

Становище

за дисертационния труд на Веселин Лазаров Шахпазов
на тема „Изследване на модели за прогнозиране на капиталовите пазари
с невронни мрежи” по докторска програма „Информатика”,
професионално направление 4.6. „Информатика и компютърни науки”
от доц. д-р Вася Красиминова Атанасова, секция „Биоинформатика и математическо
моделиране”, Институт по биофизика и биомедицинско инженерство – БАН

Със заповед 262 / 31.10.2019 г. на Директора на ИИКТ-БАН съм определена за член на Научното жури в процедурата за придобиване на образователната и научна степен „доктор” на Веселин Лазаров Шахпазов по професионално направление 4.6. „Информатика и компютърни науки”, докторантска програма по научна специалност 01.01.12 „Информатика”. На 27.11.2019 г. получих по електронен път документите по процедурата, а именно: дисертационен труд, автореферат на български език, автореферат на английски език, автобиография, Справка за изпълнение на минималните изисквания на ИИКТ и копия на 6 публикации. В съкратените срокове, които са били определени на първото заседание на Журито на 04.11.2019, на което поради участие на конференция не можах да присъствам, успях да се запозная само с дисертационния труд и автореферата на български език, както и с публикациите на докторанта, като английския вариант на автореферата не съм прегледала и няма да коментирам.

Впечатленията ми от представените материали са разнопосочни. От една страна, много добро впечатление например правят част от публикациите, особено самостоятелната статия „Високочестотната търговия” на български в научно-информационното списание „Техносфера” от 2016 г. и добре цитираният през годините доклад „Design and Application of Artificial Neural Networks for Predicting the Values of Indexes on the Bulgarian Stock Market” от SPS’2013. От друга страна базираните върху тези публикации дисертационен труд и автореферат оставят впечатлението, че са подготвени прибързано и без необходимото внимание към детайла. Прави впечатление и експерименталната работа с данни от периода 2010-2013 г.

С изключение на неотдавнашния доклад на конференция в Пловдив, който е под печат, всички останали публикации по дисертационния труд са от периода 2013-2016 г. и с изключение на тази в „Техносфера”, която е на български, всички останали са доклади от научни форуми и в съавторство. Наличие на публикация в рецензирано научно списание, поне с SJR ранг, би било още по-силен атестат, но и предоставените данни в Справката за изпълнение на минималните изисквания на ИИКТ за доктор по 4.6. Информатика и компютърни науки за Веселин Шахпазов показват, че той покрива минималните изисквания, предвидени от Правилника на ИИКТ-БАН.

По отношение на дисертационния труд и автореферата, трудно – а и ненужно – е да се изброят всички неточности от терминологичен и фактологичен характер, от които те изобилстват. Ще посоча само част от тях:

