

Становище

от

доц. д-р Димитър Иванов Василев
ръководител група "Биоинформатика", АгроБиоИнститут
бул. "Драган Цанков" №8, София 1164,

относно

дисертационния труд

на

Милко Красномиров Крачунов

Изкуствен интелект в биоинформатиката: автоматизиран анализ и класификация на данни от паралелно секвениране

представен за придобиване на образователна и научна степен "Доктор"
по направление 4.6 "Информатика и компютърни науки",
научна специалност 01.01.12 "Информатика" (Изкуствен интелект)

Въз основа на издадена заповед № РД38-122/06.02.2015 от ректора на Софийския Университет "Св. Климент Охридски", проф. дин Иван Илчев съм избран за член на научно жури по процедура за придобиване на образователната и научна степен "доктор" от Милко Красномиров Крачунов, редовен докторант към катедра "Компютърна Информатика" при Факултет по Математика и Информатика на същия университет с дисертация на тема "Изкуствен интелект в биоинформатиката: автоматизиран анализ и класификация на данни от паралелно секвениране".

В качеството си на член на научното жури ми бяха предоставени: дисертацията и автореферата по нея - в напечатан и подвързан вид, професионалната биография на докторанта, както и копия от публикациите, свързани с дисертацията.

Настоящото становище е съобразено с изискванията на Закона за развитие на академичния състав в Република България (ЗРАСРБ), Правилника за неговото прилагане (ППЗ), както и Правилника за условията и реда за придобиване на научни степени и за заемане на академични длъжности във Факултета по Математика и Информатика към Софийския Университет "Св. Килмент Охридски".

Представеният дисертационен труд е съобразен с основните изисквания според ЗРАСРБ, ППЗ и правилника на ФМИ, като се състои от уводна глава, четири глави по съдържанието на дисертационния труд, глава за приносите и перспективите, списък на използваната от автора литература, списък на авторските публикации по дисертационния труд, списък на фигурите и списък на таблиците в дисертацията, речник на термините и съкращенията.

Като актуалност на проблема, смисъл и изпълнение на поставените цели и задачи дисертационната работа на Милко Крачунов има несъмнена стойност в контекста на представянето на нови знания от областта на изчислителната биология, свързани с използването на *in silico* методи за автоматизиран анализ и класификация на данни от паралелно секвениране на метагеномни съобщества. Тези методи и приложението им в предствения за защита дисертационен труд включват анализ на данни от ново поколение секвениране във всичките си измерения, като претърсване на данни, откриване и корекция на грешки,

съпоставяне на секвенции. Още по-важно като иновативен подход или група от подходи, е използването в дисертацията на различни класификационни методи, както и методи на изкуствен интелект за получаване на по-точни резултати за нивото и отличаването на наличните грешки в секвенционни масиви данни. Това е особено важен аспект при определяне на биоразнообразието и мутационната активност в изследваните извадки от микроорганизми. Метагеномиката е част от т.н. модерни "омикс" науки и технологии в контекста на дешифрирането и изследването на различни геноми, която придобива особен смисъл и потенциал за развитие именно с интензивното използване на високопродуктивни секвенционни технологии. Тези очевидни факти правят избраната тема за дисертационен труд значима и актуална.

Дисертацията съдържа 166 страници текст, 16 фигури и 15 таблици като е структурирана в 6 глави с увод и заключение. В дисертационния труд авторът е отбелязъл използването от него на 170 литературни източника, които преимуществено са на английски език. Представен е интересно оформен превод с определена биоинформатична семантика на речник на използваните в дисертацията термини. Представен е и списък на използваните съкращения,

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

Представният за рецензиране труд от докторант Милко Крачунов има определена стойност като актуалност на проблема, избор и изпълнение на поставените цели и задачи. Дисертацията надхвърля далеч рамките на сравнителен анализ на известни методи, но и има определени достойнства като носител на нови знания в биоинформатиката, свързани с използването на *in silico* методи за автоматизиран анализ и класификация на данни от паралелно секвениране на метагеномни съобщества.

Биоинформатика е интердисциплинарна научно-приложна изследователска област, която съчетава в себе си нови технологии, като на съвременните молекулярни и био-технологии, така и в голямата си част почти всичко ново от областта на съвременните информационни технологии. Предоставената за анализ информация, която е обект на биоинформатиката съдържа в себе си много възможности и потенциал за използване на методи на изкуствен интелект, класификационен анализ, теория на информатиката, статистика, които от своя страна предлагат различни възможности за оптимизации и получаване на иновативни и оригинални решения. Дисертацията на Милко Крачунов включва именно такива методи и технологии на изследване, ориентирани към анализ на метагеномни данни. В тази работа е направено не малко за развитие на иновативни и оригинални идеи, които несъмнено са способствали за генерирането на нови знания.

