

РЕЦЕНЗИЯ

от проф. дмн Галя Ангелова, ИИКТ-БАН
на дисертационен труд за придобиване на образователната и научна степен „доктор“
по професионално направление 4.6 „Информатика и компютърни науки“,
докторска програма „Информатика“
на тема „Интегриране на хетерогенни данни от разпределени IoT устройства“
с автор Кристина Иванова Динева

Съгласно заповед 211/2.11.2020 г. на директора на ИИКТ-БАН участвам в Научното жури за присъждане на ОНС „доктор“ на Кристина Динева, редовен докторант в ИИКТ-БАН през 2017-2019 г., отчислена с право на защита. Темата на дисертацията е насочена към предизвикателствата на реалните съвременни приложения на ИКТ в практиката: събиране на големи данни от различни устройства с цел наблюдения на обекти и процеси, интегриране на масиви от хетерогенни данни и обработка (най-често чрез машинно самообучение) на данните за нуждите на практически приложения. Тази тема е измежду най-актуалните направления в зараждащата се индустрия на данните и обект на най-голям интерес от страна както на изследователите по целия свят, така и на бизнеса. Избраната област на приложение, в която е построена прототипна система, е „интелигентно земеделие“ – един от фокусите на редица български стратегически документ за иновационно развитие и внедряване на цифрови технологии през следващото десетилетие. Така че първото положително впечатление от дисертацията идва от амбициозния избор на актуална съвременна област, в която ще се развиват приложения през сравнително дълъг период в бъдещето.

Правилникът за специфичните условия за придобиване на научни степени и за заемане на академични длъжности в ИИКТ-БАН изисква кандидатите за получаване на ОНС „доктор“ по специалност 4.6 „Информатика и компютърни науки“ да покриват минимални изисквания от 50 точки по показател А (представяне на докторски труд) и 30 точки по показателите група Г (сума от индикатори 5-10). Кристина Динева представя списък от общо 12 научни статии, като тя е съавтор в 10 от тях, а в две статии на български език е единствен автор. Статиите са публикувани в периода 2017-2019 г. във връзка с резултатите от дисертационния труд, от тях 8 са индексирани от Скопус, а 6 имат SJR-ранк, при което едната е в квартал Q2 на Скопус. Представен е също списък от 19 цитата на публикуваните статии, като 13 от тях са в научни публикации от чуждестранни автори извън професионалния колектив, в който работи кандидатката. Точките събрани от публикационната дейност (над 190) и цитиранията многократно надхвърлят изискванията на НАЦИД за минимален праг от 30 точки по показателите група Г при защита на докторска дисертация. С това формалните условия за

удовлетворяване на националните критерии са изпълнени и може да се пристъпи към рецензиране и защита на представения дисертационен труд.

Целта на разработката е „да се предложи система и инструменти за интегриране на хетерогенни данни от разпределени IoT устройства, които да позволяват тяхната обработка, моделиране и интеграция“. Поставени са следните задачи: да се предложи методология за обработка, моделиране и интеграция на хетерогенни данни; да се предложат архитектура и метод за комуникация на модулна IoT хардуерна система, както и архитектура на софтуерна платформа и подход за организация на услугите за интелигентна обработка на хетерогенни данни от IoT системи; да се създадат валидни модели за машинно обучение за експериментално потвърждение и да се покажат възможни полезни приложения на разработката.

Дисертационният труд съдържа 166 страници и е организиран в увод, пет глави и заключение. Цитирани са 175 литературни източника. Изложението е илюстрирано с 55 фигури и 16 таблици. Включен е речник на термини и съкращения, използвани в дисертацията, който улеснява разбирането на съкращенията на английски език.

