Journal of Theoretical and Applied Information Technology
15th May 2020. Vol.98. No 09

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DETECTION AND ELIMINATION OF DISCREPANCIES IN BIG DATA AT TRANSPORT APPLYING STATISTICAL METHODS

AZAT TASHEV[1], JANNA KUANDYKOVA[1], DINARA KASSYMOVA[1,2], AINUR AKHMEDIYAROVA[1]

[1] Institute of Information and Computational Technologies CS MES RK, Kazakhstan

[2]Doctoral student of Kazakh National Research Technical University named after K.I. Satbayev, Kazakhstan

E-mail: [1]dikakassymova@gmail.com, [2]dika.cat@mail.ru

ABSTRACT

An article herein considers the problems of discrepancies detection and elimination upon processing the big data at transport. Tasks of detecting and eliminating the discrepancies in the data has been solved by means of Grabbs method. To obtain trip time design characteristics values there have been applied statistical methods, which allow correct their prescheduled values. The given methodology is used the big data processing at transport in real time mode.

Keywords: *Smart transport system, discrepancies, big data*

1. INTRODUCTION

Using the information and communication technologies for urban management with the aim to upgrade quality of services to the urban residents is one of the main tasks upon constructing a smart city. Applying the technologies thereof reduces infrastructure and operation costs, upgrades efficiency of available resources usage, as well, improves urban residents interrelations with city transport [1]. Nowadays, such technologies are widely employed in contemporary cities transport service sphere. Today, public transport is a role-defining category in the city infrastructure formation for maintaining citizens mobility [2]. In particular, it is felt in heavily populated areas. At present transport systems do not completely meet the passengers ever-growing requirements. To solve such problems it is indispensable to have the knowledge of overall situation and information, concerning the city transport [3]. Unfortunately, the solution of city transport problems is usually based on unreliable information. Therefore, the decision taking process in reference to the transport situation is not optimal.

Rapid development of Kazakhstan economy and automobile industry has resulted in sharp increase of transportation load. As of December 1, 2019, in the Republic of Kazakhstan number of registered passenger cars constituted 3768,7 thousand units, thereat, over millions of them are in Almaty. Thus, less, than for 5 years amount of automobiles, Almaty citizens own, almost doubled. Such great automobiles possession has quickly brought to the rise of social costs – including traffic jams on the roads and road accidents in Almaty city.

Usage of intelligent transport system plays an important role in smart cities transport systems. Intelligent transport systems include: synergic, synergetic technologies, artificial intellect, engineering principles, applied to transport systems for increasing the traffic capacity, maintaining security, etc. [4]. Intelligent transport systems allow collect large-scale data on transport and means of its movement [5], and fulfill processing the obtained information to optimize traffic current and citizens transportation.

Kazakhstan, in the frame of transport and logistic spheres digitization programs, creates intelligent transportation systems (IITS). One of ITS components is special automated measuring means, being installed at vehicles and automobile transport corridors. It secures monitoring and account of traffic intensity, excluding groundless stops. The system consists of road media boards networks, providing the road traffic participants with useful information: guidance to the nearest streets and objects with estimated time en-route to them, traffic stream speed limitation on the way, information on traffic jams and bypass roads, messages about

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

emergency situations, messages about oncoming emergency ambulance, operational messages from city administration, etc. The network is managed from a single center, able to process big data. It allows predict road situations and plan changes in road traffic, dependent on various situations, such as, weather, roads under repair, striping change, change of traffic lights working mode, etc.

In the work [6] a large-scale task is broken down into plenty of small subproblems, which are solved in parallel at different computational elements, which increases processing speed. For the recent 10 years there have been elaborated various systems for the big data analysis, using distributed computations, based on cloud technologies, inclusive [7].

Previously, the authors have developed the algorithm of specifying the maximum flow upon distributing in the network [8] and considered the problem of locating minimal number of chambers in the given transportation network [9]. The scientific work herein is a logic continuation of the work on the research topic.

At present, in Transport Holding of Almaty city, mainly, the data on routes drive and by card Onai are stored in archive without processing. The first step to upgrading the performance of the city transport is big data processing, using the new information technologies. In the result there will be improved:

i. public transport operation quality, based on the analysis of buses location history;

ii. passengers mobility applying the analysis of tickets purchase with transport cards.

One of the basic tasks upon big transport data processing is detection and elimination of discrepancies.

The article herein presents an approach to analysis and treatment of public transport big data, based on applying the statistical methods with elimination of contradictory information.

2. REVIEW OF WORKS ON RELATED TOPICS

Some approaches to detecting discrepancies and big data processing and their advantages are given in the Table 1, 2.

