

# РЕЦЕНЗИЯ

относно дисертационен труд

за придобиване на образователна и научна степен „доктор”

в област на висше образование 4. Природни науки, математика и информатика,

Професионално направление: 4.6 Информатика и компютърни науки,  
докторска програма „Компютърни науки (Изкуствен интелект)”

**Автор: Сергей Александров Миланов**

**докторант към катедра „Софтуерни технологии“ на ФМИ, СУ „Св. Кл. Охридски”**

**Тема: „Извличане на зависимости в потоци от данни”**

**Изготвил рецензията: проф. д-р Силвия Христова Илиева**

Съгласно Заповед № РД38-132/21.02.2017 г. на Ректора на СУ "Св. Климент Охридски" съм определена за член на научното жури във връзка с процедурата за защита на дисертационния труд на *Сергей Александров Миланов* на тема „*Извличане на зависимости в потоци от данни*” за придобиване на образователната и научна степен „Доктор”, област на висше образование: 4. Природни науки, математика и информатика, професионално направление: 4.6. Информатика и компютърни науки, Докторска програма „Компютърни науки (Изкуствен интелект)“. С решение на научното жури (Протокол № 1 от 27.02.2017г.) съм определена за рецензент на дисертационния труд

Като член на научното жури съм получила необходимите документи, които отговарят на нормативните изисквания както на ЗРАСРБ, така и на Правилника за прилагане на ЗРАСРБ на Министерския съвет на Република България и този на СУ "Св. Климент Охридски", което ми дава основание да рецензирам настоящия дисертационен труд.

## **1. Актуалност на тематиката и целесъобразност на поставените цели и задачи**

Разработеният от Сергей Миланов дисертационен труд е в безспорно актуална област, по-конкретно анализ и извличане на зависимости от потоци данни, при отчитане на

времевата характеристика. Това е проблем в световен мащаб и придобива все по-голяма значимост със стремителното нарастване на обема от данни и ползите от тях. Докторантът предлага задълбочен анализ на проблемите и съществуващите решения в областта. Темата на дисертационния труд е логично изведена.

Представената работа си поставя като **основна цел** „Да се разработи методология за автоматизирано извличане на съществуващи зависимости в потоци от данни, съставени от количествени динамични записи, натрупвани в дълъг период от време, която да се сравни и валидира чрез практически изследвания. Методологията да е приложима към разнообразни източници на количествени данни“. Свързани с основната цел са формулираните **четири задачи** – обзор на темата за анализ и извличане на закономерности при потоците данни; разработване на метод за представяне на потоците от данни в нова структура, посредством извличане и анализ на техни характеристики; разработване на методология, представена като систематичен подход, за извличане на зависимости от потоци данни; приложения на предложената методология за анализ и извличане на зависимости в различни бази от динамични данни: ЕЕГ сигнали и времеви редове от UCR колекция данни.

Следваната методология на научно-изследователска работа е подходяща за решаване на поставените цел и задачи. Докторантът първо **анализира** текущото състояние на изследванията в областта на *обработка и анализ на времеви потоци данни и извличане на зависимости от данни*. В резултат на **синтез** авторът предлага **нов подход** за представяне на потоците от данни в нова структура и **методология** за автоматизирано извличане на зависимости в потоци от данни. За да бъде по-убедителна работата, предложената методология е **валидирана** чрез експерименти с два вида потоци от данни - ЕЕГ сигнали и времеви редове от UCR колекция данни. По този начин основните за софтуерното инженерство дейности – анализ, синтез и валидиране – са следвани в изследователската работа.

## **2. Познание на състоянието на проблема**

Няма съмнение, че докторантът е навлязъл сериозно в тематиката на дисертацията. В дисертационният труд са цитирани 107 източника, като всички са на английски език. Дори да се извадят четирите собствени публикации на докторанта, остават достатъчно източници, доказващи познанието на проблемите в областта. Преобладаващите литературни източници са съвременни и са на английски език, а 2 са на български език.

## **3. Характеристика и оценка на дисертационния труд**

Представеният дисертационен труд се състои от 152 стр. и е структуриран по следния начин: съдържание, списък на фигури, списък на таблици, включително речник на термините и използвани съкращения, въведение, 6 глави и използвана литература.

