УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ
Бр. 1038/3-22
13.01. 20 23 год.
Београд, Студентски трг 16
ТЕЛ. 20 27 801, ФАКС: 26 30 151

Универзитет у Београду
Математички факултет
Наставно–научном већу

Извештај Комисије за оцену научне заснованости теме докторске дисертације кандидаткиње Биљане Стојановић

Одлуком Наставно-научног већа Математичког факултета Универзитета у Београду, донетој на седници одржаној 23.12.2022. године (број 1038/2), именовани смо за чланове Комисије за оцену научне заснованости теме докторске дисертације „Примена метода истраживања података на употребу кодона ради класификације протеинских секвенци и одређивања међусобне сличности вируса" кандидаткиње Биљане Стојановић.

Кандидаткиња Биљана Стојановић је 12.01.2023. године, у оквиру Семинара Катедре за рачунарство и информатику, пред члановима Комисије одржала усмено излагање и одбрану теме дисертације. Након одржаног излагања је вођена дискусија кроз коју су чланови Комисије и други чланови Катедре изнели своја запажања и сугестије за даљи рад. Комисија подноси Наставно-научном већу Математичког факултета следећи извештај.

1. Подаци о кандидату

Кандидаткиња Биљана Стојановић је рођена 6. семтембра 1982. године у Лесковцу. Гимназију у Лесковцу је завршила 2001. године, на природно-математичком смеру, као носилац Вукове дипломе. Исте године је уписала основне академске студије на студијском програму Математика, усмерење Рачунарство и информатика, на Математичком факултету Универзитета у Београду. Дипломирала је 2007. године са просечном оценом 9,61.

Докторске студије на Математичком факултету Универзитета у Београду, студијски програм Информатика, уписала је у школској 2009/10. години. Након краће паузе у студирању, поново је уписала докторске студије у шк. 2017/18. години. Положила је све предмете предвиђене студијским програмом са просечном оценом 10.

Од октобра 2007. године је била запослена на Математичком факултету Универзитета у Београду у звању сарадника у настави за ужу научну област Рачунарство и информатика. Након уписа на докторске студије, 2009. године је изабрана у звање асистента. Током свог рада на Математичком факултету држала је вежбе из предмета: Програмирање 1, Програмирање 2, Увод у организацију рачунара, Објектно-оријентисано програмирање, Увод у организацију и архитектуру рачунара 1, Трансакциона обрада, Дистрибуиране и објектне базе података.

Од 2020. године запослена је на Математичком институту Српске академије наука и уметности у звању истраживача приправника.

Учествовала је као истраживач на научним пројектима Министарства просвете, науке и технолошког развоја Републике Србије: Математички модели и методе оптимизације са применама (бр.144007, 2009-2010) и на потпројекту Примене ИКТ у дигитализацији и научно-историјско-архивском третману културног наслеђа (бр.ИИИ-044006, 2012-2019).

Коаутор је 9 техничких решења везаних за развој и унапређење техничких решења Дигиталних легата (http://legati.matf.bg.ac.rs/) и Виртуалне библиотеке Математичког факултета Универзитета у Београду (http://elibrary.matf.bg.ac.rs/). Члан је техничког уредништва националног часописа „Преглед националног центра за дигитализацију". Била је члан организационих одбора више међународних и националних научних скупова.

Главна област научног интересовања је истраживање података са применама у биоинформатици.

Списак научних радова

Радови у часописима националног значаја

1. Maljković M., Stojanović B., Mijajlović Ž.: *Digital Legacy of Mihailo Petrović Alas*. Review of the National Center for Digitization, Faculty of Mathematics, University of Belgrade, 31, pp. 10 – 17, 2017.

2. Maljković M., Stojanović B., Mijajlović Ž.: *Digital Legacy of professor Slaviša Prešić*. Review of the National Center for Digitization, Faculty of Mathematics, University of Belgrade, 30, pp. 1 – 6, 2017.

3. Stojanović B.: *Digitizing the Archive of Student Services of the Faculty of Mathematics*. Review of the National Center for Digitization, Faculty of Mathematics, University of Belgrade, 29, pp. 33 – 41, 2016.

Саопштења са међународних скупова

1. Mitić N., Pavlović-Lažetić G., Beljanski M., Malkov S., Maljković M., Stojanović B., Veljković A., Kapunac S.: *Large scale clustering in structural and evolutionary analysis of SARS-CoV-2 proteins*. Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022), The 13th International Multiconference, 04–08 July, 2022, Novosibirsk, Russia, pp. 888-889, DOI 10.18699/SBB-2021-000, 2022.

2. Stojanović B., Veljković A., Malkov S., Beljanski M., Pavlović-Lažetić G., Mitić N.: *Codon usage polymorphism in SARS-2 CoV protein coding sequences*. Biologia Serbica, Book of Abstracts, Belgrade BioInformatics Conference 2021, Department of Biology and Ecology, Faculty of Sciences, University of Novi Sad, Serbia, 43(1), pp. 97 – 97, 2021.

