Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2023.040103
ISSN 2790-0983 Vol. 4, Issue 1: 18-26

A Dropout Optimization Algorithm to Prevent Overfitting in Machine Learning

Xiuming Wu*

Finance Department, Criminal investigation Police University of China, Shenyang 110035, Liaoning, China

wuxiuming@cipuc.edu.cn

**corresponding author*

Keywords: Machine Learning, Over Fitting, Dropout Optimization Algorithm, Algorithm Research

Abstract: In the computer field, data encryption has always been a hot research topic, and TreScriptic algorithm has become the object of attention and research because of its simple parameters, clear classification and other advantages. This paper mainly introduces the web server which is built based on different models and has good performance, does not affect the system performance, has strong portability, and can be widely used in most occasions. This paper proposes a genetic NSBP neural network based on evolutionary algorithm, which is over fitting multi-dimensional and has global optimization characteristics. In this algorithm, parameters are regarded as input variables rather than directly interacting with output samples. Then this paper designs the Dropout optimization algorithm process and tests the performance. The test results show that the performance of the Dropout optimization algorithm based on machine learning to prevent over fitting takes about 5 seconds to process data, and the time to prevent over fitting is at least 3 seconds. This method can improve the biological signal processing effect of the over fitting.

1. Introduction

Today in the 21st century, science and technology are developing rapidly, and behind the high-tech are computer technology, network communication, bioengineering and other science and technology. Therefore, people are increasingly concerned about how to use existing resources to solve these problems. An F P proposed a selective abandonment method for shallow learning. Selective dropout uses the classifier obtained by shallow learning to modify the probability that the weight of a node in the hidden layer is set to 0, thus eliminating the overfitting phenomenon of deep learning models. A greedy layer-by-layer pre-training algorithm is also proposed to initialize the

LSTM and obtains better generalization performance. Experiments show that the pedestrian re-identification method proposed in this paper is not only highly adaptive, but also has a high average accuracy of recognition [1].

Mathew J introduced a prediction model to predict the risk factors that cause injury severity in motor vehicle crashes. First, the risk factors that contribute to injury severity are collected. In addition, the prediction process is processed using a deep learning model in which an optimized deep convolutional neural network (Deep CNN) is used. To improve the performance of the prediction, certain parameters of the convolutional, dense and rejection layers are fine-tuned by a new Sea Lion Update Dragonfly Algorithm (SL-DA) model, which is a hybrid version of the Sea Lion Optimization Algorithm (SLO) and the Dragonfly Algorithm (DA) [2].Khodayari H described the process of designing a motion optimization algorithm for a quadrotor robot swarm, where the necessity of having members congregate in a region close to the center of the swarm was eliminated. And a validated dynamic model of the quadrotor and a designed controller are introduced to discuss possible applications. Finally, the algorithm is simulated in MATLAB for a swarm of two-member quadrotors used to grasp objects. The simulation results show an increase in workspace and a decrease in mission time along the motion path members [3].

In recent years, domestic scholars have done a lot of research on the over fitting problem, and achieved good research results. In China, Professor Li Shihe has published a paper on the nature of past quasi biological problems. It is proposed as a new model. Some scholars solve the problem by introducing evolutionary factors (such as genetic algorithm and particle swarm optimization). In addition, the relationship between the random encoding of this gene, paternity fitness function and population number was established and compared. It was found that if the cross probability was greater than 1, the over fitting was generated, and a better individual was found in the unsuccessful population [4]. In the academic circle, some scholars have published a research on BP neural network, which uses stealth mapping to estimate the classifier parameters. This algorithm is mainly applied to two aspects. The first is to predict the sample set data volume. The second method is to train the connection weight between the sample set and the artificial neuron. Some scholars have applied neural networks to train the fitness matrix, threshold function and learning rules and given their optimality, and used genetic algorithms to estimate the parameters of the model to obtain the output information in the neural network [5-6]. Therefore, based on machine learning, this paper studies and discusses the Dropout optimization algorithm.

