)} \text{ (3)} $$

The neural network is initialized by using the Restricted Boltzmann Machine. The multi-layer structure is able to powerfully model the nonlinear relationship in the input signal. However, the backpropagation algorithm doesn’t always converge to the training sets. Initialization is very important for the convergence and successful training of multi-layer neural networks. Therefore, Restricted Boltzmann Machine (RBM) is an effective method to implement unsupervised clustering in every layers, and it also provides a preferable initial position for starting the backpropagation algorithm. Figure 3 shows the relationship between Restricted Boltzmann Machine clustering and backpropagation training.

The neural network structure usually set by means of observation and experience. For general image and speech data modeling, four to six layers of neural network is enough. The number of each layer nodes is related to the characteristics of the input data. The larger the characteristics size, the bigger the number of nodes in hidden layers. When choosing the size of the neural network, we also have to consider the capacity of data storage and cost of computation. The typical setting for the learning rate is about $0.00001$ to $0.005$ for the specific problems. In traditional back propagation iterations, the parameters in every NN layers are optimized.
and they will easily converge with good RBM initialization. It is very important for us to find out more effective DNN training optimization algorithms.

With a well-trained Restricted Boltzmann Machine, we assume that the weight matrix as $W$:

$$ W = \begin{bmatrix} w_{1,1} & w_{2,1} & \cdots & w_{M,1} \\
 w_{1,2} & w_{2,2} & \cdots & w_{M,2} \\
 \vdots & \vdots & \ddots & \vdots \\
 w_{1,N} & w_{2,N} & \cdots & w_{M,N} \end{bmatrix} \tag{4} $$

Where $M$ and $N$ is the number of visible unit and the hidden unit respectively. For any input feature vector:

$$ x = (x_1, x_2, \cdots, x_M) \tag{5} $$

According to Restricted Boltzmann Machine, hidden nodes will be activated or not. The formula to calculate activation is given below:

$$ h = Wx \tag{6} $$

The calculation formula for the open state (value 1) is:

$$ P(h_j = 1) = \sigma(h_j) = \frac{1}{1 + e^{-h_j}} \tag{7} $$

This function takes the form of logistic function:

$$ f(x) = \frac{1}{1 + e^{-x}} \tag{8} $$

Define every hidden nodes and calculate value 0 probability as:

$$ P(h_j = 0) = 1 - P(h_j = 1) \tag{9} $$

Whether to activate the node finally depends on the uniform distribution:

$$ u \sim U(0,1) \tag{10} $$

$$ h_j = \begin{cases} 1, & P(h_j = 1) \geq u \\ 2, & P(h_j = 1) < u \end{cases} \tag{11} $$

The RBM training is based on G. Hinton.’s Contrastive Divergence (CD) algorithm. For each item in training sets, the probability of activating relevant hidden units is calculated by:

$$ P(h_j^{(0)} = 1 \mid v^{(0)}) = \sigma(W_j v^{(0)}) \tag{12} $$

Select a random sample from P:

$$ h^{(0)} \sim P(h \mid v^{(0)}) \tag{13} $$

Use $h^{(0)}$ to rebuild the visible layer:

$$ P(v_i^{(1)} = 1 \mid h^{(0)}) = \sigma(W_i^T h^{(0)}) \tag{14} $$

Randomly select a visible sample:

$$ v^{(1)} \sim P(v \mid h^{(0)}) \tag{15} $$

Finally, hidden units activation probability can be calculated as:

$$ P(h_j^{(1)} = 1 \mid v^{(1)}) = \sigma(W_j v^{(1)}) \tag{16} $$

The weight matrix is updated to:
\[ W \leftarrow W + \lambda (P(h^{(0)}) = 1\mid v^{(0)})v^{(0)T} - P(h^{(1)}) = 1\mid v^{(1)})v^{(1)T} \] (17)

2.2 Shuffled frog leaping algorithm

SFLA is a meta-heuristic optimization algorithm inspired by swarm intelligence which has become a popular topic and is increasingly attracting many research attentions. We adopt the modified shuffled frog leaping algorithm [19,20] in our article to optimize parameters of NN.

The modified shuffled frog leaping algorithm consists of the following steps which is shown in Figure 4:

1) Initial frog population as candidate parameters: the number of frogs ‘N’, the dimension ‘d’ of a single frog, the number of frogs in the subpopulation ‘m’, the number of subpopulations ‘n’, the number of iterations ‘g’, the basic temperature \( T_b \), temperature coefficient ‘c’.

2) We calculate the fitness by using different parameter settings for each frog individual.

3) Implement the immune vaccination operation for each individual frog and this is the difference between our algorithm to the original shuffled frog leaping algorithm operation.

4) N frogs were divided into n sub populations according to their fitness from good to bad.

