

Резюме на трудовете
на доц. д-р Борислав Петров Юруков
(относно конкурс за професор)

Разработките на кандидата са представени в: 1 обзорец труд под общата редакция на кандидата; 11 научни публикации в рецензирани списания и сборници на български (една) и английски (десет) език, от които 10 в международни издания (четири с импакт фактор) и 7 абстракта в рецензирани сборници на международни научни конференции.

ОБЗОРЕН ТРУД

1. Нина Синягина, **Борислав Юруков**, Гергана Калпачка, Людмила Танева, Стела Русева, Павел Джунев. *Отказоустойчиви компютърни системи (Fault-tolerance Computer Systems)*. Университетско издателство „Неофит Рилски“, Благоевград, ISBN 978-954-00-0089-3, 2016. 272 страници. Под общата редакция на **Борислав Юруков**.

Съществуват два основни подхода за осигуряване на висока надеждност на компютърните комплекси и мрежи: предпазване от откази и повишаване на отказоустойчивостта. Предпазването от откази предвижда редица процедури за откриване и отстраняване на повреди и грешки в апаратната и програмната части на системата и избягване от въвеждането на нови такива. Отказоустойчивостта, от друга страна, гарантира коректно изпълнение на потребителските задачи при поява на грешки или повреди в съответната система, което се реализира чрез допълнително въвеждане на излишък (времеви, схемен, структурен или апаратен).

Тъй като моментът за появяването на отказ, времето между два отказа, а също така и времето за възстановяване на системата са случайни величини, то и показателите за надеждност се определят като вероятностни показатели. Основна количествена характеристика за надеждността е т. нар. функция на надеждността $P(t)$. Според приетата дефиниция надеждността на една компютърна система (КС) се определя като вероятността за безотказна работа на системата, т.е. вероятността да не възникне отказ при зададен режим на условия на експлоатация в границите на определена продължителност на работа на системата. Математическото моделиране на системите за отказоустойчивост използва основно теорията на вероятностите. Описани са техники за оценка на отказоустойчивостта,

като Марковските и Поасоновите процеси, които успешно могат да бъдат приложени и в много области на биоинформатиката, като например лиганд-таргет взаимодействията при лекарствения дизайн.

ПУБЛИКАЦИИ

2. Apostolov P., **Yurukov B.** Stefanov A., A Study of Spatial Filter Banks with Luneburg lens. IET Microwaves, Antennas & Propagation, 2018, (Submitted) (Impact factor: 1.187).

В публикацията се обсъжда приложимостта на антена за комуникации от пето поколение. Дефинирани са изисквания и са представени параметрите на антената. Използвайки Matlab е реализирана симулация в тримерното пространство на 37-лъчева антена с векторна диаграма, която обхваща цялата азимутална полуравнина. Приносът се изразява в постигането на възможността в разработения теоретичен модел да се реализира излъчване и приемане на електромагнитни вълни по 37 лъча в пространството от една единствена антена за технологията 5 G. Разработената теория може сравнително лесно да бъде приложена за практически реализации.

3. Valev V., Yanev N., Krzyżak A., Yurukov B. A New Geometrical Approach for Solving the Supervised Pattern Recognition Problem, 24th International Conference on Pattern Recognition (ICPR), Beijing, China, August 2018.

Задачата за класификация в термините на machine learning, pattern recognition се състои в разпознаване на класа, към който ново наблюдение принадлежи, на базата на обучаващо множество от наблюдения (instances) с известна класова принадлежност. Математическата дефиниция е: апроксимация на функция $f: X \rightarrow Y$, където X е множество от instances, а Y е дискретно множество. Предложен е алгоритъм, основаващ се на нова задача на геометричната оптимизация: дадено множество от оцветени в различен цвят (най-малко два) точки да се покрият с минимален брой едноцветни правоъгълници (в двумерния случай) или хипер-правоъгълници (в n -мерния случай). Предложена е оригинална трансформация на тази задача в задача за минимално кликово покритие на съответен граф. Класификационният алгоритъм се основава на намиране на най-близката клика до зададено ново наблюдение и

класифицирането му в нейния цвят. Приложените числени експерименти демонстрират конкурентостта на алгоритъма с най-известния (използвания) SVM (Support Vector Machine). В следващата статия е показана и доминация по отношение на бързодействието в случаите на много класове (броя на класовете (цветовете) по-голям от 2).

