Neural network data processing for financial insolvency forecasting of the water supply enterprises

G O Krylov¹,², A S Vorobeva¹ and M V Nikolaeva¹,³

¹ NRNU MEPhI, 31, Kashirskoye sh., Moscow, 115409, Russia
² Financial University under the Government of the Russian Federation, 49, Leningradsky prospect, Moscow, 125167, Russia
³ Email: mn.marianikolaeva@gmail.com, op50@mail.ru

Abstract. This work is devoted to creation of neural network methods with the increased stability and reliability for bankruptcy forecasting. In this paper, we investigate the performance of four different neural network (NN) methods for bankruptcy prediction of the resource provisioning system enterprises. Two methods based on NN-ensembles of classifiers (boosting-ensembles and bagging-ensembles) and another two tested methods are stand-alone classifiers (multilayer perceptron (MLP) and network of radial basic functions). We defined that bagging-ensembles of neural networks have advantage in comparison with classical - stand-alone classifiers. Developed bagging-ensemble of neural networks is currently used to the sphere of financial monitoring for identification of signs of financial insolvency of the water supply enterprises and assessment of risk of social tension in regions.

1. Introduction

In this work, the new subject domain is investigated: use of neural network models for bankruptcy forecasting of the water supply enterprises.

Among a wide class of approximating models, a specific place is the neural networks allocated with set of unique properties. It should be noted the following advantages of NN:

- ability generalization and adaptations to data, that is a playback capability from data deep patterns and the hidden dependences,
- in the smallest way are vulnerable to "curse of dimensionality".

Thus, by creation of neural network model it is required to retrieve new knowledge from the available data.

The most important results which are scientific novelty:

- NN-models are developed for the solution of problems of a new sphere - financial monitoring;
- four methods of training of neural networks are investigated and it is defined that bagging-ensembles of neural networks have advantage in comparison with classical approaches.
The purpose of work is improvement of quality of the made management decisions on identification of the destabilizing financial and economic processes connected with unstable functioning of the water supply enterprises.

During a practical part of the solution of an objective of classification the bagging-ensemble showed the best results, at the top level meeting all requirements to the developed model: reliability and stability of definition of a class of an object on new data.

At the present stage of development of national economy to digital continuous monitoring of a situation and timely reaction are bases of its stable growth. Creation of the mathematical model allowing to estimate a financial condition of the systemically important enterprises and to predict their possible insolvency is caused by need of adoption of management decisions for improvement and effective development of various sectors of the economy.

The need for creation of a mathematical model for support of decision-making on experimental data often arises both in fundamental, and in application studies. At creation of a mathematical model, as a rule, first of all the problem of identification most of which often is coming down to definition of such vector of parameters at which the reliability of model meets requirements imposed to it is solved. Problems of financial monitoring belong to tasks of this kind, in particular, in mathematical aspect diagnostics of bankruptcy of the enterprises belongs to a solution of a problem of classification: as basic data available characteristics of subjects to modeling are set, and it is required to approximate the financial patterns veiled in data. Thus, it is required to set by means of a mathematical model new knowledge from available data.

2. Methods

2.1. Problem Statement

This work is devoted to creation of the model capable on the example of the training selection to solve a problem of multiclass classification on the basis of neural networks. We considered it to choose as the most expedient training with the teacher. Main objective of such method: to restore (to approximate) dependence, i.e. to construct function (the decisive rule)

\[ f: X \rightarrow Y, \]  

(1)
on new objects of \( x \in X \) the predicting \( y \in Y; y = f(x) \). Here \( X \)-set of objects; \( Y \)-of group of objects.

At the same time, it is especially important that it is necessary to be able to predict \( y \) not only for objects from the training selection, but also for new objects.

The neural network is capable to work in difficult conditions of modeling and to approximate the wide range of expected models at the minimum requirements to structure of model and assumptions. However, the neural network as the independent qualifier is inclined to retraining.

As a rule, several various neural candidate networks are considered. Each network studies, and then the assessment of quality of its working capacity are carried out. Eventually, there is one neural network which showed the best results of testing, and the others are eliminated. Such approach to selection of networks is rather general practice. However, at are mute us actual we waste that time and resources which were spent for training of those neural networks which did not show the best results of testing. Not to mention the rejected neural network models. Besides, if for the choice of the best trained neural network Bayesian approach is not used, then the chosen network cannot have the high generalizing ability and to be just most successfully arranged to this test selection. But even when using the methods of selection based on Bayes's theory there is no guarantee of the exact choice of network with the maximum ability to generalization of true data as the training data, as a rule, happen predictably. The most logical exit from this situation is the choice not of one best trained neural network, and several, shown the best results of testing. This idea is the cornerstone of creation of ensembles of neural networks where the final decision is formed not by the unique network but set of networks.

