A Neural Network Classification Optimized by Genetic Algorithm to Predict the Success of Bank Telemarketing

Qiuyue Wang

Research School of Computer Science, Australian National University u6342378@anu.edu.au

Abstract. This paper uses Neural Network and optimizes through genetic algorithm to design a classification to investigate using a real-world dataset. The classification predicts whether the customer can subscribe long-term deposits through telemarketing based on the customer's personal information and funding conditions. There are two metrics, area of the Receiver Operating Characteristic curve (AUC) and the ratio of the number of predicted correct results and the number of all results (Accuracy) to show the performance. Because of the limitation of my ability, these two metrics (myAUC = 0.66, myAccuracy = 90.18%) are not as good as those given in the research paper. These still prove that the predictive model makes sense for bank telemarketing. This paper also proves that Neural Network classification optimized by genetic algorithm is better than using ordinary Neural Network classification. This paper also describes the limitations of the technology and the direction of future work.

Keywords: Neural Network, Classification, Genetic Algorithm, Bank telemarketing

1 Introduction

Sometimes the customer's personal information and economic status can be used to determine the possibility of his/her deposit. By using machine learning methods, based on the data information provided in the database, a Neural Network classification model can be trained to predict whether a customer instance will subscribe deposits.

Neural network is an algorithm and mathematical model that simulates the brain behavioral characteristics of animal neural networks and classifies the data. By adjusting the interconnection of a large number of internal nodes, the aim of information processing information is achieved. It is an adaptive nonlinear system in which a large number of simple basic elements -- neurons are interconnected [1].

Genetic algorithm is inspired by Darwin's theory of evolution, a heuristic search algorithm by referring to the process of biological evolution [2]. The genetic algorithm simulates the problem to be solved into a process of biological evolution. The solution of the next generation is generated through operations such as copying, crossover, and mutation. The solution with a low fitness function value is gradually eliminated, and the solution with a higher fitness function value is increased. After evolution for several generations, it is very likely that individuals with high fitness function values will evolve.

This paper compares the results of Neural Network with the results of the Neural Network optimized by the genetic algorithm. The Neural Networks have the same topological structure, but the latter's related parameters are optimized by the genetic algorithm. The latter method improves the accuracy of the prediction results.

1.1 Dataset Description

This study is conducted on a dataset provided by the UC Irvine Machine Learning Repository [3]. The dataset is about direct marketing campaigns of a Portuguese banking institution and the marketing campaigns are based on phone calls [3]. The aim is to predict whether the client will subscribe a long-term deposit or not. I choose an older version of this dataset with less input, which contains 4521 instances and 17 attributes. Some attributes are personal information of customers, others are related with the last contact of the current campaign. The last attribute is output, which represents if the client subscribes a long-term deposit. This dataset can better measure the ability of neural network classification.

Banking institutions introduce financial products that involve a huge amount of money to customers through telemarketing. The telemarketing campaigns are very expensive and consume a lot of manpower, material resources and financial resources. If the customer's relevant information can be used to predict the probabilities of customers subscribing long-term deposits, banks can make more contact with customers who have a high probability of subscribing, thereby greatly reducing costs and expenses.

1.2 Existing Research Results

According to a research [4], by comparing logistic regression (LR), decision trees (DTs), neural network (NN) and support vector machines (SVM). Traditional statistical models such as LR and DTs have the advantage of fitting models. They are often easily understood by humans and provide good predictions in classification tasks. In contrast, NN and SVM are more flexible and exhibit learning capabilities ranging from linear to complex nonlinear mappings. Due to this flexibility, NN and SVM tend to provide accurate predictions, but the models obtained are difficult to understand. Different problems fit different models, and we can use metrics to judge their performance.

Table. 1. Compare AUCs for different methods in research paper					
Metric	LR	DT	SVM	NN	
AUC	0.715	0.757	0.767	0.794	

From the above table, we can see that the performance of NN is the best (AUC=0.794). Such AUC corresponds to a very good prediction results.

2 Method

2.1 Pre-processing

First I load the dataset and disorganize data to prevent a sample is often selected in each batch of training. The classification requires that the inputs must be numbers and be normalized to similar ranges. So I convert nominal attributes to numeric attributes. For output, I use 1 to replace 'yes' and 0 to replace 'no'. Then I map these numeric attributes to the range of [0, 1] in order to normalize. In the multi-index evaluation system, due to the different nature of the evaluation indicators, they usually have different dimensions and magnitudes. When the levels of the indicators are very different, if the original index values are directly used for analysis, the role of the higher-valued indicators in the comprehensive analysis will be highlighted, and the effect of the lower-level indicators will be relatively weakened. Therefore, in order to ensure the reliability of the results, it is necessary to standardize the raw indicator data.

