

СТАНОВИЩЕ

на дисертационен труд

Извличане на зависимости в потоци от данни

за присъждане на образователна и научна степен "Доктор"
по докторска програма *Компютърни науки (Изкуствен интелект)*
професионално направление 4.6 *Информатика и компютърни науки*

автор: Сергей Александров Миланов
задочен докторант към катедра „Софтуерни технологии“
Факултет по математика и информатика, СУ „Св. Климент Охридски“

рецензент: доц. д-р Олга Георгиева
катедра “Софтуерни технологии”
Факултет по математика и информатика, СУ „Св. Климент Охридски“

Настоящото становище е изготвено в отговор на решение на Факултетния съвет на Факултета по математика и информатика (ФМИ) от 30.01.2017 г. за избор на научно жури, определено със Заповед № 38-132/21.02.2017 г. на Ректора на Софийския университет „Св. Климент Охридски“. То е изготвено съобразно критериите и показателите за придобиване на образователната и научна степен „доктор“ в ЗРАСРБ и неговия Правилник, както и в съответствие с Правилника на СУ „Св. Климент Охридски“ и Факултета по математика и информатика на СУ за условията и реда за придобиване на научни степени и за заемане на академични длъжности.

1. Съдържателен анализ на научните и научно-приложни приноси на дисертацията

Изследователската цел в представения за становище дисертационен труд е насочена към един особено актуален въпрос на съвременния свят. Въпросът е, как непрекъснатите и многообразни потоци от данни да се преобразуват в знания, които да бъдат предоставени навреме, точно да отразяват съответните процеси и явления, както и да служат за оперативни управленски решения и валидни прогнози. Дисертационният труд дава едно решение, което реализира цялостна процедура за реструктуриране на суровите данни, избор на подходящо пространство на данните, прилагане на конкретни алгоритми и анализиране на резултати за откриване на нови знания за изследваните процеси.

В изпълнение на целта дисертантът си поставя и решава конкретни научни задачи като първата от тях е свързана с необходимостта от критична оценка на известните решения. Изследвани и анализирани са голям брой от съществуващите методи за обработка на потоци от данни, базирани на различни подходи и приложени към широк спектър на данни. Представеният аналитичен обзор е задълбочено изследване, което позволява на автора да

направи обобщение на съществуващите методи и на тази база да предложи уникална класификация на методите по отношение на начините за оценка на сходството на потоци от данни. Добре обоснована е нуждата от нов систематичен подход за решение на задачата за извличане на зависимости от данни, натрупвани във времето.

Във втора глава на дисертационния труд се предлага нова методология, която оригинално съчетава различни алгоритми, включително и разработени от докторанта, които се прилагат в отделните етапи на процеса на откриване на знания от потоци от данни. Методологията е детайлно описана като последователност от етапи за решение на отделните подзадачи за групиране и класификация на данните като позволява да се отчетат особеностите на отделните случаи на използване. Обоснован е универсалният характер на приложимост на процедурата като неограничена от предметната област.

Безспорен научен принос на дисертационния труд е разработеният нов алгоритъм за представяне на потоците от данни на основата на сегментиране и извличане на атрибути за дефиниране на тяхно сходство. Представеният алгоритъм за представяне на потоци данни чрез честота на срещане на прототипи е едно ефективно решение за предварително структуриране на данните, което интензифицира откриването на съществуващи сходства при това с минимизиране на ресурсите и инициализиращите параметри.

С научно-приложна стойност са изследванията, представени в трета и четвърта глава. В трета глава дисертантът прилага разработената методология върху ЕЕГ данни като аргументирано от общата база данни формира отделни бази за откриване на специфични зависимости, както по отношение на мозъчните канали, така и по отношение на хората, реагиращи на дразнителя. От една страна са направени значими изводи за способността на различни клъстеризационни алгоритми да групират този вид сигнали. От друга страна, резултатите имат висока стойност предвид направения невро-биологичен анализ на значимите атрибути на данните, както и характеристиките на резултатите от клъстеризацията и класификацията. Силен резултат на това изследване е посочването на конкретните канали на мозъка, способни да дискриминират зрително възприятие.

В четвърта глава е показана универсалността и потвърдена ефективността на предложената методология. Систематизирано са изследвани 20 различни колекции от динамични данни и резултатите са сравнени с резултатите от други методи, прилагани към тези данни. За целите на този анализ дисертантът е предложил нов алгоритъм за автоматично създаване на прототипи, чрез което се повишава нивото на автоматизация при прилагане на методологията.

Дисертационният труд демонстрира последователност и цялостност на изследванията, което е показано и чрез описаното в пета глава софтуерно приложение *Data Expert*. То е разработено и използвано за целите на дисертацията. Следван е системен работен процес на софтуерна разработка с прилагане на съвременни стандарти и технологии. Възможността за оптимизация на вече вградените алгоритми, както и за разширяване с нови методи за анализ на данни, го прави ценен инструмент за бъдещо използване - самостоятелно или чрез вграждане в нови системи.

Дисертационният труд демонстрира дълбоко и обхватно познаване от автора на подходите за анализ на потоци от данни, вижда се стремежът да отговори обосновано и подробно на поставените задачи като има собствен оригинален принос в представените решения. Считам, че посочените от автора научни, научно-приложни и приложен принос са основателни и ги приемам напълно. По силата на обстоятелствата на мен не ми е вменено да задавам въпроси към дисертанта, но мога категорично да свидетелствам за неговата задълбочена, последователна и самостоятелна работа в разработването и описването на този труд.

2. Общи характеристики на дисертацията, автореферата и публикациите, които отразяват дисертацията

Дисертационният труд е представен в обем от 152 страници с уводна, пет съдържателни и една заключителна глави. Той е илюстриран с 32 фигури и 17 таблици. Цитирани са 107 литературни източници в областта на дисертацията, които в голямата си част са публикации, представени след 2000 година в реномирани научни издания. Те представят достатъчно пълна извадка от литературни източници, която да валидира резултатите на дисертацията.

Във връзка с дисертационния труд са представени пет публикации като и петте са публикувани в рецензирани издания. Публикациите обхващат съществените части на изследванията. Две от публикациите са в чуждестранни списания, едно от които е с импакт фактор (IF1,492 за 2015). Други две са представени и публикувани в сборниците на IEEE конференции, а петата е публикация от докторантска конференция в България. Като съавтор мога да потвърдя, че участието на докторанта в всички тях е значително, а в публикации [1,4] от списъка на публикациите дисертантът има основен принос.

Авторефератът е подготвен в съответствие с изискванията на закона и отразява правилно и вярно съдържанието на дисертационния труд. В него точно са резюмирани и представени приносите на труда.

Считам, че представеният дисертационен труд отговаря на изискванията на Закона за развитие на академичния състав в Република България и отговаря на специфичните за ФМИ на СУ изисквания за дисертационен труд за получаване на степен „Доктор“.

3. Заключение

Като имам предвид обема, начина на изследване и актуалността на представените разработки, както и достигнатото ниво на научна новост на теоретичните анализи и приложни резултати имам пълно основание и убедено препоръчам на уважаемото жури да гласува Сергей Александров Миланов да придобие образователната и научна степен „Доктор“ в **професионално** направление **4.6 Информатика и компютърни науки**.

25.04.2017 г.
София

Член на научното жури:
/доц. д-р Олга Георгиева/