So, it is necessary to develop the neural network ensemble capable on the example of the training selection to solve a problem of multiclass classification. As objects of a research were chosen 401
enterprises of the water supply system. As an initial data set, we have the training selection that is sign space which consists of a set of the indicators reflecting a financial condition of a water utility purchasing activity and signs of doubtful activity of the enterprise. And also, at on there are four groups to which the model trained by us has to carry objects of selection:

- operating / goes reorganization,
- invalid,
- it is liquidated,
- is in a condition of bankruptcy.

2.2 Ensembles of neural networks

The main advantage of ensemble is that it can improve considerably results of work on new data, do not demand big additional computing expenses. Actually, results of work of ensemble in general can even be better, than results of work of the best of the used networks separately. It can be shown for the following most obvious form of ensemble. It consists of the L trained neural network models forming output values \( y_i(x) \), where \( i = 1, 2, ..., L \). These networks can have various quantities of the hidden layers and neurons in them or to be same, but trained to the states corresponding to various local minima of criterion of an error of training. Perhaps also use of networks of various type or even their associations with other computing modules. Let’s designate the true modeled function as \( h(x) \). Then neural network transformation can be written down as the sum of value of required function and some deviation (an error of reproduction of the given function by network):

\[
y_i(x) = h(x) + \varepsilon_i(x),
\]

The average square of a deviation of the value modeled by network from ideal is defined by expression [1]

\[
E_i = M\left[(y_i(x) - h(x))^2\right] = M[\varepsilon_i(x)^2],
\]

where \( M[.] \) - operation of calculation of population mean. Thus, if to consider work of networks separately, then the average error of a set of networks will be defined as

\[
E_{AV} = \frac{1}{L} \sum_{i=1}^{L} E_i.
\]

Returning to ensemble in which output \( y_{COM} \) value is formed as an average of values of exits \( y_i, i = 1, L \), the L networks which are its part

\[
y_{COM} = \frac{1}{L} \sum_{i=1}^{L} y_i(x),
\]

it is possible to write down that the mistake of committee represents value

\[
E_{COM} = M\left[\left(\sum_{i=1}^{L} y_i(x) - h(x)\right)^2\right] = M\left[\frac{1}{L^2} \left(\sum_{i=1}^{L} \varepsilon_i(x)\right)^2\right].
\]

Considering the assumption that the mistake \( \varepsilon_i(x) \) has zero population mean and is uncorrelated \( M[\varepsilon_i] = 0 \) and \( M[\varepsilon_i\varepsilon_j] = 0 \) for \( i \neq j \) using the above-stated equalities it is possible, it is possible to establish the following interrelation between an error of work of ensemble and an average error of work of networks separately [1]:

\[
E_{COM} = 1/L^2 \sum_{i=1}^{L} M[\varepsilon_i^2] = \frac{1}{L} E_{AV}.
\]
Expression (2) reflects the remarkable fact – the size of a mistake of ensemble of work of networks \( L \) times less, than an average error of work of networks separately if output values of ensemble are formed by means of simple averaging of output values of the networks which are its part. In fact, improvement, unfortunately, is not so considerable. The matter is that reasonings were based on the assumption that errors of work of networks \( \varepsilon_i(x) \) uncorrelated. In practice they are, as a rule, strongly correlated because for training of separate models the general selection of data so a condition is used \( M[\varepsilon_i\varepsilon_j] = 0 \) for \( i \neq j \) it is not carried out. Using Cauchy's inequality in shape

\[
\left( \sum_{i=1}^{L} \varepsilon_i \right)^2 \leq L \sum_{i=1}^{L} \varepsilon_i^2,
\]

becomes possible to establish that [1]

\[
E_{\text{COM}} \leq E_{\text{AV}},
\]

that is use of ensemble does not lead to increase of a mistake in comparison with results of work of networks separately.

Development of the neural network classifying model took place in several stages:

- collecting and analysis of data,
- preparation of data,
- development of NN-models,
- NN-models assessment.