I split dataset into training set and testing set. The training set contains 2/3 of the original dataset while testing set contains 1/3. This division ratio is the same as mentioned in Moro's paper [4]. There are also many other ways to partition datasets, such as leave-one-out, cross-validation, bootstrapping, and so on. The approach taken in this paper is to be consistent with the research paper and control variables.

2.2 Define a Neural Network

Sklearn is an important machine learning library for Python, which encapsulates a large number of machine learning algorithms, such as classification, regression, dimension reduction, and clustering. It also includes three modules: supervised learning, unsupervised learning and data transformation. Sklearn is a simple and effective tool for data mining and analysis and relies on NumPy, SciPy, and matplotlib [5]. Sklearn has adequate documents which makes it easy-to-use. Therefore, it has become the most popular tool for studying and practicing machine learning. The focus of this paper is to explore how genetic algorithm can optimize and improve neural network rather than how to construct neural network. So I choose the sklearn tool to construct neural network classification. This reduces the workload.

A multilayer neural network structure is defined. The topology of the neural network for classification can be described as follows: The neural network is consisted of three layers, input layer, hidden layer and output layer. After testing the relationship between the different combinations of attributes and the target attribute, I find that the best case is to set all the attributes as inputs. That is why input layer contains 16 neurons, representing the number of input attributes while the output layer contains 2 neurons, representing the type of output results. The number of neurons of hidden layer could be adjusted until the best performance is reached.

Neural network models can achieve better performance by modifying parameters, such as the number of hidden layer neurons, the number of hidden layers, initial learning rate and the maximum number of iterations. There are many activation functions to choose from, without using activation function, each layer of the neural network only performs linear transformation, and the multi-layer input is also linearly transformed. The linear model has insufficient expressive power, but the activation function can introduce nonlinear factors.

In terms of the optimizer, I provide stochastic gradient descent (SGD), L-BFGS and Adaptive Moment Estimation (Adam). When using SGD, the gradients calculated for each iteration contain relatively large noise. In essence, Adam has a momentum term, which dynamically modifies the learning rate of each parameter using first order moment estimation and second order moment estimation of gradient. The main advantage of Adam is that after offset correction, the learning rate for each iteration has a certain range, making the parameters more stable [6]. The L-BFGS algorithm is

suitable for use in large-scale numerical calculations. It has the characteristics of fast convergence, and saves a lot of space and computing resources.

2.3 Genetic Algorithm

As described above, the genetic algorithm is a search algorithm for solving optimization in computational mathematics and it is a kind of evolutionary algorithm. We can use genetic algorithm to get the best parameters in a neural network model. The approximate steps of a genetic algorithm are:

- 1. Initialization the number of evolution, population size, crossover probability, mutation probability;
- 2. Encode the population using real number and use the accuracy between the prediction data and the expected data as a fitness function;
- 3. Cycle through selections, crossovers, mutations, and calculate fitness until the number of evolutions is reached to obtain the optimal parameters;
- 4. The best parameters obtained will be used to build a BP neural network.

The flow of the algorithm can refer to the following figure.



Fig. 1. Genetic algorithm flow chart.

Through genetic algorithms, we can obtain a set of parameters that can obtain the highest accuracy, including activation function, optimizer, the number of hidden layer neurons, threshold value and so on. These parameters will be directly applied to the classification model, and its performance will be discussed later.

2.4 Reference Technology

In order to improve the classification, I refer to a paper [7]. In that paper, it improved accuracy by modifying the threshold. The default initial threshold is usually 0.5, but in different situations this does not necessarily lead to the highest accuracy. The threshold is a critical value. Results greater than the threshold are classified as positive, and results less than the threshold are classified as negative. In order to get the best initial threshold, I still use the genetic algorithm. Genetic algorithms can help us to get the optimal initial threshold. The specific operation steps are similar to those described in the previous section, and the description will not be expanded here. And in the process of training the model, the threshold will be automatically updated and adjusted. The best threshold will eventually be obtained.

3 Results and Discussion

Through genetic algorithms, I obtain a set of current optimal parameter combinations. That is {'solver': 'adam', 'max_iter': 50, 'activation': 'logistic', 'learning_rate_init': 0.1, 'hidden_layer_sizes': (1000,)}. This is just an example. Because the testset data is randomly selected each time, the optimal combination of parameters obtained after each execution of the code is different. I will compare the performance of classification built using this set of parameters and classification built using other parameters.

After passing testing data to the built neural network, we need some metrics to show the performance of the classification. The popular metrics are the confusion matrix [8] and the Receiver Operating Characteristic (ROC) curve [9]. Through confusion matrix and ROC curve, we get the Accuracy and AUC of the classification. Each column of the confusion matrix represents the prediction category. The numerical value of each column indicates the number of real data predicted to be the category; each row represents the true attribution category of the data, and the total number of rows represents the number of data instances of the category. It helps people better understand the errors in the classification will be obtained. From Figure 2, we can see that there are 1518 instances in the test set, and the number of correct result predicted using a common neural network classifier is 1348, while the number of correct results predicted using the optimization, the accuracy rate can reach 90.18%. The number of True Positive, True Negative, False Positive, False Negative values can also be obtained from confusion matrix. Recall, precision and other metrics can be calculated from these values.