Table 1: Examples of techniques for transport data processing

Method	Advantage	Employment
k- nearest neighbors[10]	Prediction precision, efficiency and sustainability	Traffic state forecasting
Random forest, Bayesian inference [11]	Traffic improvement and safety, active traffic control	Security maintenance at urban freeways
Bayesian classifier, support vector regression (SVR)[12]	Experiment outcomes have shown, that an approach, using evaluation, based on SVR, has higher precision, than linear regression	Traffic flow prediction in real time
k- nearest neighbors, Gaussian process[13]	Processing time cut for 69%, an offered method can accurately predict traffic stream speed with a mean error, less than 2 miles per hour	Traffic speed prediction
k- nearest neighbors [14]	Upgrade of performance and scaling of transport streams short-term scaling, comparing to existing approaches	Traffic speed prediction
k- nearest neighbors, Bayesian inference, algorithm MOcell[15]	Employed multicriteria honeycomb genetic algorithm MOcell for optimizing the bus depot schedules with various busload	For improving the system of public transport

Table 2: Examples of training methods, used in anomalies identification

Method	Usage
Training method: controlled	
Hidden Markov Model, HMM	Controlled statistic Markov model, in which the system under simulation is considered to be Ma rkov process with hidden states: employed for anomalies detection [16].
Support Vector Machine (SVM)	Presentation of data points in space, displayed in such a way, that separate categories are divided, forming close-cut separation between them: special class SVM, namely, one class of SVM (OCSVM) is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	widely used for anomalies detection [17].
Gaussian regression (GR)	General controlled training technique, designated for solving the regression and probabilistic classification of the problem: used for anomalies detection from video [18],[19].
Convolutional Neural Networks (CNN)	Class of in-depth neural networks, applied conventionally for visual images analysis: owing to its applicability to retrieval of semantic level functions from the input, it has become popular in plenty of applications, including the anomalies detection [20].
Training technique: uncontrolled	
Latent Dirichlet Allocation (LDA)	Thematic model, using statistic analysis for obtaining the topics main distribution in documents: used to model video vivid words for anomalies detection [21]
Probabilistic Latent Semantic Analysis (pLSA)	Model for presenting the information about concurrent incoming in probability structure: used in [22] for anomalies detection.
Hierarchical Dirichlet process (HDP)	Nonparameteric Bayesian approach, designed, based on LDA, for data clustering: used at modeling data to detect anomalies [23].
Gaussian Mixture Model (GMM)	Probabilistic model, assuming, that all data points are generated from finite numbers mixture of Gaussian distributions with unknown parameters: used for anomalies detection [24].
Principal component analysis (PCA)	Orthogonal transformation statistic procedure for obdervations set transforming, possible, for correlated variables into values set of linearly uncorrelated variables: used for dimensionality cutoff [25].
Training technique: Hybrid	
HDP + HMM	Hybrid model: used for presenting the sub-trajectories in [26] for detecting anomalies, using MIL
CNN-LSTM	Hybrid model: detecting anomalies based on forecast, by means of CNN-LSTM [27]

3. PROBLEM STATEMENT

There is following initial data:

- Aggregate of random magnitudes actual values, representing headways, public transport running time, etc.
- Mathematical expectation a priori values and random magnitudes mean-square deviation. For instance, there prescribed public transport headways and their permissible deviations.

The task consists in tentative detecting and eliminating the discrepancies in the aggregate of random magnitudes actual values, as well, in estimation of mathematical expectation and random magnitude mean square deviation and adjustment of their a priori values.

4. THEORETICAL PART

4.1. General algorithm discrepancies detection and elimination.

Considerable role in the research plays the information quality, which is affected with contradictions, missings, bad values, ejections, etc. They might be defined and eliminated with various methods [28], computer-aided learning, inclusively. Fig.1 shows the general process of data pre-processing.

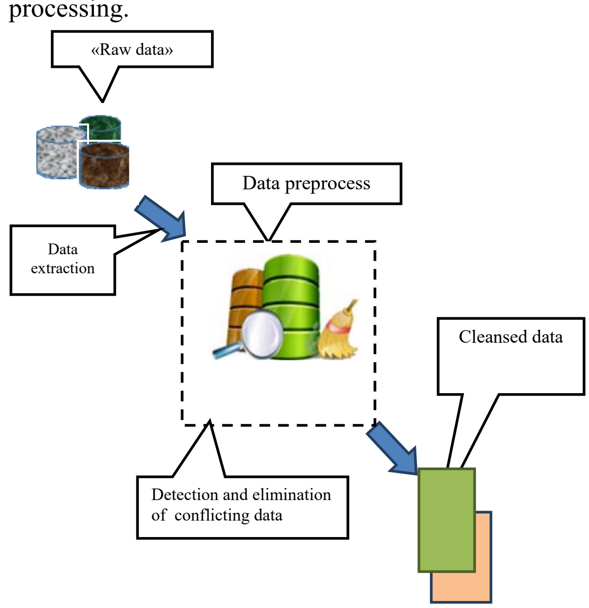

Figure 1: General process of data pre-processing

Figure 1 demonstrates, that the incoming data is subject to pre-processing (classified, cleaned, transformed, checked) and transmitted to subsequent analysis.