2.3. Data collection and analysis
Formation of sign space happened as follows: by the RCEAP code from a reference system SPARK-Interfax necessary data on water utilities from which 33 indicators characterizing a financial condition of a water utility were unloaded: index of due discretion (IDD), index of financial risk (IFR), index of payment discipline (IPD), coefficient of fast liquidity, coefficient of the current liquidity, coefficient of concentration of equity (autonomy), coefficient of security with own current assets, index of a constant asset, general profitability, turnover of assets, turnover of accounts payable, turnover of receivables, period of repayment of receivables, period of repayment of accounts payable, profitability of assets, profitability of the capital, non-current assets, stocks, receivables, short-term financial investments, assets (all), capital and reserves, accounts payable, income of future periods, liabilities (all), revenue, gross profit (loss), profit (loss) before taxation, net profit (loss), number of government contracts, sum of government contracts, share of the code of internal control, the number of refusals in carrying out financial transactions.

2.4. NN-models development
For the most effective use of the available set of experimental data it is offered to use the following technique of formation of the training, test and test (validation) selections for creation of neural network models. The initial data set in a random way breaks into three parts: training, test and check (figure 1).

![Figure 1. Illustration of a method of formation of selections.](image-url)
In our case division happens in a proportion 50:25:25. The allocated test part will be used only for assessment of an error of work of already trained ensemble and does not influence the course of process of training of the networks which are its part. Training of networks is provided with use of a part of the examples which entered the training part of set.

Test selection is applied to control of the attributes called by hyper parameters which control as the model studies. Efficiency indicators on a test set should not be used as assessment of real efficiency of model as the network was ready, using test data. Sets of the training examples created thus can be crossed for various networks which are a part of ensemble but cannot coincide completely. This technique of formation of selections is based on the idea of cross-validation (cross-validation) and allows reaching smaller correlation between errors of work of separate networks than in case they study at the same data sets.

3. Results

3.1. Assessment of quality of NN-models

In search of the most rational model of the solution of an objective of classification in the environment of IBM SPSS Modeler several NN_models were developed. Among which a multilayer perceptron, RBF-function \[2-16\], a boosting \[17-18\] and a bagging. The reliability of the developed NN-models is reflected in figure 2.

As independent qualifiers it is necessary to specify as a lack of neural network models that they tend to retraining, that is in the course of training we "too adjust" model to the available data, but on new data it shows will show bad results of prediction. Instead of spending time and resources on training of candidate NN which did not show good results it is quite logical to use ensembles of neural networks where the final decision is formed not by the unique network but set of networks. Besides, such methods of creation of ensembles as a boosting and a bagging, are directed to increase in reliability and stability of models respectively.

Unlike a boosting which demands enough big training selections the bagging is capable to build qualitative NN-models on rather small data sets.

![Figure 2. Reliability of the developed NN-models.](image)

During a practical part of the solution of an objective of classification the bagging-ensemble showed the best results, at the top level meeting all requirements to the developed model: reliability and stability of definition of a class of an object on new data.

In this regard there is a need in detail to consider the developed bagging-ensemble and results of its work.

3.2. Bagging

The (bootstrap aggregating) uses parallel training of basic qualifiers (speaking to language of mathematical logic, a bagging – the improving association, and a boosting – the improving crossing). During a bagging there is a following:

- from a set of basic data, it is in a random way selected several subsets containing the quantity of examples corresponding to quantity of examples of an initial set;
- as selection is carried out in a random way, a set of examples will always be a miscellaneous: some examples will get to several subsets, and some will not get to one;
• on the basis of each selection the qualifier is under construction;
• conclusions of qualifiers are aggregated by vote.

Thus, 14 neural network models constructed on the basis of a multilayer perceptron which studied in parallel entered into ensemble (figure 3). Each NN contains 1 hidden layer, from 77 to 194 synapses (the model size), and, therefore, from 154 to 388 neurons. At their construction on an entrance 33 predictors moved. For receiving a full picture of results of work of ensemble it is necessary to consider in more detail some schedules reflecting construction details.

![Table](image)

**Figure 3.** Details of models of components a bagging - ensemble.

![Chart](image)

**Figure 4.** Quality of work of ensemble.
In figure 4 we see that the reliability of prediction of the ensemble received by aggregation of results of separate qualifi ers is 97.9% that above, then reliability basic (95.9%) and naive (89.7%) models. The provided schedule clearly demonstrates efficiency of use of ensemble the NN-models for a solvable task.

Apparently from figure 5, the ensemble constructed by means of a bagging revealed 14 operating enterprises which actually are at a stage of bankruptcy (8) or are liquidated (6).

4. Discussions

4.1. Interpretation of results of work of the developed bagging-ensemble

On the basis of expert check results the correctness of the developed model forecast was confirmed. For example, on graphics (figure 5) the negative value of size of own working capital at all revealed water utilities is visually shown, involves wear of municipal water supply systems and water disposal, and, as a result, emergence of municipal accidents.

