

Рецензия

на дисертационния труд за придобиване на
образователна и научна степен „доктор“
по направление 4.6 „Информатика и компютърни науки“
научна специалност 01.01.12 „Информатика“ (Изкуствен интелект)
на Милко Красимиров Крачунов
на тема “Изкуствен интелект в биоинформатиката: автоматизиран анализ и
класификация на данни от паралелно секвениране“

Рецензент: доц. Стефка Фиданова

Утвърдена съм със заповед No РД38-122/06.02.2015 на Ректора на СУ „Св. Климент Охридски“ проф. дин. Иван Илчев за член на Научно жури във връзка с процедура за придобиване на образователна и научна степен „доктор“ по специалност 01.01.12 „Информатика“, професионално направление 4.6 „Информатика и компютърни науки“ (Изкуствен интелект) от Милко Красимиров Крачунов с дисертация на тема “Изкуствен интелект в биоинформатиката: автоматизиран анализ и класификация на данни от паралелно секвениране“, научни ръководители , проф д-р.Мария Нишева, доц. д-р Димитър Василев.

Като член за Научното жури съм получила:

1. Дисертация за придобиване на образователна и научна степен „доктор“ по направление 4.6 „Информатика и компютърни науки“;
2. Автореферат;
3. Професионална биография;
4. Копия от публикациите на автора свързани с дисертацията.

При оценяването на дисертационния труд се взема под внимание изискванията на Закона за развитие на академичния състав в Република България (ЗРАСРБ), Правилника за неговото прилагане (ППЗ) и Правилника за условията и реда за придобиване на научни степени и за заемане на академични длъжности във ФМИ на СУ. Основните норми, които трябва да бъдат спазени са:

1. Съгласно чл. 6(3) от ЗРАСРБ „дисертационният труд трябва да съдържа научни или научно приложни резултати, които представляват оригинален принос в науката. Дисертационният труд трябва да показва, че кандидатът притежава задълбочени теоретични знания по съответната специалност и способности за самостоятелни научни изследвания“.

2. Според чл. 27(2) от ППЗ дисертационният труд трябва да се представи във вид и обем, съответстващи на специфичните изисквания на първичното звено. Дисертационният труд трябва да съдържа: заглавна страница; съдържание; увод; изложение; заключение – резюме на получените резултати; библиография.

Дисертацията на Милко Красимиров Крачунов се състои от Увод (глава първа), три глави, анализ на резултатите (глава пета), приноси и перспективи (глава шеста), библиография, списък на авторските публикации по дисертационния труд, списък на фигурите и списък на таблиците в дисертацията, списък на съкращенията и речник на термините.

1. Актуалност на проблема и целесъобразност на поставените цели и задачи

Биоинформатиката е сравнително ново интердисциплинарно научно направление, което комбинира познания от биологията, химията, математиката, информатиката и инженерните науки. Това е област на изследвания, които генерират нови знания и нови теории. Анализът на данни от паралелно секвениране в геномните изследвания е актуален проблем, с широко приложение – от изучаването на еволюцията на живите организми, до здравеопазването. Метагеномиката се занимава с изследване на местообитания и съобщества от микроорганизми. Различните метагеномни изследвания са затруднени от липсата на добри решения за елиминация на шум в данните, предизвикан от грешки на апарата за секвениране.

Представеният за рецензиране дисертационен труд е насочен основно към методи за различаване на грешки от мутации. На страница 21 в Увода е формулирана целта на дисертационния труд:

Основната цел на настоящата работа е разработването и изследването на нови подходи за откриване и корекция на грешки в метагеномни данни от паралелно секвениране, използвайки методи от информатиката и изкуствения интелект.

За постигането на така поставената цел се решават следните по-специфични задачи:

- Разработване на аналитичен подход за откриване на грешките;
- Приложение на оценките на грешките;
- Валидация на методите;
- Оценяване на методите с невронни мрежи;
- Създаване на софтуер за автоматизирано изпълнение на разгледаните процеси.

2. Познаване състоянието на проблема от страна на дисертанта

Без съмнение, дисертантът е навлязла много добре в научната проблематика. Тематиката е сравнително нова. Списъкът на цитираните литературни източници е актуален – публикациите от последните 10 години представляват 40% от всички цитирани заглавия, но от друга страна в него присъстват и позовавания на по-стари източници. Общият брой цитирани източници е 170, като всички са от чуждестранни автори, 166 са на английски език и 4 – на руски.

Познаването на проблема от страна на дисертанта е много добре илюстрирано в глава 2 на дисертационния труд. Там са представени теоретичните основи и състоянието на изследвания проблем. Направен е подробен литературен преглед на съществуващите методи и алгоритми за анализ и класификация на данни от паралелно секвениране.

3. Методика на изследването

Методиката за провеждане на изследването, избрана от дисертанта, произтича от поставената цел и съответства на произтичащите от целта задачи. Тя е напълно адекватна на целта на работата. Предложен е аналитичен подход за корекция на грешки в метагеномни данни. Разработен е индиректен подход за валидация на различни методи за корекция на грешки върху данни от паралелно секвениране. Въведен е размит индикатор за оценка на грешките. Използвани са невронни мрежи за класификация на потенциални грешки. Предложен е комбиниран метод съчетаващ в себе си аналитичен подход и невронни мрежи. Това дава възможност за постигане на по-голяма точност.