Оформлението е много добро, като е илюстрирано с 32 фигури и 17 таблици. Всяка глава съдържа кратко въведение и завършва със секция „обобщение и изводи“, което значително улеснява възприемането на съдържанието и доказва способността на докторанта за задълбочен критичен анализ. Като цяло прави добро впечатление стилът на изложение и липсата на граматични и пунктуационни грешки.

Във въведението се мотивира актуалността на проблема, обосновават се и се формулират целта и задачите и се обобщава структурата и съдържанието на дисертационния труд по глави.

**В Глава 1**, озаглавена „Теоретични основи на извличане на зависимости от потоци от данни“, са въведени основните понятия, от значение за труда в областите *обработка и анализ на времеви потоци данни и извличане на зависимости от данни*. Предложена е цялостна класификация на методите за анализ на времеви данни. При представяне на методите за обработка и анализ на времеви данни по новата класификация авторът се придържа към следната структура – предимства, недостатъци и приложимост. Обстойно е разгледан систематичен подход за извличане на зависимости от данни, както и отделните му етапи.

**В глава 2**, озаглавена „Извличане на зависимости от потоци данни“, е същностната част от дисертационния труд. В нея се представя нов метод за представяне на потоци от данни. Методът комбинира аспекти от два съществуващи подхода и представя времевите редове посредством честота на срещане на прототипи. Основава се на изследване на локални свойства и последващо създаване на краен брой съществени характеристики на потоците данни. Методът удовлетворява поставените в дисертационния труд изисквания и са оценени предимствата му. Другият съществен принос на дисертационния труд е предложената нова методология за извличане на зависимости от потоци данни. На практика тя конкретизира стъпките в известния процес на откриване на знания в данни KDD, като се предлагат нови решения и алгоритми, както и комбинации от съществуващи на всеки от четирите етапа. Методологията е разширяема и улеснява процеса на автоматизация на анализ на потоци от динамични данни.

**Глава 3** проследява приложението на предложената методология при ЕЕГ (електроенцефалограма) потоци от данни. Дефинирани са 3 задачи и са приложени и оценени различни техники за извличане на данни, като са направени изводи за най-подходящите. Получените резултати са теоретична база, която да се използва на практика за откриване на модели и автоматизиране на прогноза на човешкото състояние.

**Глава 4** проследява приложението на предложения подход при разнородни времеви данни от физическия свят на Университета в Калифорния (20 бази данни от физическия свят и 2 синтетични колекции). Обемът на данните е значително по-голям в сравнение с ЕЕГ данните в глава 3. Извършено е сравнение с други методи в областта и то показва превъзходство на предложената методология в голяма част от случаите.

Експериментите потвърждават широката приложимост на предложените от докторанта подход и методология. Получените резултати са анализирани, като докторантът показва умения за критичен анализ и формулиране на възможности за оптимизиране.

**Глава 5** е посветена на разработеното софтуерно приложение за извличане на зависимости от потоци данни, следвайки предложения в глава 2 нов метод. Въпреки ограничения обем, докторантът демонстрира отлично владение на съвременните технологии и инструменти.

Заключението е оформено като **Глава 6** и обобщава водещия научен въпрос, обекта на изследване и постигнатите резултати, в съответствие с поставените основни цел и задачи на дисертацията. Формулирани са приносите на дисертационния труд и са очертани насоки за бъдещо развитие.

#### **4. Приноси и значимост на разработката за науката и практиката**

Приемам разделянето на приносите от докторанта в три групи – научни, научно-приложни и приложни. Полезно е и обобщението им в табличен вид с реферането към публикациите, съответстващи на отделните приноси.

*Научните приноси* могат да се обобщят както следва:

- Предложен е **нов начин за класификация на методите за анализ на времеви данни**, чрез който съществуващите подходи се разграничават по по-голям брой критерии с по-голяма степен на детайлност и с описание на йерархични връзки в отделните групи.
- Предложен е **нов подход за представяне на потоците от данни в нова структура** от по-високо ниво посредством краен брой характеристики, получени след изследване на локалните им свойства.
- Предложена е **нова методология извличане на зависимости от потоци данни**, съдържаща подходящи алгоритми и решения за всеки етап, с акцент върху етап 3 – анализ.