![Figure 5](image-url)

**Figure 5.** An indicator of own working capital for 2017.

![Figure 6](image-url)

**Figure 6.** Geographic location of the revealed enterprises which appealed to court with the statement for bankruptcy.
Also, in figure 6 we see geographic location of the enterprises which appealed to court with the statement for bankruptcy. One of them of North Caucasus federal district is already declared by bankrupt.

On the located data set 4 neural network models are constructed: the multilayer perceptron (MLP), network of radial basic functions, a boosting - ensemble of neural networks on the basis of a multilayer perceptron and bagging-ensemble of neural networks on the basis of a multilayer perceptron. Indicators of reliability of the constructed neural network models are presented in table 1.

|        | MLP | RBF-network | Boosting | Bagging |
|--------|-----|-------------|----------|---------|
| 92.3 % | 90.3 % | 100 %      | 97.9 %   |

As it is stated above, creation of NN boosting-ensemble for identification of functional dependences and forecasting of financial insolvency of the water supply enterprises is inexpedient in connection with the small volume of selection of basic data and also high level of reliability of basic model.

5. Conclusion
The aim of this paper was to create a neural network methods with the increased stability and reliability for bankruptcy forecasting. We compared and combined four different NN-methods (MLP, RBF-network, Boosting-ensemble and Bagging-ensemble) to identify the method which will be capable to distinguish most authentically the status of the enterprises of a system of resource provisioning, and that is especially important, to predict financial insolvency of the enterprises.

Among the above NN-methods, the best is bagging-ensemble of neural networks on the basis of the "multilayer perceptron". This model allowed to gain the following advantage in comparison with classical neural network approximators: increase in reliability on average by 5.6% in comparison with a multilayer perceptron, for 7.6% in comparison with radial basic function.

During practical activities interpretation of the received estimates of possible insolvency of the water supply enterprises is made. By results of the revealed signs of bankruptcy and also risks of destabilization of the water supply system in a number of regions measures for their minimization and decriminalization of the industry were taken.

References
[1] Bishop C M 2006 Pattern recognition and machine learning Springer
[2] Lancaster P and Salkauskas K 1986 Curve and Surface Fitting (London, Great Britain Academic Press)
[3] Franke R 1987 Recent Advances in the Approximation of Surfaces from Scattered Data. Topics in Multivariate Approximation (New York, USA: Academic Press)
[4] Kansa E J 1990 Multiquadrics – a scattered data approximation scheme with application to computational fluid dynamics I Computers Math. Applic 19(8-9) 127-45
[5] Kansa E J 1990 Multiquadrics – a scattered data approximation scheme with application to computational fluid dynamics II Computers Math. Applic 19(8-9) 147-61
[6] Poggio T and Girosi F 1990 Networks for approximation and learning Proceedings of the IEEE 78(9)
[7] Dyn N 1991 Interpolation and Approximation by Radial and Related Functions. Approximation Theory VI (New York, USA: Academic Press) p 211-34
[8] Powell M J 1992 The Theory of Radial Basis Function Approximation. Advances in Numerical Analysis Wavelets, Subdivision Algorithms and Radial Basis Functions II (Oxford University Press) p 105-210
[9] Girosi F 1992 On some extensions of radial basis functions and their applications in artificial intelligence Computer Math. Applic 24(12) 61-80
[10] Girosi F 1993 Regularization theory, radial basis functions and networks (From Statistics to
Neural Networks) Theory and Pattern Recognition Applications Springer-Verlag

[11] Buhman M D 2003 Radial Basis Functions: Theory and Implementations (Cambridge: Cambridge University Press)

[12] Franke R 1982 Scattered data interpolation: tests of some method Math. Comp. 38(5) 181-200

[13] Micchelli C A 1986 Interpolation of scattered data: distance matrices and positive definite functions Constructive Approximation 2 11-22

[14] Powell M J 1987 Radial Basis Functions for Multivariable Interpolation: a Review. Algorithms for Approximation (Oxford: Claredon Press)

[15] Dyn N 1987 Interpolation of Scattered Data by Radial Functions. Topics in Multivariate Approximation (New York: Academic Press)

[16] Broomhead D S and Lowe D 1988 Multivariate functional interpolation and adaptive networks Complex Systems 2 321-55

[17] Buhmann M D 1990 Multivariate cardinal interpolation with radial basis functions Constructive Approximation 6 225-55

[18] Freund Y and Seapire R E 1999 A short introduction to boosting Journal of Japanese Society for Artificial Intelligence 14(5) 771-80