5) According to the fitness level from high to low, N frogs are divided into n sub populations.

a) For each subgroup, perform the following steps g times.

b) Renew the best individual \( X_{w} \) in each sub-population;

c) Renew \( X_{w} \) using the basic shuffled frog leaping algorithm strategy, \( X'_{w} = X_{w} + \Omega \), where \( \Omega = \text{rand}(N) (X_{b} - X_{w}) \);

d) Execute Gaussian mutation on individuals;

e) Execute Tent chaotic mapping on individuals.

6) After the basic depth search is completed and if the reference temperature \( T_{b} \) is reached then output the global optimum and end the iteration otherwise, let \( T = c \times T_{b} \) and go to 3).

Shuffled frog leaping algorithm based parameter optimization can replace the traditional initialization method and the initialized NN will use sufficient training data for retraining to achieve fine adjustment of parameters.

![Figure 4. Algorithm flowchart of modified SFLA](image-url)
The random leapfrog algorithm can jump out of the local optimum and find the global optimum. Although generations of mutations, random search methods can normally approximate the target function.

Table 1 shows the mapping between immune principle and computational optimization. Its remarkable advantages are high efficiency of searching and easy to implement. Through combine the 4 operations, the ability to avoid local optima is enhanced. As suggested in the past research report [21,22], it is suitable to optimize the NN parameters and can effectively avoid local optimization.

| Immune Principles | Computational Optimization |
|-------------------|---------------------------|
| Antigen           | Optimization Problem      |
| B Cell            | Candidates                |
| Memory Cell       | Preferable Solution       |
| Antibody          | Target Function           |
| Clone Selection   | Selection among population|

2.3 the dataset

The big dataset used in our article includes three main data sets, namely breast cancer dataset, wine dataset and vowel dataset. From a classification perspective, these datasets contain a lot of training and test samples, so they are suitable for evaluating classification algorithms [23-26]. It also covers varieties of different samples and is appropriate for verifying the generalization ability of the algorithm.

UCI big datasets are used for training the network and testing it. This is one of the three fields provided by the Institute of Oncology, which has repeatedly appeared in the machine learning literature. The dataset includes 2000 instances of one category and 800 instances of another category. These instances are described by nine attributes, some of which are nominal and some are linear.

The dataset is composed of image, acoustic and medical data samples. So, our proposed system is verified by extending it to different fields. Because our optimization method does not depend on specific data structure so it shows good generalization ability in practical application.

3. Experimental results

In the experiment, the datasets are divided into training dataset and test dataset. Through cross validation experiments, we can obtain reliable results. Use tenfold cross-validation, 90% of the data is used for training networks and the rest 10% is used for test. The platform based one 1.5GHz CPU and 4GB RAM [27,28].

Tables 2 to 4 shows the classification results below. The NN, SVM and SFLA algorithm based on NN are compared. From the test results, we can observe that the shuffled frog leaping algorithm based classifier gives the best results. The random optimization algorithm is appropriate for classifying big datasets. Verified by the cross-validation method, the NN parameters are optimized, and the generalization ability is acceptable. To avoid overfitting we used 10 fold cross validation to demonstrate this capability.

| Recognition Rate | Convergence Speed | Generalization Ability |
|------------------|------------------|------------------------|
| NN 88.1%         | 42.3s            | Poor                   |
| SVM 89.2%        | -                | Poor                   |
| SFLA 93.3%       | 30.2s            | Good                   |

| Recognition Rate | Convergence Speed | Generalization Ability |
|------------------|------------------|------------------------|
| NN 82.3%         | 33.4s            | Poor                   |
| SVM 87.1%        | -                | Good                   |
| SFLA 94.1%       | 29.3s            | Good                   |
4. Conclusions
This article studies the latest stochastic optimization algorithm and applies the parameter optimization method based on SFLA to the classifier based on neural network. Experiments on large datasets show that the results are advantageous and superior to traditional SVM and NN.

The improved SFLA algorithm has powerful random search ability and suits NN modeling. Additionally, SVM is limited to smaller datasets and more dependent on kernel selection. Our experimental results show that the optimization method has great potential in mining the big data. In future research, we will study more datasets and compare SFLA with other algorithms of random search.

Acknowledgments
This work was financially supported by National Natural Science Foundation of China (No.61976101), New generation information technology innovation project of China University industry university research innovation fund of science and technology development center of Ministry of Education(2019ITAO1041), Industry-University-Research Cooperation Project(No. RD/PY-01-2019, RD/ZD-01-2019, RD/YW-01-2019, RD/PY-01-2020, RD/BC-01-2020), and National innovation and entrepreneurship training program for College Students(202010373050).

