

РЕЦЕНЗИЯ

**по процедура за защита на дисертационен труд на тема:
„ПОДПОМАГАНЕ ВЗИМАНЕТО НА РЕШЕНИЯ ЗА ОПТИМИЗИРАНЕ НА
ОБЩЕСТВЕН ТРАНСПОРТ С ПОМОЩТА НА ГОЛЕМИ ДАННИ“**

**за придобиване на
образователна и научна степен „доктор“**

от

кандидат: Георги Калинов Йосифов,

Област на висше образование: 4. Природни науки, математика и информатика /

Професионално направление: 4.6. Информатика и компютърни науки

Докторска програма: „Софтуерни технологии“ - Софтуерно инженерство,

Факултет по математика и информатика (ФМИ),

Софийски университет „Св. Климент Охридски“ (СУ).

**РЕЦЕНЗЕНТ: д-р Борис Благовестов Шишков, доцент по Контекстно-базираны информ-
формационны системы в ИМИ-БАН, доцент по Бизнес информационны системы в
УниБИТ.**

Избран съм за рецензент по обявената процедура на 08.11.2022 г., на заседание на Научното жури, свикано съгласно Заповед № РД 38-599/28.10.2022 г. на Ректора на Софийския университет. Като рецензент съм получил документите по конкурса (на електронен носител): дисертационен труд, автореферат (на български и английски език), списък с публикации на кандидата, копия от дипломи и др.

1. Обща характеристика на дисертационния труд и представените материали

Дисертационният труд „Подпомагане взимането на решения за оптимизиране на обществен транспорт с помощта на големи данни” е в обем от 180 страници, с включени 6 глави: Увод (6-15), Сравнителен анализ на текущото състояние на изследванията в областта на методи за събиране на данни и определяне на нивата на трафика на общественя транспорт (16-30), Индекс на натовареност на градския трафик, базиран на позиционни данни от периодичен обществен

транспорт (31-91), Описание на създадените софтуерни инструменти за провеждане на експериментите (92-110), Предсказване на индекса на натовареност на трафика (111-141), Заключение и бъдещо развитие (142-146), като са представени 76 броя фигури, 17 броя таблици и пет приложения. Библиографията има над 100 източника на български и английски език. Приносът на дисертационния труд е в няколко направления; предложени са: (i) методология за класификация и анализ на текущото състояние на методите за събиране на данни и определяне нивата на натовареност на движението (разглеждайки в частност гр. София), както и обзор на съответни методи за анализ и прогнози; (ii) алгоритъм за определяне нивата на трафик, валидиран посредством експериментални сценарии; (iii) инструменти, подпомагащи обработката и управлението на данните от алгоритъма в експерименталните сценарии.

2. Данни и лични впечатления за кандидата

От автобиографията, представена във файла "23.Автобиография.pdf", която е само една страница, не можах да добия достатъчно релевантна информация за кандидата, а лични впечатления от него нямам, тъй като не го познавам. От други документи, подадени от него (свързани с настоящата процедура) е видно, че кандидатът е: (i) получил висшето си образование в България: Бакалавър по Компютърни науки (ФМИ - СУ) и Магистър по Информатика (ФМИ - СУ), с много добър успех, като дипломната работа е защитена с отличен; (ii) задочен докторант към катедра „Софтуерни Технологии" (СУ); (iii) представил 2 Springer - LNNS публикации с Q4 и индексирани в СКОПУС, както и една ACM публикация, индексирана в СКОПУС; (iv) със СКОПУС h-index 1, отразяващ трите споменати публикации.