[[1292 48]		[[1312	28]]
[[111	56]]	[[110	57]]

Fig. 2. Confusion matrix of original neural network (left) and optimized neural network (right).

The ROC curve and AUC are shown in Figure 3. The ROC curve shows the False Positive Rate (fpr) versus True Positive Rate (tpr) and the area of the curve is AUC. The ROC curve is a comprehensive index reflecting sensitivity and specificity, and each point on the ROC curve corresponds to a threshold. Probability is greater than or equal to the threshold is positive, less than the threshold is negative. As the threshold decreases, more and more instances are classified as positive, but these positive classes are also doped with negative instances, that is, TPR and FPR increase at the same time. When the threshold is maximum, the corresponding coordinate point is (0, 0), and when the threshold is the minimum, the corresponding coordinate point (1, 1). According to [9], the AUC value is equivalent to the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example. Classification with a larger AUC works better.

From Figure 3, we can see that the AUC of the common classification is 0.65, and after optimization is 0.66. The AUC difference is not too large. The more convex the ROC curve is and the closer it is to the upper left corner, the greater its predictive value. The difference between the two ROC curves is also small. However, it can still be seen that the optimized classification performs better than the ordinary classification.



Fig. 3. ROC curve and AUC of original neural network (left) and optimized neural network (right).

In the relevant paper of the dataset provider, the AUC of NN can reach about 0.8. Specific values can be seen in Table 1. Obviously, the performance of my classification is not as good as the classification in the paper. The reason may be that I choose the old dataset containing fewer attributes or my code is not optimized enough. These metrics show that the

forecasting model is very meaningful for the banking institutions to select the target customers. By predicting the possible outcomes of customers subscribing long-term deposits, banks can have more contact with customers whose output is true.

4 Conclusion and Future Work

In this paper, I use neural network algorithm to build a classification and use genetic algorithm for optimization. The classification correctly predicts long-term deposit results for most customers by entering customer-related information. Neural network classification has good performance in practice and is easier to implement. Selecting different thresholds, activation functions and optimizers can also affect the performance of the classification. Experiments show that the genetic algorithm can help us to choose the optimal combination of parameters to improve the accuracy of the prediction.

Genetic algorithm is a kind of global optimization algorithm, which can quickly search the whole solution in the solution space without falling into the fast fall trap of the local optimal solution. By using its inherent parallelism, it can conveniently perform distributed computing. But it has some limitations. Programming is more complicated. First, the problem needs to be coded. After the optimal solution is found, the problem needs to be decoded. The algorithm has a certain dependence on the initial population selection. The local search ability of genetic algorithm is poor, which results in the simple genetic algorithm being time-consuming and the search efficiency is low in the later stage of evolution.

In the future work, when bank institutions collect data, more customers' personal information should be collected to help with the classification, such as income. The class of each attribute should also be improved and optimized, such as the current classes of occupations are too rough. The class of the customers' profession should be refined so that the possibility of customers' long-term deposits in different fields of work can be better derived. For scholars who are studying data mining, they can make full use of the extensibility of genetic algorithm and combine it with other algorithms. In practical applications, the genetic algorithm is prone to premature convergence. What kind of selection method should be adopted not only to keep good individuals but also to maintain the diversity of the populations can be an important direction for future research.

References

- 1. J J, Hopfield: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. vol.79, pp. 2554-2558. NatL. Acad. Sci of USA(1982)
- 2. Davis, L. (ed.): Handbook of genetic algorithms. New York(1991)
- 3. UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/Bank+Marketing
- 4. S. Moro, P. Cortez, P. Rita: A Data-Driven Approach to Predict the Success of Bank Telemarketing. Guimaraes, Portugal (2014)
- 5. Ketkar N: Introduction to PyTorch. In: Deep Learning with Python. Berkeley(2017)
- 6. Kwang Gi Kim: Deep Learning. Chapter 6. Korea(2016)
- 7. L.K. Milne1, T.D. Gedeon1, A.K. Skidmore: Classifying Dry Sclerophyll Forest from Augmented Satellite Data: Comparing Neural Network, Decision Tree & Maximum Likelihood. Sydney
- 8. Kohavi.R. and Provost, F: Glossary of Terms: Machine Learning 30, No.2-3, pp. 271-274 (1998)
- 9. Fawcett.T: An introduction to ROC analysis.: Pattern Recognition Letters 27, No.8, pp. 861-874 (2005)