4. Yanev N., Traykov M., Milanov P., **Yurukov B.** A new classifier for protein fold class recognition, C. R. Acad. Bulg. Sci., 2018, (Under review) (impact factor: 0.25).

Класифицирането на нагъванията на протеините е важен подход за откриване на структури на протеини без да се разчита на подобност в аминокиселинната последователност. Това все още е предизвикателство при проблема за разпознаване на модели. Всички описани в литературата подходи се опитват да покажат превъзходството на SVM над други класификатори. SVM използват методи за машинно самообучение, които от аминокиселинната последователност генерират характеристичен вектор. Този характеристичен вектор може да се базира на аминокиселина последователност, вторичната структура на протеините, хидрофобност, полярност и др.

В тази публикация е използван подхода за генериране на характеристичен вектор, а именно преобразуване на даден стринг над 20-буквена азбука (протеиновата аминокиселина последователност) в 20^k вектор с компоненти - относителна честота на поднизове с дължина k . Предложен е класификатор, базиран на обобщение на диаграми на Вороной, в които точките на Вороной са хипер-правоъгълници. Показано е, че те са изпъкнали обвивки на едноцветни точки в Manhattan метриката. Класификационният алгоритъм строи минимален (или почти минимален) брой такива обвивки, разстоянието до които се използва за класифициране на новите наблюдения (instances). Числените експерименти върху данни, които се използват в над 100 публикации по тематиката, демонстрират по-добра точност на класификация в сравнение с всички цитирани класификатори, като доминацията в бързодействие е значителна (за размерност 20 на feature space и брой класове 27 посоченото време е повече от 2 дни при 5 мин. за новия класификатор). Времева сложност на предлагания класификатор позволява да се получат резултати за размерност $20^3=8000$. Това е и основното препятствие пред съществуващите класификатори (включително и SVM), които са бързодействащи само за бинарна класификация.

5. Traykov M., Yanev N., Mavrevski R., **Yurukov B.** Algorithm for protein folding problem in 3D lattice HP model. 20th International Conference on MATHEMATICAL and COMPUTATIONAL METHODS in SCIENCE and ENGINEERING (MACMESE '18), Florence, Italy, March 24-26, 2018 (Accepted Full Papers).

Предсказването на третичната структура на даден протеин по неговата аминокиселинна последователност от дълго време е важен и предизвикателен обект на изследвания в различни научни области. Този проблем възниква от наблюдението, че взаимодействия между хидрофобните (H) аминокиселините са главната сила, която предизвиква сгъването на протеините в тяхното природно състояние. Основен инструмент за предсказване на пространствената структура на протеините е решетъчният и извън-решетъчният Hydrophobic-Polar (HP) модел. HP моделът за нагъването на протеини в решетка се основава на наблюдението, че в полярна среда, протеините се нагъват по начин, при който в ядрото на полученото нагъване има повече хидрофобни аминокиселини, образувайки контакти (H-H контакти) между тях, а повечето полярни (P) аминокиселини са разположени по периферията на нагъването в контакт с полярната среда.

Проблемът за нагъване на протеини в 3D решетъчен HP модел е NP-пълна задача. В тази публикация е предложено разширение на евристичен алгоритъм за решаване на проблема в 2D решетъчен HP модел, който е докладван на International Congress on Mathematics (MICOM), Атина, Септември, 2015 [16] и е описан в публикация от Traykov, Angelov и Yanev (Traykov, M., Angelov, S., Yanev, N., 2016. A new heuristic algorithm for protein folding in the hp model. *Journal of computational biology*, vol. 23(8), pp. 662-668). Разширеният евристичен алгоритъм решава проблема за нагъване на протеини в 3D решетъчен HP модел. За изчислителните експерименти са използвани HP протеинови последователности, които са известни в литературата бенчмарк последователности за 3D решетъчни HP модели. Получените резултати са сравнени с резултати, получени от алгоритми за решаване на 3D проблема, такива като алгоритми от генетичен тип, оптимизационния алгоритъм на мравките и алгоритъм тип Монте Карло.

6. Stoilov A., **Yurukov B.**, Milanov P. Analysis of docking algorithms by HPC methods generated in bioinformatics studies. ITM Web of Conferences, 2018, 16, <https://doi.org/10.1051/itmconf/20181602009>.