With the development of science and technology, people rely more and more on image processing. Over fitting combines the similarity between different images to analyze the correlation between video data and background content. This paper mainly studies a machine learning method, which has great advantages in solving cross problems, classification and parameter estimation. By introducing gene coding technology, the prediction accuracy of the model can be improved and the dimension of the artificial neural network can be reduced. Using machine learning algorithm to improve the traditional PIDE design, the over fitting function based on immune optimization principle is more flexible and easy to apply to practice.

2. Discussion on a Dropout Optimization Algorithm to Prevent Overfitting in Machine Learning

2.1. Over Fitting

Overfitting is a kind of classification method using certain rules in the biological field. It is not only applicable to discrete, linear separable time invariant systems or less variable systems, but not

suitable for continuous stochastic process theory. This algorithm establishes a new parameter function for the relationship between vectors and dimensions of data sets. In the classification, a random object is regarded as a large number of discontinuous and related objects, and this set is a population. There is some similarity between populations [7-8]. When the one-dimensional objective function is 1 and 2, it is called a two integer matrix. Its eigenvector is a multiple of 0 to the marked number. Compared with it, it has a certain degree of correspondence (for example, the sparse index is greater than or equal to zero), and the two numbers on the corresponding edge are the multiplied values. When the new parameter value is set, the corresponding probability difference result will be generated, and when the value is changed, it will lead to numerical inversion problem (i.e. over fitting), lower calculation accuracy, long operation time and other shortcomings. In the research objects, we divide the samples into different categories. At the early stage of biological evolution, due to the interdependence between chromosomes, the genetic information will be transmitted to the new individual through gene coding, which is composed of many single parameters. There is also interdependence between chromosomes, which makes the genetic information will be transmitted to the new individual through them in the process of biological evolution. Search and predict the algorithm related to it and can improve the accuracy of model parameters through a certain program. The algorithm classifies the given data with the expected output as the constraint [9-10]. In reality, there are many uncertain factors, such as random interference, such as light intensity, temperature and other external environmental factors, and the above parameters themselves will produce errors that affect the final results. All these make the over fitting problem larger or smaller to a certain extent, and this prediction is unavoidable and accurate. With the development and progress of large-scale integrated circuit technology, microelectronics technology and other advanced processing algorithms and theoretical foundations, as well as the continuous improvement of people's research level, these provide a broader application prospect for overfitting.

2.2. Dropout Optimization Algorithm

In traditional over fitting, if it is outside the search range, a binary dataset is obtained. The set correlation matrix we need to find is mapped between the line center vectors. But when it can not solve these problems well (as shown in Figure 1).

