THE PREDICTION OF CRIMINAL PHENOMENA BY USING NEURAL NETWORK

PETRUCHA, Jindrich – VESELA, Radomira – KOVARIK, Pavel, ČR

Abstract: The contribution is based on the fact that crime is a specific human behaviour that can be measured in part, and also to predict. The article discusses the use of artificial neural networks in predicting crime and their trends over time.

Primary data on criminality were obtained from official police statistics. After adjustment of statistical data in the required format model was developed for the processing of data by type of crime, identified and explained both in the entire Czech Republic and in selected regions. The data used to create patterns for simulator artificial neural network. Results prediction for the next period are compared with the real situation. Application of this model, allows sequential monitoring of criminal activity and its evaluation from different perspectives. In conclusion, the data are compared with the data obtained by statistical calculation, which allows a comparison with the results of artificial neural network simulator.

Key words: neural network, neural architecture, PHP program, economic crime, statistic model, regression

1 Introduction

Application of artificial intelligence finds application in many sectors of human activity. These tools use the opportunity to extract data from the Internet and then convert them into information that is used for decision making. The importance of these tools is the possibility of continuous learning and creating data models, which are stored for later use knowledge in practice. One of these tools as neural networks, which are implemented through programs in various programming languages for creating simulators of artificial neural networks.

The article deals with using your own simulator created in PHP language as the client application server on the Internet using prediction in crime in the Czech Republic. Data processing these criminal phenomena are used on the Internet in the form of statistical data of the Ministry of Interior. Statistical data are in the form of spreadsheets and processing simulator requires an adjustment to the format of the series. Statistical data are in the monthly periods in incremental system. The building of data time series requires preprocessing and data cleaning, this stage is difficult because of the different variants of crime and the calculation of solved crimes.

Criminality is the most dangerous socially pathological phenomenon. The content of this dynamic phenomenon changes depending on time. In the past quarter-century, there have been significant qualitative and quantitative changes which are the result of the dynamic social development. There is a need to react to these massive changes regarding its prevention.

The prevention of criminality is understood as a collection of various activities of non-criminal nature oriented toward removal, weakening or neutralization of criminogenic factors aiming to stop the increase of criminality or achieve its reduction. The current situation of criminality in the Czech Republic, its changes in comparison with the past, and the prediction of changes in the future in particular are evaluated based on the statistical data from the official police records. The analysis of these data and its subsequent

prediction of the development of crime are significant in terms of prevention of threats of the increase of criminality or decrease of its detention rate.

2 Proposed simulator neural networks

The simulator artificial neural network is constructed using objects that represent each phase of the neural network as training and testing of artificial neural networks. This is the architecture of feedforward neural networks with the use of three layers, where the input layer is used for the input vector time series. The hidden layer contains neurons with sigmoid transfer function and is used to adjust the weights between the joints of individual neurons. The output layer contains one neuron with sigmoid transfer function and represents a prediction for the next period.

In terms of design concept, the user can select the parameter the number of neurons in the input layer and the number of neurons in the output layer. Another important parameter which the user can change the number of periods for learning. These parameters most affect the outcome of the learning process and are very important for prediction accuracy. Outcome prediction is shown by the graph in which the time series of input values and the values of the prediction.

2.1 The algorithms used for neural network simulator

For learning algorithm is used backropagation algorithm backpropagation signal and calculating weights between neurons. Individual values are stored in memory as more dimensional array of individual layers. Data for the input layer are read from the input browser window and further processed according to the size of the input layer artificial neural networks. This creates a series of patterns that are submitted to an artificial neural network in the learning phase of the network. Individual values for input to a simulator are normalized to the interval of values from 0 to 1.

2.2 Create simulator of neural network as WWW page of internet

Web simulator is designed as a Web page using the PHP language in which they are written algorithms in object simulator variant using classes that manipulate data. Other parts are entry forms, which are in HTML using CSS types for creating graphical elements on a Web page.

An important part is the use of an output chart in which are shown the data after learning the simulator artificial neural networks, and output values of prediction. This graph is created using GOOGLE API GRAPH object. For these reasons, under certain circumstances slower response when working with the simulator with a larger amount of data time series. The advantage is that the graph himself adjusts the image size by number of used data that are entered by the user to input a web page form.

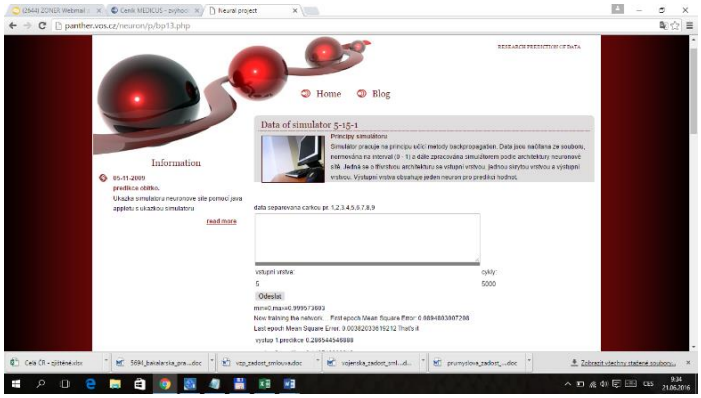

Fig. 1: Input window www simulator artificial neural networks

3 Data preparing for neural network simulator

Preparing the data base is one of the most important parts of the preparation process. Of the various data has been selected characteristic data to the Zlín region and the overall data for the Czech Republic. These data are in the form of excel tables where individual values detected and solved crimes with numerical values for individual months.