- На стр. 8: „генетически алгоритми... „мравешки“ алгоритъм“.
- На стр. 15: „... динамични редове от данни“ – вероятно става дума за времеви редове.
- На стр. 26: „Ерата на Камелот“ е определена от Еберхарт и Добинс в интервала 1890 – 1969, а не до 1926 г, както е написано в дисертационния труд, както е видно дори от абстракта на статията на Еберхарт и Добинс: <https://ieeexplore.ieee.org/document/59207>).
- На стр. 30 се твърди „С времето, изследванията в областта нарастват експоненциално“, което освен фриволно подбран идиом не отговаря и на литературния анализ от стр. 81-82 и на личното мнение на докторанта от стр. 13 за „нишата, в която този труд се е прицелил“ и „малкото, по скромното мнение на автора, изследвания в областта“.
- На стр. 35, 61, 75 и 81 възприетият иначе в българската научна литература превод на термина *feedforward neural network* като „невронна мрежа с право разпространение на сигнала“ се среща под екстравагантните вариации „невронна мрежа с разпространение на информацията направо“, „мрежите с разпространение на информацията напред“, „мрежи с прав поток“.
- На стр. 38 от математическа гледна точка, формула (3) е грешна: първоначално, функцията f е двуаргументна, а след това става едноаргументна. Така, както е записана формула (6) на стр. 40, показва, че λ е функция, а не параметър на функцията, както е обявено.
- На стр. 39 формули (8) и (9) не са консистентни: параметърът α изчезва някъде между двете.
- На стр. 43 интервалът $[-1; 1]$ е разкъсан между редовете по абсолютно недопустим начин.
- На стр. 48 (и на други места) пише: „Невронни мрежи с радиална базирана функция“, а няколко реда по-долу (и на други места) пише: „невронната мрежа с радиална базова функция“. Обичайно срещан и коректен термин е „радиална базисна функция“, както е уцелено само на две места в труда.
- Отново на стр. 48 изречението „Въпреки че мрежата осигурява многовариантен вектор на изхода, който не губи способност да обобщава, описанието на логиката по която действа мрежата с обща регресия е опростена до случай с единствен статистически изход.“ е непонятно.
- На стр. 52 се твърди „Алгоритъмът на най-малките квадрати е предложен от Уидрол и Хоф, като е известен и като правилото Уидрол-Хоф [65] през 1960 г.“ Това е невярно: алгоритъмът на най-малките квадрати *Least Squares* е предложен от Карл-Фридрих Гаус през 1795 г. Въведеното от Уидрол-Хоф понятие *Least Mean Squares* следва да се назовава „метод на най-малките средни квадрати“. Макар и близко звучащи и на английски и на български език двете понятия са натоварени с различен математически смисъл и смесването им е недопустимо.
- На стр. 61 пише „където Y е изхода на невронната мрежа“, но никъде в конкретния текст няма символ Y .

- На много места в текста, например стр. 72, 81, 85, 86 се срещат изразите „спират на цифрата 20“, „където цифрата е 19“, „около цифрата от 350“, „За България, цифрата е около 4%.“ Такава журналистическа и псевдонаучна лексика няма място в дисертационен труд, още по-малко такъв в Научна област 4 „Математика и информатика“.
- На стр. 78: „ИНМ не успяват да надвишат значително способността за прогнозиране на наивните предсказатели“. Без да съм проучвала цитиране [29] предполагам, че става въпрос за наивни Бейсови мрежи / класификатори и техните предиктори.
- На стр. 89-90, от формули (29) и (30) не става ясно как ще се изчислява първият ден (т.е., при $t = 1$).
- На стр. 89, 90, 94, долните индекси в поясненията след формулите са непонятно negliжирани.
- На стр. 94 се срещат „многослойни пепетрирони“ (?!); на други места многослойният перцептрон се разнообразява още като „многослоев“, „многослоест“, „мултислоев“.
- На стр. 107 се споменава „т. нар. [sic!] проклятие на измерението“ вместо утвърдения и лесен за превеждане математически термин „проклятие на размерността“ (*curse of dimensionality*).
- На стр. 111: „Изкуствените невронни мрежи с обща регресия са сравнително нечувствителни към външни стойности (диви точки)“. За термина от статистиката „outliers“ няма възприет добър превод на български, но „външни стойности“ и още по-малко екзотичните „диви точки“ не са подходящи кандидати. Бих предложила варианти като „нетипични стойности“ или „извънредни стойности“.
- Някои неносещи смисъл изречения, като например от стр. 111, 112, 114: „Изкуствената невронна мрежа с обща регресия изпълнява регресия, при която целевата променлива е непрекъсната, която извършва класификация, при която целевата променлива е категорична“, „Подходящата цифра може лесно да бъде избрана чрез експеримент, като се избере число, което води до ниска грешка при подбора,...“, „Най-големите предимства на мрежите с обща регресия са фактът, че резултатът е вероятностен, което прави интерпретацията на изхода лесна и скоростта на обучение“, „...Средно квадратичната грешка. В математиката, терминът се обяснява, като статистическа мярка за големината на променлива величина.“
- На стр. 113, 115, 117 „k – средство (K-Means)“ е очебиен машинен превод.
- На стр. 118, изразът „мрежови архитектури“ в областта на компютърните науки е натоварен с по-различен смисъл.