Свързването на малки молекулни лиганди с големи рецептори е от основно значение за множество биологични процеси. Точното прогнозиране на начина на свързване между лигандата и рецептора (проблема с докинга) е от основно значение за съвременния структурен дизайн на лекарства. В тази публикация е направен анализ на алгоритми за докинг чрез HPC методи, генерирани в биоинформатични изследвания. За целта е използвано бенчмарк множество разработено от Cambridge Crystallographic Data Center (CCDC) за докинг софтуера GOLD.

7. Todorin I., Yanev N., Traykov M., **Yurukov B.** A new off-lattice HP model with side-chains for protein folding problem. ITM Web of Conferences, 2018, 16, <https://doi.org/10.1051/itmconf/20181602007>.

Тук е разгледан нов модел за решаване на проблема за нагъване на протеини в 3D извън-решетъчен HP модел. Моделът представя белтъчната молекула чрез алфа въглеродния атом и центъра на страничната верига на всяка аминокиселина, като разстоянията между съседните елементи могат да варират в малки отклонения, за да се използва нов метод за целенасочено нагъване, съобразно физичните принципи за постигане на енергиен минимум. Моделът отчита различния размер на радиусите на страничните вериги, (радикали) и се състои от две стъпки, а именно генериране на начална случайна структура и последващ целенасочен процес на нагъване на протеини според физическите сили и енергии. Числените експерименти демонстрират адекватността на модела и ефективността на алгоритъма.

8. Apostolov P., **Yurukov B.** Stefanov A., An Easy and Efficient Method for Synthesizing Two-Dimensional Finite Impulse Response Filters with Improved Selectivity. IEEE Signal Processing Magazine, 2017, 34(5), 180-183 (Impact factor: 9.654).

Много съвременни технически устройства работят с двуизмерни (2D) сигнали в процес, наречен цифрова обработка на изображения. 2D цифровите филтри с крайно импулсно отражение (FIR) са основни технически средства за обработка на изображения. Също така, FIR филтрите се използват широко в цифровата телевизия, радиоастрономията, радиото, биомедицината и т.н. В тази публикация е разработен лесен и ефективен метод за синтезиране на двумерни филтри с крайно импулсно отражение с подобрена селективност. Като резултат, филтърът е с много прост дизайн и по-голямо бързодействие. Резултатите от разработеният алгоритъм и самият метод са публикувани в списание IEEE Signal Processing Magazine (Impact factor: 9.654).

9. Yanev N., Traykov M., Milanov P., **Yurukov B.** Protein Folding Prediction in a Cubic Lattice in Hydrophobic-Polar Model, Journal of Computational Biology, 2017, 24(5), 412-421 (impact factor: 1.032).

В тази публикация е представен нов математически модел (оптимизационна задача за максимизиране на Н-Н контактите) за предсказване на третичната структура на протеините, основаващ се на решетъчен HP модел. Предложени са един точен и два евристични алгоритми за решаване на проблема за нагъване на протеини в тримерното пространство. Изчислителните експерименти са проведени с известни в литературата бенчмарк последователности за 3D решетъчни HP модели. Получените резултати подобряват описаните в литературата резултати за някои от тестовите последователности в сравнение с други известни алгоритми.

10. Krlev V., Krleva R., **Yurukov B.** An event grouping based algorithm for University course timetabling problem. International Journal of Computer Science and Information Security, 2016, 14(6), 222-229.

Представен е оптимизационен алгоритъм, базиран на групиране на събития за решаване на проблема за учебния разпис на университета (NP-hard problem). Разпределянето на събития в групи с равен брой събития във всяка група, не е приложимо за всички входни множества от данни. По тази причина в тази публикация е предложен универсален подход,

който е приложим към всички възможни групи от събития, съизмерими по размер. Също така е представено прилагане на алгоритъм, базиран на този подход. Описани са методологията, условията и целите на експеримента. Експерименталните резултати са анализирани и са направени заключенията от тях. Представени са бъдещите насоки за по-нататъшни изследвания.

11. Stoilov A., **Yurukov B.** An estimate of the short-term local climatic changes by measuring the air temperature, air pressure and humidity, Proceedings of the Third International Conference on Advances in Bio-Informatics and Environmental Engineering - ICABEE 2015. Rome, Italy 2015, 18-21.