Съдържание и приноси на дисертацията

В първа глава е направен анализ на състоянието на свързани с темата подходи и изследвания. Поради широката гама от задачи, близки до разглежданата проблематика, са включени кратки обзорни описания на Индустрия 4.0, Интернет на нещата IoT (технологии за изграждане – хардуер, софтуер, комуникации, платформи), приложения на IoT, предизвикателства пред IoT, хетерогенни данни и предизвикателства пред интегрирането им, както и подходи за интегриране на хетерогенни данни. Направен е изводът, че е необходимо да се създаде единен и систематичен подход за интеграция и обработка на хетерогенни данни.

Във втора глава се предлага методология за обработка, моделиране и интеграция на хетерогенни данни. Показана е стратегия за организация на работата в четири етапа (дефиниране на проблема; подготовка на данни – събиране, почистване, трансформации, анализ на данните; моделиране на задачата чрез машинно самообучение – избор на алгоритъм, конфигуриране, обучение и валидиране; интегриране на обучението модел в работна среда). На пръв поглед, за всеки учен с дългогодишен опит в изкуствения интелект, описваните стъпки за подготовка и обработка на данните са обичайни, но представената концептуална схема има определени достойнства. Тя се характеризира с ясно разграничаване на отделните етапи и с декомпозиране на задачата до по-малки модули, за които се предлагат решения базирани на различни технологии. В тази глава проличава старателната подготовка на докторантката с цел идентифициране на разнообразни подходи, алгоритми, технологии и готови програмни средства, подходящи

за реализацията на различните етапи от обработката на данните и създаването на модел. Изложението в дисертационния труд показва познанията на Кристина Динева в областта на „науката за данните“ (data science) и натрупания опит с различни среди за обработка на данни.

Трета глава представя разработената система за събиране, обработка и моделиране на данни, състояща се от отдалечени IoT устройства (инсталирани в пчелни кошери) и разпределено облачно-базирано софтуерно приложение, което позволява наблюдения на резултати от обработката на данните чрез потребителски интерфейси. Данните се събират в реално време или през определени интервали. Комуникацията между хардуерните компоненти е безжична, в центъра на системата е разположен рутер. Архитектурата на софтуерната система е адаптивна, понатоящем основана на осем вида микросървиси с възможност за разширяване. Потребителският интерфейс, реализиран чрез платформата Angular, позволява на потребителите в описателната част да се въвежда и обновява информация за техните пчелини и кошери, а в контролната част – да се извършва мониторинг в реално време на състоянието в кошерите за тези потребители, които са инсталирали IoT системата в техните пчелини. В тази глава са демонстрирани уменията на Кристина Динева като професионален информатик и експерт по информационни и комуникационни технологии, който се справя с разработката на хардуерни и софтуерни компоненти, предлага оригинални начини за адресация и решения свързани с енергийна ефективност на системата. Изграден е ефективен прототип, подготвен за реално внедряване, който благодарение на своята модулност позволява добавяне на допълнителни устройства и нови функционалности.

В четвърта глава е представен експеримент над реални данни с цел валидиране на представената методология за обработка и интеграция на хетерогенни данни, получени от IoT устройства. Събраните данни от наблюдения на медоносни пчели се използват за предсказване на две ситуации: „пчелното семейство няма достатъчно условия за оцеляване“ или „пчелното семейство има достатъчно условия за оцеляване“. Над учебни данни за вътрешна и външна температура на пчелните кошери и атмосферни условия е обучен бинамен класификатор, като сравнителните изследвания за успеваемостта на модела са извършени с четири алгоритъма за класификация в облачната среда Microsoft Machine Learning Azure Studio. Получени са резултати за точност, прецизност и чувствителност за обучените модели, с хармонично средно 0.845. Чрез регресионен анализ са изследвани връзките между измерваните индикатори. Показано е, че най-важните определящи фактори с различен коефициент на влияние върху промяната на количеството мед в пчелен кошер са: температурата, атмосферното налягане, влажността, скоростта на вятъра и часа. Направени са прогнози за очаквани количества в 10-дневен срок, с цел да се улесни планирането на дейностите на пчеларите, които трябва да посещават своите пчелини. Подходът за извършване на експеримента е впечатляващ със своята цялостност, дълбочина и изчерпателност на анализа на данните, както и с неговата ориентация към нуждите на потребителя при реално приложение.