Видимо, част от терминологичните неточности са резултат от боравенето основно с англоезични източници и използването им в дисертационния труд след (необработен) машинен превод, въпреки наличието на понятиен апарат и достатъчно ресурси по невронни мрежи и на български език. За друга част от грешките вероятното обяснение е недостатъчно доброто познаване от страна на докторанта на онази част от

проблемната област, която се отнася до математическото моделиране с невронни мрежи. Грешките от правописен и пунктуационен характер няма да споменавам, въпреки че те правят не само лошо впечатление предвид нивото, на което следва да е трудът – за най-високата от образователните степени – ОНС „доктор“, но и защото тези елементарни грешки често затрудняват непосредственото разбиране на смисъла на написаното.

Структурирането на дисертационния труд и автореферата е друг аспект, към който могат да се отправят сериозни забележки и препоръки. На практика трудът съдържа около 80 от около 130 страници литературен обзор на модели за прогнозиране на капиталовите пазари с изкуствени невронни мрежи. Действително, намирам за информативно и увлекателно краткото представяне на избраните за анализ 53 научни труда, но не мога да не отбележа, че делът с обзора е несъразмерен спрямо частта с оригиналните авторски резултати, която се губи на фона на огромно количество баластна информация от разнообразен характер. На финала на литературния анализ, т. 2.7.2. на стр. 80-82 г., както текстът на раздела, така и Фигури 7–9 (впрочем преповтарящи информацията от текста в графичен вид и неносещи допълнителна семантика) следваше да съдържат цитиранията към съответните анализирани 53 публикации – но тази слабост е компенсирана от принципно информативната Таблица 5.

При направения литературен анализ някои от изводите на авторите на анализираните публикации, които докторантът е взел наготово, е можело и следвало да бъдат критично оценени с оглед изминалото от тези публикации време и съобразно ревизирани. Визирам например твърдения като това от стр. 74 по отношение цитирания дисертационен труд [49] *„Проучвания на авторите показват, че в България засега не се използват подобни модели. Една от причините за това е ограничения брой специалисти в областта на Изкуствения интелект и недостатъчното познаване на мощните технологии в тази сфера.“* Това твърдение може и да е било в сила към 2008 г., но определено не мисля, че важи и през 2019 г., както от тъй написаното може да се заключи, че дисертантът смята.

Особено хаотична е глава III, конкретно т. 3.4, в която липсва каквато и да е логическа връзка между подточките и последователност на изложението. В добавка, в представянето на понятието експертна система в т. 3.4.3 освен грешки от терминологичен характер (*„районът“*, *„капанът на безкрайна примка“*) е даден и пример в Таблица 12, който е едновременно ненужен за информираните читатели и неинформативен за онези, за които би бил нужен; много повече смисъл от пример би имало, ако, вместо ниски температури, хлъзгав път и зачервени уши, се разглеждаше добре подбрана илюстративна ситуация от капиталовите пазари.

По същество, към предложения в тази глава хибриден модел и резултатите на дисертационния труд е трудно да се формулират конкретни критики. Авторът се е насочил към *„слабо изследвания български пазар“*, но и сам отчита (стр. 98) *„особеностите на българския пазар на акции и по-специално неговата неликвидност, която позволява отделни сделки, не изпълнени на чисто пазарен принцип да влияят върху цялостната стойност и изменение на индекса“*.

Още по отношение на структурирането, смятам, че въвеждането на римска номерация за главите и арабска за – дефакто – по едната-единствена точка на първо ниво в рамките на глава, освен че е ненужно, създава объркване и у автор, и у читател. Номерирането на формули, които не се реферират в текста, също е излишно. На много места се виждат формули, които не са набрани в текста, а са изрязани изображения от (неуказани) други места, което освен за уменията на дисертанта да борави с математически текстове, поставя въпроса и дали с валидно въведени и консистентни в рамките на целия труд означения ли е боравено навсякъде (а отговорът на немалко места е – не).

При изготвянето на автореферата е трябвало да бъдат запазени номерациите на формулите и фигурите от дисертационния труд с оглед лесната проследимост между двата документа. Това би спестило аномалии като появата на „Фигура X“ на стр. 16 от автореферата.

Много от цитатите в библиографията са описани некоректно.

