Operational Forecasting of Road Traffic Accidents via Neural Network Analysis of Big Data

Oleg Golovnin
Department of Information Systems and Technologies
Samara National Research University
Samara, Russia
golovnin@ssau.ru

Ekaterina Sidorova
Department of Information Systems and Technologies
Samara National Research University
Samara, Russia
sidoroekaterina@gmail.com

***Abstract*—The paper proposes an approach to the operational forecasting of traffic accidents with the separation of accident types based on the multilayer Rumelhart perceptron. The approach is applied to analyze Big Data collected from external heterogeneous data sources and traffic control systems or Smart City solutions. The approach increases the accuracy of determining the accident possibility by simultaneous analysis of multiple parameters covering weather conditions, conditions of the road and control devices, seasonal traffic fluctuations, traffic flow, individual vehicle speed, organizational factors, and events. The software implementation of the approach uses the TensorFlow framework and the Keras library. The experiments showed that the approach provides a 90% accuracy in recognizing situations. The forecast results are useful within an hour from the calculation moment, which is enough to react to an emergency situation or notify the drivers. The software is intended to function as part of accident prevention systems and, in this case, could reduce an accident rate and severity and increase the awareness of traffic participants.**

Keywords— traffic accident, forecasting, TensorFlow, Keras

I. Introduction

A tangible problem for the transport complex of modern urban agglomerations is road traffic accidents that damage drivers, pedestrians, vehicles, cargos, and transport infrastructure, which, in turn, leads to economic and social costs. According to the data of the State traffic inspectorate, only in October 2019, 17.0 thousand traffic accidents occurred on federal roads of the Russian Federation, in which 3.9 thousand people were registered dead and 25.6 thousand people injured [1]. Digitalization of management processes, development of intelligent technologies and big data processing methods have led to the emergence of new solutions that can be used in the task of operative forecasting the occurrence of accidents for taking preventive countermeasures to prevent accidents [2].

Actual problems of modern traffic that can be detected or predicted before an accident can be occurred [3]:

- inappropriate speed limits for vehicles;
- extreme weather conditions;
- violation of traffic rules;
- damage to the roadbed and technical means of traffic management;
- dangerous behavior: aggressive driving, obstructing overtaking, failure to maintain a safe distance between vehicles, sudden braking, pedestrians entering the roadway.

The development of active and passive means of ensuring road safety in recent years has significantly reduced the number of accidents and the severity of their consequences, however, in some situations, the measures introduced are insufficient, for example, when a vehicle is drifted on a slippery road or if the driver is inattentive [4]. Ensuring road safety and reducing damage from a predicted accident can be achieved through directive and indirect impact on-road behavior by actively controlling traffic lights and road signs with variable information, prompt notification of special services, as well as informing road users.

This paper proposes an approach to predict the possibility of a road accident through a neuro network analysis of Big Data from different traffic control systems.

II. State-of-the-art

In this study, let an accident is a road traffic incident that occurred with the participation of at least one vehicle during its movement on the road network, in which people were injured or killed, or damage was caused to the vehicles, cargos, transport infrastructure and facilities [5].

At the moment, the active development of methods and tools is underway to detect [6], predict [7], inform [8], and prevent accidents [9]. There are a number of solutions based on measuring sensors [10, 11]. In [12], an approach based on infrared sensors is described, which ensures operation in a two-phase mode: accident detection, accident prevention. The implementation of the approach operates with indicators of traffic congestion but does not take into account other factors that may affect the modeling of a dangerous situation [13].

In [14], a model of short-term traffic flow forecasting taking into account spatial and temporal channeling is presented. The model was implemented using the Apache Spark framework based on the MapReduce distributed computing model, thereby achieving a high speed of operation sufficient for online prediction but a functional block was not implemented for analyzing the possibility of road accidents.

An intelligent approach based on a neural network that automatically detects an accident that has already occurred according to indirect traffic data is presented in [15]. The approach is based on the assumption that the average speed of the traffic flow is changing in the case of an accident. The proposed approach does not predict the possibility of an accident. In [16], geoinformation models for managing traffic flows in the event of an accident are detected, but reliable data on the fact of an accident are not obtained.

In [17], several controlled training methods were analyzed to classify the degree of damage resulting from an accident: fatal injuries, severe injuries, minor injuries, and a car accident. The solution proposed in this work cannot be used to monitor the situation on the road network and, accordingly, cannot be used as part of an accident prevention system.