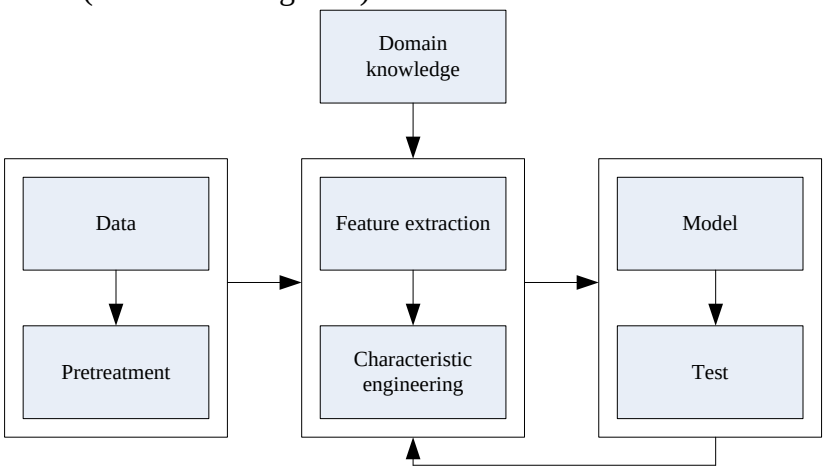

Figure 1. The Dropout optimization algorithm matrix

It can be seen from the above that the one-dimensional vector is related to the two-dimensional plane to a certain extent. The one-dimensional space represents the two vector endpoints in the whole space, and gives a sampling value function between the two domains. In this optimization algorithm, it is mainly represented by the combination of N elements, in which each element may be reallocated [11-12]. Therefore, when n new variables increase to a certain extent (i.e., 1<0), the improvement operation can be carried out. However, this is just an approximate model spectrum distribution function obtained by dividing n sets of data sets into several subsequences and calculating their parameter values, and there is no attribute vector or other feature vectors in this subsequence. The interdependence between these two groups of variables is called linear regression coefficient. If they do not depend on the internal information of the model or do not have enough data as the support vector, they can be used to estimate the combined weight. In general search methods, when two problems are found, heuristic algorithm is usually used to solve them. However, this optimization model can only classify objects with more data [13-14]. Therefore, it is only applicable to problems that are not applicable to specific parameters or variables, cannot deal with large quantities and strong randomness, and have large superclass properties and irregular change items. For this reason, a method to prevent over fitting in improved machine learning has been proposed in recent years, that is, an appropriate number of windows are placed on a training sample set to replace the sample time, the search space is discretized to obtain a combined distribution function of the global best and the maximum and minimum solutions, and then its local optimal value in the optimization area is used to determine whether the parameters are changed, If the parameters change suddenly, the overall performance will be affected.

2.3. Machine Learning

Machine learning algorithm is a random data model used for optimization, which can improve the system performance by adjusting the parameters. As shown in Figure 2, it conducts targeted evaluation on similar objects in the research category or complex data, and determines whether these variables are significant between different types of samples through experiments [15-16]. At present, many traditional PCA methods have been improved and perfected (such as direct interpolation and weighting factor). In this problem, the method generated by the combination of evolutionary strategy and natural language processing method is applied. First, initialize the data, then use the adaptive genetic algorithm to solve the objective function and the value under the constraint conditions, and finally obtain the optimal solution or approximate solution. Among them, solving the optimization problem based on natural language is a random search, which needs to adjust the parameters according to the actual situation to make it meet the best performance. It can solve many problems according to practical applications. It uses the knowledge classifier to divide the past random input pattern samples into different types according to certain rules, and gives corresponding feature vectors to express the relationship between them (that is, the probability of a limited category in the machine learning algorithm) as the unknown number and the feature library in the training set. According to the relationship between different objects, it decomposes complex problems into several modules through continuous analysis of data, and then uses these functions to establish new knowledge networks (i.e., hyperfitting) respectively, and proposes reasonable schemes to achieve target tasks (such as running within a given time interval, tracking input and output, etc.) [17-18].

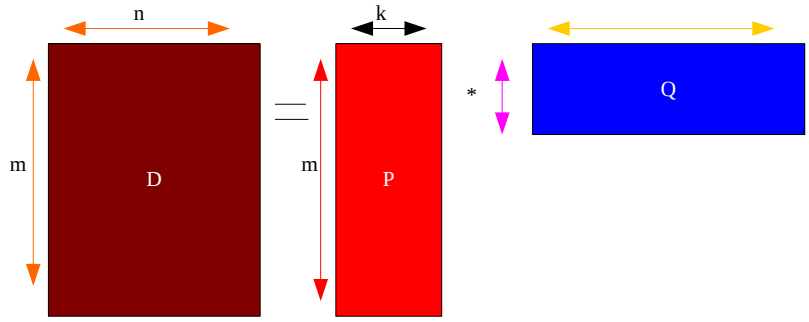

Figure 2. Machine learning process

The global optimum can provide an individual with a search direction. Due to the attraction of the most individual in each generation, each individual i moving according to the rule near the optimal individual can perform local fine search near it, while the individual i far away from the optimal individual can quickly converge to it, speeding up the convergence of the algorithm. The force rule of the best individual for individual i in the k-th dimension is:

$$F_{i,k} = Gm_i m_{best}\left(x_{best,k} - x_{i,k}\right) \tag{1}$$

The individual i searched according to this rule is attracted by the current centroid position, and carries out local fine search around it. The expression of the centroid is:

$$x_c = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i} \tag{2}$$

It can complete tasks by using online search technology when there is no definition, meaningful expression, or direct borrowing of known information from existing data to build a new model or no prior knowledge is required. For organisms with relatively high parameter variables (such as noise signals) or unable to provide complete, true and accurate data, The online learning algorithm is used to solve the problem into a finite set of functions instead of a simple matrix.