В заключение, отбелязвам всички тези терминологични неточности най-вече по две причини. От една страна, безспорна е актуалността на тематиката на дисертационния труд и с висока степен на увереност може да се твърди, че получените резултати от докторанта ще представляват интерес за практиците в областта на капиталовите пазари. От друга страна, съгласно процедурата по защита, заложен в ЗРАСРБ, на автореферата се осигурява публичност и съдържанието му на свой ред може да се превърне в литературен източник за други докторанти, учени и практики в областта, които да го ползват и реферират в бъдеще, и така да мултиплицират в литературата грешките, които трудът съдържа. В този смисъл, въпреки че вероятно процедурно не е предвиден такъв вариант, искрено препоръчвам на докторанта да ревизира поне онлайн достъпната версия на автореферата си, в която да отрази колкото се може повече от забележките.

По отношение на забелязаните цитирания на публикацията „*Design and Application of Artificial Neural Networks for Predicting the Values of Indexes on the Bulgarian Stock Market*“, ще отбележа, че не мога да приема цитирането

Samit Bahnja, Abhishek Das, *Impact of Data Normalization on Deep Neural Network for Time Series Forecasting*, 2018.

което е непубликуван ръкопис в arXiv, но за сметка на това при проучването си открих още едно цитиране на същата публикация от 2014 г., достъпно в SCOPUS:

Shi, H., Liu, X. (2014). *Application on stock price prediction of Elman neural networks based on principal component analysis method*. 11th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2014, 7073438, pp. 411-414.

Предвид че не е изискване по процедурата докторантът да има цитирания, оценявам много положително наличието на шест цитирания на публикация по дисертационния труд.

Въпроси към докторанта

- Каква е разликата между метода на най-малките квадрати (Least Squares) и метода на най-малките средни квадрати (Least Mean Squares)?
- На стр. 82 е написано „Предизвикателството да се печелят пари от инвестиции в финансови инструменти все пак изисква усъвършенствани методи и правилен подбор на различните компоненти от моделите, като входните данни, архитектурата на мрежата, методите за обучение и в крайна сметка начина на материализиране на продуцираните резултати.“ Съгласявайки се с това принципно твърдение, въпросът ми към докторанта е да конкретизира по какъв начин са подбрани параметрите на архитектурите на невронните мрежи от Таблица 8 на стр. 113 и Таблица 10 на стр. 118. Как са били определени броят елементи на входа, на скрития слой и на изхода на избраните три вида невронни мрежи? Особено любопитно е как са избрани параметрите на невронната мрежа с обща регресия.
- На стр. 94 е написано „...като в случая е използвана онлайн версията на алгоритъма на обратно разпространение на грешката,“ и на стр. 112 е написано „Софтуерния продукт използван за имплементирането им в експеримента използва невронни мрежа с обща регресия, заедно с вероятностни невронни мрежи в алгоритмите за избор на функции, които автоматично търсят полезни входове.“ Може ли да се конкретизира за коя „онлайн версия“ и кой „софтуерен продукт“ става дума?
- На стр. 75 в литературния анализ е анализирана публикация [47], свързана с прогнозиране на цените на девет акции и индекса Nifty 50 преди и след демонетизацията в Индия. При (хипотетично) въвеждане на еврото в България, възможна ли е прогноза с хибридният модел, предложен от докторанта, на разглежданите в труда два индекса и какви биха били резултатите?

Отчитайки факта, че практически всички така отбелязани грешки е възможно да бъдат отстранени, т.е. не са непреодолими слабости на труда; отчитайки безспорната актуалност на тематиката на дисертационния труд и възможностите за развитие на проблемната област и пред докторанта, както и факта, че дисертационният труд удовлетворява изискванията на ЗРАС и Правилника на ИИКТ–БАН, давам положително становище и препоръчам на уважаемите членове на Научното жури да гласуват за присъждането на Веселин Шахпазов на образователната и научна степен „доктор“ по професионално направление 4.6 “Информатика и компютърни науки“, докторантска програма „Информатика“.

3 декември 2019 г.

Изготвил становището:

**NOT FOR
PUBLIC RELEASE**

Доц. д-р Вася Атанасова

Секция „Биоинформатика и математическо моделиране“, ИБФБМИ-БАН