4. Характеристика и оценка на приносите на дисертационния труд

Дисертацията се състои от пет глави (включващи увод и анализ на експерименталните резултати), списък на цитираните публикации, списък на авторските публикации по дисертацията и списък на научните форуми на които са докладвани резултатите включени към дисертацията. Приложени са таблици и графики илюстриращи разглежданите задачи и постигнатите резултати. Много добро впечатление ми направи, че дисертацията започва със списък с използваните съкращения и с речник на използваните термини. Това много улеснява четенето.

Глава първа представлява увод в предметната област на изследването. Там е дадена мотивация за избор на проблема и обекта на изследването. Направен е кратък преглед на използваните в дисертацията методи. Посочени са целите и задачите на дисертацията.

В Глава 2 са изследвани теоретичните основи на изследвания проблем. Разгледани са съществуващите решения и алтернативни методи. Направен е подробен литературен преглед на направеното до сега по тематиката.

Глава 3 дава постановка на задачата и решение на проблема за анализ и класификация на данни от паралелно секвениране. Разработен е аналитичен подход за корекция на грешки и индиректен подход за валидация. Използвани са невронни мрежи за откриване на зависимости между данните и наличието на грешки. Предложена е комбинация на аналитичен подход с невронни мрежи. Така се подобрява точността на получения резултат.

В Глава 4 е описана разработената софтуерна реализация на предложените в дисертацията методи. За целта е предложен език за описание на работни потоци на основата на езика YAML.

В Глава 5 е направен анализ на резултатите от проведените експерименти. Тя играе ролята на заключение. Направен е анализ на резултатите от прилагане на отделните методи както и на комбинираното им прилагане. Направено е сравнение между предложените методи, както и с методи предложени от други автори и е показано, че предложените в дисертацията методи дават по-добър резултат. Разработените методи са приложени при изследване на генома на пшеницата.

Дисертантът претендира за следните приноси:

Научно-приложни:

1. Разработен е аналитичен подход за корекция на грешки в метагеномни данни от паралелно секвениране.
2. Разработен е индиректен подход за валидация на различни методи за корекция на грешките върху данни от паралелно секвениране.
3. Въведен е размит индикатор на увереност за оценка на грешката при анализ на данни от секвениране без корекция на грешката.
4. Предложен е модел на невронна мрежа за класификация на потенциалните грешки.
5. Предложен е подход за обучение на модели на машинно самообучение за откриване на грешки без използване на обучаващи данни с действителни грешки.
6. Създаден е модел на система за извършване на сложен анализ на данни от секвениране, чрез език за описание на управляеми работни потоци.

5. Значимост на разработката за науката и практиката

Извършената от дисертанта работа е достатъчна по обем и задълбоченост на изследването. Показана е възможност за практическа реализация. В този смисъл намирам работата за значима както в научно, така и в практическо отношение.

6. Преценка на публикациите по дисертацията

Кандидатът има общо над 20 публикации. Във връзка с дисертацията, той е представил 6 публикации, две от които в международни списания (едното Journal of Computational Science е с IF 1.567 и SJR фактор 1.038). Останалите публикации са в томовете на специализирани международни конференции.

Дисертантът е участвал на 4 конференции с доклади, на други 4 конференции с постери, както и в три научни проекта: два финансирани от Фонда за научни изследвания и един от програмата за развитие на човешките ресурси съфинансирана от ЕС.

7. Преценка на публикациите по дисертацията

Кандидатът не е представил справка за личното си участие в колективните публикации, но от дългогодишната му работа по темата и общият брой публикации оставам с впечатлението, че неговото участие е съществено.

8. Автореферат

Авторефератът правилно отразява съдържанието на дисертацията.

9. Критични бележки

Като цяло дисертацията е добре написана и оформена. Имам някои критични бележки от технически характер, които и биха били от полза на дисертанта в работата му в бъдеще.

1. Макар и не много, но се срещат някои технически грешки като неправилно членуване и съгласуване на частите на речта. Има изпускане на предлози, най-често това е предлогът „на“.

2. На стр. 85 формула 3.19 за разстояние на Левенщайн е пропуснато да се напише стойността при $i=0$ и $j=0$. Коректно е да го има, въпреки че е очевидно.

3. Приносите в дисертацията са разделени на „Теоретико-методични“ и „Експериментални“. Обикновено делението е „Научни“, „Научно-приложни“ и „Приложни“. Мисля, че приносите на тази дисертация могат да се опишат като „Научно-приложни“ (така наречените в дисертацията „Теоретико-методични“ приноси) и „Приложни“ (разработения софтуер).

4. Според мен експериментите не могат да бъдат приноси. Те са средство за потвърждаване на теорията.

10. Лични впечатления

Не познавам лично дисертанта, но съм участвала в две конференции, на които той е докладвал. Останала съм с отлично впечатление от качеството на докладваните резултати и от начина им на докладване.

Заклучение

Като следствие на изложеното по-горе, може да се констатира, че са изпълнени всички изисквания на Закона за развитие на академичния състав (ЗРАСРБ), Правилника за неговото прилагане (ППЗ) и Правилника за условията и реда за придобиване на научни степени и заемане на академични длъжности във ФМИ на СУ.

Посочените от мен критични бележки засягат предимно техническото оформление на дисертацията и не намаляват научната ѝ стойност.

Всичко това ми дава основание за положителна оценка и предлагам на почитаемото Научно жури да присъди образователната и научна степен „доктор“ по специалност 4.6 „Информатика и компютърни науки“ на Милко Красимиров Крачунов.

20.02.2015 год.
гр. София

(доц. д-р. Стефка Фиданова)