In [18], a method was proposed for determining the temporal characteristics of accidents based on a high-speed thermogram, but it provides low-quality indicators. In [19], it was proposed to use wavelet spectrograms to assess the characteristics of traffic flows, but the determination of the factors leading to an accident is possible only by indirect signs with an insufficient time accuracy of the event binding.

This study proposes an approach to forecasting accidents by type of accident, using the Rumelhart multilayer perceptron [20] as applied to Big Data coming online from external heterogeneous data sources providing, for example, weather conditions, road network and road traffic characteristics, events on the road network, etc.

III. The Neural Network Model

A. Input Data Description

For training and testing the neural network, a model of training with a teacher is used, therefore n-dimensional vectors describing the input data are required. Input data consist of the weather, road, and organizational factors.

The data on the road network includes:

- type of motion control device, TRAFDEV ϵ {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, x1;
- state of the motion control device, TRAFFUNCT ϵ {0, 1, 2, 9}, x2;
- speed limit, SPEEDLIMIT ϵ {0, 24, 25, ..., 119, 120, 121, 999}, x3;
- type of road, RELTOJUNCT ϵ {0, 1, 2, 3, 4, 5, 9}, x4;
- type of pavement, SURFTYPE ϵ {1, 2, 3, 4, 5, 8, 9}, x5;
- the number of lanes, RDLANES ϵ {1, 2, 3, 4, 5, 6, 7, 9}, x6;
- type of dividing strip on the right, LINERIGHT ϵ {0, 1, 2, 3, 4, 5, 9}, x7;
- type of dividing strip on the left, LINELEFT ϵ {0, 1, 2, 3, 4, 5, 9}, x8.

The data on the weather includes:

- weather condition, WEATHER ϵ {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 98, 99}, x9;
- road surface condition, SURFCOND ϵ {1, 2, 3, 4, 5, 6, 7, 8, 9, 98, 99}, x10;
- lighting condition, LIGHTCOND ϵ {1, 2, 3, 4, 5, 9}, x11.

Date and time data are described as the following:

- time, CRASHTIME ϵ Time, x12;
- day of the week, DAYOFWEEK ϵ {1, 2, 3, 4, 5, 6, 7}, x13;
- month, CRASHMONTH ϵ {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}, x14.

Vehicle and event detected data includes:

- type of vehicle, BODYCAT ϵ {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 99}, x15;
- event on the road, CRITCAT ϵ {1, 2, 3, 4, 5, 6, 8, 9}, x16;
- vehicle speed, DVEST ϵ {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, x17.

B. Output Data Description

The output data are a vector with the following possible values:

- no accidents, y1;
- head-on collision, y2;
- side collision, y3;
- rear collision, y4;
- rollover, y5;
- collision with an object off the road, y6;
- collision with an object on the road, y7;
- another type of accident, y8.

Since the type of accident is encoded as an integer, one-hot coding is used to solve the multiclass classification problem for the received categories: 0 → [1,0,0,0,0,0,0,0,0], 1 → [0,1 , 0,0,0,0,0,0,0], ..., 7 → [0,0,0,0,0,0,0,0,1].

C. Neural Network Topology

Accident forecasting from the point of view of classical machine learning refers to the problem of multiple classifications. Thus, in accordance with predetermined input and output data, the number of input and output neurons is determined: a vector of 17 values is input to the neural network, and a vector of 8 values is output. The neural network is based on the Rumelhart perceptron with 1 hidden layer.

The topology of the used artificial neural network is shown in Fig. 1.

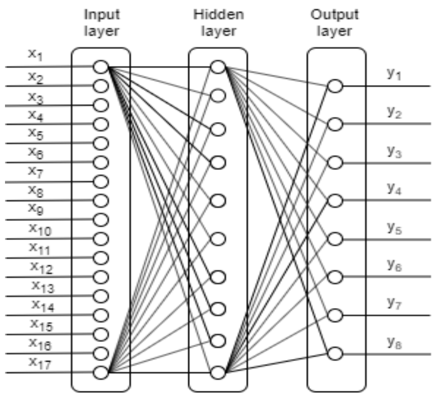

Fig. 1. Neural network topology.

The number of hidden layer neurons is determined by the rule of the geometric pyramid:

$$k = \sqrt{nm}, \tag{1}$$

where k is the number of neurons in the hidden layer; n is the number of neurons in the input layer; m is the number of neurons in the output layer.

Thus, the number of neurons in the hidden layer is 12.

IV. Software Implementation

The proposed neural network approach for accident prediction is implemented in the form of a software profiling subsystem designed to function as part of an accident prevention system (Fig. 2). Data streams entering the profiling subsystem are logically combined into data sources. The profiling subsystem is implemented in Python in the PyCharm environment; the TensorFlow framework and the Keras library are used to implement neural networks.

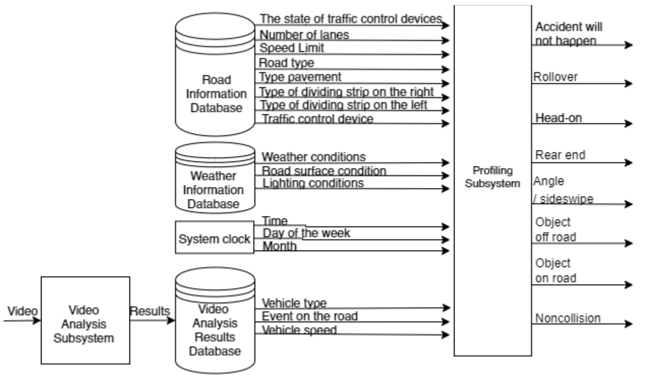

Fig. 2. Data streams.

The resulting traffic accident analytics is stored in the database. For access and data management, the psycopg2 library and the PostgreSQL, which provides spatial-temporal binding of accidents, were used. The incoming accompanying data on the road, weather, vehicles are also logged in relation to an accident (Fig. 3).

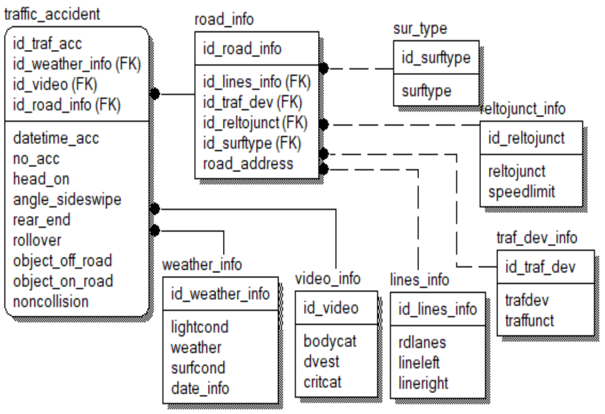

Fig. 3. Data model.

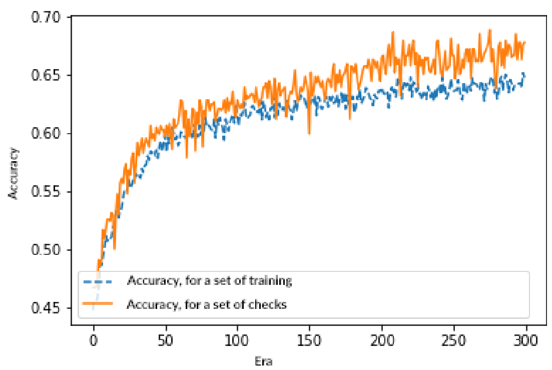

Fig. 4. The graph of the number of correct classifications of the neural network from the era of training. The final era of 300.

The profiling subsystem does not have a graphical interface for the end-user, so it provides an API for external subsystems connections.

A graphical interface has been separately developed using the API, to train and configure the neural network, to force the start of forecasting for the indicated data sets, and to view notifications of possible accidents in the situation center mode.

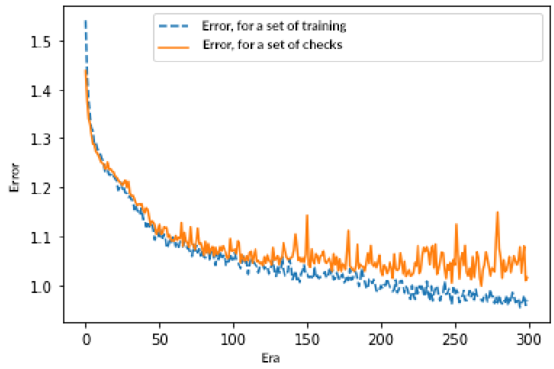

Fig. 5. Graph of the number of erroneous classifications from the era of training. The final era of 300.