Разработването и усъвършенстването на тези методи в дисертационния труд на Милко Крачунов са способствали за получаването на нови резултати с помощта на методите на изкуствения интелект при откриването на грешки и класифицирането им с по-голяма точност при анализ на данни от метагеномно паралелно секвениране. Метагеномиката, като обособен дял от геномиката се развива именно с широкомащабното използване на високопроизводителните "омикс" технологии - целящи главно определяне на биоразнообразието в изследваните извадки от микроорганизми. Оценката на биоразнообразието се основава на два базови фактора - точност на определяне на мутационната активност и на оценка на грешките в секвенционната информация. Втората задача, която определя биологическия смисъл на работата на Милко е не само добре представена и решена от докторанта, но и е допълнена от представените алтернативни методи за подобряване на нейното качество, основани на съвременни подходи от информатиката, изкуствения интелект, класификационния анализ. Всичко това прави избора на темата, целите и задачите като и самата работа на докторант Милко Крачунов значими и актуални.

Определяйки спецификата на работата като оригинална и иновативна, тя е насочена

главно към 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 и е предложено решение на следните следните по-специфични задачи:

- Разработване на аналитични подходи за откриване на грешки;
- Класифициране на оценките на грешки;
- Валидация и сравнимост на методите за откриване на грешки;
- Усъвършенстване на предложените аналитични подходи с използване на методи на изкуствен интелект;
- Създаване на софтуерни решения за автоматизирано изпълнение на разгледаните процеси.

Всички изложени задачи са решени от докторанта и публикувани в реферирани списания, доклади и постери от научни конференции.

2. Познаване състоянието на проблема от страна на дисертанта

Оригиналното и обхvatно преставяне на състоянието на проблематиката по темата на дисертацията е едно от главните достойнства на работата на Милко Крачунов. Задълбоченото познване на проблемите, както и представянето на съществуващите методи и алгоритмични решения, придружени със съответната дискусия поставят в много добра светлина написаното от докторанта в глави втора и трета на дисертационния труд. Използването на съвременна литература, колаборацията с известни центрове по отношение на темите на работа, допринасят също за разбирането на състоянието на проблемите от докторанта от една страна. Иновативността при представяне на определени проблеми (каквото е проблема с валидирането на методите за откриване на грешки) допринасят от друга страна за много добрата подготовка на докторанта и неговата склонност към самостоятелна работа. Без да се старя да хвърлям и най-малка сянка върху достойнствата на написаното в двете галви от автора бих отправил известна забележка за някои небалансирани представяния на определени методи и дискутирането им, което отдавам повече на намирането на информация за тях, както и на ролята им за постановката на самия дисертационен труд

3. Методика на изследването

Методиката за провеждане на изследването, избрана от дисертанта, се основава както на класически методи така и главно на негови оригинални идеи, свързани с поставените цел и задачи на работата. Така предложените методи и средства като аналитични подходи, алгоритмични решения от различни области на информатиката са абсолютно адекватни по отношение на работата, както и имат определено практическо приложение особено при получаване на по-добри резултати. Авторът предлага група от нови методи за анализ на секвенционни данни от метагеномно секвениране с цел откриване на грешки. Тези методи биха могли да се определят в две групи - аналитични и методи с помощта на използване на машинно обучение. Така, класификационният анализ на грешките до голяма степен се основава на използването на методи на изкуствения интелект. Едно от главните достойнства и оригиналност на работата е, че тези методи се използват съвместно като това дава определени положителни резултати при точността на откриване и корекция на грешки в данни от паралелно метагеномно секвениране. Валидирането на използваните методи е разработено с помощта на индикатор за оценка на грешките основан на тория на размити множества. Като цяло, посредством използването на тези оригинални подходи е постигнато качествено оптимизиране на методиката за откриване на грешки, чрез съвместното използване на аналитични методи и методи на машинното обучение.

4. Характеристика и оценка на приносите на дисертационния труд

Дисертацията се състои от шест глави, списък на цитираните литературни източници, списък на публикациите на докторанта свързани с дисертацията, графики и таблици които много добре илюстрират консеквентостта на разработването на тезата на докторанта. Речникът на използваните съкращения е много оригинален и своевременен за биоинформатиката която се преподава и практикува у нас.

В глава 1 на дисертацията са представени добре и ясно предметната област и методите на формулирани са целите и задачите на работата. Глава първа е уводна в предметната област на изследването. Определено добре е формулирана представената мотивация за избор на проблема и обекта на изследването.

Теоретичните основи, литературния преглед на състоянието на изследваните в дисертацият проблеми са представени много добре в глава 2. Много добре са дискутирани теоретичният апарат на съществуващите методи на изследване, особено в контекста на размити множества, изкуствен интелект, анализ на данни от паралелно секвениране в метагеномика - методи и алгоритми, съществиващи решения и софтуерни програми. .