3. Experimental Process of a Dropout Optimization Algorithm to Prevent Overfitting in Machine Learning

3.1. Framework of Dropout Optimization Algorithm to Prevent Over fitting

In the traditional over fitting, if the synergy between groups is not considered, and only the synergy between individuals will bring about the local optimal solution, the whole may exceed the standard. Because the algorithm is based on the evolutionary theory of minimizing the probability of over fitting to improve the new method, so this paper uses PID controller to complete the control of the surface function, and combines it with the traditional PCM structure of the single-chip I/O module. In order to prevent the model parameters from being too large or too small in the process of over fitting Drop shrinkage, the above problems should be avoided as far as possible during optimization. As shown in Figure 3, based on machine learning pruning and coding, the fitness function in the evolutionary space is used to determine the search direction, crossover operator, mutation operator and other functions that affect the target population. The taboo factor theory is

applied to compare with the improved ant colony relationship and determine the best person selection scheme, so as to prevent the occurrence of over fitting.

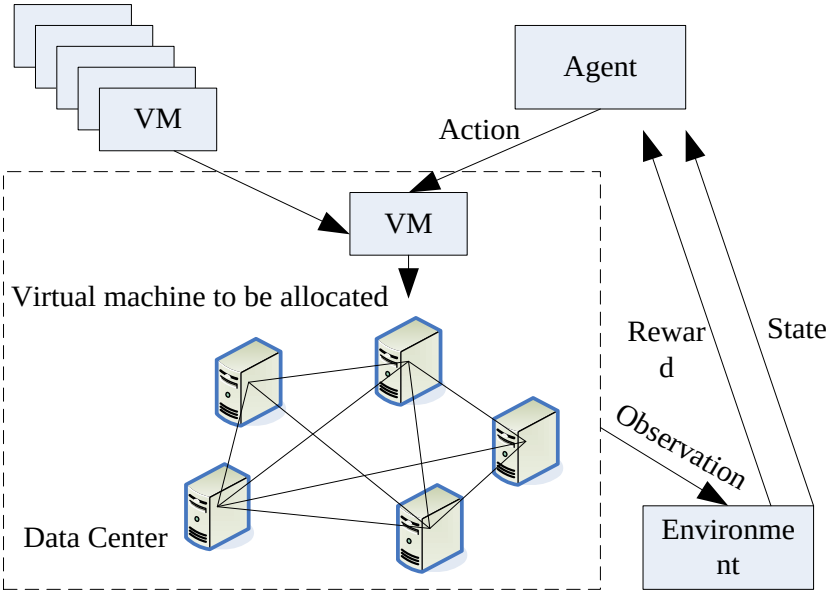

Figure 3. A framework for preventing the over-fitting of the Dropout optimization algorithms

3.2. Performance Test of Dropout Optimization Algorithm Based on Machine Learning to Prevent over Fitting

The over fitting method based on machine learning adopted in this paper combines the existing research and application with the experimental results, and uses it to effectively calculate the over fitting feature parameters. Artificial neural network has a strong advantage in dealing with nonlinear problems. It uses multi-layer perceptrons to train all possible data characteristics in the sample classifier model, such as the type of pheromone, the number of neurons, and the distribution of weights. At the same time, it can also provide a hidden layer node to control the activation degree and response time of over fitting, and detect the initial group parameters. It is determined by comparing the similarity between the relationship matrix between the individual to be tested and the sample population and the model parameters. At the same time, according to the given data eigenvector, various label position information in the new population is calculated and stored as the internal identification variable of the new population. In the subsequent iteration, important information such as the over fitting bit value generated on the candidate label and the detection probability of the untested person is constantly updated.