Тази публикация представя оценка на местните климатични промени в Югозападна България, особено в Благоевград. В това проучване са извършени изследвания и измерване на климатични параметри като температура на въздуха, влажност и налягане на въздуха. Анализът е направен за месец юли през периода 2011-2015 г. Метеорологичната станция, използвана за научните изследвания, е WS-2355. Климатичните промени се анализират и дават оценка на изменението на климата в периода 2011-2015 г. Също така, получените резултати са анализирани с цел откриване на нови тенденции и идентифициране на зависимости за предсказване на следващите климатични параметри. Данните, получени от измерванията, не потвърждават наличието на глобално затопляне на времето в краткосрочен план за Югозападна България. Относително малкото налична информация за такива промени все още не може да даде недвусмислена оценка на изменението на климата, но събирането на местните данни е стъпка към решаването на този научен проблем.

12. **Yurukov B.**, Aleksov S., Dalgacheva V., Sinyagina N. Developing of an input data stream to simulate the dynamics of forest fires. Втора научна конференция с международно участие "Компютърни науки и технологии", Варна, България, 1, 2014, 107-113.

Дадени са насоки за създаване и актуализация на необходимите данни за топографията, горимите материали и метеорологичните условия за симулация на горските пожари. Софтуерът на ArcGIS ArcMap представлява инструмент, използван за подготовка на

данни за получаване на формат използван от модела Fire Area Simulator (FARSITE) и за формиране на (LCP) файла. Предложени са сайтове за събиране на геопространствени данни и друга необходима информация. Разгледани са особеностите на необходимите параметри за модел на динамиката на горските пожари, характерни за Българските условия. Моделите за симулиране на горски пожари могат да бъдат от голяма полза на ръководителите и екипите за борба с огъня при вземането на решения, за управлението и контрола над пожара. В публикацията са предложени инструменти, подпомагащи екипите при борба с горските пожари.

АБСТРАКТИ НА ДОКЛАДИ В КОНФЕРЕНЦИИ

13. Stoilov A., **Yurukov B.** Bioinformatics measurements with High performance computing. (abstract) Biomath Communications - Featuring International Conference BIOMATH2016, 2016 3(1), pp. 62.

Взаимодействията между рецептора и лигандата са с квантов механичен характер, но поради сложността на биологичните системи квантовата теория не може да бъде приложена директно. Следователно, повечето методи, използвани при докинг и изчислително откриване на лекарства, са по-емпирични по природа и обикновено нямат общоприет характер. Използвайки високопроизводителни изчисления са направени различни измервания и анализи в Биоинформатиката, чрез HPC методи, генерирани в биоинформатични изследвания. За целта е използвано бенчмарк множество разработено от Cambridge Crystallographic Data Center за докинг софтуера GOLD.

14. Mavrevski R., Milanov P., **Yurukov B.** Comparison and assessment of commonly used model selection criteria in modeling of experimental data. (abstract) Biomath Communications - Featuring International Conference BIOMATH2016, 2016, 3(1), pp. 46.

Перспективите за развитие на математически модели в медико-биологичните изследвания и биоинформатиката почиват на използването на информационните технологии. Задачата за избора на модели, които „най-добре“ описват данните е важна част от приложните и научните изследвания в биоинформатиката. Ако се приеме, че съществува

един основен модел („най-добър“), който описва данните, то критериите за избор на модел се опитват да намерят този модел.

В този труд е изследвана ефективността на различни критерии за избор на „най-добър“ модел чрез генерирани експериментални данни и са сравнени два често използвани в литературата критерия за избор на модел: информационен критерий на Акайке (AIC) и информационен критерий на Бейс (BIC). Използвани са различни модели с нарастваща сложност за да се провери дали моделът, използван за генериране на данните, би могъл да бъде правилно идентифициран чрез критериите AIC и BIC и да се оцени и сравни емпирично представянето на двата критерия. Генерирането на изкуствени експериментални данни и фитването (апроксимирането) на кривите е извършено чрез използването на софтуера GraphPad Prism.

15. Milanov P., Trenchev I., Mirchev I., **Yurukov B.** Investigation of “optimal” properties of the contemporary genetic codes according to simple mutation. (abstract) International Congress on Mathematics MICOM, Athens, September, 2015, pp. 17.