Data processing can be obtained from the address of the Ministry of the Interior
http://www.policie.cz/statistiky-kriminalita.aspx
www.policie.cz/clanek/statisticke-prehledy-kriminality-za-rok-2016.aspx

The abovementioned data from internet resources include not only crimes detected in the individual months from January 2015 to March 2016, but also crimes solved in these period. Besides the total rate of criminality in the Czech Republic, the data also contain economic crimes in the Czech Republic as each attack on the country's economy has a social and economic impact on its internal stability. The same indicators are observed in the Zlín region where the University resides.

3.1 Chart of prediction from neural network simulator

The picture Fig. 2 shows the graphical output of the simulator, artificial neural networks, which displays the input time series data using the first line (green) and predicting a second line (yellow) in a graph that represents the prediction using neural network algorithm.

The first part of this series prediction starts at zero value, which represents the data used for the input layer of the neural network. The first predictive value is then up to the next step after the input data to the input layer. The patterns are created in PHP language program, which after learning neural network processes the data into a format suitable for graphical display using a graphic object.

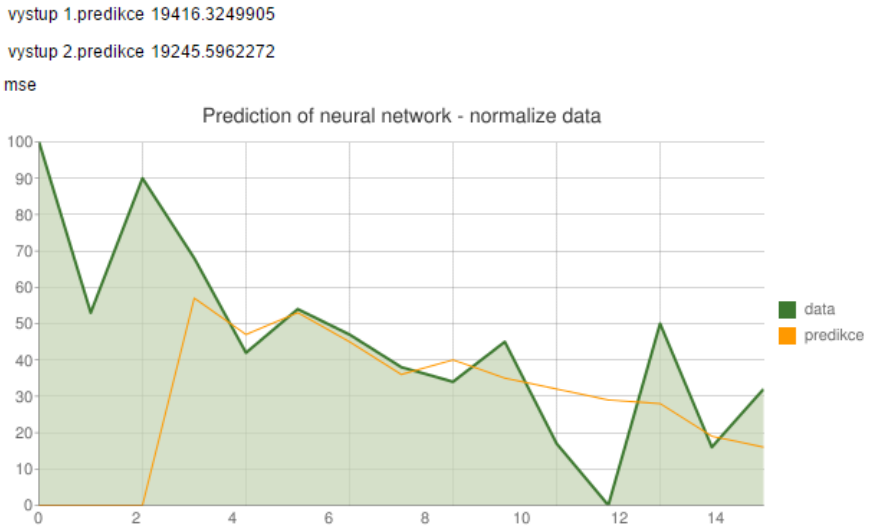

Fig. 2: Chart prediction simulator neural network for total rate of criminality

After learning neural network simulator is made predictions for two periods these values are at the top of the Fig 2.

4 Comparison the results neural networks and statistical methods

Statistical methods allow for processing time series regression analysis using the least squares method. Least squares method is used to approximate the input values selected functions to create an approximation model. In the case of processing time series for criminal phenomena line is used for linear approximation, further parabola second order polynomial and a polynomial of the third degree and fourth-degree polynomial.

Processing of individual time series was done in Excel using processing functions regression analysis.

Individual coefficients of regression functions are listed in the following items regression functions

(1) y=22475,714 - 252,6393x

(2) y=23543,64 - 629,553x + 23,55713x^2

(3) y=23261,26 - 446,61x - 4,12663x^2 + 1,15349x^3

(4) y=24027,53 - 1191,26x + 190,0267x^2 - 17,2981x^3 + 0,576611x^4

Each prediction using regression functions are listed in the following table.

Value x	value y	line	parabolic	kubic
January 15	23467	22223,1	22937,6	22811,7
Februar 15y	20879	21970,4	22378,8	22360,8
March 15	22884	21717,8	21867,0	21915,4
April 15	21694	21465,2	21402,3	21482,6
May 15	20219	21212,5	20984,8	21069,2
June 15	20936	20959,9	20614,4	20682,2
July 15	20530	20707,2	20291,1	20328,4
August 15	20029	20454,6	20014,9	20014,9

September 15	19790	20202,0	19785,8	19748,4
October 15	20418	19949,3	19603,8	19536,0
November 15	18873	19696,7	19469,0	19384,5
December 15	17909	19444,0	19381,2	19300,9
January 16	20669	19191,4	19340,6	19292,2
February 16	18821	18938,8	19347,1	19365,1
March 16	19701	18686,1	19400,7	19526,7
April 16	18942	18433,5	19501,4	19783,8
May 16		18180,8	19649,2	20143,4

Table 1: values of the time series and prediction using regression functions

Graphical output lines and values of the time series in the chart below.