Figure 2 shows general block-scheme of detecting and eliminating the conflicting data in real time scale.

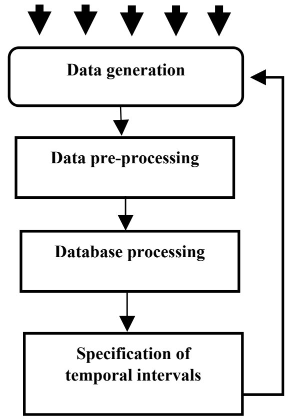

Figure 2: General block-scheme of detecting and eliminating the conflicting data (extra-bold arrows denote entering the new information into the base to form the database)

It follows from the Figure, that the data enters the system uninterruptedly and it is processed as and when it is received. The next step is the big data pre-processing. Hereby, we consider merely temporal contradictions, linked with public transport running in Almaty city.

Conflicting data processing consists of the following stages:

1) identifying conflicting data;
2 conflicting data processing

Conflicting data definition in the article herein is fulfilled by means of Grabbs method, and processing might be executed with one of the following techniques:

1) detected conflicting data is eliminated;
2) detected conflicting data is corrected (for example, replaced with mathematical expectation estimation).

In the result of the process thereof, the data is reduced to the normalized form, which is applied the statistical methods of analysis and big data processing.

Grabbs method for detecting conflicting data consists in the following.

At first there is assessed the sampling's arithmetic mean $\hat{y}$ and mean square deviation $\hat{\sigma}$:

$$\hat{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad (1)$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n-1}} \qquad (2)$$

To specify abnormality y_i there is computed the parameter

$$\lambda_i = \frac{|y_i - \hat{y}|}{\sigma_y} \qquad (3)$$

and compared with permissible λ_{per} [29].

If $\lambda_i > \lambda_{per}$, then y_i is considered to be contradictory.

Figure shows 3 the block-scheme of discrepancies detection and elimination, using Grabbs criterion.

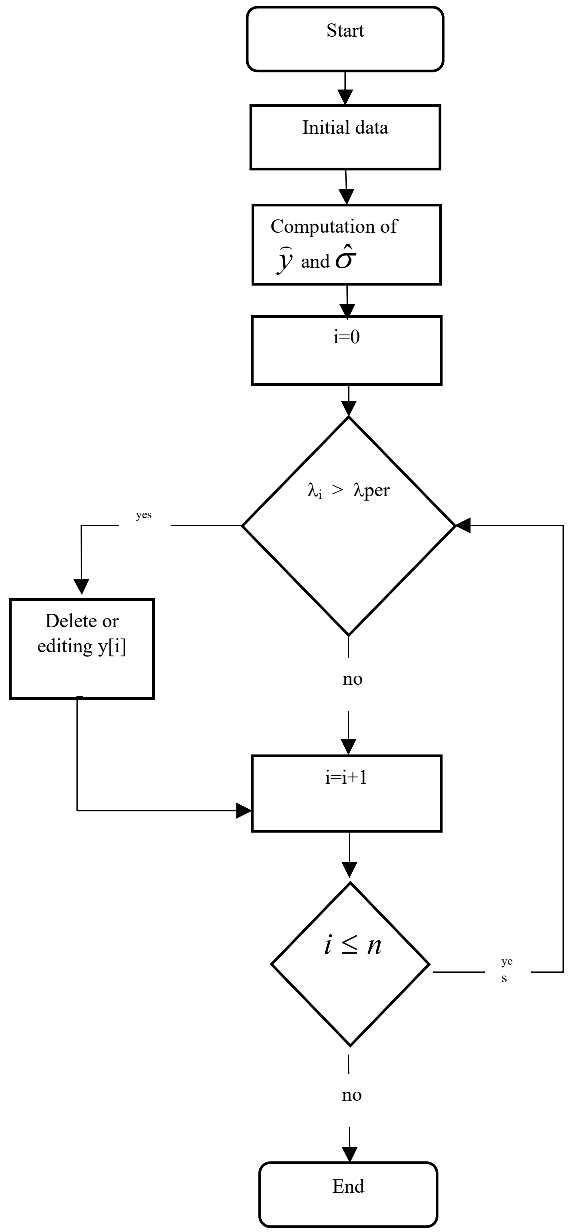

Figure 3:Block-scheme of discrepancies detection and elimination, using Grabbs method

Journal of Theoretical and Applied Information Technology
15th May 2020. Vol.98. No 09

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Grubb's test is applied to assessing cross errors (parasitic errors) of sampling doubtful values from a random magnitude, having normal distribution. The most known and often applied criterion variety is the case, when the parameters of normal distribution – mathematical expectation and general dispersion – are unknown and assessed according to sample mean and sample variance, and for parasitic error there is assessed only one sample value – maximum or minimal. Grubb's test specifies one outlier per one iteration. That outlier is excluded from the data set and the test is repeated until there are detected all outliers.

