Analysis of Marketing Campaigns with Neural Network

Chuqiao Song

School of Computer Science
College of Engineering and Computer Science
Australian National University
ACT 0200 Australia
u6512204@anu.edu.au

Abstract. In this report, we try to analyze a problem of a banking institution with a cascade algorithm. It is a kind of supervised training algorithm that can grow in the training process, which is close to the minimum network structure [1], it is suitable for classification problems. The data we use is related with direct marketing campaigns gained by calling clients of a Portuguese banking institution and the ultimate goal is to classify the variable y and select the customers who will order the term deposit. By using cascade algorithm, comparing it with the traditional algorithm and determining the benefit to use it. Based on UCI machine learning database of bank marketing data set, to make a study of the data set from the data preprocessing and the characteristics of the engineering, to study evaluation and selection of the model, a more complete demonstrates the general process to solve the problem of classification. Some common problems are addressed, such as missing values.

Keywords: cascade networks，marketing campaigns，Back Propagation，Neural network, preprocessing

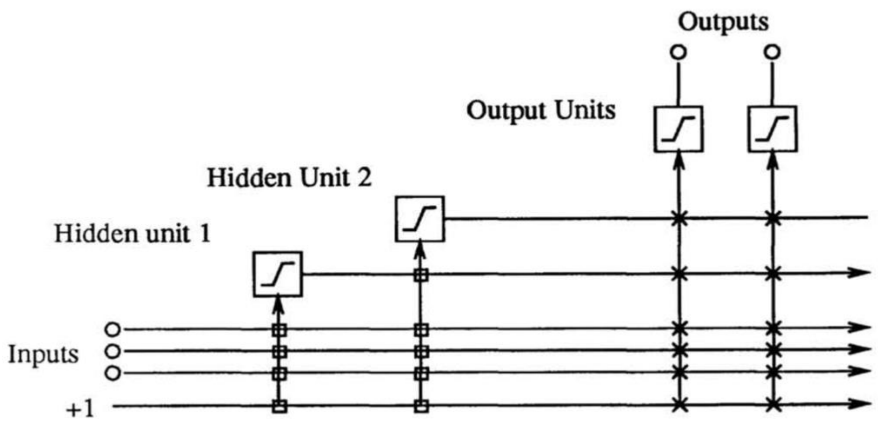
1 Introduction

The data related to Portugal direct marketing activities of Banks. These direct marketing campaigns are based on phones. Typically, the customer service staff at the bank will need to contact the customer at least once to see if the customer will subscribe to the bank's products (time deposits). Therefore, the task corresponding to this data set is the classification task, and the classification objective is to predict whether the customer is (yes) or not (no) subscribing to time deposits (variable y). This neural network classification exercise will help us understand its important role in the business world.
As for Cascade-Correlation there are two key ideas: the first is the cascading architecture, where the hidden unit is added to the network at once and will not change after it is added. The second is the learning algorithm, which creates and installs new hidden units. For each new hidden unit, we try to maximize the correlation between the output of the new unit and the residual error signal we are trying to eliminate. The cascade structure is shown in figure 1[1].

Neural network (NN) is a complex network composed of a large number of simple processing units connected to simulate the structure and behavior of human brain neural networks. It reflects many basic characteristics of the human brain function, but it is not a true portrayal of the human brain, but only some kind of simplification, abstraction, and simulation [2]. The three elements of the NN are the direction of information flow, the topology of the network, and the learning style. According to the information flow and the topology of the network, the NN model can be divided into two categories: feedforward network and feedback network. The neurons of the feed-forward network are arranged in layers [3]. Each neuron receives input from the previous layer and outputs it to the next layer. There is no information exchange between neurons in each layer. The feedforward network includes Perceptron and MuItipIayer Feedforward Neural Networks (MFNN). The perceptron consists of an input layer and an output layer with no hidden layer in between. It can only solve the problem of linear separability. MFNN consists of one input layer, several hidden layers, and one output layer [4]. It can be used to solve nonlinear classification problems. Back Propagation (BP) network is a multi-layer feed-forward neural network, which is the most common and widely used neural network. The learning algorithm is very important to the entire network.

Among various learning algorithms, the error back propagation algorithm (Error Back-propagation A190 rithm) is the most widely used, and the corresponding feedforward network is called BP network. The BP algorithm was first proposed by Werbos in 1974 and Rumelhart et al. developed the theory in 1985 [5]. The BP network employs a tutor's learning method. The learning consists of the following four processes: the input mode is the error mode of the difference between the expected output and the actual output of the network. The "error back propagation" process of connection weights is corrected layer by layer by the output layer via the hidden layer, and the network "memory training" process is repeated by "pattern smooth propagation" and "error reverse propagation", and the network tends to converge, ie, the overall network. The error converges to a minimal "learning convergence" process.