Посредством математическо моделиране на процеса на декодиране на информацията на съвременния генетичен код са разгледани и получени резултати относно устойчивостта на генетичния код по отношение на прости мутации и свързаните с тях проблеми. Съвременният генетичен код е важен за правилното декодиране на информацията записана в гените. С решението на оптимизационни задачи с подходящо избрани критерии са изследвани свойствата на генетичния код. Като оптимизационни критерии са използвани вероятността за появата на една аминокиселина в белтък с усреднен брой аминокиселини, вероятностите за появата на прости мутации в първата, втората или третата позиция на кодона. Предложено е експлицитно описание на множеството на всички теоретични генетични кодове като изпъкнал многостен и върховете му са всички теоретични генетични кодове. Това дава възможност да се решават широк кръг от оптимизационни задачи с различни целеви функции и да се разкрият по добре свойствата на съвременния генетичния код. Като допълнителен факт е доказано, че многостенът е сечение на два полиматроида.

16. Yanev N., Milanov P., **Yurukov B.**, Trenchev I., Traykov M. Heuristic algorithm for hp folding in hp model. (abstract) International Congress on Mathematics (MICOM), Athens, September, 2015, pp. 31.

Опитът да се преодолеят съществуващите трудности при решаване на проблема за нагъване на протеини е употребата на опростени модели, като 2D HP решетъчни модели, където протеините се моделират като последователности, чиито елементи са означени като хидрофобни или полярни (хидрофилни) и получената верига се поставя в решетка.

Предложен е евристичен алгоритъм, който може да бъде приложен за решаване на проблем за нагъване на протеини в 2D HP решетъчен модел. Алгоритъмът разглежда последователност от аминокиселини в квадратна решетка. Протеиновата последователност се разделя на сегменти с предварително зададен размер и се генерират всички възможни нагъвания за всеки един от сегментите. Избират се най-добрите нагъвания и се поставят в квадратна решетка. Тази концепция позволява да се достигне до решение за протеини с всякаква дължина.

17. Stoilov A., **Yurukov B.** High performance computing on Bioinformatics. (abstract) International Congress on Mathematics MICOM, Athens, September, 2015, pp. 34.

Всяка докинг програма използва един или повече специфични алгоритми за търсене, които са методите, използвани за прогнозиране на възможните конформации на двоичният комплекс. Тук е представен обзор на текущите техники за докинг и е направено описание на приложението им, включително бенчмарк за докинг на различни платформи за високопроизводителни изчисления, като също така е представен математически алгоритъм. Настоящият референтен показател е направен от съществуващ набор от тестове (CCDC / Astex Validation Set) на типична HPC система. Избрани примери са доквани със софтуера GOLD. Подробно описание на направения обзор и проведените експерименти е представено от Stoilov и Yurukov (2016) в Bulletin of the hellenic mathematical society (<http://bulletin.math.uoc.gr/vol/59/59-99-105.pdf>).

18. Trenchev I., Mavrevski R., Stoilov A., Traykov M., **Yurukov B.** Computer graphics's application in Bioinformatics. (abstract) International Congress on Mathematics MICOM, Athens, September, 2015, pp. 35.

Представено е използването на компютърна графика и математически модели за създаване на 3D обекти в биоинформатиката. Тук са разгледани различни начини за визуализиране на биологични структури, превръщайки файловите формати директно в 3D графики. Също така е представен компютърен софтуер за 3D визуализация и как компютърната графика може да бъде адаптирана към нуждите на биоинформатичните изследвания. Разгледан е също, процеса за представянето на 3D структура на молекула.

19. Mavrevski R., Koroleova G., Pencheva N., Milanov P., **Yurukov B.** The Hausdorff distance as a criterion for the optimal model selection in torque-angle relationships. (abstract) International Congress on Mathematics MICOM, Athens, September, 2015, pp. 45.

Хаусдорфовото разстояние се използва като мярка за близост между две множества от точки. Целите на този труд са: (1) да се провери възможността за използване на Хаусдорфовото разстояние като критерий за избор на оптимален модел за фитване (апроксимиране) на зависимост въртящ момент-ъгъл на лакътните стави и (2) да се сравни ефективността му с други често използвани критерии: AIC и BIC, анализирани в [14]. Представен е ефикасен алгоритъм за изчисляване на Хаусдорфовото разстояние между множеството от точки съдържащо експерименталните данни и множеството от точки получено от изчисления с оптималния фитващ модел в дадените класове. Изчислителните експерименти показват, че Хаусдорфовото разстояние може да бъде използвано като критерий за оценка на "оптимално" фитване с полиноми при този вид зависимости.

доц. д-р Борислав Юруков