Table 3 shows the results of a comparison of the criteria for the detection of gross errors.

Table 3: Gross Error Detection Results

Name of criterion	Error detection in small samples	Error detection in large samples
1.Irwin Method	+	-
2. Student criterion	+	-
3. The criterion of the largest absolute deviation	-	+
4. Maximum relative deviation criterion	-	-
5. Romanovsky criterion	+	-
6. Variational range method	-	+
7. 3 Sigma Criterion	-	+
8. Wright criterion	-	+
9. Grubbs criterion.	-	+
10. Q-test (Dixon)	+	-
11. Lvovsky criterion	+	-
12. Chauvinet criterion	+	-
13. David criterion.	-	+
14. Hoglin-Iglevich criterion	+	+
15. L-test (Titien-Moore test)	+	-
16. Smolyak-Titarenko criterion	-	+
17. Brodsky-Batsan-Vlasenko criterion	-	+
18. Kimber criterion	-	+

4.2. Apriori information adjustment, using statistical methods

Preliminary disambiguation leads to mathematical expectation change and mean square deviation. Therefore, there happens deviation of those values from a priori ones. Accordingly, there occirs the task of mathematical expectation estimation and mean square deviation, as well, of empiric distribution function determination. Solution of the tasks thereof consists of the following basic stages:

1) Computation of expectation estimation and mean square deviation;
2) Determination of the interval length;
3) Detecting of distribution empiric functions;
4) Computation of theoretical hitting frequency into the interval;
5) Checking the correspondence hypothesis to distribution empiric and theoretical functions, employing the selected criterion (Pearson χ2).
6) Calculating the class marks and hitting frequency into the interval;

Computation of expectation estimation and mean square deviation is executed according to a formula:

$$\hat{m} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{4}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{m})^2}{n-1}} \tag{5}$$

Let's find the interval value:

$$\Delta x = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 * \lg n} \tag{6}$$

Let's hypothesize the random magnitude distribution function with probability distribution density p(x) with mathematical expectation $\hat{m}$ and MSD $\hat{\sigma}$.

Formula of empiric and theoretical distribution functions [30].

$$F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} I(x_i < y) \tag{7}$$

$$I(x_i < y) = \begin{cases} 1, & if \quad x_i < y, \\ 0 & otherwise \end{cases} \tag{8}$$

$$p(x) = \frac{1}{\sqrt{2\pi} * \hat{\sigma}} * e^{-\frac{(x-\hat{m})^2}{2\hat{\sigma}^2}} \tag{9}$$

Journal of Theoretical and Applied Information Technology
15th May 2020. Vol.98. No 09

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In the work herein, to compare empiric and theoretical distribution functions, we use a criterion χ^2:

$$\chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i} \tag{10}$$

where $k = \dfrac{x_{\max} - x_{\min}}{\Delta x}$ - intervals number; m_i – experimental hitting frequency of variate value into i – interval; np_i – theoretical frequencies.

5. PRACTICAL PART

Every five seconds the employees of Central dispatcher service of Almaty city Transport Holding receive the information about buses location and driving speed. Moreover, the public transport is watched by means of video cameras, installed at every street. Dispatchers trace the situations at roads 24-four hours. All data on traffic congestion, intervals of bus running, dead time at bus stops are analyzed in real mode. That data allows adjust routes and bus number in compliance with citizens' needs.

We have received the data archive for the big data analysis and processing from Almaty city Transport Holding. The latter manages about 114 city buses. All buses and trolleybuses are equipped with GPS-trackers for transport tracing and information receiving in real time mode (Figure 4).

The installed GPS-tracker is directly connected with the terminal ONAI. The data enters the transport holding dispatcher section via satellite GLONASS every second and it is renewed with minimal delay.

Figure 4: GPS-tracker frotcom gv65

If a bus is delayed in the jam, or it has fallen out from the route due to any other reason, the changes will be displayed on the map within several seconds. Application CityBus (Figure 5) in Almaty city is connected to the Holding's base system and it shows the location and routes of the buses at the moment. Complete daily information about buses and routes is recorded and transmitted to the data archive. Figure 5 presents a map of watching over public transport on-line, obtained by means of application CityBus of Almaty city.