3. Veljković A., Stojanović B., Malkov S., Beljanski M., Pavlović-Lažetić G., Mitić N.: *Codon Usage-based SARS-CoV-2 protein classification*. Biologia Serbica, Book of Abstracts, Belgrade BioInformatics Conference 2021, Department of Biology and Ecology, Faculty of Sciences, University of Novi Sad, Serbia, 43(1), pp. 92 – 92, 2021.

Саопштења са скупова националног значаја

1. Mitić N., Malkov S., Stojanović B., Maljković M.: *Nacionalni centar za digitalizaciju - prošlost, sadašnjost i budućnost*. Book of Abstracts: XVII National Conference Digitization of Cultural Heritage, Old Records from the Natural and Social Sciences and Digital Humanities, Belgrade, Serbia, pp. 6 – 6, 2019.

4. Mitić N., Malkov S., Stojanović B., Maljković M.: *Digitalizacija sabranih dela Nedeljka Parezanovića*. Book of Abstracts: XVI National Conference Digitization of Cultural Heritage, Old Records from the Natural and Social Sciences and Digital Humanities, Belgrade, Serbia, pp. 5 – 5, 2018.
5. Maljković M., Stojanović B., Mijajlović, Ž.: *Digitalni Legat Mihaila Petrovića Alasa*. Book of Abstracts: XV National Conference Digitization of Cultural Heritage, Old Records from the Natural and Social Sciences and Digital Humanities, Belgrade, Serbia, pp. 21 – 21, 2017.
6. Stojanović B., Mijajlović Ž., Maljković M.: *Digitalni Legat profesora Slaviše Prešića*. Book of Abstracts: XIV National Conference, Digitization of Cultural Heritage, Old Records from the Natural and Social Sciences and Digital Humanities, Belgrade, Serbia, pp. 5 – 5, 2016.
7. Mitić N., Malkov S., Vukša N., Stojanović B.: *Digitalizacija arhive studentske službe Matematičkog fakulteta u Beogradu*, XIII Conference Digitization of Cultural Heritage and Digital Humanities, Belgrade, Serbia, pp. 7 – 8, 2015.

2. Предмет и садржај докторске дисертације

Научна област из које је тема

Предложена тема припада научној области Рачунарство и информатика, а ужој научној области Биоинформатика.

Предмет истраживања и докторске дисертације

Предмет истраживања је изучавање могућности примене метода истраживања података на податке о употреби кодона и разматрање могућности дефинисања мере за одређивање међусобне сличности и класификовање вируса или појединих вирусних протеинских секвенци.

Генетски код почива на употреби 64 триплета нуклеотида (кодона) за кодирање 20 различитих аминокиселина (АК) од којих се састоје протеинске секвенце. Већина АК се кодира са више различитих кодона. Заступљеност различитих кодона који кодирају исте АК се разликује како у различитим протеинским секвенцама, тако и у различитим организмима. Иако се разматрање заступљености различитих кодона обично означава термином *употреба кодона*, заправо постоји већи број различитих мера, тј. метода да се изрази употреба кодона.

Ово истраживање се бави првенствено протеинским секвенцама РНК вируса. Вируси су инфективни агенси који се репликују само унутар живих ћелија сва три суперкраљевства (Архее, Бактерије, Еукарие). Пошто не поседују све одлике живих организама, вируси се дефинишу као репликатори, или (квази)/организми, на граници живог-неживог. У раду ће бити анализирано више фамилија вируса, при чему ће посебна пажња бити поклоњена вирусима из фамилије коронавируса. Биће анализиране постојеће методе за карактеризацију употребе кодона.

Процењује се да ће материјал обухватити између 750.000 и милион вирусних нуклеотидних секвенци и између 10 и 20 милиона протеинских секвенци. Такав обим материјала захтева примену метода истраживања података. Сви подаци ће бити прикупљени из јавно доступних база биоинформатичких података, пре свега из базе података Националног центра за биотехнолошке информације (*National Center for Biotechnology Information – NCBI*).

3. Научни циљ докторске дисертације

Основни циљ истраживања је разматрање и утврђивање могућности и метода за аутоматизацију одређивања сличности и класификације вирусних протеинских секвенци, па и читавих вирусних протеома.

Из расположиве литературе и до сада познатих резултата види се да су истраживања до сада рађена али на мањим узорцима, обично са фокусом на појединачне протеине и мере употребе кодона или само поједине фамилије вируса. Овде је план да се анализа изврши на великом узорку, уз употребу различитих мера употребе кодона и примену различитих метода истраживања података. Такође, биће испитивано и постојање мутација у сличним протеинима вируса и утицај тих мутација на евентуалну промену карактеристика протеина или организма.

Очекује се да добијени рачунарски модели допринесу бољем разумевању карактеристика вируса, предвиђању сродних региона у вирусима, као и бољем резумевању утицаја мутација на нуклеотидном нивоу на промену карактеристика самог вируса.

