Procedia Computer Science

Volume 88, 2016, Pages 463–468

7th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2016

Model of the forecasting cash withdrawals in the ATM network

Sergey Nemeshaev and Alexander Tsyganov

The National Research Nuclear University MEPhI (*Moscow Engineering Physics Institute*), *Russian Federation*

sergey.nemeshaev@gmail.com, a2tsy-kaf22@yandex.ru

Abstract

To increase the economic efficiency of the self-service devices networks operated by banks and processing companies, possibly due to the use of mathematical models and algorithms in the cash-in-transit forecasting process. In this article the issues are considered to improve the cash drawing forecast accuracy using the model for the neural networks. Special focus is on the description of data pre-processing.

Keywords: bank, replenishment, forecasting, statistics

1 Introduction

Nowadays, the banks are developing and offering actively the services to the public in connection with the card payment system. Not only the services range is expanding but also the network of the bank branches and outlets, and the ATMs. One of the forms of the customer service offering by the banking systems is a cash withdrawal through the ATMs network [1]. Nowadays, the self-service devices already have a wide range of different functions: cash withdrawals (various currencies at the multi-currency ATMs), bank account replenishment, payment of various services (cellular service, utilities payments, etc.). All of these services mean the constant ATMs replenishment and emptying, or rather different cartridges installed in them. In this regard, sooner or later you have to fill up the cartridges with different currencies sorted par value. It is not effective just to send CIT vehicle and replace the cartridges randomly. You need to determine the optimal date and amount of the filled funds par value, and only then to hire an armored car delivery. As the number of ATMs increased the work for the department responsible for the cash delivery vehicle is increased too. The problem that has appeared is to computerize the ATM cycles research, the processing the resulting statistical information on withdrawals and depositions, and to make decisions on the timing and amount of future cash collection.

Selection and peer-review under responsibility of the Scientific Programme Committee of BICA 2016

doi:10.1016/j.procs.2016.07.465

2 Applicability of neural network technology for the business forecasting problems

The neural networks are the adaptive systems to process and analyze the data, which are the mathematical structure that simulates some aspects of the human brain's operations, and demonstrate such brain's possibilities as the ability to non-formal learning, generalizing, and clustering the unclassified information, the ability to make forecasts on the basis of presented time series. Their main difference from other methods, such as expert systems, is that the neural networks do not require pre-known model in principle, and build it with their own recourses only on the basis of presented information. This is precisely why the neural networks and genetic algorithms come into practice there where you need to solve the problem of forecasting, classification, management - in other words, in the field of human activity where there are the bad algorithmic problems whose solution required a permanent work of the group of qualified experts, or adaptive automation systems which the neural networks are [2].

The neural networks are being used increasingly frequently in the real-world business applications. In some field, such as fraud detection and risk assessment, they became the undisputed leaders among the methods used. Their use in the forecasting systems and market research systems is constantly growing.

It should be noted that as far as the economic, financial and social systems are very complex and are the result of actions and reactions of different people, it is very hard (and may be impossible) to create a full mathematical model taking into account all the possible actions and reactions. It is almost impossible to approximate in detail the model based on such traditional parameters as utility or profit maximization.

In the systems with such complexity it is the natural and the most effectively to use the models simulating directly the behavior of society and the economy. And it is just the thing that the neural networks methodology is able to offer.

Also the neural networks can be used to solve other problems. The main predetermining conditions for their use are the presence of the "historical data", using which the neural network that can be trained, as well as the impossibility or failure of the use of more formal methods [3].

3 Selecting a network architecture

Based on the specifications of the basic models for the neural networks and to solve the problem of the demand forecasting the multilayer neural network model was selected.

The multilayer neural networks are flexible in solving various problems, and consume low resource for training. As well as they are fairly easy to understand their structure and working principles. Choose the network architecture with 2 hidden layers, and this will reduce the total number of neurons in the hidden layers. The number of neurons in them shall be determined by the experimentally obtained formula for the first hidden layer: $(inputs + outputs)*2 - outputs$, for the second hidden layer $(inputs + outputs)*2 - outputs$, where $inputs$ are the number of the neural network inputs, and the $outputs$ are the number of outputs (in our case - 1, since it is necessary to receive the unique predicted value of the next day withdrawals).

As the activation functions of the hidden layers neurons and the output functions we select the hyperbolic tangent.