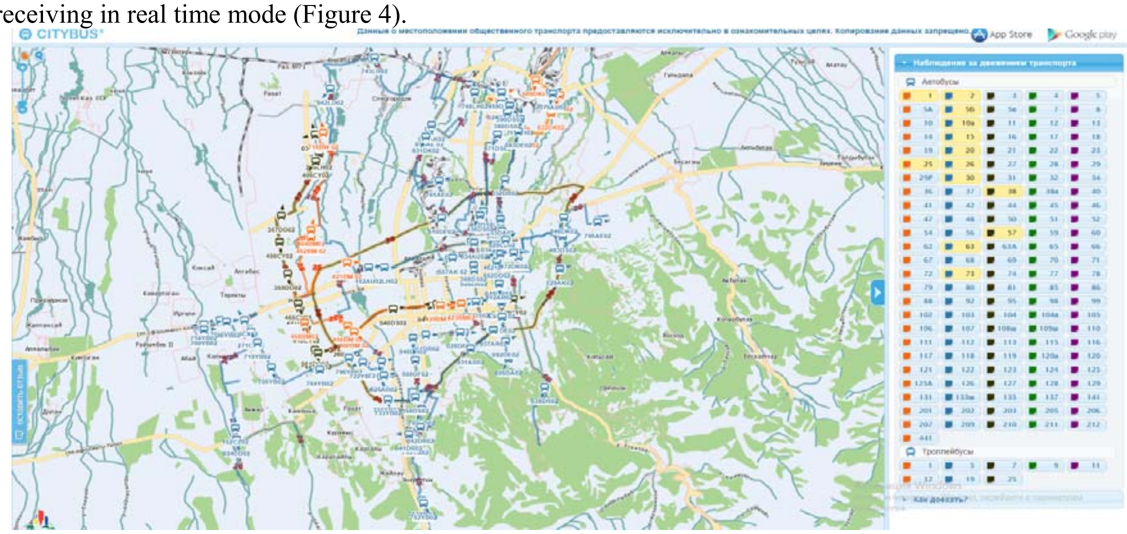

Figure 5: Almaty city public transport routes map

Figure 6 shows the architecture of data collection, transmission, analysis and visualization process. Arrows direction denotes the data and information flows.

The data thereof have been collected, united into CSV files, using protocol OPC (one CSV file per day), then it is transmitted to the file server. The

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

data, being stored in the database, is obtained by the software for high level analysis

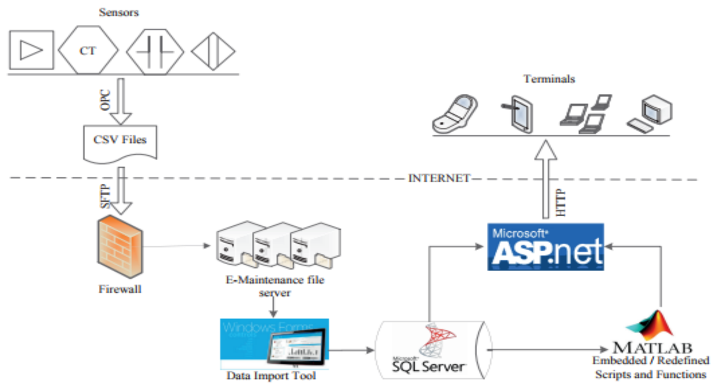

Figure 6: Process of data collection, transmission, analysis and visualization

Obtained from the sensors data is stored in the form of a file csv. The given database encloses complete information on the buses state number plates, bus stops names on a certain route, actual time of arrival to a stop and idle time at a stop.

Total volume of data achieves, obtained from Almaty city Transport Holding is 3 TB.

Figure 7 presents an example of the report on actual and scheduled route traffic in time, denoting stops and idle time (A-starting stop, Б-end stop).

Gov.number	A-B	Microdistrict No. 6				Altynsarin / Abay			
	B-A	Railway station "Almaty-1"				Seyfullina / Sholokhov			
		plan	fact	deviation	plain	plan	fact	deviation	plain
582DS02	A-B		06:15 (S)		0:07:01		06:21 (S)		0:02:49
	B-A		07:20 (S)		0:12:43		07:23 (S)		0:01:17
	A-B		08:30 (S)		0:06:39		08:34 (S)		0:00:21
	B-A		09:48 (S)		0:01:10		09:50 (S)		0:00:28
	A-B		10:46 (S)		0:10:03		10:48 (S)		0:00:35
	B-A		12:31 (S)		0:44:58		12:36 (S)		0:02:34
	A-B		13:42 (S)		0:08:12		13:45 (S)		0:00:42
	B-A		15:12 (S)		0:31:35		15:15 (S)		0:01:24
	A-B		16:22 (S)		0:06:53		16:28 (S)		0:01:59
	B-A		17:30 (S)		0:02:34		17:32 (S)		0:01:17
	A-B		18:54 (S)		0:02:06		18:56 (S)		0:00:28
	B-A		19:49 (S)		0:05:14		19:51 (S)		0:00:42

Figure 7: Actual and scheduled route traffic in time, denoting stops and idle time

As an example, let's take the travel time from the starting stop to an end stop of the route 72 (Table 4). Table 4 presents an actual daily trip time of the bus route 72.

TABLE 4: Data on the route №72 per one day (N- route number, t_{att} - actual trip time).

N	t_{att}	№	t_{att}	№	t_{att}	№	t_{att}	№	t_{att}	№	t_{att}	№	t_{att}
1	56	**21**	70	**41**	49	**61**	69	**81**	58	**101**	62	**121**	59
2	59	**22**	58	**42**	30	**62**	77	**82**	62	**102**	58	**122**	37
3	58	**23**	62	**43**	60	**63**	49	**83**	63	**103**	60	**123**	55
4	69	**24**	61	**44**	63	**64**	52	**84**	63	**104**	68	**124**	62
5	59	**25**	70	**45**	53	**65**	49	**85**	68	**105**	61	**125**	**45**
6	55	**26**	52	**46**	58	**66**	59	**86**	75	**106**	61	**126**	63
7	61	**27**	62	**47**	64	**67**	66	**87**	49	**107**	62	**127**	65
8	59	**28**	77	**48**	69	**68**	60	**88**	60	**108**	76	**128**	78
9	61	**29**	47	**49**	64	**69**	63	**89**	62	**109**	67	**129**	55
10	63	**30**	60	**50**	57	**70**	50	**90**	68	**110**	50	**130**	55

Journal of Theoretical and Applied Information Technology
15th May 2020. Vol.98. No 09

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

11	69	**31**	62	**51**	53	**71**	55	**91**	62	**111**	59		
12	52	**32**	58	**52**	57	**72**	61	**92**	64	**112**	72		
13	49	**33**	62	**53**	51	**73**	63	**93**	62	**113**	49		
14	55	**34**	65	**54**	59	**74**	67	**94**	61	**114**	92		
15	61	**35**	83	**55**	76	**75**	62	**95**	64	**115**	59		
16	62	**36**	49	**56**	47	**76**	64	**96**	65	**116**	63		
17	61	**37**	58	**57**	60	**77**	58	**97**	74	**117**	57		
18	69	**38**	58	**58**	58	**78**	59	**98**	48	**118**	69		
19	60	**39**	62	**59**	79	**79**	72	**99**	62	**119**	63		
20	50	**40**	59	**60**	64	**80**	60	**100**	66	**120**	58		

The software part has been implemented in Python 3.8. Used scripts: sympy, numpy, matplotlib.pyplot as plt, math, tkinter, faker, time. The interface "Identification and elimination of contradictions methods" has been created, as shown in Figure 8. Three methods are considered herein: k means, Grabbs criterion, and Statistical Processing for identifying and eliminating conflicting data.

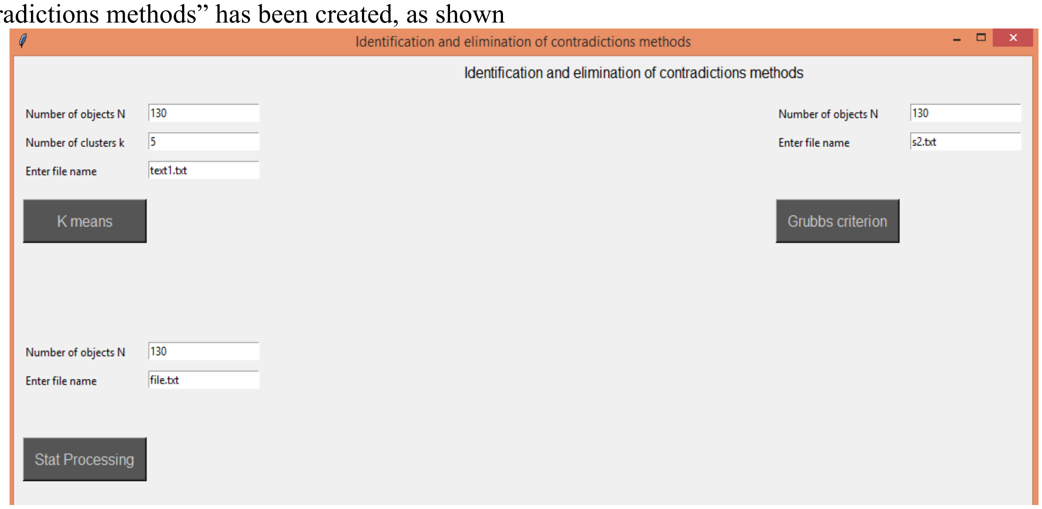

Figure 8: Interface «Identification and elimination of contradictions methods»

To apply the k means method [31], you need to enter the number of objects and clusters, as well as select a file and click the "k means" button.

The result of dividing the data into 5 classes is presented in Figure 9. In the figure an axle x denotes bus trip actual time from starting to an end stop, and an axle y - number of bus routes.

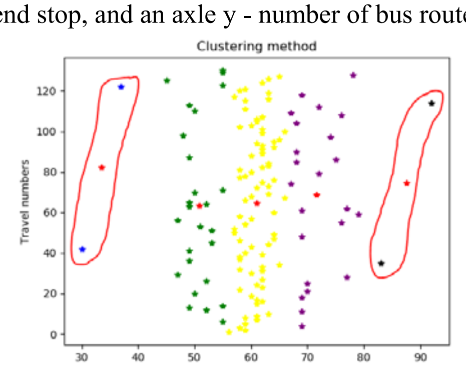

Figure 9: The result of splitting data into 5 classes

When the "Grabbs criterion" button is clicked, we get the mean square deviation, average value and outliers of the first and second iteration according to the Grabbs criterion in the results window (Figure 10).