Figue1. The Cascade architecture, initial state and after adding two hidden units. The vertical lines sum all incoming activation. Boxed connections are frozen, X connections are trained repeatedly



This report will evaluate whether consumers will use the bank account to deposit. Cascade combination algorithm is that each layer includes a classifier, each member classifier is executed in sequence, and the input of the first layer classifier is the initial training set. After each layer is performed on the basis of the previous step, the attributes of the initial training set are performed. To expand, the corresponding class probability estimates obtained from the previous layer are taken as new attributes, and the last layer is output as the final decision result. In the Cascade combination algorithm, the attributes of the initial data set are expanded in each stage. The bank account information of the high-level classifier is increased, and the error bias is significantly reduced, so that a good classification effect can be obtained.

2 Method

2.1 reading dataset

Table 1. The information and feature’s type of the Bank Direct Marketing Data Set

Number	Feature	Type	Number	Feature	Type
1	Age	Numeric	10	Day	Numeric
2	Job	Nominal	11	Month	Numeric
3	Marital	Nominal	12	Duration	Numeric
4	Education	Nominal	13	Campaign	Numeric
5	Default	Nominal	14	Pdays	Numeric
6	Balance	Numeric	15	Previous	Numeric

7	Housing	Nominal	16	Poutcome	Nominal
8	Loan	Nominal	17	y	Nominal
9	Contact	Nominal			

The dataset contains four CSV files:

1) bank-additional full-csv: contains all samples (41,188) and all feature inputs (20), sorted by time;
2) bank - additional CSV: randomly select 10% of the sample (4119);
3) bank-full.csv: contains all samples (41,188) and 17 feature inputs, sorted by time;
4) bank.csv: 4119 samples of 10% were randomly selected from bank-full.csv.
5) Number of Attributes: 20 + output attribute.

2.2 preprocessing

2.2.1 missing value processing

Unknown values may exist for non-numeric variables. Use the following code to see the number of values for the character variable unknown.

```
for i in data.colums;
            if type(data[i][0]) is str;
                  print("unknow value count is" + i + ":\t" +
str(data[data[i] == 'unknow']['y'].count()))
```

then we can easily get:

Unknow value count in	Number
Job	330
Marital	80
Education	1713
Default	8697
Balance	990
Housing	990
Loan	0
Contact	0

Missing value processing usually has the following methods:

For variables with a small number of unknown values, including job and marital, deleting these variables with the missing value (unknown).

If the variable is expected to have little effect on the learning model, the unknown value can be assigned. In this case, it is considered that all variables have a great influence on the learning model.

A complete row of data can be used as a training set to predict the missing values of missing values, variables housing, loan, education and default. Since sklearn's model can only deal with numerical variables, the classified variables need to be numeralized first and then predicted.

2.2.2 Numerical classification of variables

In order to enable classification variables to participate in model calculation, we need to numeralize classification variables. Classification variables can also be divided into two categories of classification variables, ordered classification variables and unordered classification variables. There are also differences in encoding methods of different classification variables.

The variables default, housing, and loan can be considered as binary classification variables, which can be encoded by 0 and 1.

You can think that variable education is an ordered classification variable, and the size of the variable is "in the name of the" ", "" basic, 4y", "basic, 6y", "basic, 9y", "the high school", "professional,", "the university, degree," and the variable affects the order of the variables to be 1, 2, 3…

Variables job, marital, contact, month, day_of_week can be considered as unordered classification variables. Note that the variables month and day_of_week are ordered from a time perspective, they are unordered for the target variable. For unordered classified variables, dummy variables can be used to encode. Generally, n categories require n-1 dummy variables. For example, the variables divorce, married, and single are encoded using two dummy variables V1 and V2.

marital	V1	V2
Divorced	0	0
Married	1	0
Single	0	1

2.2.3 Numerical feature preprocessing

One advantage of discretization of continuous feature is that it can overcome the hidden defect in data effectively: make the model result more stable. For example, extreme values in the data are an important factor affecting the effects of the model. Extreme values cause the model parameters to be too high or too low, or cause the model to be "confused" by false phenomena, and take the previously non-existent relationship as an important model to learn. However, discretization, especially equidistant discretization, can effectively reduce the influence of extreme value and abnormal value.

By observing the statistical information of the original data set, it can be seen that the variable duration of maximum value is 4918, and 75% quantile for 319, far less than the maximum value, and the standard deviation of the variables for 259, relatively is also big. So duration is discretized. . In particular, the use of pandas.qcut () function to discretize continuous data, it USES the quantile of data partition, can get the basic equal to the size of the box (bin), in the form of interval said. Then use the pandas.factorize () function to convert the interval to a value.

2.3 Training

Because the training model is required to predict the unknown value, the time cost of the pretreatment process is relatively high. Therefore, the preprocessed data is persisted and stored in the file. The learning model after that reads the file data directly for training and prediction without preprocessing.