4. Основне претпоставке и претходни резултати

Основне хипотезе од којих се полази у овом раду су:

- Употреба кодона је специфична за појединачни протеин у оквиру вируса. Такође, употреба кодона за исте протеине/протеине са сличном функцијом

у различитим или истим фамилијама вируса може да се разликује у зависности од (под)врсте вируса. Ова различитост може да се односи и на регионе протеина, односно на његове подсеквенце.

- На основу употребе кодона може да се формира карактеристични профил секвенце који може да се представи вектором вредности сваког кодона. Могуће је формирати више различитих вектора, у зависности од мере која се користи за изражавање употребе кодона.
- Добијени профили (вектори вредности) могу да се искористе као материјал за прављење класификационих модела помоћу којих би се одредила припадност (а самим тим и оквирна функционалност) улазне (непознате) ниске.
- Кластер анализом материјала може да се одреди хоризонтална и вертикална блискост појединих вируса или њихових делова.
- У случају мутација, промењени генетски материјал задржава употребу кодона из оригиналног вируса.

Претходни резултати у области

1. Tan PN, Steinbach M, Karpatne A, Kumar V.: *Introduction to Data Mining*. 2nd ed. Pearson Education, 2019.
2. Aggarwal C. (eds.): *Data Classification Algorithms and Applications*. Chapman and Hall/CRC, 2014.
3. Chen JY, Lonardi S (eds.): *Biological Data Mining*. Chapman and Hall/CRC, 2017.
4. Abu-Jamous B, Fa R, Nandi AK: *Integrative Cluster Analysis in Bioinformatics.* John Wiley & Sons, Ltd, 2015.
5. Giovanni S, Elder JF: *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions.* Morgan & Claypool, 2010.
6. Xu R, Wunsch DC: *Clustering*. IEEE Press, 2009.
7. Aggarwal C.: *Machine Learning for Text.* Springer International Publishing, 2018.
8. NCBI вирусне секвенце, https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/
9. Dmitrij Frishman and Manja Marz, (Eds.), *Virus Bioinformatics*, Chapman and Hall/CRC, 2021, ISBN: 978-1-003-09767-9
10. Jane Flint, Vincent R. Racaniello, Glenn F. Rall, Theodora Hatziioannou, Anna Marie Skalka *Principles of Virology*, ASM Press, 2020, ISBN: 978-1-683-67358-3

11. Malik YS, Ansari MI, Kattoor JJ, Kaushik R, Sircar S, Subbaiyan A, Tiwari R, Dhama K, Ghosh S, Tomar S, Zhang KYJ. *Evolutionary and codon usage preference insights into spike glycoprotein of SARS-CoV-2*. Briefings in Bioinformatics. 2021;22(2):1006–1022. doi: 10.1093/bib/bbaa383.

12. Sharp PM, Li W-H (1987) *The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications*. Nucleic Acids Res 15(3):1281–1295

13. Sun X., Yang Q., Xia X. *An improved implementation of effective number of codons (nc).* Mol Biol Evol. 2013;30(1):191-6. doi: 10.1093/molbev/mss201.

5. Научни методи који ће бити коришћени у раду

Основне научне методе које ће бити коришћене у истраживању засноване су на моделима и алгоритмима у области истраживања података.

Припрема података подразумева да се за сваку појединачну посматрану протеинску секвенцу одреди вектор који описује употребу кодона у секвенци. Затим се на тако припремљене податке примењују различити методи истраживања података (класификације, кластеровања и правила придруживања). Модели ће бити засновани на одређивању правилности појављивања континуалних или дисконтинуалних ниски кодона у зависности од карактеристика протеина/подсеквенци које се обрађују.

Квалитет добијених резултата биће приказан у зависности од примењених математичких модела и мера који се користе у дефинисаним моделима. На основу резултата поређења ће бити процењени доприноси новодефинисаног приступа.

6. Закључак и предлог

Истраживачка област, која је предмет ове докторске дисертације, је релевантна и актуелна. Значај планираног истраживања се види из великог броја радова који се баве употребом кодона, али и из чињенице да до сада није било резултата овако широких и темељних истраживања.

Кандидат је у претходном периоду самостално или у сарадњи са ментором и другим колегама учествовао у истраживањима која су посвећена анализи различитих биоинформатичких база података, при чему је посебна пажња била посвећивана управо употреби кодона.

На основу свега наведеног, Комисија констатује да је тема докторске дисертације „**Примена метода истраживања података на употребу кодона ради класификације протеинских секвенци и одређивања међусобне сличности вируса**" кандидаткиње **Биљане Стојановић** научно заснована и предлаже Наставно-научном већу Математичког факултета да донесе позитивну одлуку о прихватању теме и одреди за ментора **проф. др Ненад Митића**.

У Београду, 13.01.2023. године

Прилози:

1. Подаци о предложеном ментору

_____

др Саша Малков, председник комисије

ванредни професор Математичког факултета Универзитета у Београду

_____

др Марија Гавровић-Јанкуловић, члан комисије

редовни професор Хемијског факултета Универзитета у Београду

_____

др Мирјана Маљковић, члан комисије

доцент Математичког факултета Универзитета у Београду