Figure 10: Grubbs test results

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

At applying Grabbs method there has been adjusted the data 42 (with value 30), 114 (with value 92) and 122 (with value 37), 35 (with value 83).

The results of the first and second iteration using the Grubbs method are presented in Figures 11 and 12.

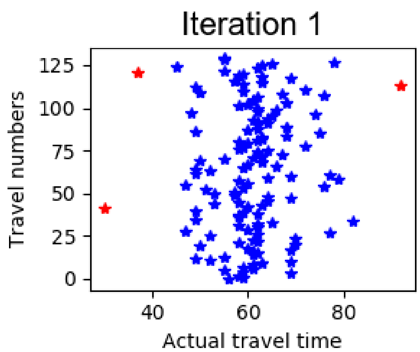

Figure 11: The result of the first iteration

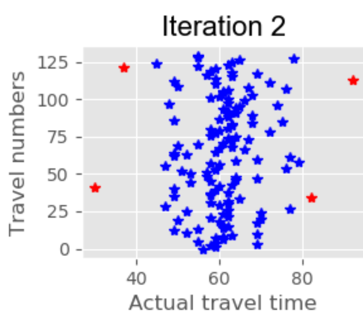

Figure 12: Result of the second iteration

The numbers of the conflicting data from the last iteration are consistent with the conflicting data, obtained by the k means method (Figure 9).

Subsequent to data adjustment according to the above described methodology, we obtain theoretical and experimental distributions of buses trip time (Fig.13).

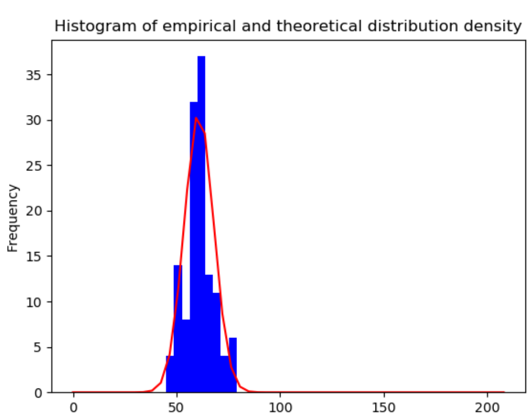

Figure 13: Column diagram of empiric and theoretical density distribution

In the Figure an axle x denotes the bus trip time, and an axle y - scale of events hitting frequency.

Computed value $\chi^2 = 6.63$, which is less, than 9.21 for significance level 0.1. It means, that with probability 0.9 it might be confirmed, that theoretical and experimental distribution functions coincide.

Thereat, estimation of expectation has constituted - 60, and mean square deviation - 7.28.

Trip passing time according to a priori data (per plan) amounts to 50 minutes, and computation results - 60. Therefore, it is necessary to correct the schedule for 10 minutes.

CONCLUSION

We have carried out analytical review of existing works on the thematic under study and demonstrated work's actuality.

In the work we have offered the methodology of discrepancies detection and elimination in large-scale transport data, as well, the statistical method for adjusting a priori information. For that aim there is first fulfilled preliminary detecting and eliminating of discrepancies, and further, cleaned data is used for adjusting the scheduled data.

Preliminary detection and elimination of discrepancies in big transport data has been executed employing Grabbs method and clustering (k-average method). Outcomes, obtained with those methods for the being considered example, coincide.

Received data subsequent to pre-processing has been used for getting the trip time statistical characteristics, which have been applied to their adjustment. At that, it has been shown, that the trip time distribution corresponds to normal law with a probability of 0.9. In the result of statistical processing the trip time of the selected route has come to 60 minutes, and scheduled time composes 50 minutes. It means, that there is required the schedule adjustment for 10 minutes.

Proposed methodology has been implemented in Python software medium.

The work herein has been executed in the frame of the project AP05133699 «Research and development of innovative, information and telecommunication technologies, employing modern cyber-engineering means for the city's smart transport system».