3. Съдържателен анализ на научните и научно-приложните постижения на кандидата, съдържащи се в представения дисертационен труд и публикациите към него, включени по процедурата

С развитието и достъпността на (безжичните) телекомуникации, а и на сензорните технологии, и с миниатюризацията на многофункционални хардуерни устройства, стана възможно (още преди две десетилетия) използването на мощни компютърно-сензорни мрежи за предоставянето на „нов“ тип услуги на крайния потребител, а именно – такива услуги, които не се ограничават до приемането на потребителските нужди като статични. Напротив, доставчиците на IT услуги възприемат потребителят в неговия/нейния контекст, адаптирайки предлаганите услуги към ситуацията му/й, отразяваща местоположение, време и т.н. Преносимите устройства поставят потребителя и в ролята на източник на актуална информация. В допълнение, сензо-

рите правят възможно обективното установяване състоянието на потребителя. И накрая, компютърните мрежи и (безжичните) телекомуникации способстват за ефективният поток от данни, касаещи доставчика на услугата, потребителя и релевантни трети страни. Развитието на DATA технологиите през последното десетилетие, базирано на интер-дисциплинарен и мулти-дисциплинарен подход по отношение основно на три класически научни дисциплини, а именно: математика, статистика и информатика, но адресиращо морето от данни, които ни „запливат“ в днешно време (от камери, интернет „логове“ и др.), открива нови хоризонти: (i) Ние говорим вече не само за бази данни (и OLTP), но и за складове от данни (и OLAP), разглеждайки „големи“, ненормализирани данни, с цел анализи и открояване на трендове; (ii) Класически статистически инструменти като например регресионен анализ и Бейс моделиране, се поставят в перспективата на компютърните науки с цел обогатяването на IT алгоритмизации с вероятностни аспекти; (iii) Машинното обучение (и изкуственият интелект) и невронните мрежи „преживяват“ нов бум, касаещ обучаването и тренирането на IT системи, базирано на т.н. training data. Всичко това се съчетава изключително добре с контекстната ориентираност (дискутирана по-горе), тъй като данните (касаещи състоянието на потребителя), придобити от сензори (и/или от други източници), могат да бъдат анализирани и използвани с цел прогнозиране на бъдещи състояния. Това намира широко приложение в теле-медицината, умните градове/сгради, дрон технологиите и др. Аспекти от гореизложеното касаят функционирането на един голям съвременен град какъвто е София – една от столиците в ЕС и един от големите градове в общността. Ключов проблем за един такъв град е публичният транспорт, а София се характеризира с мулти-модалност в това отношение, доколкото жителите на града могат да използват и автобуси, и трамваи, и тролейбуси, и метро. За разлика от малки градове, като например нидерландския град Делфт (където транспортната оптимизация следва традиционно установени транспортни „линии“ и това „работи“, доколкото не се допуска пренаселване и свързаността със съседните населени места е ключова), София изпитва предизвикателствата на бързо увеличение на населението, динамика в развитието на кварталите, липса (извън централната част на града) на традиционно установени комуникационни „линии“ и т.н. Следователно, оптимизирането на обществения транспорт в София (и други (европейски) големи градове) е актуално предизвикателство, касаещо качествено медицинско обслужване, придвижването на хората между дома и работното място, достигането до културни и спортни обекти, развлечения и т.н.

Видно от дисертационния труд на Георги Калинов Йосифов, както и от гореспоменатите 3 научни публикации, той се фокусира върху този актуален проблем като визията му за подпомагане решаването на проблема касае дискутираното по-горе, особено в частта му, насочена

към анализ на данни. Това прави неговата работа актуална и съотносима по отношение на проблеми, свързани със съвременното обществено развитие и в частност, касаещи логистика, вериги за доставки, пациентски потоци и т.н.

Разглеждайки гореспоменатия дисертационен труд, аз бих представил (софийската) система за градски транспорт (от гледна точка на желаното оптимизиране на потоците) като една **КОНТЕКСТНО-БАЗИРАНА** информационна система, която следва да **АДАПТИРА** своето „поведение“ в зависимост от състояния, касаещи **ПЪТНИЦИТЕ, ТРАНСПОРТНИТЕ СРЕДСТВА и ТРАНСПОРТНАТА ИНФРАСТРУКТУРА**. Информация за тези състояния се получава от преносими устройства (на пътници), **GPS** позициониране (на превозни средства), **ССТV** и др. От тази гледна точка, доколкото се разглеждат тези неща, както и самия анализ на данни, намирам за **пропуск** липсата на (експлицитно) рефериране към **DATA FUSION** алгоритмите, които по мое мнение са от фундаментално значение при работа с „големи“ данни, получавани от разнородни източници.