V. Results

For training the neural network, reliable data of a special format were used, which are freely available on the data.gov.uk server under the OGL (Open Government License) license [21].

To evaluate the results, we used a graphical interpretation of the results and the metric roc_auc_score of the sklearn package. To improve the accuracy of training, thinning of 20% was used. The best results were shown by the number of training examples for one training of 7. Figures 4 and 5 show graphs of the dependence of accuracy and error indicators for each epoch with a final 300 epoch when using 7 examples for training at a time.

Starting from the 200th era, indicators remain approximately in the same aisles, while the metric roc_auc_score shows a result of 0.90.

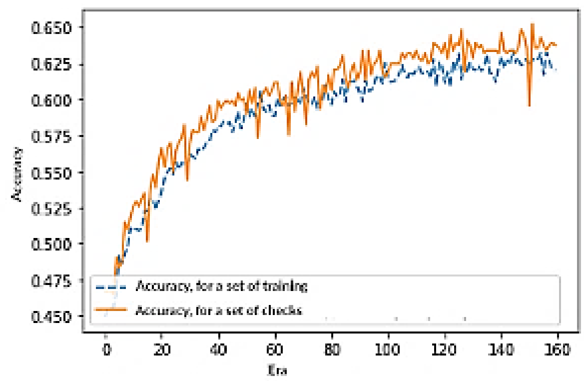

Fig. 6. The graph of the number of correct classifications of the neural network from the era of training. The final era of 160.

In order to prevent the retraining of the neural network, the number of epochs was reduced. The graphs in figures 6 and 7 show the performance of learning in 160 eras; the metric roc_auc_score shows the result of 0.91. Thus, with the same learning parameters and network topology, 160 epochs are effective for learning.

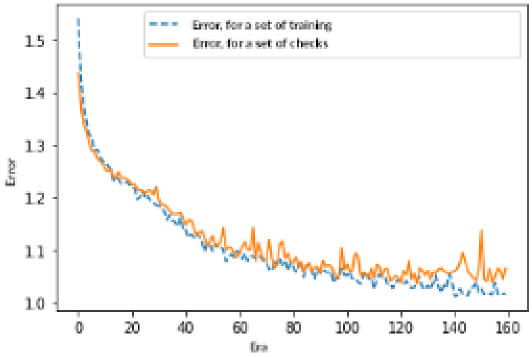

Fig. 7. Graph of the number of erroneous classifications from the era of training. The final era of 160.

Since forecasting is carried out taking into account the indicators of events recognition, according to the results given above, it is possible to estimate the time period when solving the forecasting problem. In a stationary mode, the situation on the road is constantly monitored and indicators that change every second are taken into account, which leads to a maximum value of the time interval of one minute. However, due to the fact that relatively constant indicators are taken into account, such as weather conditions and road network conditions, the forecast results can be useful within an hour from the moment they are calculated, for example, in the form of indicative data for drivers.

Therefore, we can conclude that the results obtained make it possible to notify drivers or emergency services in advance about the dangerous conditions on the road. The simultaneous analysis of multiple parameters using the proposed approach let consider almost all the factors affecting road safety.

VI. Conclusion

Improving road safety with the latest achievements of science and technology is an obvious way for a developed society to reduce the number of incidents and accidents. Intelligent transport systems, systems for the Smart Cities, advanced technical means of ensuring passive and active safety are constantly being improved. The introduction of technologies for processing Big Data and machine learning seems effective in many areas of the transport industry, including predicting the possibility of an accident.

The approach presented in this work increases the accuracy of determining the possibility of an accident by analyzing classes of parameters covering such important factors as weather conditions, conditions of the road and control devices, seasonal traffic fluctuations, traffic flow, and individual vehicle speed. Classification by accident type provides the most effective measures aimed at preventing a specific type of accident. The experiments carried out identified the most effective parameters of the neural network and achieved the accuracy of situations recognition in 90%. The software implementation of the proposed approach in integration with accident prevention systems can achieve a reduction in accident rate, reduce the severity of the consequences of an accident, and increase the awareness of traffic participants.

References

[1] State traffic inspectorate [Online]. URL: http://stat.gibdd.ru.