REFRENCES:

[1] Deakin, M., & Al Waer, H. (2011). From intelligent to smart cities. Intelligent Buildings International, 3(3), 140-152.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] Grava, S. (2003). Urban transportation systems. Choices for communities.
[3] Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M.: (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation Research Part C: Emerging Technologies 68, 285–299.
[4] Sussman, J. S. (2008). Perspectives on intelligent transportation systems (ITS). Springer Science & Business Media.
[5] Figueiredo, L., Jesus, I., Machado, J. T., Ferreira, J., & de Carvalho, J. M. (2001). Towards the development of intelligent transportation systems. In Intelligent transportation systems (Vol. 88, pp. 1206-1211).
[6] Foster I. (1995). Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
[7] White T. (2009). Hadoop: The Definitive Guide (1st ed.). O'Reilly Media, Inc.
[8] Akhmediyarova A.T., Kassymova D.T., Utegenova A.O, Utepbergenov I.T. Development and research of the algorithm for determining the maximum flow at distribution in the network // Open Computer Science. – 2016. – Vol.6, №1. – P. 213-218.
[9] Waldemar Wójcik, Akhmediyarova A.T., Mamyrbayev O., Kassymova D.T., Utepbergenov I.T. Problem of placement of the minimal number of cameras at a given transport network // Przegląd Elektrotechniczny. – 2017. – Vol.93, Issue 6. – P.137-140
[10] Oh, S., Byon, Y. J., & Yeo, H. (2016). Improvement of Search Strategy With K-Nearest Neighbors Approach for Traffic State Prediction. IEEE Transactions on Intelligent Transportation Systems, 17(4), 1146-1156.
[11] Shi, Q., & Abdel-Aty, M. (2015). Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. Transportation Research Part C: Emerging Technologies, 58, 380-394.
[12] Ahn, J., Ko, E., & Kim, E. Y. (2016). Highway traffic flow prediction using support vector regression and Bayesian classifier. In 2016 International Conference on Big Data and Smart Computing (BigComp) (pp. 239-244). IEEE.
[13] Chen, X. Y., Pao, H. K., & Lee, Y. J. (2014). Efficient traffic speed forecasting based on massive heterogenous historical data. In Big Data (Big Data), 2014 IEEE International Conference on (pp. 10-17). IEEE.
[14] Xia, D., Wang, B., Li, H., Li, Y., & Zhang, Z. (2016). A distributed spatial–temporal weighted model on MapReduce for short-term traffic flow forecasting. Neurocomputing, 179, 246-263.
[15] Pena, D., Tchernykh, A., Nesmachnow, S., Massobrio, S., Drozdov, A.Y., and Garichev, S.N., Multiobjective vehicle type and size scheduling problem in urban public transport using MOCell, IEEE International Conference Engineering and Telecommunications, Moscow, Russia, 2016, pp. 110–113
[16] S. Biswas and R. V. Babu. Short local trajectory based moving anomaly detection. In ICVGIP, 2014.
[17] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C.Williamson. Estimating the support of a high-dimensional distribution. Neural computation, 13(7):1443–1471, 2001.
[18] K. Cheng, Y. Chen, and W. Fang. Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. IEEE Transactions on Image Processing, 24(12):5288– 5301, Dec 2015.
[19] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette. Real-time anomaly detection and localization in crowded scenes. In CVPRW, 2015.
[20] X. Hu, S. Hu, Y. Huang, H. Zhang, and H. Wu. Video anomaly detection using deep incremental slow feature analysis network. IET Computer Vision, 10(4):258–265, 2016.
[21] H. Jeong, Y. Yoo, K. M. Yi, and J. Y. Choi. Two-stage online inference model for traffic pattern analysis and anomaly detection. Machine vision and applications, 25(6):1501–1517, 2014.
[22] R. Kaviani, P. Ahmadi, and I. Gholampour. Automatic accident detection using topic models. In ICEE, 2015
[23] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis. Multiple hierarchical dirichlet processes for anomaly detection in traffic. Computer Vision and Image Understanding, 169:28–39, 2018.
[24] Y. Li, W. Liu, and Q. Huang. Traffic anomaly detection based on image descriptor in videos. Multimedia Tools and Applications, 75(5):2487-2505, Mar 2016.
[25] L.-L. Wang, H. Y. T. Ngan, and N. H. C. Yung. Automatic incident classification for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

large-scale traffic data by adaptive boosting svm. Information Sciences, 467:59–73, 2018.
[26] W. Yang, Y. Gao, and L. Cao. Trasmil: A local anomaly detection framework based on trajectory segmentation and multi-instance learning. Computer Vision and Image Understanding, 117(10):1273–1286, 2013.
[27] J. R. Medel and A. Savakis. Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv preprint arXiv:1612.00390, 2016.
[28] PetrenkovV.I., Kopytov V.V., Sidorchuk A.V., Antonov V.O. Working out on technology of anomalous values detection in in big data streams. Proceedings of the Sixth All-Russian scientific conference "Information technologies of smart decision taking support", May 28-31, Ufa--Stavropol, Russia, 2018, p.p. 23-29.
[29] Lemeshko B.Yu. Data statistical analysis, modeling and studying probabilistic regularities. Computer approach: Monography / Lemeshko B.Yu., Lemeshko S.B. Postovalov N.S., Chimitova Ye.V.- Novosibirsk: Publisher NSTU, 2011. – 888 p.
[30] https://nsu.ru/mmf/tvims/chernova/ms/lec/node4.html
[31] https://blog.floydhub.com/introduction-to-anomaly-detection-in-python/?utm_source=aidigest&utm_medium=email&utm_campaign=featured