

СТАНОВИЩЕ

върху дисертационен труд на тема:

Иконометричен анализ на големи данни

Научна област: 3 „Социални, стопански и правни науки“

Професионално направление: 3.8 „Икономика“

Научна специалност: Аналитични изследвания върху данни (Data Science)

Катедра: „Статистика и иконометрия“ при Стопански факултет на Софийски университет „Св. Климент Охридски“

Име на дисертанта: Борислава Петрова Толева

Автор на становището: доц. д-р Константин Николов Колев, Лесотехнически университет – София

Становището е изготвено на основание на решение от 15. 07. 2021 г. на научното жури, формирано със Заповед № РД-38-303/08. 07. 2021 г. на Ректора на Софийски университет „Св. Климент Охридски“.

1. Кратки биографични данни

През 2011 г. Борислава Петрова Толева завършва средното си образование в гимназия с преподаване на чужди езеци в гр. Плевен. След това е приета за студент в Стопански факултет на Софийски университет „Св. Климент Охридски“, където през 2015 г. и 2017 г. придобива последователно образователно-квалификационни степени „Бакалавър по икономика“ и „Магистър по икономика – статистика, иконометрия и актюертство“. През периода 2014 – 2019 г. докторант Борислава Толева е хоноруван асистент по дисциплината „Статистика“ към катедра „Статистика и иконометрия“ при Стопанския факултет на СУ „Св. Климент Охридски“. Понастоящем работи като анализатор на големи данни в Euromoney Institutional Investor.

2. Обща характеристика на дисертационния труд

Представената за рецензиране докторска дисертация е в обем от 161 страници и съдържа увод, три глави, заключение и използвана литература. В изложението са

поместени 32 таблици и 2 фигури. Библиографската справка съдържа 113 източника (3 на кирилица и 110 на латиница).

Актуалността на дисертационния труд е безспорна. В условията на разширяващата се глобализация нарастват възможностите да се събира огромно количество информация. Това налага намирането на нови способности за бързо и ефективно обработване на големите множества наблюдения, от които да се извличат модели на поведение.

Предмет на дисертацията е буутстрап процедурата и нейните приложения, а **обекти** са панелните данни (в първа глава) и моделът на поддържащите вектори, като представител на класификационните модели (втора и трета глава).

Целта на дисертационния труд е да разшири използването на алгоритмичното учене от данни до панелните иконометрични модели и да усъвършенства вече съществуващи модели от алгоритмичното учене от данни за големи множества от наблюдения. Поставената цел намира широко приложение в икономиката, бизнеса, маркетинга, биологията, медицината и др.

Изследователските хипотези са правилно формулирани и моделират логиката на дисертационното изследване и приложения аналитичен инструментариум.

3. Оценка на получените научни и научно-приложни резултати

Първа глава на разработката се основава на хипотезата, че съществува интердисциплинарен подход за моделиране на панелни данни, обединяващ иконометричен подход и алгоритмично учене от данни. Хипотезата се верифицира в контекста на панелен казус за моделиране на индекса на имуществените права, който съдържа 32 държави, разгледани през периода 1999 – 2014 г. Чрез класически иконометричен подход е построен базов панелен модел от статистически значими променливи, оказващи влияние върху индекса на имуществените права. Към базовия модел се добавят последователно три контролни променливи (сива икономика, полово неравенство и неравенство на доходите) и с иконометрични тестове се установява тяхната статистическа значимост. Панелните модели се валидират чрез робастна ковариационна матрица или т. нар. sandwich estimator, панелен генерализиран метод на моментите (GMM) и тестове за излишни и пропуснати променливи. Резултатите от валидираните панелни модели се използват за сравнение с резултатите от моделите за избор на променливите от

алгоритмичното учене на данни. Във връзка с последното за установяване на статистически значимите променливи, влияещи на индекса на имуществените права, в дисертацията се използват методите ridge, lasso и adaptive lasso. Резултатите от тяхното приложение са незадоволителни. Основната причина е, че трите метода не идентифицират видовете ефекти в панела. Този недостатък се преодолява чрез използването на модела nonnegative garrote с буутстрап при фиксирани регресори. Неговото приложение осигурява резултати сходни с тези на валидираните панелни модели. Това дава основание на докторанта да предложи време-спестяващ и автоматизиран алгоритъм за избор на статистически значими променливи при малки множества и панелни данни, който съчетава общоприетата иконометрична теория и модели за избор на променливите от алгоритмичното учене от данни.

Във **втора глава** на разработката се формулира хипотезата, че буутстрап процедурата, като метод за разделяне на наблюденията в тренировъчно и тестово подмножество на големи множества от наблюдения, е ефективна алтернатива на десеткратната крос валидацията. Хипотезата се проверява чрез провеждане на експерименти в девет множества с различна размерност, като се изпълнява модела на опорните вектори чрез четири метода за формиране на тренировъчното и тестовото подмножество – три класически (train/test split и десеткратна крос валидация, train/test split и Leave-one-out крос валидация и repeated random train/test split) и предложен от докторанта буутстрап алгоритъм с реализация в Python 3.6. Сравнителният анализ на резултатите от проведените експерименти дава основание да се направят следните изводи: първо, в повечето случаи буутстрап процедурата запазва прогнозната точност, получена чрез крос валидацията или я повишава; второ, буутстрап процедурата съкращава изчислителното време на модела на опорните вектори във всички случаи – от малки до големи по размер множества; трето, буутстрап процедурата запазва или подобрява класификационните метрики (precision, recall, f_1 -score) спрямо десеткратната крос валидация.

Хипотезата, която определя съдържанието на **трета глава** на дисертационния труд гласи, че предложените нови буутстрап модификации на модела на поддържащите вектори подобряват ефективността на класическия модел с ANOVA. Верификацията на хипотезата се осъществява чрез използване на множества данни, върху които са правени експерименти и от други автори. В дисертационния труд върху тези множества се прилагат алгоритмите ANOVA-CV-SVM и ANOVA-Bootstrap-SVM в среда на Python 3.6.

Получените резултати се сравняват с тези от изследванията на други автори. Установява се, че буутстрап модификациите могат да повишат както прогнозната точност, така и AUC точките, получени от модификации на модела на опорните вектори, комбиниран с алгоритъма на целочисленото линейно смятане. Освен това буутстрап модификациите могат да подобрят прогнозната точност, резултат, както от други версии на модела на опорните вектори, така и от други класификационни модели като модела на най-близките съседи, random forest, невронни мрежи и др.

В **заключението** са направени обобщаващи изводи. Те са резултат от адекватно решените изследователски задачи и потвърждават дефинираните в началото на всяка глава хипотези.

Разработката е написана на правилен и достъпен научен език, което допринася за убедителност на изложението и непосредственост при възприемането.

4. Оценка на научните и научно-приложни приноси

В дисертационния труд са налице достатъчно доказателства за научно-приложни приноси. По-съществените от тях биха могли да бъдат представени по следния начин:

- В контекста на панелни данни с фиксирани ефекти е доказана хипотезата на Брайман, че моделът nonnegative garrote може да бъде използван с буутстрап процедурата в данни с фиксирани регресори. На тази основа е предложен ефективен интердисциплинарен подход за избор на статистически значими променливи при малки множества и панелни данни, който комбинира традиционната иконометрична теория и модели за избор на променливите от алгоритмичното учене от данни.

- Обосновано е приложението на буутстрап процедурата на Ефрон като алтернатива на десеткратната крос валидация при модела на опорните вектори за разделяне на наблюденията на тренировъчно и тестово подмножество.

- Доказано е, че десет итерации на буутстрап процедурата на Ефрон (авторът препоръчва 100, 1000 и т.н. итерации) са достатъчни за да бъде тя прилагана като алтернатива на крос валидацията.

- Предложена е модификация на буутстрап процедурата на Ефрон. Тя е спестяваща време и намира приложение при големи множества от наблюдения.

- Обосновано е приложението на бутстрап процедурата на Ефрон като алтернатива на десеткратната крос валидация при ANOVA модела на опорните вектори. Освен това е предложена модификация ANOVA-Bootstrap-SVM, която може да бъде по-ефективна от класическия ANOVA модел.

5. Оценка на публикациите по дисертационния труд

Докторантът е представил две самостоятелни публикации и две в съавторство. Тези публикации са неразделна част от съдържанието на докторската дисертация и с тях докторантът представя публично своите изследователски резултати в научната литература. Трябва да се подчертае, че една от публикациите в съавторство е индексирана в Скопус.

6. Оценка на автореферата

Авторефератът е в обем от 43 страници, вкл. приноси и списък на публикациите по дисертационния труд. Авторефератът съдържа актуалността на темата и степента на нейната разработеност, целта и задачите на изследването, издържан е в структурно отношение и коректно отразява логиката и последователността на дисертационния труд.

7. Критични бележки, препоръки, въпроси

- Съществени критични бележки към дисертационния труд не се откриват, с изключение на такива от техническо естество, които биха подобрили неговото качество: първо, уводът на дисертационния труд би могъл да се представи по-стегнато; второ, в текста има повторения на издигнатите хипотези на стр. 10-11, съответно на стр. 25, 71 и 114, което би могло да се избегне.

- Вземайки под внимание значимостта на иновативното научно изследване, реализирано от докторант Борислава Толева си позволявам да ѝ препоръчам да публикува дисертационния си труд като самостоятелен научен продукт. Това ще даде възможност на по-широк кръг от специалисти да се запознаят с него.

8. Личен принос на докторанта

Приносът на докторанта е неоспорим и се доказва от приложената методика и получените резултати. Освен това се потвърждава и от 4-те публикации към

дисертационния труд, две от които са самостоятелни, а в две Борислава Петрова Вригазова е първи автор.

ЗАКЛЮЧЕНИЕ

Представеният дисертационен труд на тема „Иконометричен анализ на големи данни“ е добросъвестно осъществено научно-приложно изследване. Поставям положителна оценка на дисертационния труд. Препоръчвам на уважаемите членове на Научното жури да присъдят образователна и научна степен „ДОКТОР“ на Борислава Петрова Толева по професионално направление 3.8 „Икономика“, научна специалност „Аналитични изследвания върху данни /Data Science/“.

Август, 2021 г.

Изготвил становището:



(доц. д-р Константин Колев)