References
[1] Xiao Z, Wu D, Zhang X, et al. Speech emotion recognition cross language families: Mandarin vs. western languages[C]// International Conference on Progress in Informatics & Computing. IEEE, 2016.
[2] Siegert I, David Philippou-Hübner, Hartmann K, et al. Investigation of Speaker Group-Dependent Modelling for Recognition of Affective States from Speech[J]. Cognitive Computation, 2014, 6(4):892-913.
[3] Chaudhary G, Srivastava S, Bhardwaj S. Feature Extraction Methods for Speaker Recognition: A Review[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2017, 31(12):1750041.1-1750041.39.
[4] Gu H Y, Tsai S F. A Voice Conversion Method Combining Segmental GMM Mapping with Target Frame Selection[J]. Journal of Information and Engineering, 2015, 31(2):609-626.
[5] Zhang J, Yin Z, Wang R. Pattern Classification of Instantaneous Cognitive Task-load Through GMM Clustering, Laplacian Eigenmap, and Ensemble SVMs[J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2017:947-965.
[6] Pravena D, Govind D. Development of simulated emotion speech database for excitation source analysis[J]. International journal of speech technology, 2017, 20(2):327-338.
[7] Lian Z, Li Y, Tao J, et al. Speech Emotion Recognition via Contrastive Loss under Siamese Networks[C]// the Joint Workshop of the 4th Workshop. 2018.
[8] Li Z, Cao J, Zhao X, et al. Swarm intelligence for atmospheric compensation in free space optical communication—Modified shuffled frog leaping algorithm[J]. Optics & Laser Technology, 2015, 66:89-97.
[9] Singh R R, Conjeti S, Banerjee R. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals[J]. Biomedical Signal Processing & Control, 2013, 8(6):740-754.
[10] Balconi M, Grippa E, Vanutelli M E. What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing[J]. Brain Cogn, 2015, 95(apr.):67-76.

[11] Jiyoung K, Jongkuk L, Changho K. Emotion collector, a wearable multi-sensor band to recognize fear[J]. Journal of Intelligent and Fuzzy Systems, 2018, 35(1):1-7.

[12] Xiujun Z, Keqin L, Zhi-Ping L, et al. NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference[J]. Bioinformatics, 2013, 29(1).

[13] Sarkheyli A, Zain A M, Sharif S. The role of basic, modified and hybrid shuffled frog leaping algorithm on optimization problems: a review[J]. Soft Computing, 2015, 19(7):2011-2038.

[14] Lippmann R. Book Review: "Neural Networks, A Comprehensive Foundation", by Simon Haykin[J]. International Journal of Neural Systems, 2011.

[15] Zhang Z, Geiger, Jürgen, Pohjalainen J, et al. Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments[J]. Acm Transactions on Intelligent Systems & Technology, 2017, 9(5):1-28.

[16] Miao Y, Zhang H, Metze F. Speaker Adaptive Training of Deep Neural Network Acoustic Models Using I-Vectors[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23(11):1938-1949.

[17] Yang L, Pong T K, Chen X. Alternating Direction Method of Multipliers for A Class of Nonconvex and Nonsmooth Problems with Applications to Background/Foreground Extraction[J]. Mathematics, 2015, 10(1):74-110.

[18] Tzirakis P, Zhang J, Schuller B. End-to-End Speech Emotion Recognition Using Deep Neural Networks[C]// 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

[19] Sarkheyli A, Zain A M, Sharif S. The role of basic, modified and hybrid shuffled frog leaping algorithm on optimization problems: a review[J]. Soft Computing, 2015, 19(7):2011-2038.

[20] Mossaiby F, Bazzpach M, Shojaei A. Extending the method of exponential basis functions to problems with singularities[J]. Engineering Computations, 2015, 32(2):406-423.

[21] Lu Y, Carin L, Coifman R, et al. Quantitative Arbor Analytics: Unsupervised Harmonic Co-Clustering of Populations of Brain Cell Arbors Based on L-Measure[J]. Neuroinformatics, 2015, 13(1):47-63.

[22] Sun Q, Zhang J, Zhu P, et al. Robust Covariance Representations with Large Margin Dimensionality Reduction for Visual Classification[J]. IEEE Access, 2018, 6:5531-5537.

[23] Nitta K. Study of Discussion Record Analysis Using Temporal Data Crystallization and Its Application to TV Scene Analysis[J]. Bmc Medical Research Methodology, 2014, 15(1):1-12.

[24] Abbayapala T D, Gupta A. Spherical Harmonic Analysis of Wavefields Using Multiple Circular Sensor Arrays[J]. Audio, Speech, and Language Processing, IEEE Transactions on, 2010, 18(6):P.1655-1666.

[25] Rakotomamonjy A, Gasso G. Histogram of gradients of Time-Frequency Representations for Audio scene detection[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 23(1):142-153.

[26] Jiang W, Liu P, Wen F. An improved vector quantization method using deep neural network[J]. AEU - International Journal of Electronics and Communications, 2017, 72:178-183.
[27] Vafeiadis A, Kalatzis D, Votis K, et al. Acoustic scene classification: from a hybrid classifier to deep learning[C]// Detection & Classification of Acoustic Scenes & Events. 2017.

[28] Ren Z, Qian K, Zhang Z, et al. Deep Scalogram Representations for Acoustic Scene Classification[J]. IEEE/CAA Journal of Automatica Sinica, 2018, 5(3).