Предсказването на ситуации, свързани с трафика (отнасящи се до публичния транспорт) е актуално предизвикателство, открито в дисертацията на Георги Йосифов, както и възможността да се използват **исторически данни** (например по отношение на задръствания), чиито анализ да помогне за идентифицирането на проблемни „точки“ и съответно предлагане на оптимизации.

Дисертационният труд е основно фокусиран върху самото **събиране на данни**, „мерки“ за **натовареност, софтуерни инструменти за провеждане на експерименти** и гореспоменатото **прогнозиране**.

Научно-изследователската цел е определена така: **“да се подпомогне взимането на решения за оптимизация на обществения транспорт чрез определяне, изследване и прогнозиране на нивата на натовареност на движението с помощта на данни,,**

С оглед на поставената цел, научният принос, касаещ дисертационния труд на Георги Йосифов е в следните направления:

- Извършен е (макар и само на 15 страници) **сравнителен анализ на state-of-the-art изследвания, касаещи събирането на данни, съотнасяйки това към натовареността на трафика (по отношение на публичния транспорт)**, отразен в Глава 2.
- Предложен е **модел за изчисление на индекс за натовареност на трафика**, както за цял град, така и за отделен пътен сегмент, на база изследването на реални данни от публичен транспорт в гр. София и шотландския град Единбург – виж Глава 3.

- Идентифицирани са изисквания (**requirements**), дефинирани са **модули** и са представени **класове** по отношение на **разработен софтуер** (на C#, в .NET среда), изпълняващ различни аспекти на подпомагане взимането на решения за оптимизиране на обществен транспорт с помощта на „големи“ данни, обработването, създаването, *изчислението и визуализацията на данните, съпътстващи или съставлящи съответни експерименти* – виж Глава 4.
- Предложена е **симулация** на разписания, базирана на параметри, дефинирани от потребителите.

Що се отнася до валидацията (proof-of-principle / proof-of-concept) на това, което е представено в дисертационния труд, аз отчитам разглежданите експерименти, но само с ЧАСТИЧНА роля, доколкото те са по-скоро инструмент за постигане на целите на дисертационния труд, а не толкова ДОКАЗАТЕЛСТВО за валидността на това, което се предлага в дисертацията.

4. Аprobация на резултатите

Както е споменато по-горе, Георги Йосифов е представил 3 публикации; и в трите той е първи съавтор. Тези публикации имат пряко отношение към съдържанието на дисертационния труд:

- „Review of Urban Traffic Detection Approaches with Accent of Transportation in Sofia, Bulgaria” – Глава 2.
- „Traffic Flow City Index Based on Public Transportation Vehicles Data” – Глава 3, Глава 4.
- „Predicting Traffic Indexes on Urban Roads based on Public Transportation Vehicle Data in Experimental Environment“ – Глава 5.

Както е споменато по-горе, представените публикации, в които Георги Йосифов е съавтор са индексирани в СКОПУС, с което се покриват и изискванията на ЗРАСРБ, и тези на СУ „Св. Климент Охридски“. И в трите публикации Георги Калинов Йосифов е съавтор, а не единствен автор – в тези случаи аз приемам еднакъв принос за всеки от съавторите.

Няма доказано по законно установения ред плагиатство в научните трудове.

5. Качества на автореферата

По мое мнение, авторефератът отговаря общо взето на изискванията за изготвянето му и представя коректно резултатите и съдържанието на дисертационния труд, въпреки че например представеното във втория параграф на автореферата (касаещо „чувствителни“ данни и GDPR)

няма отражение под формата на съответна експлицитна дискусия в дисертационния труд. По-нататък, бих искал да отбележа, че считам уводната част на автореферата за недостатъчно убедителна, доколкото Георги Калинов Йосифов не е обвързал достатъчно задълбочено целта на дисертационния труд (виж стр. 3 на автореферата) с идентифицирания проблем (виж стр. 2). Също така, не е направено достатъчно задълбочено изложение за това какъв е оригиналният научен принос на дисертационния труд; на стр. 43 от автореферата са изброени реализираните анализи, изготвения алгоритъм и т.н., но остава недостатъчно ясно какво прави тези неща оригинален научен принос и как те корелират с идентифицирания проблем и формулираните цели. Отчитайки тези пропуски, аз съм като цяло удовлетворен от съдържанието на автореферата.