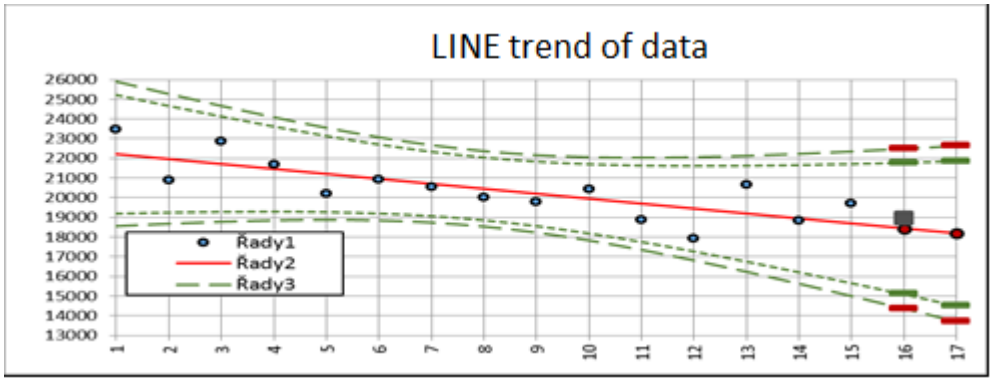

Fig 3: Graphical input with approximations using the line

The results achieved by simulation of artificial neural networks:

Number epochs 15000

The number of neurons in the hidden layer 15

The number of neurons in the input layer is varied from 3 of 7, and is listed in the table as the first column

Overall crime CR detected crimes and crimes solved

the number of neurons in the output layer	The first prediction detected crime	The second prediction detected crime	The first prediction crimes solved	The second prediction crimes solved
3	19416	19245	9668	9827
5	19139	19152	9412	9499
7	19174	19050	9398	9399

Tab 4 outputs the prediction from the neural network, total crime Czech republic

the number of neurons in the output layer	The first prediction detected crime	The second prediction detected crime	The first prediction crimes solved	The second prediction crimes solved
3	2582	2548	1625	1617
5	2490	2420	1571	1569
7	2174	2451	1595	1590

Tab 4 outputs the prediction from the neural network, economic crime Czech republic

the number of neurons in the output layer	The first prediction detected crime	The second prediction detected crime	The first prediction crimes solved	The second prediction crimes solved
3	89	90	66	64
5	89	87	66	66
7	90	89	64	64

Tab 4 outputs the prediction from the neural network, economic crime Zlin region

Individual tables contain data for Detected Crimes and crimes solved, identifying optimal variants prediction. The tables show that the optimal value of the prediction is achieved at the value of 5 for the number of neurons in the input layer. The values of the time series shows, a declining trend over the past year in the area identified and solved crimes.

Conclusion

In conclusion we can say that criminal phenomena in the Czech Republic are declining character, which can be well approximated by a statistical method using the straight-line equation. If we observe the dynamic nature of the development of criminal acts, it is advisable to use artificial intelligence tool that better captures the movement data for individual months. An important role for neural network plays a neural network architecture that can be controlled with the size of the input layer. This value should be chosen in the range of 3-10 neurons and test the output from the simulator. Statistical method shows the long-term trend, and its predictions are suitable for annual monitoring of the overall trend in crime in the Czech Republic.

Bibliography

[1] JELÍNEK, J.*Trestní zákon a trestní řád s poznámkami a judikaturou* 6. aktualizované vydání podle stavu k 1. 2. 2016. Praha: Leges, 2016. 1280 s. ISBN 978-80-7502-106-9

[2] NOVOTNÝ, O., ZAPLETAL, J. a kol., *Kriminologie* Praha: ASPI – Wolters Kluwer, 2008. 528 s. ISBN 978-80-7353-376-8

[3] ZHANG, P. G. *Neural networks in business forecasting*. LONDON: IDEA GROUP PUBLISHING, 2004. ISBN 1-59140-176-3.

[4] Policie.cz. 2016. *Statistické přehledy kriminality* [on-line] [cit. 2016-05-25]. (Online). Available http://www.policie.cz/statistiky-kriminalita.aspx.

Contact address:

Jindřich Petrucha. Ph.D.,
Department of Informatics , European Polytechnic Institute, Kunovice 699 Czech Republic
Phone: +420 572 549 018 , e-mail: petruchap@edukomplex.cz
Radomíra Veselá. Ph.D.,
Department of Informatics , European Polytechnic Institute, Kunovice 699 Czech Republic
Phone: +420 572 549 018 , e-mail: vesela@edukomplex.cz
Pavel Kovařík. Doc.,
Department of Informatics , European Polytechnic Institute, Kunovice 699 Czech Republic
Phone: +420 572 549 018 , e-mail: kovarik@edukomplex.cz