The problem to select the network architecture has been solved experimentally in a medium Matlab. For different input data sets the different architecture MNS have been generated. The described above architecture had the smallest margin in the test sample

4 Selecting the input data model

Development of the input data model is a method to convert the output data before submitting them to the inputs of the neural network. It is needed in order to the NN defines the relationship between the correlated variables, not between the values depended on each other randomly.

The time series analysis is based on the presumption that the measurable time changes are a display of certain deterministic laws that are not known to us, but at the same time they exists. For example, the value of withdrawals in a certain ATM may be influenced by the weather in some degree. According to certain valuable's changes in the process you can conclude on the curves in the phase space, and that will help to reconstruct the curves in the chaotic attractor. In other words, with the behavior of one variable you should draw conclusions about the behavior of other variables, even if the exact number of them may be initially unknown. Such conclusions can be obtained by the method called the attractor reconstruction. This transformation helps to input of the neural network not a single value (the value of withdrawals per day) but a vector of values. In our terminology, the values number corresponds to the inputs index.

Suppose that at some time point t_1 the initial value x is determined, a second value x_2 is determined at the timepoint displaced from the first one by a constant value T . In the bottom of the figure the both the values found are laid along the coordinate axes, whereby we obtain a point on the subspace. Thus, all axes points are processed that gives the path shown in the figure. The plot shown in phantom is an extrapolation in the future when the measurements are not yet available. In the case of chaotic attractors the attractor reconstruction requires at least a three-dimensional coordinate system (or even a system of higher dimension). At this time axis replacement the appropriate new coordinates obtain not only by a constant value T but also by $2T$ etc.

Two important problems appear:

1) Select constant T

2) Select the dimension of the coordinate system.

Here is an option to solve them.

Let us consider a smooth deterministic dynamical system. Let the X is the time series generated by the system, i.e., the value X_i is an arbitrary function of such a system state. Then, by the theorem of Takens there is a diving depth d (approximately equal to the spectacular number of degrees of freedom of the dynamic system), which provides an unambiguous prediction of the next value of the time series. Thus, by selecting sufficiently big d we can ensure an unambiguous dependence of the series future value of its previous values d : $X_i = f(X_i,...,X_{i-d})$, i.e. time series forecasting is depended on the interpolation function of many variables.

In practice, the d - dimensional measurements are not obtained usually. However, there are ways to restore the phase space in the presence of a smaller number of dimensions. In this case of the time series forecasting there is only one dimension – the demand time series. Next, we will consider the reconstruction methods based on the only one dimension.

The reconstruction of the phase space with the time delays embedding is a method of generating a d - dimensional space equivalent to the original d - dimensional space, using a vector coordinates delays matrix.

Consider the column vector of the time series $x(i)$. Define the d - dimensional matrix of the vector- columns coordinates delays by summing delay coordinate shifted copies of time series,

$X = [x(i), x(i+\tau),..., x(i+(m-1)\tau)]$. Such matrices $X[n-(m-1)\times\tau), m]$ are called the attachment matrixes. The τ and m parameters should be optimally assessed. The first option is the τ delay time, i.e., the time to move between successive delays of coordinates vectors. The second option is the embedding dimension m, the number of such coordinates vectors delays.

Solving the problem of the optimal selection for the time delay, two different approaches can be found based on the nearest neighbor search procedure. One approach suggests using the first minimum delay of the mutual information τ for all dimensions. The second approach uses the first minimum the distance ratio to the nearest neighbor in order to set the delay time for each measurement.

In the following we will consider the second approach, because it allows a more accurate input selection.

The dimension of the phase space embedding is estimated as the share of F tending to zero.

Here is the implementation of data preprocessing procedure:

1) τ_i index search:

a. An initial time series column-vector $x(i)$. For each τ, $\tau = 1,..., \frac{1}{10n}$, build a temporary attachment of the matrix $T = [x(i), x(i+\tau)]$

b. For each two-dimensional point, that is, for every line of the matrix T look for its two-dimensional nearest neighbor. Calculate the Euclidean distance $dE1$ between them.

c. Consider two points shifted forward by a predetermined step, and calculate new Euclidean distance $dE2$ between them.

d. Calculate $\frac{dE2}{dE1}$ and count the number of cases when the $\frac{dE2}{dE1}$ is more than 10. We call this share as N.

e. Choose at least the first $N(\tau)$, which is optimal time delay for the first investment cycle, τ_1.

f. Estimate the shares F depending on the dimension of the embedding matrix (the evaluation procedure F is described below).

g. Consider now the matrix $X = [x(i), x(i+\tau_1)]$ as a starting point for the second embedding cycle. For each τ, make the temporary attachment matrix $T = [x(i), x(i+\tau_1), x(i+\tau)]$.

h. Repeat the steps b-h while the share F does not drop to 0.