6. Критични бележки и препоръки

Научно-изследователската работа на Георги Йосифов е актуална и интересна, а дисертационният труд е обвързан с три публикации, индексирани в СКОПУС. Всичко това прави много добро впечатление, но аз имам критични бележки, които се отнасят най-вече до това как е написан дисертационният труд, отчитайки в същото време, че основните му компоненти (глави) са ясно обособени като съдържание и адекватно свързани помежду си:

- Считам, че уводът е неубедителен, подобно на уводната част на автореферата, дискутирана по-горе, тъй като съдържанието не е добре фокусирано и липсват достатъчно добре обосновани връзки между идентифициран проблем, цели и научен принос. Аз получих добра ориентация в това направление, тъй като първо прочетох трите гореспоменати статии и след това започнах да чета дисертационния труд. Но ако би се разчитало само на увода на дисертацията, то читателят не би получил (по мое мнение) достатъчно добра ориентация за научния принос, обвързан със съответни идентифицирани проблеми.
- На различни места в дисертацията има дълги и недостатъчно ясни изречения; считам, че едно внимателно редактиране на текста би направило някои параграфи по-кратки и по-ясни. Също така, не винаги има ясно разграничение между неща, които са оригинални (от автора) и неща, които могат да бъдат намерени в литературата.
- Липсва СИСТЕМАТИЧЕН литературен анализ относно релевантни дисциплини (като Statistics, Data Analytics и Machine Learning), както и достатъчно задълбочен анализ на релевантни изследвания на други колективи (related work) с оглед на това да се покаже какво от съществуващите научни познания и опит вече „решава“ някои от идентифицираните проблеми, мотивирайки по този начин още по-добре целите на дисертацията, показвайки, че се адресират неща, които все още не са решени в пълнота.

- Не са разгледани т.н. социални ценности (public values), като защита на лични данни, отчетност, прозрачност и т.н., въпреки беглото им споменаване в автореферата, но не в дисертацията. Това по мое мнение е пропуск, тъй като всякакви решения, касаещи публичните процеси в един град следва да са ценностно-ориентирани (value-sensitive).
- Авторът не е представил визия за мястото на data fusion алгоритмите в събирането и използването на данни, касаещи публичния транспорт.
- Много пъти се споменава за невронни мрежи, но те не са разглеждани задълбочено в дисертационния труд, а по скоро са маркирани, бе за става ясно как точно предлаганите решения се обвързват с невронни мрежи.

7. Заключение

След като се запознах с представените в процедурата дисертационен труд и придружаващите го научни трудове и въз основа на направения анализ на тяхната значимост и съдържащи се в тях научни и научно-приложни приноси, **потвърждавам**, че представеният дисертационен труд и научните публикации към него, както и качеството и оригиналността на представените в тях резултати и постижения, отговарят на изискванията на ЗРАСРБ, Правилника за приложението му и съответния Правилник на СУ „Св. Климент Охридски“ за придобиване от кандидата на образователната и научна степен „доктор“ в област на висше образование: 4. Природни науки, математика и информатика / професионално направление: 4.6. Информатика и компютърни науки. В частност кандидатът удовлетворява минималните национални изисквания в професионалното направление и не е установено плагиатство в представените по конкурса научни трудове.

Въз основа на гореизложеното, **препоръчвам** на научното жури да присъди на Георги Калинов Йосифов образователна и научна степен „доктор“ в област на висше образование: 4. Природни науки, математика и информатика / професионално направление: 4.6. Информатика и компютърни науки.

20.01.2023 г.

Изготвил рецензията:

(доц. д-р Борис Шишков)