

## Рецензия

от доц. д-р инж. Ивайло Пламенов Пенев, кат. „Компютърни науки и технологии“, Технически университет – Варна

на дисертационен труд на тема „Използване на семантични технологии в един клас софтуерни приложения“ за придобиване на ОНС „Доктор“,

разработен от Даниела Иванова Петрова, докторска програма „Автоматизирани системи за обработка на информация и управление“ към професионално направление:  
5.3. „Комуникационна и компютърна техника“

Представеният дисертационен труд разглежда проблема за класифициране на мнения в потребителски коментари, като работата на докторанта е концентрирана върху текстове на български език.

1. Актуалност на разработвания в дисертационния труд проблем в научно и научно-приложно отношение. Степен и мащаб на актуалността на проблема и конкретните задачи, разработени в дисертацията.

Считам, че актуалността на разгледаната в дисертационния труд предметна област е безспорна. Разпознаването на потребителски мнения и класифицирането им като позитивни или негативни е задача, която е обект на активна научно-изследователска работа. Множество технологични решения на тази задача са внедрени в софтуерни приложения и системи, които се използват ежедневно от голям брой потребители. В същото време все още липсват достатъчно надеждни решения за класифициране на мнения в текстове на български език. Основна причина за това е недостатъчна база от подходящи текстове на български език, които да бъдат използвани за създаване, обучение и тестване на модели за разпознаване и класифициране на мнения. В настоящия труд докторантът се опитва да запълни тази празнина. По този начин може да осигури значително развитие на софтуерните системи, базирани на класифициране на потребителски коментари, напр. препоръчващи системи. Наред с това постигането на целта и свързаните с нея задачи в дисертационния труд могат да осигурят ефективен преход към развитие и на друг тип системи, базирани на български текстове, напр. системи за анализ, обобщаване и извличане на информация на български език.

2. Познава ли дисертантът състоянието на проблема и оценява ли творчески литературния материал.

В дисертационния труд са цитирани 115 източници. Основната част от тях са на английски език. Десет от цитираните източници са на български език.

В изложението си докторантът прави подробен анализ на трудовете на множество български и чуждестранни автори, които работят в предметната област на дисертационния труд. Докторантът коректно посочва приносите на цитираните автори, като в същото време акцентира върху недостатъци, които са провокирали разработването на дисертационния труд.

Считам, че дисертантът познава актуалното състояние на проблема и демонстрира способност за собствено тълкуване на проучения литературен материал.

3. Избраната методика на изследване може ли да даде отговор на поставената цел и задачи на дисертационния труд?

В дисертационния труд са приложени емпирични и теоретични методи, които са подходящи за постигане на целта и решаване на формулираните задачи.

4. Кратка аналитична характеристика на естеството и оценка на достоверността на материала, върху който се градят приносите на дисертационния труд.

Основен принос на дисертационния труд са създадените две бази от данни с коментари на български език. Всяка база съдържа над 100000 коментари от потребители на хотелски услуги както и на потребители на услуги в различни сфери, напр. различни културни и развлекателни събития, ресторанти, козметични и спа процедури, медицински услуги, магазини за хранителни и нехранителни стоки. Двете бази от данни са публично достъпни в GitHub. Докторантът е обяснил подробно в дисертационния труд от какви източници и как са създадени тези бази.

Този принос служи като основа, върху която са изградени следващите приноси в дисертационния труд. Всички приноси са подкрепени с данни от експериментални изследвания.

Считам, че приносите са изградени върху достоверен и богат материал.

5. В какво се заключават научните или научно-приложните приноси на дисертационния труд?

Създадените бази от данни с български коментари се използват за създаване, обучение и тестване на моделите за класифициране на потребителски мнения с различни методи и алгоритми от машинно обучение. Освен това дисертантът е споделил тези бази публично, с което помага на други учени да работят в областта на разпознаване и анализ на български текст. Авторът претендира, че в резултат от работата по дисертационния труд е създаден най-голямата база от данни с коментари на български език. Това е важно постижение, защото то безспорно може да даде тласък в развитието на софтуерните системи, обработващи текст на български език.