Пета глава представя практическото приложение на разработената IoT система, наречена SmartBeeHives. Системата съчетава „основните компоненти на ИКТ“ – софтуер, хардуер, транзакции, комуникации, данни, интернет достъп и облачни изчисления. Извършва се цялостна услуга за потребителя – събиране, предаване, обработка, анализ, моделиране, визуализиране на данни и предсказване на събития. Областта на приложение е т.нар. „прецизно пчеларство“, в което цифровизацията навлиза на първо място като технология за отдалечено наблюдение и контрол. Представен е кратък обзор и сравнителен анализ на четири системи за интелигентно пчеларство, които са съпоставени с авторската система SmartBeeHives. Показани са преимуществата на SmartBeeHives. Те до голяма степен се дължат на модулността и гъвкавостта на системата, която може да бъде изградена над различни (вече съществуващи) устройства.

Обсъждане

Приемам научно-приложните приноси на автора така, както са декларирани в заключението на дисертационния труд, но за мен най-голямото постижение е амбициозният подход докторантът да започне от теоретични постановки и за 3-4 години да достигне до продукт, подготвен за пазарна реализация. Задълбочеността на анализа на данните, базиран на солидни статистически методи, е гаранция за изчисление на възможно най-добро предвиждане относно процесите, протичащи в кошерите, в контекста на една IoT система. Освен това е възможно софтуерът да бъде многократно обучаван със събирането на нови данни и предвиждащото ядро на системата да се усъвършенства, без да се прекратява функционирането на средата. Концептуалната простота на хардуерния компонент на системата и нейната гъвкавост са доказателство за експертизата на автора в разработката на съвременни приложения, базирани върху ИКТ.

Текстът на дисертацията е стегнат и конкретен, подходящо организиран с ясно разделяне по глави и тематика, с удобна за възприемане последователност на изложението, достатъчно ниво на детайлност и добре илюстриран с фигури и таблици. Полезно е включването след всяка глава на списък от авторски публикации, в които се представени резултатите. Авторефератът отразява коректно съдържанието на труда.

Бих отправила забележка за използването на терминологията: например, на български се употребява „машинно самообучение“ (така е озаглавена и съответната статия за Machine Learning в българската Wikipedia), а английският термин confusion matrix/ matrix of errors обикновено се превежда като „матрица на грешките“. Друг важен въпрос е дали разглежданите числови данни в експеримента са „истински“ хетерогенни, но считам, че не е редно тук да дискутираме дефинициите за хетерогенност в областта на IoT. Във всеки случай предложената във втора глава на дисертационния труд методология за интеграция и обработка на данните би изглеждала по принципно подобен начин и при много по-сложни многообразия от различни по характер данни.

Заклучение

Законът за развитието на академичния състав в Република България изисква докторският труд да съдържа научни или научно-приложни резултати, които представляват оригинален принос в науката. Дисертацията трябва да показва, че кандидатът притежава задълбочени теоретични знания по съответната специалност и способности за самостоятелни научни изследвания. Тези изисквания са изпълнени за дисертационния труд на Кристина Динева и научните публикации, в които са апробирани получените резултати. Големият брой цитирания е доказателство за международен отзвук и интерес от страна на общността, свързана с прецизно пчеларство. Дисертацията прави впечатление с амбицията си да произведе реално практическо приложение с високо качество, което да се позиционира на световно ниво и да се сравнява с най-добрите системи в тази област.

На тези основания ще гласувам положително за присъждане на степента и убедено предлагам на уважаемото Научно жури да **присъди на Кристина Иванова Динева образователната и научна степен "доктор"** по професионално направление 4.6 „Информатика и компютърни науки“.

30 ноември 2020 г.

Член на Научното жури:

