



РЕЦЕНЗИЯ

на

**дисертационен труд за придобиване на образователната и научна степен доктор по
докторска програма Информатика, професионално направление направление: 4.6**

„Информатика и компютърни науки“

Автор на дисертационния труд: Емилиано Максим Манколли

Тема на дисертационния труд (на английски език): Optimization Methods for Machine Learning Applications

Автор на рецензиията: проф. д-р Кирил Иванов Симов, ИИКТ-БАН

Със заповед № 319 /06.12.2023 г. на директора на ИИКТ съм определен за член на журито по защитата за дисертационен труд на Емилиано Максим Манколли за присъждане на образователната и научна степен доктор. Представеният набор от материали от кандидата включва всички необходими документи по процедурата.

Докторантът е представил пет свои публикации, които са публикувани в сборници от конференции. От тях публикации 4 са в съавторство и 1 е самостоятелна. Една от тях е публикувана в Communications in Computer and Information Science, vol 1316. Springer (SJR (0.188), Computer Science - Quartile Q3). Тези публикации са достатъчни за защитата на дисертацията.

Авторефератът (38 страници) е на английски език и отразява пълно и правилно структурата и съдържанието на дисертационния труд. Авторефератът е преведен на български.

Представени са пет научни приноса и три научно-приложни приноса – според българския превод на автореферата, които в английския текст са представени общо като шест приноса. Може да се каже, че добавените два приноса в българския превод доразширяват в правилна посока приносите на дисертацията. Така те вече правилно отразяват представените разработки в дисертацията. Аз бих добавил като принос и създаването на множества от данни за решаването на поставените в дисертацията проблеми, тъй като тяхното създаване

е достатъчно сложно и освен това те може да се използват в бъдещи разработки както на дисертанта, така и на други изследователи на сходни проблематики.

Темата на дисертацията „Методи за оптимизация за приложения на машинно обучение“ е от актуалните теми при прилагане на машинно обучение и е добре представена в дисертацията от гледна точка на приложната област - *набиране на персонал* на базата на различни по тип документи, описващи качествата, знанията и уменията на кандидатите. При тази задача унифицираното представяне, над което да се приложат различни подходи от машинното обучение за класификация, ранкиране и оценка, представлява голямо когнитивно натоварване на специалиста по управлението на човешки ресурси. Поради този факт много важни са резултатите от дисертацията, които демонстрират теоретичната и практическата полза от прилагането на интегрирани системи, включващи в себе си елементи от Изкуствения интелект – методи от обработката на естествен език, векторизацията на думи и текстове и различни методи от машинното самообучение.

Дисертационният труд на Емилиано Максим Манколли (общо 128 стр.) е написан на английски език. Той съдържа увод, три глави, заключение, списък с приносите на дисертацията, публикациите на дисертанта и декларация за оригиналност на резултатите, както и използвана литература. Списъкът с използваната литература включва 131 заглавия на английски език. Те съдържат фундаментални и съвременни теоретични трудове за обработката на естествен език, за машинното самообучение, както и трудове от областта за набиране на персонал и за оценка на успешността на човек. Много приятно впечатление прави и фактът, че има списък на съкращенията в началото, а в края на всяка глава са изведени заключения за главата, включващи резюме на резултатите, ограничения на изследването, както и насоки за бъдеща работа.

Уводът представя в детайли проблемите, с които се сблъскват работещите в областта на набиране на персонал, и как методите на машинното обучение и обработката на естествен език могат да революционизират тази област. По този начин дисертантът мотивира в пълнота необходимостта и важността на представените в дисертацията изследвания и получените резултати. Освен при подготовката на информацията, която се съдържа в документите на кандидатите, се отбелязва и намаляването на личностния фактор на специалистите по

управление на човешките ресурси при оценките и решенията, които правят на базата на тази информация.

Първа глава „Обзор на състоянието на съвременните изследвания: машинно обучение, обработка на естествен език и оптимизация“ дава много детайло представяне на развитието, методите и приложенията на технологиите, които се използват в дисертационния труд: обработката на естествен език; машинното самообучение и оптимизационните проблеми при машинното самообучение. Обработката на естествен език е описана на няколко взаимосвързани нива: задачи, които се решават; методи и обработки, които се използват, и приложения в областта на набиране на персонал. Обзорът е пълен, като описва също така методи и задачи, които не са използвани в дисертационния труд. Това показва едно добро ниво на познание на дисертанта по тази тематика. По подобен начин е представена и областта на машинното обучение, като са обхванати всички важни подходи в областта. Освен двата основни подходя – с учител (Supervised Learning) и без учител (Unsupervised Learning) - са представени и методи като Reinforcement Learning, Semi-Supervised, Active Learning Techniques, Deep Learning and Neural Networks, Transfer Learning и Ensemble Methods. Тези методи играят важна роля при обработката на естествен език. Особено при създаването на големи езикови модели, които в момента се разглеждат като основа за напредъка на изкуствения интелект. Дискутиирани са моделите, свързани с машинното самообучение и оптимизационните проблеми.

Главата завършва с формулиране на целите на дисертацията и задачите, които се решават в дисертацията.

Втора глава „Рационализиране на набора от кандидати: Оптимизиране на точността и ефикасността при сходството на длъжностите“ предлага няколко решения на проблема с прегледа на информацията за голямо количество кандидати. Задачата е представена като търсене на подобни длъжности в рамките на множество от описания на длъжности (job titles) (80 000) и описания на индустрии (700). Дадена е обосновка защо този проблем е важен в областта на набиране на персонал. Предложеното решение използва комбинация от векторизация чрез използване на word2vec (който правилно е описан като статичен подход) и комбинация от SVM (Support Vector Machines) и k-NN (k -Nearest Neighbors). Те са детайлно описани в дисертацията. Процедурата е, както следва: word2vec

се използва за векторизация на описанията на длъжностите и индустриите, които се използват за създаване на данни с етикети. Тези данни се използват от SVM и k-NN за решаването на проблема. Предложеният подход е тестван на базата на събрани данни. Един въпрос, който е отворен, е как се получава векторното представяне на текстовото описание на длъжностите и на индустриите, т.к. word2vec работи само с една дума. Но може би това е в посочените референции и е използвано наготово.

Вторият подход към същата задача, представен в тази глава, използва друг метод за векторизация на думи и текстове – BERT. Самият BERT е добре описан в дисертацията. При този втори подход също се използва XGBoost (Extreme Gradient Boosting) за подобряване на изпълнението на модела. Приложението на тези модели е подобно на това в първата част на главата. BERT използва свързване на описанията на длъжности с етикети на базата на близост до описания на индустрии. След това се използва XGBoost за намиране близост на длъжности. След което те се филтрират чрез директно изчисление на близостта. Също е даден конкретен пример.

Трета глава „Прогнозни модели за успех в работата. Оптимизиране на стратегията за подбор на кандидати“ описва метод, който да оцени предварително вероятността даден кандидат да е успешен дългосрочно, ако бъде нает за дадена длъжност. Главата първо описва теоретичните основи за тази оценка. Данните, които се използват в изследването, представено в главата, са описания на професионалното развитие на кандидата: професионален опит, умения, предишни работни места и други. Тези описания се представят като набор от характеристики за кандидата, чрез нормализация на числовия вход или векторизация с помощта на BERT. След създаването на множествата от характеристики, те се използват за класифициране на кандидатите като успешни или не с помощта на XGBoost. Показано е предимството на този подход в сравнение с предишни подходи.

Заключението накратко резюмира труда, обобщава резултатите, с които се подкрепят представените в началото цели. Показва се, че получените резултати оптимизират обработката на информацията за кандидата и ефективното ѝ сравнение с информацията за съответните длъжности.

Представеният труд има ясно изразени научни и научно-приложни приноси. Приносите са много добре представени в разработените алгоритми и архитектури за работа с описания на длъжности, индустрии и кандидати. Те показват също така оптимизиране на ефективността на решенията на тези задачи от областта на набиране на персонал. По този начин целите на дисертацията са постигнати напълно.

Въз основа на всичко казано по-горе по отношение на качествата на представената дисертация, бих искал да заявя, че убедено предлагам на уважаемото жури да присъди на Емилиано Максим Манколли научната и образователна степен „доктор“.

19.02.2024 г.

Рецензент:

