

СТАНОВИЩЕ

от доц. д-р Иван Койчев, СУ "Св. Кл. Охридски"

на дисертационния труд на асист. маг. Дорина Петрова Кабакчиева,
на тема „Изследване на Data Mining модели за класификация“
представен за придобиване на образователната и научна степен „доктор“
в професионално направление 4.6 „Информатика и компютърни науки“

Представената дисертация е в актуалната област: откриването на знания в данни (Data Mining). Избраните приложни области също са актуални, а именно: откриване на знания в данни за студенти, които да се използват за предсказване на тяхното представяне по време на обучението; и класификация на морски цели.

Дисертацията е много добре структурирана. Състои се от:

- Увод с кратка мотивация и ясно дефинирани цели с списък от пет произтичащи от тях конкретни задачи. Обем: 4 страници.
- Глава първа прави обзор на методите и алгоритмите от областта на откриването на данни, които се използват за учене на модели за класификация, както и методите за тяхното оценяване и сравняване. Направен е и обзор на изследванията прилагачи тези алгоритми и методи в двете избрани приложни области. Обем: 51 страници.
- Глава втора представя избраната методиката за провеждане на изследванията. Прави сравнителен анализ на системи за откриване на знания в данни и съответния мотивиран избор на софтуери инструменти за провеждане на експериментите. В тази глава са представени предварителните анализи и подготовката на данните от приложните области. Обем: 50 страници.
- Глава трета подробно представя резултатите от проведените експерименти с данни от двете приложни области. Получените резултати са оценени, проверени и сравнени с използването на съответните методи и мерки представени в глава първа. Обем: 71 страници.

- Заключение, което обобщава постигнатите приноси от докторантката и предлага идеи за продължаване на изследвания. Обем: 11 страници.
- Списък с литература, съдържащ 71 заглавия на научни статии и книги.
- Направен е и списък на термините на английски с техните преводи и дефиниции на български, който е поместен веднага след съдържанието. Намирам го за добро решение в случая, тъй като дава яснота в доста разнообразната, а по някога объркваща терминология дори на английски. Това важи още повече и за български, където често за един английски термин има по-няколко превода на български.
- Има приложени списъци на таблиците и фигурите.

Докторантката демонстрира задълбочени знания в областта на методите и алгоритмите за откриване на знания в данни и по специално на тези за учене на класификатори. Познава методите и метриците за оценка и сравняване на научените класификатори. Много добре е запозната със съвременното състояние на изследванията свързани с прилагането им в избраните приложни области. Направените обзори са систематични и изчерпателни. Всяка секция завършва със систематизирани изводи. За всеки от разгледаните методи и алгоритми са систематизирани неговите предимства и недостатъци. Добре е мотивирана нуждата от използването на откриването на знания в данни за приложните области. Списъка с използвана литература е представителен за областта и достатъчен за докторска дисертация.

Избраната методологията на изследване включва: подробен аналитичен обзор на областта; мотивиран избор на подходящи методи и алгоритми за избраните приложни области; систематичен анализ на досегашните резултати за приложните области и избор на подходяща изследователска ниша. Направени са предварителен анализ и обработка на две нови множества от данни и последващо прилагане на различни алгоритми за откриване на задания и зависимости в данните. Получените резултати („научени” модели) са оценени с подходящи мерки и сравнени по между си. Направени съответните изводи и

препоръки за най-подходящите алгоритми за откриване на знания в данни от този тип. Можем да обобщим, че приложената изследователска методология е подходяща за поставените цели в дисертацията и е традиционна в тази научна област.

Много добро впечатление прави и оформянето на самият текст, който е добре структуриран, естетически формен и добре форматиран. Издържан е и от гледна точка на правопис, граматика и стилистика. Важните единици (определения, обобщения и изводи) са обозначени и почертани с подходящо форматиране, което ги прави открояващи се от останалият текст. Това прави четенето на дисертационният труд по-ефективно и приятно. Фигурите и таблиците са надлежно номерирани, имат обяснителни надписи и са цитирани в текста. Графиките са с наименувани оси, добре визуализират данните и са лесни за разбиране.

Също така добро впечатление прави, че термините са надлежно преведени на български, като съответният терминът на английски е даден в скоби за сравнение и яснота. Трябва да отбележим, че в тази насока забележките ми от предзащита са вземати в предвид.

Представената дисертация цели да подбере и приложи подходящи методи и алгоритми за откриване на знания в данни за две конкретни приложни области: предсказване представянето на студентите в университета на базата на техни персонални характеристики преди и след приемане в университета; и за класификация на морски обекти. Тази цел е ясно формулирана в увода и са дефинирани пет основни задачи за постигането ѝ. Според мен поставените задачи и целта са изпълнени.

Приносите на докторантката са от научно-приложен характер и са достатъчни за дисертационен труд за образователна и научна степен доктор. Публикувани са 6 статии, като на 5 от тях докторантката е водещ автор. Две от публикациите са в международни списания, а на останалите са в рецензирани конференции. Приемането на статиите, докладващи резултатите от изследванията на специализирани рецензирани конференции в съответните приложни области, е едно убедително доказателство за тяхната значимост. Има две забележани цитирани: едното в статия публикувана в списание с имакт фактор, а другото в статия на международна конференция.

Авторефератът е добре направен, като вярно и точно отразява дисертационния труд.

По време на пред защита направих някои забележки към дисертацията. Констатирам, че те са вземати под внимание и аз съм удовлетворен от резултата.

Към представената дисертация нямам съществени забележки. Като неизпипах детайл по представянето мога да отбележа, че на някои от фигурите не е обърнато достъпно внимание: на някои от графиките няма имена на координатните осите, други са прекалено малки и надписите не са четими; а някои от цветните фигури са трудно четими в разпечатаното черно-бяло копие.

Можем да **обобщим**, от представеното дотук е видно, че представената работа има съществени научно-приложни приноси, напълно достатъчни за присъждане на научно-образователна степен „доктор”.

В заключение, давам **положителна** оценка на представената дисертация и предлагам на уважаемото научно жури, по обявената процедура за защита на докторска дисертация в ИИКТ-БАН, да даде образователната и научна степен „доктор” в професионално направление 4.6. „Информатика и компютърни науки“ на автора на дисертационния труд Дорина Кабакчиева.

Дата 14.02.2013 г.