

Становище

от

академик Атанас Ив. Атанасов

управител на Съвместен Геномен Център
при Софийски Университет "Св. Климент Охридски"
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)

Със заповед № РД38-122/06.02.2015 на ректора на Софийски 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 дисертанта, 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. Климент Охридски".

Представеният дисертационен труд е съобразен с основните изисквания според ЗРАСРБ, ППЗ и правилника на ФМИ, като се състои от увод (глава 1), четири тематични глави, приноси и перспективи (глава 5), библиография, 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, заключаваща се 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 се отнасят принципно към дисциплината "Биоинформатика" и са основа на настоящия дисертационен труд, включващ анализ на данни от ново поколение геномно секвениране във всичките си измерения, като претърсване на данни, откриване и корекция на грешки, съпоставяне на секвенции. Като иновативен подход или група от подходи, кандидатът е използвал в дисертацията си различни класификационни методи, и такива на изкуствен интелект за получаване на по-точни резултати за отличаването на наличните грешки в секвенционни информация. Това е особено важен аспект в

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. Тя 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 е написана на 166 страници, състои се от 6 глави с увод и заключение, съдържа 16 фигури и 15 таблици, използвани са 170 литературни източника. Съставен е списък на използваните съкращения, както и много съдържателен и добре преведен речник на използваната терминология.

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

Биоинформатика е област от науката, която съчетава в себе си нови технологии, както и от областта на съвременните молекулярни и био-технологии, така и в голямата си част почти всичко ново свързано от областта на съвременните информационни технологии. Без нея е невъзможно развитието на съвременните омикс технологии. Предоставената за анализ и изследване информация, съдържа в себе си много възможности и потенциал за използване на методи на изкуствен интелект, класификационен анализ, теория на информатиката, статистиката, които от своя страна предлагат различни възможности за оптимизации и получаване на иновативни и оригинални решения. В дисертацията на Милко Крачунов е направно не малко за развитие на иновативни и оригинални идеи, които несъмнено са способствали за генерирането на нови знания.

Представеният за рецензиране дисертационен труд е насочен основно към методи за автоматизиран анализ и класификация на данни от метагеномно секвениране, като основна цел на работата според формулиривката на докторант Милко Крачунов 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 аналитични подходи за откриване на грешките;

Класифициране на оценките на грешките;

Валидация на методите;

Усъвършенстване на предложените аналитични подходи с използване на методи на изкуствен интелект;

Създаване на софтуерни програми за автоматизирано изпълнение на разгледаните процеси.

Всички изложени задачи са решени с оригинални предложения на докторанта и публикувани в реферирани списания и материали от научни конференции.

2. Познание състоянието на проблема от страна на дисертанта

Няма съмнение, че докторант Милко Крачунов е навлязъл много добре в тази нова научна проблематика. Тематиката е нова. Списъкът на цитираните литературни източници е актуален – преобладават публикации от последните 10 години, те са от списания имащи много широко разпространение и цитируемост в изследванията и разработките по биоинформатика в световен мащаб.

Всичко това е много добре представено в глава 2 на дисертацията, където са изложени за пръв път някои важни проблеми и обяснения от теория на биоинформатика като алгоритми за секвениране, алгоритми за класификация, методи за анализ на метагеномни 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. Авторът предлага метод за обработка на секвенционни данни с цел откриване на грешки. Представен е в

няколко варианта на работа аналитичен метод за откриване и корекция на грешки в данни от паралелно метагеномно секвениране. Този метод е усъвършенстван спомощта на методи на изкуствения интелект, които видимо са подобрили резултатите от откриването и корекцията на грешки. Валидирането на използваните методи е разработено с помощта на индикатор за оценка на грешките основан на методи на размити множества. Класификационният анализ на грешките до голяма степен се основаа на използването на методи на изкуствения интелект. Постигната е оптимизация 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 по дисертацията. Представените графики и таблици много добре илюстрират разглежданите задачи и постигнатите резултати. Речникът на използваните съкращения е един от най-добрите, които са правени у нас на биоинформатични теми.

Глава първа е уводна в предметната област на изследването. Определено добре е формулирана представената мотивация за избор на проблема и обекта на изследването. Много добре е направено описанието на използваните в дисертацията методи. Целите и задачите на работата са добре формулирани.

В Глава 2 е направен критичен литературен обзор на изследваните проблеми. Много добре са разгледани съществуващите решения и възможности за алтернативни методи.

Постановката на задачата и начините за решение на поставените цели и задачи са преставени в Глава 3. Всички методи и алгоритмични решения са описани много добре и създават впечатление на неоспорима иновативност.

Софтуерната реализация на разработените в дисертацията методи е представена в Глава 4“ резултатите от използването им потвърждават доброто им качество

Дискусията и изводите по получените резултати са направени в Глава 5. Прави добро впечатление добре представената сравнителна характеристика на резултатите от разработените методи и софтуерната им реализация. От особен интерес е потенциалното развитие на тези методи за изследване и дешифриране на сложния геном на хексаплоидната пшеница.

Приносите, перспективите, оригиналността на изследванията в дисертацията, наред с публикационните им характеристики, оригиналността им, и потенциала за бъдещо развитие са преставени подробно в Глава 6, която приемам напълно.

5. Значимост на разработката за науката и практиката.

Представената в дисертационния труд работа по обем, задълбоченост, иновативност и приложимост открива възможности не само за усъвършенстване на методите, които е използвал кандидата, но и тяхното реално приложение.

6. Преценка на публикациите по дисертацията.

Докторантът има общо над 20 публикации. Във връзка с дисертацията, той е представил 6 от тях, които са публикувани в реферирани издания. Има две статии в международни списания, едната от които с импакт фактор IF 1.567. Кандидатът е участвал в 4 конференции с доклади, и в 4 конференции с постери. Няма данни за цитирания. 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 десейнот

Докторантът не е представил информация за за личното си участие в колективните публикации, но от дългогодишната работа по темата оставам с впечатлението, че неговото участие е съществено.

8. Автореферат

Авторефератът правилно отразява съдържанието на дисертацията.

9. Критични бележки

Като цяло дисертацията е написана и оформена много добре. Имам някои критични бележки от технически характер:

1. Има определни колебания при изпозлването на пълен и кратък член на частите на изречението.
2. Има не навсякъде правилно построяване на изречението - но това може би е свързано с начина на представяне.
3. Софтуерната реализация би могла да се представи в по-опростен вид, който да е достъпен за работа с тези програми и на не много посветени в ИТ биолози, химици и др.
4. В последната глава (Глава 6) според мен би могло в известна степен лескически да се коригира и обобща представянето на приносите и изводите

10. Заключение

В представената от него работа, докторантът демонстрира, че е изпълнил 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.

Представените от мен критични бележки се отнасят предимно до техническото оформление на дисертационния труд и в никакъв случай не намаляват много добрата научна стойност на работата

Всичко това ми дава основание да дам положителна оценка на дисертационния труд, с което предлагам 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)

17.03.2015 год.

гр. София

Акад. Атанас Ив. Атансов