Освен създадените бази от данни докторантът е разработил списък от стоп думи, които могат да бъдат използвани за предварителна обработка на български текстове. Това е още един полезен резултат от дисертационния труд, който е в състояние да подкрепи работата на други колеги и да допринесе за прогрес в проблемната област.

Считам, че предложеният алгоритъм за разпознаване на мнения в потребителски коментари на български език не е нов, а е комбинация от известни методи от машинното обучение – метод на опорните вектори с включен механизъм на внимание, рекурентни невронни мрежи, случайни гори. Това не намалява значимостта на приноса, защото авторът е формулирал последователност от стъпки, на всяка от които се прилага съответен метод от машинното обучение. Успеваемостта на този адаптиран алгоритъм е обоснована с данни от експерименти, които показват висока точност на разработения класификатор (над 94%).

6. Може ли да се оцени в каква степен дисертационния труд и приносите представляват лично дело на дисертанта?

От изложението и от представените публикации проличава, че дисертационният труд и приносите са лично дело на дисертанта.

7. Преценка на публикациите по дисертационния труд: брой, характер на изданията

По дисертационния труд са представени седем публикации. Три от тях са самостоятелни, четири са в съавторство с ръководителя на докторанта. Четири от публикациите са индексирани в SCOPUS. Тези публикации са докладвани на международни научни конференции. Една от публикациите е в международно списание (American Journal of Engineering Research), една публикация е в списание „Компютърни науки и технологии“, издавано от ТУ-Варна, и една е в Годишник на ТУ-Варна. До момента е отразено едно цитиране от чуждестранни автори.

Считам, че представените публикации коректно отразяват резултатите от дисертационния труд.

8. Резултатите от дисертационния труд използвани ли са вече в научната и социалната практика? Има ли постигнат пряк икономически ефект и пр.? Документи, на които се основава твърдението.

Не са ми предоставени документи за практическо внедряване на резултатите от дисертационния труд.

9. Мотивирани препоръки за бъдещо използване на научните и научно-приложните приноси: какво и къде да се внедри.

Съществуват множество публично достъпни приложения за класифициране на потребителски коментари, които са предимно на английски език, а липсват такива приложения на български език. Препоръчвам да бъде предоставен достъп до описаните класификатори (напр. като уеб приложение), чрез който те да могат да бъдат тествани от различни потребители с непознати за моделите текстове.

10. Авторефератът направен ли е съгласно изискванията, правилно ли отразява основните положения и научните приноси на дисертационния труд?

Представеният автореферат се състои от 40 страници. В него са посочени целта и задачите, обект и предмет, методи на изследването. Обосновани са научната новост и практическа ценност. Следва кратко изложение на най-важните моменти от всички глави на дисертационния труд.

Считам, че авторефератът е подготвен според изискванията на ТУ-Варна, отразява коректно основните положения и научните приноси на дисертационния труд.

11. Критични бележки по дисертацията, включително и по литературната осведоменост на кандидата.

Имам следните критични бележки:

- На стр. 85 авторът посочва, че изборът на алгоритъм за оптимизация има силно влияние върху точността на използвания метод за машинно обучение. Авторът посочва, че експерименталните изследвания са проведени с три конкретни оптимизационни алгоритми,

без да направи обосновка на този избор. В резултат не става ясно дали може да се очаква, че използването на други алгоритми може да доведе до по-добри резултати при обучение и тестване на класификационния алгоритъм (или на цялостния подход).

- Липсва информация за времето на обучение на класификационните модели, построени чрез представените алгоритми. Това време може да бъде важен параметър за сравнение с други класификационни алгоритми, които съществуват или ще бъдат предложени в бъдеще.

12. Други въпроси, по които рецензентът счита, че следва да вземе отношение.

Имам отлични впечатления от представения дисертационен труд.

13. Заключение с ясно становище да се даде или не научна степен.

Считам, че въз основа на представения дисертационен труд докторантът Даниела Иванова Петрова заслужава придобиване на ОНС „Доктор“.

15.01.2024 г.

/...../

доц. д-р инж. И. Пенев