REVIEW REPORT

by Prof. Dr. Boyan Paskalev Bontchev

on a PhD dissertation on a topic

"Aided decision making for public transportation optimizations using Big Data"

authored byGeorgi Kalinov Yosifov –

PhD student at the Department of Software Technologies

at the Faculty of Mathematics and Informatics of Sofia University "St. Kliment Ohridski",

for the acquisition of an educational and scientific degree "Doctor",

professional direction 4.6 *"Informatics and computer sciences"*

(doctoral program "Software Technologies" - Software Engineering)

This review was prepared in connection with the procedure for the defense of Georgi Kalinov Yosifov's dissertation on the topic "*Aided decision making for public transportation optimizations using Big Data*" for the acquisition of the educational and scientific degree "Doctor" in the field of higher education : "*4. Natural sciences, mathematics and informatics*", professional direction: "*4.6. Informatics and computer science*", doctoral program "Software Technologies" - Software Engineering), according the Order of the Rector of SU "St. Kliment Ohridski" with number RD 38-599/28.10.2022 and Protocol No. 1 of the first meeting of the scientific jury hold on 8.11.2022.

As a member of the scientific jury I have received the necessary documents that meet the regulatory requirements of both ZRASRB and the Rules for implementation of ZRASRB of the Council of Ministers of the Republic of Bulgaria and that of Sofia University "St. Kliment Ohridski", which gives me reason to review this dissertation work.

1. Relevance of the problem

The dissertation work developed by Georgi Kalinov Yosifov is in a dynamically developing field - that of optimizing public transport in large cities, carried out by means of determining, researching and forecasting the levels of loading of motor vehicle movement of periodic urban transport. This optimization uses methods processing large datasets that are collected from positional coordinates of periodic public transport used as a traffic sample. This makes the work particularly relevant and gives significance to the achieved results regarding their practical use.

The main гоал of the thesis (p. 10) is "*to support public transport optimization decision-making by determining, investigating and predicting traffic congestion levels using data collected from positional coordinates of periodic public transport used as a sample in traffic*". In order to achieve this goal, the doctoral student sets himself adequate tasks arising from the goal, providing for the preparation of:

1. Methodology for classification and analysis of the current state of data collection methods and determining traffic levels.

2. Overview of different methods for analyzing and predicting traffic levels at a future point in time.

3. Algorithm based on the compiled methodology, aimed at determining traffic levels in an urban environment.

4. Experimental scenarios for studying the qualities and limitations of the created algorithm

5. Software tools supporting algorithm data processing and management in experimental scenarios

6. Instruments to predict traffic levels in an urban environment at a future time.

7. Comparative analysis of the performance of the various forecasting methods in order to determine the most appropriate one for the purpose.

2. Knowing the state of the problem

It is clear from the dissertation work and the presented publications that the doctoral student is very well acquainted with the issues under consideration. The work contains citations of a sufficient number of literary sources - a total of 109 in number, of which 6 are in Bulgarian and the rest - in English. Most of them are contemporary (since 2015), with only 25 sources cited before that year, which testifies to the topicality of the researched problem.

Regarding the applied problems solved in the work, the PhD student used his practical experience in the field of designing software systems, as well as in the use of different types of neural networks and software tools for machine self-learning.

3. Research methodology

The methodology of research and practical work is appropriately chosen in accordance with the topic of the work. After conducting a thorough comparative analysis of the current state of research on data collection methods and determination of public transport traffic levels, city traffic load indices are determined using an algorithm using positional data from intermittent public transport in Edinburgh and Sofia. Based on the presented algorithm, software solutions have been developed for collecting, processing, calculating and visualizing the level of traffic load. Ways to predict future traffic load levels have been analyzed using two types of machine learning models – single-step models (predicting a single traffic load index value in the future) and multi-step models (predicting several steps ahead in time series) using convolutional and recurrent neural networks.

4. Characteristics and evaluation of the content of the dissertation

Georgi Yosifov's dissertation is an original scientific study in the field of optimizing public transport in large cities. Its content and structuring are determined by the tasks set and following the chosen methodology. The work has a volume of 180 typewritten pages and includes 76 figures and 17 tables. The bibliography includes 87 titles, of which 18 are sources available only on the Internet.