So we get a vector of values of $\tau = [\tau_1, \tau_{(m-1)})]$

Final embedding matrix is $X = [x(i), x(i+\tau_1),..., x(i+\tau_{(m-1)})]$

2) m index search:

The algorithm to estimate the share F is similar to the estimation the share N.

Adding each new measurement for each pair of nearest-neighbor distance $R1$ of the previous dimension it should be consider the distance $R2$ between them in the new space of higher dimension and evaluate the value of $\frac{R2}{R1}$. Count the number of cases where the $\frac{R2}{R1}$ is more than a predetermined smallness threshold (ie, the share of false nearest neighbors). When the share of false neighbors tends to zero, you can stop the procedure and adopt the embedding dimension as m.

To improve the quality of training it should be to strive for the inputs statistical independence, ie the lack of correlation. In addition, in order to reduce the level of correlation you need to select as NN inputs and outputs the more statistically independent quantities than daily withdrawals. Therefore, as input variables, you should choose not the value of withdrawals but, for example, changes in the amounts of withdrawals ΔC_t or the fractional change logarithm $\log(\frac{C_t}{C_{t-1}}) \approx \frac{\Delta C_t}{C_{t-1}}$.

The last option is good for the long time series, when the inflation influence is significant. In this case, the simple differences in different parts of the series can have different amplitude because they are measured actually in different units. On the contrary, the relationship of the successive quotations does not depend on the units and are of the same size, despite the change in the inflation measurement units. As a result, a large stationary of the line enables to use for training a larger history and will provide the best training [4].

In addition, before input the obtained data to the neural network their normalization should be conducted, bringing all data to the range [-1; +1]. This normalization allows the neural network to interpret the values in the same way and allocate inputs in the attribute space better. In this case it is carried out by dividing each member of the series for the maximal series member by the modulus on the specifications.

5 Results

To test and check the model it has been used the cash withdrawals statistics in the ATM network of the city of Yekaterinburg (Russia). Initial data to examine has been the value of daily withdrawals in 50 units for the period from July 01, 2014 till February 01, 2015. The control period to compare the predicted and actual values was from February 02, 2015 till March 01, 2015. That is based on 7 months of statistics the model has made automatically a daily forecast for the control period, then data were compared with the statistics for the same period. The mean absolute percentage error (MeanAbsolutePercentageError or MAPE) was used as the estimate of the forecast accuracy:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_i^{'}|}{y_i}$$

Forecast error is on average 32%.

6 Conclusion

This study focused on the pre-processing input statistical series, namely it is offered the method of data pre-processing based on the reconstruction of the phase space. To make the forecast the multi-layer neural network is used that is trained by the back-propagation method. As the result of the analysis of the neural network types and architectures, as well as methods to train them, there has been developed a theoretical model according to which the module to forecast the cash withdrawals from ATMs was implemented.

Based on the obtaining test results the next conclusion can be drawn. If we have a representative training sample consisting of a sufficient number of training examples, the network is cope very well with the problem of forecasting the demand for 5-6 days forward (forecast error is on average 32%). Then by inputing the predicted data the prediction error increases but saving the dynamic changes. Thus, NN is well suited to forecast the demand for 5-6 steps forward, if there is sufficient amount of the training examples (in this case there are 900 statistical values, ie. 900 values of the quantity of daily cash withdrawals from an ATM, or in other words 3 months of daily data).

Acknowledgements

This work was supported by Competitiveness Growth Program of the Federal Autonomous Educational Institution of Higher Professional Education National Research Nuclear University MEPhI (Moscow Engineering Physics Institute).

References

[1] Gospodinova, E.A. Features of the market of replenishment services, mathematic and statistical analysis of social and economic processes. Interuniversity Digest of scientific papers – MESI, 2008. – Number 5.

[2] Voronenko D.I., Neural networks - the pros and cons, Kharkiv 2004.

[3] Danko, T.P., Hodimchuk M.A., Systems of artificial intelligence in the development of corporate marketing strategies, Marketing in Russia and abroad - 2000. - № 5. - pp. 26-36.

[4] Ezhov A.A., Shumskij S.A., Neurocomputing and its application in economics and business. Moscow: Mephi 1998, p. 222.