2.4 Cascade network

In first part it initializes the particles and lets the first particle be the global best, and we input first code into cascade training part.

In cascade training, CC starts with a minimal network consisting only of an input and an output layer which are fully connected.

Generate the so-called candidate units. Every candidate unit is connected with all input units and with all existing hidden units. Between the pool of candidate units and the output units there are no weights.

Train all the connections ending at an output unit with a usual learning algorithm until the error of the net no longer decreases.

Try to maximize the correlation between the activation of the candidate units and the residual error of the net by training all the links leading to a candidate unit. Learning takes place with an ordinary learning algorithm. The training is stopped when the correlation scores no longer improves.

Print results (self, inputs, targets).

Then after cascade training finished, the data can be loaded from data set and test how is going.

3 Result and Discussion

Tests show that this model can achieve this effect. In general, the model has a good judgment result. From the results of training, testing and application, the accuracy rate of judgment is more than 90%. One of the reasons for the high rate of judgment is that the neural network has good knowledge discovery and feature extraction capabilities. Another reason may be that all data are interpolated data. Many experiments show that the neural network model is less effective in extrapolating data. In addition, we set up files such as initial weights, learned weights, training data sets, test data sets, and application data sets, respectively, to facilitate program calls. bank account learning only needs to be performed once. After successfully learning, the learned weights are saved to the file. When testing and actual application, it is not necessary to learn. Directly calling the learned weights and data files can give the result. If there are major changes in the system, and there are new differences in the financial ratio characteristics of different credit companies, then the bank account can be re-selected, and after successful learning, the model can be put into use again. In this sense, this model is a dynamically adjustable model with good adaptability.

Cascade-Correlation algorithm is compared with existing algorithms, learning very fast, higher calculation accuracy, wider range of application, the training of network to determine the size and topology itself, it is an artificial neural network with strong generalization ability and easy to train.

5 Conclusion and Future Work

There are three reasons why Cascade-Correlation algorithm is faster than the other algorithms:

1、It only does forward propagation, while Back Propagation algorithms require forward and reverse propagation operations.

2、In most cases, the algorithm training network is smaller than the resulting network.

3、We can make full use of the computer's cache to speed up the network training. BP network training speed is very slow, in the process of practice is acceptable for the unit inside a constantly changing network, making it difficult to determine the optimal result, the cc algorithm, each adding a hidden unit with a hidden layer, each hidden unit has strong characteristics of function, enhance the network operation ability.

But the algorithm is: faced with the problem of network topology changes will cause the oscillation of the network performance, in order to make up for this shortcoming, we always put the new node is the initial value of very small (not 0), later on how to get more accurate results and guarantee the stability of the network also need further research and discussion.

Direct marketing of Banks is important in today's business world. The most recent KDnuggets public opinion survey shows that Banks and direct marketing are ranked 4th and 11th in the voting questions, respectively [6]. Due to the low cost of communication and the rapid development of database technology, it has developed rapidly in recent years. With a large amount of customer information gathered from various communication channels, Banks can sell their products and services to potential customers according to their needs. Data mining technology can achieve this function. [7] Neural network algorithms are playing an increasingly important role in banking and other commercial activities.

References

1. Fahlman, E.C. & Lebiere, C. (1997). The Cascade-Correlation Learning Architecture. Advances in Neural Information Processing Systems. 2, http://papers.nips.cc/paper/207-the-cascade-correlation-learning-architecture.pdf
2. Kane, G. (2014). Neural networks -- neural computing: Theory and practice by philip wasserman / neural networks: A tutorial by michael chester / neural networks: Algorithms, applications, and programming techniques by james freeman and david skapura / and others. JAMA, 271(12), 956. Retrieved from https://search.proquest.com/docview/211265259?accountid=13151
3. Lee, K. C., Han, I., & Kwon, Y. (2016). Hybrid neural network models for bankruptcy predictions. Decision Support Systems, 18(1), 63. Retrieved from https://search.proquest.com/docview/206600041?accountid=13151
4. Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. Research in Higher Education, 59(3), 382-400. doi:http://dx.doi.org/10.1007/s11162-017-9473-z
5. Yang, H., Kitamura, R., Jovanis, P. P., Vaughn, K. M., & Abdel-Aty, M. (2013). Exploration of route choice behavior with advanced traveler information using neural network concepts. Transportation (1986-1998), 20(2), 199. Retrieved from https://search.proquest.com/docview/211980839?accountid=13151
6. KDnuggets Polls: Where did you apply Analytics/Data Mining/Data Science? (accessed February 2014). Retrieved from http://www.kdnuggets.com/polls/2012/where-applied-analytics- data-mining.html
7. Kalid, S.N ,Khor K. C. , Ng K. H & Ting C.Y, Knowledge Management International Conference (KMICe) 2014, 12 – 15 August 2014, Malaysia http://www.kmice.cms.net.my/