Постановката на задачата, входните данни, предварителната обработка на данните, начините за анализ на данните посредством аналитични подходи, индиректни подходи за валидация на откритите грешки, използването на апарата на невронните мрежи за решението на поставените задачи решение на поставените задачи са преставени в Глава 3. Всички методи и алгоритмични решения са описани много добре и създават впечатление за много добрите знания и подготвеността на докторанта, както и подчертана сколнност към търсене на оригинални решения и иновативност

Софтуерната реализация на разработените в дисертацията методи е представена в Глава 4, като резултатите от използването им потвърждават доброто им качество. Тук бих могъл да отправя втората си забележка към работата на докторанта, засягаща завършеността на софтуерната реализация от една страна и практичността ѝ от гледна точка на използваните информационни технологии. Смятам, че авторът би направил една отлична организация на софтуерната реализация, ако от самото начало се беше възприел единен ситемен подход за организация на работните потоци и начина на преставяне на самия софтуер като завършен продукт с цел инегрирането му в големи аналитични платформи за биоинформатика. Това не намалява стойността на работата, а напротив представлява едно предизвикателство за ориентация към определена практична приложимост и интегративност на развитите софтуерни програми.

Дискусията и изводите по получените резултати въз основа на използваните методи, тяхната сравнимост и отвореност към усъвършенстване са представени и дискутирани в глава 5. Тук има много оригинални резултати, сравнителни характеристики, предложения за подобрене на методите и използваните алгоритмични подходи. Важно значение има предсатвянето на възможността за използването на методите и алгоритмите от дисертацията за откриване на грешки в метагеномни данни за приложението им в анализ на данни от *de novo* секвениране сложни геноми, като този на хексаплоидната пшеницата.

Приносите и перспективите, оригиналността на изследванията, публикационната дейност на докторанта и участието му в различни проекти, както и потенциалът за бъдещо развитие са преставени подробно в Глава 6.

5.Значимост на разработката за науката и практиката.

Представеният дисертационен труд има несъмнен научен и практически принос за създаване на нови методи и алгоритмични решения за откриване на грешки в метагеномни данни. Работата на докторанта може да бъде отчетена определено положително що се отнася до иновативност и приложимост.

6. Преценка на публикациите по дисертацията.

Публикационната десйност на Милко Карчунов по темата на дисертацията е от много добро качество, като тя се изразява в 6 публикации в реферирани издания, 4 конференции с доклади, и в 4 конференции с постери. Две от статиите са в международни списания, едната от които с импакт фактор IF 1.567. Докторантът е участвал в три научни проекта: два финансирани от Фонда за научни изследвания и един от програмата за развитие на човешките ресурси, съфинансирана от ЕС.

7. Лично участие на докторанта в публикационната десйност

Докторантът не е представил информация за за личното си участие в колективните публикации, но от дългогодишната ми съвместна работа с него като негов научен ръководител мога да кажа, че личното му участие е изключително съществено и от много високо качество.

8. Автореферат

Авторефератът е написан много добре и отразява правилно съдържанието на дисертацията.

9. Критични бележки

Като цяло дисертацията е написана и оформена много добре. Имам следните критични бележки:

1. В граматическо отношение има някои колебания при използването на пълен и кратък член на частите на изречението - но това го отдавам повече като болест на съвременото на младите хора.

2. Отделните глави имат понякога известна небалансираност и прекъснатост - така е в глава 2 и глава 4. Тук може да се изисква едно по-добро представяне и консеквентност както на използваните методи, така и на анализа и дискусията на получените резултати.

3. Както отбелязах по-горе, предатвянето на софтуерната реализация би могло да се предсатви и релаизира в по-опростен вид, който да е като достъпен за работа на биолози , така и да е интегрируем в големи биоинформатични платформи.

4. Представянето на приносите и изводите може да се оптимизира както в контекста на поставените цели и задачи на изследването, така и да се облекчи като фразеология.

10. Заключение

Като заключение на моето становище относно представения за рецензиране дисертационен труд на докторант Милко Крачунов заявявам, че той представлява една задълбочена, оригинална и иновативна работа, която съответства на всички административни и законови изисквания и разпоредби. Въз основа на това, както и на постоянната ми съвместна работа и наблюдение върху развитието на докторанта , мога с пълна увереност и чистосърдечно да предложа на почитаемото научно жури да присъди на Милко Красномиров Крачунов образователната и научна степен „доктор“ по по направление 4.6 "Информатика и компютърни науки", научна специалност 01.01.12 "Информатика" (Изкуствен интелект)

19.03.2015 год.
гр. София

доц. д-р Димитър Ив. Василев