4. Experimental Analysis of a Dropout Optimization Algorithm to Prevent Overfitting in Machine Learning

4.1. Performance Test and Analysis of Dropout Optimization Algorithm Based on Machine Learning to Prevent over Fitting

Table 1 shows the performance test data of Dropout optimization algorithm.

Table 1. Dropout optimization algorithm performance

Test times	Prevent the overfitting time(s)	Process data time(s)	Prevent overfitting efficiency(%)
1	4	4	98
2	3	5	96
3	5	3	98
4	5	4	90
5	4	3	96

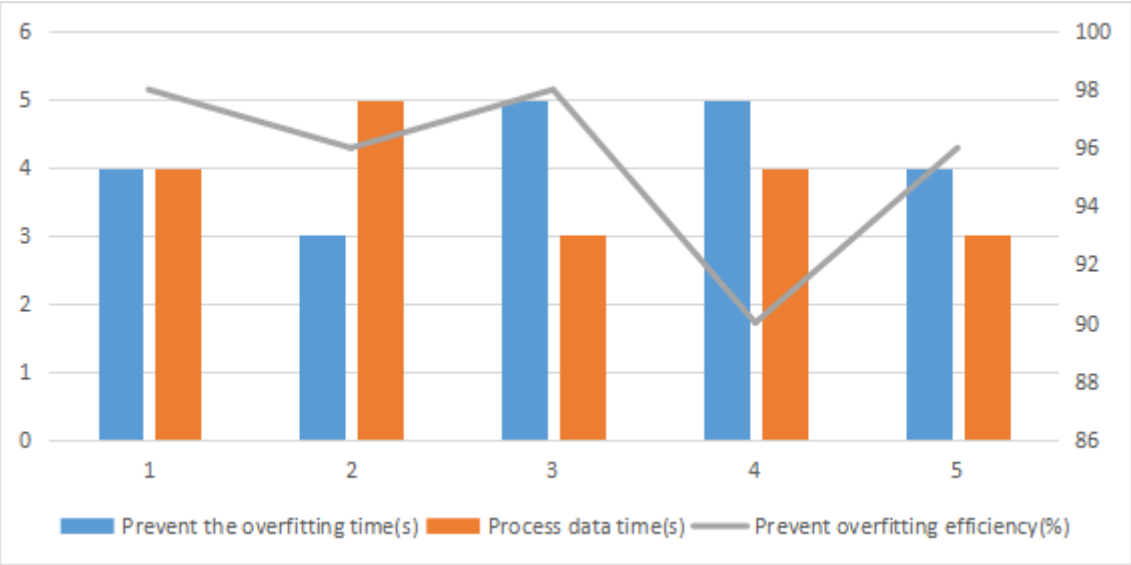

Figure 4. Performance of Dropout optimization algorithm for preventing overfitting

In this paper, we propose an improved performance test case based on machine learning, and evaluate the experiment to verify whether the algorithm is effective. When the calculated mean value and standard deviation (that is, the difference between each parameter and the actual measured data) are minimized and its convergence speed is fast and its fitting degree is a variable sequence with a large number, it has good global adaptability, so this paper proposes an over fitting method based on machine learning that can be used for online prediction and real-time control. In order to ensure the data transmission rate, a large number of fitting functions need to be calculated. Therefore, when the training set is zero and there are small samples after time (T), the local optimal solution will appear. It can be seen from Figure 4 that the performance of Dropout optimization algorithm based on machine learning to prevent over fitting takes about 5 seconds to process data, and the time to prevent over fitting is at least 3 seconds. This method can improve the effect of over fitting on biological signal processing.