*Научно-приложните приноси* са свързани с валидацията на предложената методология при различни типове потоци от данни. Препоръчвам докторантът да прецизира формулировката на научно-приложните приноси, като формулира кратко и ясно постигнатите резултати, вместо да обяснява извършени дейности. Изследванията и анализите сами по себе си не са приноси, а резултатите и изводите от тях са важни.

*Приложният принос* е реализацията на новия подход за представяне на потоците от данни в нова структура като работещо софтуерно приложение.

Като цяло дисертационният труд е завършено изследване в областта на информационните технологии. Предложеният нов подход на докторанта включва аспекти на два съществуващи подхода, а методологията е конкретизация на известен процес на откриване на знания в данни KDD.

## **5. Преценка на публикациите по дисертационния труд**

Резултатите от докторантурата са обобщени в 5 публикации, всички в съавторство с научния ръководител, като една е в съавторство и с международен колектив. Две публикации са в международни научни списания – Neural Computing and Applications с импакт фактор 1.492 за 2015 и International Journal of Reasoning-Based Intelligent, 2 в сборници на международни IEEE конференции - IEEE Intelligent Systems и IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA) и една на докторантски симпозиум. Всички публикации на докторанта са на английски език и покриват поставените в дисертационния труд задачи. Прави впечатление, че и трите научни приноса са в 1 публикация.

Към момента не са известни цитирания на публикациите към докторантурата. Докторантът е участвал в национален научен проект, финансиран от фонда на Софийския Университет, и в проект, финансиран от структурните фондове.

## **6. Автореферат**

Авторефератът е в обем от 37 страници и отразява адекватно основните резултати, постигнати в дисертационния труд.

## **7. Критични бележки и препоръки**

Към работата бих отправила следните **забележки и препоръки**:

1. При предложената нова класификация на подходите за анализ на времеви данни в категорията Compression-based е представен само един метод и по този начин тя е небалансирана, в сравнение с останалите 5 категории.
2. Би било полезно при оценката на предложения нов подход за представяне на потоците от данни освен положителните страни, да се посочат и недостатъците.
3. Би било полезно да се коментират представителността на изводите при експериментите с ограничения брой ЕЕГ данни (от 26 участници) и въпросът как големи обеми от данни биха се отразили на приложимостта на подхода и методологията.

4. Формулировката на научно-приложните приноси на дисертационния труд би могла да се прецизира в съответствие с коментарите в секция 4 от настоящата рецензия.
5. Научните приноси биха могли да са обект на повече от една публикация и по този начин да станат достояние на по-широк кръг от изследователската общност.
6. Библиографското описание на цитираните публикации № 72, 75 и 100 е непълно, като начинът на описание би следвало да е унифициран за всичките източници.

Със задоволство отбелязвам, че докторантът е отразил голяма част от забележките, отправени по време на предварителната защита в катедра „Софтуерни технологии“.

## ЗАКЛЮЧЕНИЕ

Общата ми оценка за дисертационния труд, автореферата и научните публикации на **Сергей Александров Миланов** е **положителна**. Отчитайки оригиналните научни, научно-приложни и приложни приноси на дисертационния труд и направените в тази връзка научни публикации, както и постигнатите значими практически резултати, намирам, че Сергей Александров Миланов притежава способност за задълбочени теоретични изследвания, както и за прилагането им в практиката. Докторантът показва абстрактно мислене, способност за самостоятелно формулиране и решаване на актуални проблеми в областта на информатиката. В този смисъл дисертацията отговаря на съвкупността от критерии и показатели за придобиване на образователната и научна степен „Доктор“, съгласно ЗРАСРБ, неговия Правилник и Правилниците за условията и реда за придобиване на научни степени и за заемане на академични длъжности на Софийския университет и на ФМИ. Това ми дава основание да предложа да бъде присъдена образователната и научна степен „Доктор“ на **Сергей Александров Миланов** в област на висше образование: 4. Природни науки, математика и информатика, професионално направление: 4.6 Информатика и компютърни науки, докторска програма „Компютърни науки (Изкуствен интелект)“.

28.04.2017г.

София

Подпис:

/проф. д-р Силвия Илиева/