

РЕЦЕНЗИЯ

на дисертационен труд за придобиване на образователната и научна степен „доктор“ по професионално направление 4.6 „Информатика и компютърни науки“ (Информатика – Изкуствен интелект)

с автор Гургана Ангелова Лазарова

на тема „Извличане на знания от много източници на данни“

Рецензент: проф. д-р Мария Нишева – ФМИ, СУ „Св. Климент Охридски“

1. Обща информация за докторанта

Гургана Лазарова е задочен докторант към катедра „Софтуерни технологии“ на ФМИ на СУ „Св. Климент Охридски“ от 2011 до 2015 г. Има висше образование – ОКС „бакалавър“ по компютърни науки, получена през 2008 г., и ОКС „магистър“ по информатика (изкуствен интелект), получена през 2010 г. Има около 10-годишен професионален опит в областта на проектирането и разработването на компютърни игри. Водила е упражнения във ФМИ по учебните дисциплини Изкуствен интелект и Откриване на знания в данни.

Научните интереси и досегашните изследователски резултати на Гургана Лазарова са в направленията интелигентни агенти, извличане на информация, машинно самообучение, невронни мрежи и генетични алгоритми. Тя е един от основателите и активен член на IEEE Young Professionals Affinity Group of Bulgaria

2. Анализ на съдържанието, резултатите и приносите на дисертационния труд

Дисертационният труд е посветен на изследвания в областта на машинното самообучение. Той има за цел създаване на нови подходи за смесено машинно самообучение (semi-supervised machine learning) чрез използване на много източници на данни, както и провеждане на подходящи експерименти, илюстриращи работоспособността на предложените алгоритми. Като естествено развитие на основните изследвания на автора е показано как предложените

подходи намират конкретно приложение в две съществено различни области – сегментация на изображения и оценка на чувствата в текстове на български език.

Дисертацията е написана на английски език и се състои от 118 страници текст, включващ шест глави, заключение и списък на цитираните литературни източници.

Глава 1 има уводен характер. В нея е представена областта на изследванията и са обосновани значимостта и актуалността на проблематиката, върху която е фокусирана работата на автора. Формулирани са целите на дисертационния труд и задачите, поставени за постигане на тези цели.

Глава 2 съдържа аналитичен обзор на изследваната област. Представени са различните видове машинно самообучение (МС) и са анализирани основните типове съществуващи методи за МС с учител, МС без учител и смесено МС. Специално внимание е отделено на методите за смесено МС с използване на много изгледи.

Глава 3 е посветена на разработения от автора на дисертационния труд генетичен алгоритъм за смесено машинно самообучение от много изгледи, който използва както класифицирани, така и неклассифицирани примери. В тази глава са включени: представяне на избраната методология; описание на предложения алгоритъм; описание на проведен експеримент, целящ сравнение на предложения генетичен алгоритъм и генетичен алгоритъм, който не използва неклассифицирани примери; анализ на получените експериментални резултати.

В глава 4 се съдържа въведение в областта „Оценка на чувства“ (Sentiment Analysis) и описва едно нетривиално приложение на предложения в предходната глава генетичен алгоритъм при създаване на софтуерна система за оценка на чувствата в текстове на български език.

В глава 5 е представен разработеният от автора на дисертацията алгоритъм за машинно самообучение, базиран на много изгледи. Този алгоритъм може да бъде разглеждан като модификация на базовия алгоритъм за взаимно обучение, предложен от Блъм и Митчел. Описан и анализиран е проведеният експеримент за оценка на предложения алгоритъм.

Глава 6 описва създадената като част от работата по дисертацията софтуерна система за сегментиране на изображения, която използва два изгледа и прилага по подходящ начин алгоритъма, предложен в пета глава.

Списъкът на използваните литературни източници включва 101 заглавия, преобладаващата част от които са цитирани по подходящ начин в текста на дисертационния труд. Този списък съчетава много добре класически и най-актуални източници в различни направления, свързани с отделните аспекти на темата на дисертацията.