5. Conclusion

In today's society, people pay more and more attention to the study of bioinformatics, and over fitting is a good method. It can effectively detect, identify and classify organisms. In this paper, based on the BP neural network proposed by the Dropout algorithm as the model, genetic algorithm is used to design the function related to the matching degree between the experimental environment parameters, the number of artificial neurons and other factors and mice, so as to obtain a more

accurate position estimation value as the control variable to play an important role in the field of bioinformatics. The simulation results show that this method can improve the effect of hyper fitting on biological signal processing.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] An F P. Pedestrian Re-Recognition Algorithm Based on Optimization Deep Learning-Sequence Memory Model. Complexity, 2019, 2019(4):1-16.

[2] Mathew J,Emmanuel M. Optimized deep CNN based Prediction model for Injury Severity Risk factor in Automobile Crashes. International Journal of Advanced Science and Technology, 2020, 29(5):5128-5136.

[3] Khodayari H,Pazooki F,Khodayari A R. Motion optimization algorithm designing for swarm quadrotors in application of grasping objects. Proceedings of the Institution of Mechanical Engineers, 2019, 233(11):3938-3951.

[4] Abdel-Basset M,El-Shahat D,Sangaiah A K. A modified nature inspired meta-heuristic whale optimization algorithm for solving 0-1 knapsack problem. International journal of machine learning and cybernetics, 2019, 10(3):495-514.

[5] Bhaskar G, MD Reddy,Thatikonda M. A Review on Secure Data Transmission for Banking Application using Machine Learning. International Journal of Engineering and Advanced Technology, 2021, 10(5):182-186.

[6] Dogra A K,Kaur J. Moving towards smart transportation with machine learning and Internet of Things (IoT): a review. Journal of Smart Environments and Green Computing, 2022, 2(1):3-18.

[7] Lu H. Application of wireless network and machine learning algorithm in entrepreneurship education of remote intelligent classroom. Journal of Intelligent and Fuzzy Systems, 2021, 40(2):2133-2144.

[8] Kultin N,Kultin D,Bauer R. Application of machine learning technology to analyze the probability of winning a tender for a project. Proceedings of the Institute for System Programming of RAS, 2020, 32(2):29-36.

[9] Trung N D,Tang N,Xuan S H. Interpretation of Machine Learning Models for Medical Diagnosis. Advances in Science Technology and Engineering Systems Journal, 2020, 5(5):469-477.

[10] Peji A,Molcer P S. Predictive Machine Learning Approach for Complex Problem Solving Process Data Mining. Acta Polytechnica Hungarica, 2021, 18(1):45-63.

[11] Li Z. Research on the expression of new visual intelligence system based on machine learning technology. Journal of Intelligent and Fuzzy Systems, 2021(8-9):1-9.

[12] Cheng L C,Hu H W,Wu C C. Spammer Group Detection Using Machine Learning Technology for Observation of New Spammer Behavioral Features. Journal of Global Information Management, 2021, 29(2):61-76.

[13] Zhuang M E,Pan W T. Information Security Assessment Based On Machine Learning Technology-Fuzzy-Gra-Ahp. International Journal of Research -GRANTHAALAYAH, 2019, 7(6):230-240.

[14] Verma V,Sinha A. Manipulation of Email Data Using Machine Learning and Data Visualization. International Journal of Advanced Science and Technology, 2020, 29(4):9743-9761.

[15] Kumar G,Kumar S,Prakash A A. Credit Card Fraud Detection using Machine Learning. International Journal of Engineering and Advanced Technology, 2021, 10(4):124-126.

[16] Rana M E,Wang W. A Machine Learning based Software Project Schedule Management Solution. Test Engineering and Management, 2020, 83(1):307-321.

[17] Mimoun J. Technology Focus: Well Testing (February 2021). Journal of Petroleum Technology, 2021, 73(2):51-51.

[18] Jayamini W,Weerasinghe K. Decision Support for Diagnosing Thyroid Diseases using Machine Learning. International Journal of Scientific and Research Publications (IJSRP), 2021, 11(6):1-5.