

РЕЦЕНЗИЯ

на дисертационен труд за присъждане на образователната и научна степен „доктор”
по професионално направление 4.6. Информатика и компютърни науки

Научна организация: Химикотехнологичен и металургичен университет – София

Автор на дисертационния труд: маг. Камал Смаил Мохаммед Ал-Барзнжи

Тема на дисертационния труд: „ИЗВЛИЧАНЕ НА ИНФОРМАЦИЯ ОТ ГОЛЕМИ МАСИВИ ОТ
ДАННИ С ЦЕЛ УВЕЛИЧАВАНЕ КОНКУРЕНТНОСПОСОБНОСТТА НА ПРЕДПРИЯТИЯТА”

Рецензент: доц. д-р инж. Веска Стефанова Ганчева
катедра "Програмиране и компютърни технологии", Технически университет - София

1. Кратки биографични данни и характеристика на научните интереси

Камал Смаил Мохаммед Ал-Барзнжи е завършил бакалавърска степен по "Компютърни науки" в Университета Дохук, Ирак през 2006г. и магистърска степен по "Компютърни науки" в Университета "Джавахарлал Неру" в Хидерабад, Индия през 2011г. Ръководител на катедра „Компютърни науки“ в Университета Рапарин, Иракски Кюрдистан, 2012-2015 г. Преподавател в Университета Рапарин, където води лекционни курсове и упражнения за магистри и бакалаври в катедра "Компютърни науки" и в катедра "Математика". Участвал е в разработката на системи за информационно осигуряване на студентска, библиотечна, медицинска и стопанска дейност в университета Рапарин. Зачислен за редовен докторант в катедра "Информатика" в ХТМУ-София през декември 2015г. Докторантурата му е спонсорирана от европейската програма Erasmus Marhaba. Научните интереси на Камал Смаил са в областите: бази от данни, обработката на големи масиви от данни и машинно обучение.

2. Преглед на дисертационния труд и анализ на резултатите

Дисертационният труд съдържа списъци на фигури, таблици и използвани съкращения; въведение; литературен обзор и анализ на големи данни – характеристики, класификация, области на използване; извличава на информация от големи данни и машинно обучение; системи за препоръки, архитектура за генериране на препоръки на основата на големи данни; методология и експериментални резултати; заключение;

приноси на дисертационния труд; списък на публикации по дисертационния труд; библиографска справка.

Обемът на дисертационния труд е 165 страници, съдържа 58 фигури и 13 таблици, библиографската справка съдържа 107 литературни източници. Литературният обзор е адекватен на тезата, целта и задачите на дисертацията.

Тематиката на дисертационния труд е актуална и свързана с решаване на проблеми в областта на анализа на големи данни, които са изключително важни за извличане на информация и знания, откриване на закономерности, прогнозиране на промени и разпознаване на често повтарящи се форми, схеми и модели.

В първа глава е направено въведение в големите данни, целите и техниките за извличане на информация. Дефинирани са целта и основните задачи на дисертационния труд. Цел на дисертационния труд е изграждането на разпределена софтуерна платформа, която да осигурява алгоритми за генериране на препоръки за повишаване на конкурентноспособността на фирмите. Платформата трябва да има възможност за комбиниране на два различни подхода за формиране на препоръки с цел постигане на по-добри резултати.

Във втора глава е направен литературен обзор на големите данни, извличането на информация (знание) от тях и приложението им за подпомагане на компаниите и бизнеса. Също така, са разгледани характеристиките и класификациите на големите данни. Анализирани са системите за препоръки и прогнозиране, ориентирани към големи данни. Обоснована е необходимостта от нови алгоритми в областта на откриване на често повтарящи се модели и съвместно филтриране, за да може да бъде извлечена информация от големите данни.

Трета глава е фокусирана върху извличането на информация от големи данни и характеристиките на големите данни - хетерогенност, мащабируемост, скорост, точност, достоверност, поверителност. Направен е анализ на методите, техниките и платформите за обработка и анализ на големи данни, както и алгоритмите и библиотеките за машинно обучение с големи данни.

Четвърта глава е посветена на анализ и класификация на системите за препоръки за работа с големи данни, както и някои метрики за подобие и оценка, използвани за изчисляване на препоръки на основата на големи множества от данни. Разработени са нови алгоритми и техники за съвместно филтриране (базирани на памет, базирани на потребител и базирани на продукт) за генериране на препоръки при наличие на големи количества структурирани данни с използване на Apache Mahout на платформата

Nadoor. Предложена е архитектура на система за формиране на препоръки чрез анализ на големи данни, чрез разработване на филтриращи техники, базирани на памет, базирани на потребител и базирани на продукт. Разгледани са мерки за оценка на представянето на алгоритмите и качеството на направените препоръки. Направено е сравнение между различните техники за определяне на качеството на формираните препоръки с цел подобряване на работата на системите чрез включване на предимствата и отстраняване на недостатъците на всяка от техниките.

В пета глава са представени използваната методология и експериментите. Предложена е архитектура на платформа за облачно базирана хибридна система за препоръки (Framework for Cloud Based Hybrid Recommender System (FCHRS)) за извличане на информация от големи масиви от данни. Дискутирана е реализацията на всички използвани в системата методи и алгоритми: алгоритми за съвместно филтриране и алгоритми за анализ на мнение. Софтуерната платформа комбинира различни подходи за формиране на две категории хибридни системи за препоръки. Проведени са експерименти с избрани набори от етикетирани данни за оценяване на работоспособността на всички алгоритми и качеството на предложените модели, използвани в платформата. Анализът на резултатите от оценката потвърждава, че разработените хибридни системи дават по-точни препоръки.

Чрез представените резултати в дисертационния труд са преодолени трите основни предизвикателства при системите за препоръки: студен старт, разреденост и мащабируемост. За постигането на по-добри резултати са комбинирани различни подходи и алгоритми, с което се подобрява точността на крайните резултати. По-точните резултати са предпоставка за повишаване конкурентноспособността на предприятията.

Използването на знания, извлечени от големи данни с помощта на разработената софтуерна платформа би имало значим ефект върху работата на компании и предприятия по отношение на взимане на бързи и адекватни решения, оптимизация на работата, спестяване на ресурси и като цяло би довело до повишаване на тяхната конкурентноспособност.

Представените материали и наличието на 5 публикации, в 4 от които дисертантът е първи автор ми дават основание да направя извода, че дисертационният труд е лично дело на маг. Камал Смаил Мохаммед Ал-Барзнжи. Основните постижения са публикувани в престижни международни списания и конференция и по този начин са станали достояние на научната аудитория.

3. Оценка на съответствието между автореферата и дисертационния труд

Авторефератът е в обем от 55 страници. Съдържа въведение, цел и задачи на дисертационния труд, литературен обзор на големите данни – характеристики, класификация, извличане на информация, машинно обучение, системи за препоръки, техники за филтриране, архитектура за генериране на препоръки на основата на големите данни, методология и експерименти, анализ на резултатите, препоръки за бъдеща работа, приноси и списък на публикации по дисертационния труд. Адекватно отразява основната цел на дисертационния труд и основните задачи, свързани с нея, методите за постигането им, постигнатите и описани експериментални резултати по дисертационния труд.

4. Характеристика и оценка на приносите в дисертационния труд

Приемам постигнатите в дисертационния труд приноси, които могат да се обобщят по следния начин:

Научно-приложни приноси:

- Направен е сравнителен анализ и класификация на техники, алгоритми и платформи за обработка, съхранение и извличане на информация от големи данни.
- Предложени са алгоритми за съвместно филтриране за генериране на препоръки от големи масиви данни на основата на: потребителски-базиран, предметно-базиран, както и моделно-базиран ALS алгоритъм.
- Предложени са алгоритми за анализ на текст и за генериране на прогнози от големи набори от етикетирани данни, базирани на известни алгоритми за класификация, като Naïve Bayes и SVM.
- Разработени са две нови хибридни системи за препоръки, базирани на големи данни. Първата система е базирана на каскадна хибридизация и използва алгоритми за съвместно филтриране. Втората система е базирана на смесена хибридизация и използва алгоритми за анализ на мнение.
- Предложена е софтуерна платформа, наречена облачно базирана хибридна система за препоръки (Cloud Based Hybrid Recommender System) за големи данни, включваща разработените нови алгоритми за извличане на знания от големи данни.

- Направена е експериментална оценка на качеството и производителността на предложените алгоритми и модели с набори от етикетирани големи данни, като са използвани средноквадратична грешка (RMSE) и крос-валидация, която потвърждава, че разработените хибридни системи дават по-точни препоръки.

Приложни приноси:

- Реализирана е софтуерна платформа за препоръки чрез използване на библиотеките за машинно обучение Mahout на Apache Hadoop и Spark ML на Apache Spark.
- Разработени са програмни имплементации на всички алгоритми за софтуерната платформа на език за програмиране Java.

5. Мнение за публикациите на дисертанта по темата на дисертационния труд

Списъкът от приложени публикации към дисертационния труд съдържа 5 публикации в съавторство с научния ръководител, като в 4 от тях докторантът е първи съавтор. Три от публикациите са в списания, една на международен симпозиум и една е докладвана на международен конгрес и отпечатана в списание. Публикувани са в списание „Engineering and Automation Problems“ (1/2018), Moskva, Russia с импакт фактор (IF 0.133); научен журнал на Химикотехнологичния и металургичен университет „Journal of Science, Engineering, & Education“ (No 1/2017); списание „Автоматика и Информатика“ (3/2017); международен симпозиум „Управление на енергийни, индустриални и екологични системи“, 2017; 13-ти Международен конгрес „Machines, Technologies, Materials“, 2016 и отпечатана в пълен текст във виртуалния журнал „International virtual journal for science, techniques and innovation for the industry MTM“ (12/2016).

Броят на публикациите надхвърля изискванията на Правилника за придобиване на научни степени и заемане на академични длъжности на ХТМУ. Публикациите са оформени добре, като съдържанието им отразява основните резултати от дисертационния труд.

6. Критични бележки и препоръки

Нямам съществени критични бележки към дисертационния труд. Препоръчвам на докторанта да продължи изследванията в областта обработка и анализ на големи

данни, да потърси практическо приложение на разработената софтуерна платформа и да подготви самостоятелни публикации.

7. Лични впечатления за дисертанта

Запознах се с маг. Камал Смаил Мохаммед Ал-Барзнжи на предварителната защита на дисертационния труд пред катедра „Информатика“, ХТМУ-София. Личното ми впечатление е, че докторантът има отлични познания в областта на разработвания дисертационен труд - много добра научна подготовка и умения за самостоятелно решаване на сложни научни задачи.

8. Заключение

Въз основа на направения анализ на предложения дисертационен труд, неговата значимост и съдържанието се в него приноси, считам че докторантът е демонстрирал отлично познаване на изследвания проблем, умения за изследване и успешно решаване на изследователска задача.

Представеният дисертационен труд напълно отговаря на изискванията в Закона за развитие на академичния състав в Република България, Правилника на неговото прилагане и Правилника за придобиване на научни степени и заемане на академични длъжности на ХТМУ.

Постигнатите резултати ми дават основание за **положителна оценка** на дисертационния труд и препоръка към Научното жури да бъде присъдена образователната и научна степен „доктор” на маг. Камал Смаил Мохаммед Ал-Барзнжи в област на висшето образование - 4. Природни науки, математика и информатика, професионално направление - 4.6. Информатика и компютърни науки (Информатика).

04.06.2018

София

Рецензент:

/ доц.д-р В. Ганчева /