Основните *научни приноси* в дисертационния труд могат да бъдат формулирани както следва:

- Разработен е генетичен алгоритъм за смесено машинно самообучение от много източници на данни;
- Разработен е алгоритъм за смесено машинно самообучение, базиран на много изгледи, който представлява модификация на базовия алгоритъм за взаимно обучение, когато единият изглед е по-лош от другия.

Като основни *научно-приложни приноси* в дисертационния труд може да бъдат посочени:

- Проектиране и провеждане на експерименти за проверка и изучаване на поведението на разработените алгоритми;
- Анализ на резултатите от проведените експерименти.

Приложните приноси на дисертацията може да бъдат обобщени както следва:

- Проектирана е софтуерна система за оценка на чувства в текстове на български език;
- Проектирана е софтуерна система за сегментация на изображения.

Общото ми впечатление от дисертационния труд е много добро. Изложението е добре структурирано, ясно и компетентно, илюстрирано с подходящи фигури и таблици. Избраната методика е адекватна на предмета, целите и задачите на изследването. Постигнатите теоретични и приложни резултати са оригинални и значими и напълно съответстват на поставените цели. Някои от формулираните в заключението на дисертацията насоки за развитие на получените резултати имат оригинален характер.

3. Публикации, които отразяват дисертацията. Отражение на резултатите на дисертацията в трудове на други автори

Резултатите, получени в дисертационния труд, са отразени в шест статии, публикувани както следва:

- две – в специализирани тематични сборници, издадени от Springer International Publishing;
- една – в сборник с доклади на международна научна конференция, издаден от IEEE;
- една – в научно списание, издавано от ТУ-София;
- две – в сборник с доклади на национална научна конференция.

Една от публикациите е самостоятелна, а останалите са в съавторство с научния ръководител и/или други негови докторанти. Във всички колективни публикации Г. Лазарова е първи автор. Нямам никакво съмнение в същественния ѝ личен принос в тези публикации.

Към настоящия момент е известно едно цитиране на публикацията на Гергана Лазарова по резултати от дисертационния ѝ труд.

4. Критични бележки и препоръки на рецензента

Като слабо място на дисертацията бих посочила наличието на чувствителен дисбаланс между обзорната ѝ част и частите, представящи изследванията на автора и получените резултати. В частност, изложението на най-съществените приноси глави (глава 3 и глава 5) е прекалено лаконично и до известна степен недобре аргументирано. Част от литературните източници, включени в частта “References”, фактически не са цитирани в текста на дисертацията.

Нямам други съществени забележки към съдържанието на дисертационния труд, тъй като ми беше предоставена възможност да се запозная с негова предварителна версия и повечето от направените от мен забележки по отношение на тази версия са намерили отражение в окончателния текст.

5. Автореферат

Авторефератът е изготвен в съответствие с изискванията на Правилника за условията и реда за придобиване на научни степени и за заемане на научни длъжности на ФМИ – СУ и едновременно пълно, компактно и точно отразява съдържанието и приносите на дисертационния труд. Текстът му следва да бъде подобрен в езиково отношение.

6. Заключение

Въз основа на изложеното по-горе смятам, че рецензираният дисертационен труд напълно удовлетворява и дори надхвърля изискванията на ЗРАСРБ, ПЗРАСРБ и Правилниците за условията и реда за придобиване на научни степени и за заемане на научни длъжности на СУ „Св. Климент Охридски“ и ФМИ. Убедено **препоръчвам на уважаемото научно жури да присъди на автора му Гергана Ангелова Лазарова образователната и научна степен „доктор“** по професионално направление 4.6 „Информатика и компютърни науки“ (Информатика – Изкуствен интелект).

София, 19.02.2016 г.

Рецензент:

/проф. д-р М. Нишева/