From the results presented in the work, it can be concluded that the doctoral student has successfully completed the tasks set before the dissertation work. The introduction presents the actuality of the problem, the purpose and tasks of the dissertation, as well as the motivation of the doctoral student for solving them. The second chapter is entitled "Review of scientific sources" and provides a categorization of methods for predicting traffic load levels. The third chapter describes the statistical analysis of data on the transit times of periodic urban transport vehicles, calculated from two heterogeneous sources from Edinburgh, Scotland and Sofia, Bulgaria, through selected road segments. An algorithm based on the statistical analysis is presented, with the help of which the degree of load on a road segment is indirectly determined, and the qualities of the presented algorithm are experimentally determined and verified. The fourth chapter describes software solutions for collecting, processing, calculating and visualizing the level of traffic load, by means of the algorithm presented in chapter 3, offering the possibility of modular integration to support different types of input data. The fifth chapter is devoted to a comparative analysis of the results of single-step and multi-step machine learning models with convolutional and recurrent neural networks to determine traffic load levels at a future time. The conclusion presents a summary of the implementation of the initial objectives, as well as lists of approvals of the results. Five appendices are presented at the end of the work – A: Index of figures, B: Index of tables, C: Glossary of used terms and abbreviations, D: Processed segments of the city of Edinburgh and their graphs, and D: Analysis of the distribution of data in their pure and logarithmic form

5. Contributions of the dissertation

The work presents substantial scientific-applied and applied contributions that are the original work of the author.

The scientific and applied contributions of the dissertation can be summarized as follows:

- A comparative analysis and categorization of the methods for determining the traffic load in an urban environment are presented.
- A statistical analysis of data on the transit times of vehicles from periodic urban transport, calculated from two heterogeneous sources through selected road segments in Edinburgh, Scotland and in Sofia, Bulgaria, was performed.
- On the basis of the statistical analysis, an algorithm was created to determine the level of traffic on a road segment based on an indirect sign, and the qualities of the algorithm were verified experimentally.

- A comparative analysis of the results of single-step and multi-step machine learning models was performed to determine future traffic load levels.

The developed software solutions for collecting, processing, calculating and visualizing the level of traffic load, using the presented algorithm, with the possibility of modular integration and support of different types of input data, can be defined as an applied contribution.

6. Degree of independence of the contributions of the dissertation

Considering the contributions of the work and the three publications presented, I consider that *the main results and contributions of the dissertation have a substantial degree of independence.*

7. Evaluation of publications

Three publications of the doctoral student are cited in connection with the dissertation, all co-authored with the scientific supervisor, where the doctoral student is the first co-author. One of the publications was made in the international journal Lecture Notes in Networks and Systems, and the other two - in proceedings of prestigious international conferences (International Congress on Information and Communication Technology - published in Lecture Notes in Networks and Systems, and International Conference on Computer Systems and Technologies - CompSysTech'2020). All three publications are in publications with an impact rank (SJR), which is a certificate of significance of the achieved results.

To date, the publication in the CompSysTech'2020 proceedings has been cited twice by other authors.

8. Usability of the results in practice

The successful approbation of the achieved results through sectional reports at prestigious international conferences proves the applicability and usefulness of the results achieved by the doctoral student. The achieved results can find application in the creation of an integrated traffic control system in an urban environment, as well as serve as a basis for building software applications for navigation and routing of urban traffic.

9. Opportunities for future use of the dissertation contributions

The methods proposed in the dissertation for optimizing public transport in large cities by means of determining, researching and forecasting the load levels of periodic urban transport provide opportunities for conducting additional experimental studies. Possibilities for future use of the thesis contributions are adequately summarized in the directions for future development in the concluding chapter.

10. Dissertation abstract

The abstract is in a volume of 47 typewritten pages (in its version in Bulgarian) and correctly and adequately reflects the contributions and results achieved in the dissertation.

11. Critical remarks

All critical comments of the reviewer to the content and layout of the work submitted to the internal defense of the dissertation work have been taken into account and are adequately reflected in the final version of the dissertation.

12. Conclusion

My overall assessment of Georgi Yosifov's dissertation work, abstract and scientific publications is **positive**. Taking into account the original scientific-applied and applied contributions of the dissertation work and the scientific publications made in this connection, as well as the numerous and significant practical results achieved, I find that the doctoral student has in-depth theoretical knowledge in the relevant field and abilities for independent scientific research. In this sense, the dissertation meets the requirements of the ZRASRB and the regulations for the application of the ZRASRB cited at the beginning of this review, which gives me the reason to propose to the esteemed jury to award Georgi Kalinov Yosifov the educational and scientific degree "Doctor" in the field of higher education : "*4. Natural sciences, mathematics and informatics*", professional direction: "*4.6. Informatics and computer science*", doctoral program "Software Technologies" - Software Engineering.

18/01/2023   Signature:

Valencia, Spain   /Prof. Dr. B. Bontchev/