[2] V.I. Mayorov, "Risk management in the system of road safety," Bulletin of the Ural Institute of Economics, Management and Law, vol. 3, no. 44, pp. 8-12, 2018.

[3] D.V. Sokolov and E.V. Koluzakova, "Actual problems of ensuring road safety in the Russian Federation," Science. Thought, vol. 5-2, 2016.

[4] A.P. Tarko, "Surrogate measures of safety Safe Mobility," Challenges, Methodology and Solutions, pp. 383-405, 2018.

[5] Russian Federation Law No. 196-FZ "On Road Traffic Safety," 2019.

[6] U. Khalil, A. Nasir, S. M. Khan, T. Javid, S. A. Raza and A. Siddiqui, "Automatic road accident detection using ultrasonic sensor," Int. Multi-Topic Conf. (INMIC), pp. 206-212, 2018.

[7] M. Zheng, "Traffic accident's severity prediction: A deep-learning approach-based CNN network," IEEE Access, vol. 7, pp. 39897-39910, 2019.

[8] W. Tai, H. Wang, C. Chiang, C. Chien, K. Lai and T. Huang, "RTAIS: road traffic accident information system," Int. Conf. on High Performance Computing and Communications, Exeter, pp. 1393-1397, 2018.

[9] M. Lamr, "Big Data and its usage in systems of early warning of traffic accident risks," in Int. Conf. on Enterprise Systems (ES), pp. 154-157, 2018.

[10] Y. Pan, L. Zhang, X. Wu, K. Zhang and M.J. Skibniewski, "Structural health monitoring and assessment using wavelet packet energy spectrum," Saf. Sci., vol. 120, pp. 652-665, 2019.

[11] O. Dehzangi, V. Sahu, V. Rajendra and M. Taherisadr, "GSR-based distracted driving identification using discrete & continuous decomposition and wavelet packet transform," Smart Health, vol. 14, 100085, 2019.

[12] N.T.S.A. Wadhahi, S.M. Hussain, K.M. Yosof, S.A. Hussain and A.V. Singh, "Accidents detection and prevention system to reduce traffic hazards using IR sensors," Int. Conf. on Reliability, Infocom Technologies and Optimization (ICRITO), pp. 737-741, 2018.

[13] O.K. Golovnin, "Data-driven profiling of traffic flow with varying road conditions," CEUR Workshop Proceedings, vol. 2416, pp. 149-157, 2019.

[14] A.A. Agafonov, A.S. Yumaganov and V.V. Myasnikov, "Big Data analysis in a geoinformatic problem of short-term traffic flow forecasting based on a K nearest neighbors method," Computer Optics, vol. 42, no. 6, pp. 1101-1111, 2018. DOI: 10.18287/2412-6179-2018-42-6-1101-1111.

[15] Y. Ki, J. Kim, T. Kim, N. Heo, J. Choi and J. Jeong, "Method for automatic detection of traffic incidents using neural networks and traffic data," Annual Information Technology, Electronics and Mobile Communication Conf. (IEMCON), pp. 184-188, 2018.

[16] T.I. Mikheeva, A.A. Osmushin, S.V. Mikheev and O.K. Golovnin, "GIS-based models for transport network emergency management," Journal of Physics: Conference Series, vol. 1353, no. 1, 012009, 2019.

[17] M.F. Labib, A.S. Rifat, M.M. Hossain, A.K. Das and F. Nawrine, "Road accident analysis and prediction of accident severity by using machine learning in bangladesh," Int. Conf. on Smart Computing & Communications (ICSCC), pp. 1-5, 2019.

[18] J. Zhang, J. Wang and S. Fang, "Prediction of urban expressway total traffic accident duration based on multiple linear regression and artificial neural network," Int. Conf. on Transportation Information and Safety (ICTIS), pp. 503-510, 2019.

[19] O.K. Golovnin and A.A. Stolbova, "Wavelet analysis as a tool for studying the road traffic characteristics in the context of intelligent transport systems with incomplete data," SPIIRAS Proceedings, vol. 18, no. 2, pp. 326-353, 2019.

[20] O.A. Samonina, "Methods and problems of training a multilayer neural network used to assess the characteristics of the designed information systems," Bulletin of the St. Petersburg University of Railway Engineering, vol. 4, pp. 148-159 , 2008.

[21] Road Safety Data [Online]. URL: https://data.gov.uk/dataset/road-accidents-safety-data.