



Warszawa, 29 marca 2024

Prof. dr hab. inż. Przemysław Biecek
Wydział Matematyki i Nauk Informatycznych
Politechnika Warszawska
Koszykowa 75/507
00-662 Warszawa

Recenzja rozprawy doktorskiej „Modele rekomendacyjne wspólnej filtracji w serwisach ogłoszeniowych” pana mgra Roberta Kwiecińskiego.

Promotor rozprawy: prof. UAM dr hab. Tomasz Górecki

Promotor pomocniczy: prof. UEP dr hab. Agata Jolanta Filipowska

Dyscyplina: Informatyka

Niniejsza recenzja oparta jest o otrzymaną rozprawę doktorską oraz dodatkową zewnętrzną walidację algorytmów zaproponowanych w rozprawie.



Praca doktorska powstała w ramach programu doktorat wdrożeniowy i była realizowana na Uniwersytecie im. Adama Mickiewicza w Poznaniu we współpracy z Grupą OLX. Z tego powodu w dalszej części recenzji odnoszę się zarówno do oczekiwań stawianych pracom doktorskim jak i oczekiwań stawianych doktoratom wdrożeniowym.

Główny cel rozprawy doktorskiej jest sformułowany w pierwszym rozdziale i dotyczy zaproponowania i zbadania skuteczności nowych metody rekomendacji. Cel ten jest uzupełniony o cele pomocnicze, w tym przygotowanie i opublikowanie nowego zbioru danych, przetestowania systemów rekomendacyjnych oraz opis doświadczeń z wdrożenia zaproponowanej metody u pracodawcy doktoranta, w Grupie OLX.

Struktura pracy odzwierciedla liniową sekwencję uzyskanych wyników. Rozdział 3 przedstawia pierwszy z opracowanych wyników - nowy zbiór danych bazujący na danych firmy OLX. Na tym zbiorze danych będą prowadzone dalsze analizy. Rozdział 4 przedstawia analizę skuteczności pięciu metod wyznaczania rekomendacji (ALS, LightFM, SLIM, RP3Beta i Prof2Vec) przeprowadzoną na przygotowanym wcześniej zbiorze danych. Najlepsze wyniki uzyskuje metoda RP3Beta i jest ona podstawą propozycji nowych metod opisanych w dalszej części rozprawy. Rozdział 5 przedstawia pierwszą z nowych propozycji - metodę RP3Beta real-time dostosowaną do pracy w trybie z częstą aktualizacją (w pracy nazwaną metodą czasu rzeczywistego), rozdział 6 przedstawia drugą nową propozycję - metodę RP3LTR, która uelastycznia model RP3Beta. Rozdział 7 przedstawia kolejną propozycję uelastycznienia metody RP3Beta, czyli RP3GNN. Poniżej przedyskutujemy te główne wyniki.

Zbiór danych 'OLX Jobs Interactions' przedstawiony w rozdziale 3 to zdecydowanie jedna z najsilniejszych części tej rozprawy doktorskiej. Dostyc duży (ponad 2GB) zbiór danych, o ponad 65 milionach interakcji pomiędzy ponad 3 milionami użytkowników z prawie 200 tysiącami ofert pracy. Dane zostały upublicznione przez doktoranta trzy lata temu na popularnej platformie



kaggle.com, przez co mogą być szeroko wykorzystywane do celów badawczych. Zbiór ten nie został szerzej zauważony i cytowany (współczynniki wpływu przedstawione na platformie Kaggle są raczej znikome), co prawdopodobnie wynika z braku szerszej reklamy dostępności tego zasobu. Ale z pewnością opracowanie i udostępnienie danych to przykład bardzo dobrej praktyki badawczej, umożliwiający walidację i analizę kolejnych wyników zaprezentowanych w rozprawie. Pozostaje mieć nadzieję, że będzie on w przyszłości wykorzystywany tak często jak na to zasługuje.

Metoda RP3Beta real-time, jest bardzo pomysłową acz niewielką modyfikacją metody RP3Beta, która została zaproponowana w 2015 roku i od tego czasu zdobyła pewną popularność. W metodzie RP3Beta real-time część operacji do wyznaczenia rekomendacji jest wykonywana z niską częstotliwością (w pracy jest informacja o kilku aktualizacjach dziennie), ale pozwala to na częstsze wykonanie pozostałych operacji, przez co można „prawie w czasie rzeczywistym” uwzględnić zmieniające się informacje o dotychczasowej interakcji/aktualizowanych cechach użytkownika. Potencjalnie takie częstsze uwzględnienie aktualnej informacji o użytkowniku może prowadzić do poprawienia skuteczności rekomendacji. Autor zbadał to z użyciem eksperymentu A/B, który wykazał poprawę wyników (zaskakująco, pomimo początkowo błędnej implementacji, ale z jeszcze lepszym wynikiem dla poprawnej implementacji). W wartościach względnych ta poprawa jest niewielka, ale okazała się istotna dla firmy OLX, została uznana za wartościową oraz wprowadzona do systemu produkcyjnego. Zarówno metoda, jak i wyniki testów A/B zostały opublikowane w czasopiśmie IEEE Access (to jedyna „punktowana” publikacja autora). Współpraca z firmą o skali OLX tworzy ciekawą możliwość przetestowania skuteczności narzędzia na prawdziwym systemie w postaci testów A/B co też jest ciekawym i wartościowym wyróżnikiem tej rozprawy i podkreśla wdrożeniowy i stosowany charakter uzyskanych wyników.



Metoda RP3LTR jest kolejnym, jeszcze ciekawszym rozszerzeniem metody RP3Beta, pozwalającym na bardziej elastyczne (szerzej parametryzowane) opisanie interakcji pomiędzy użytkownikami a przedmiotami. W rozdziale 6 autor przedstawia nową parametryzację oraz interpretację kluczowych parametrów tej metody, takich jak: wpływ stopni wierzchołków, wpływ czasu wykonywanych interakcji, wpływ typu interakcji, częstotliwości interakcji itp. Autor proponuje algorytm treningu tych parametrów oraz pokazuje wyniki eksperymentalne lepsze niż dla trzech alternatywnych metod. Wyniki te są opublikowane i przedstawione szerzej w pracy konferencyjnej na konferencji FedCIS 2022. Warto tu zauważyć, że w dokumencie zawierającym dodatkową zewnętrzną walidację na danych Allrecipes i BookCrossing metoda RP3LTR otrzymuje dobre wyniki, choć (nieznacznie) gorsze niż autorska implementacja RP3Beta.

Metoda RP3GNN jest propozycją dalszego rozszerzenia metody RP3Beta przez wykorzystanie sieci neuronowych jako „koderów cech”, koncepcji wprowadzonej w rozdziale 6. Metodologicznie jest to najbardziej zaawansowane i elastyczne podejście. Jednak wyniki eksperymentalne pokazują, że trudno jest te kodery dobrze wytrenować, a otrzymane wyniki są gorsze niż dla algorytmów RP3Beta i RP3LTR.

Podsumowując uzyskane wyniki, w pierwszej kolejności rzuca się w oczy ich wdrożeniowy charakter. Centralną motywacją dla opracowania i oceny kolejnych metod jest ich sprawność na danych przygotowanych w ramach współpracy z Grupą OLX. Z tej perspektywy doktorat bardzo dobrze wpisuje się w program doktoratów wdrożeniowych, a uzyskane wyniki nie tylko zostały opracowane podczas pracy w ww firmie, ale też prawdopodobnie są tam wykorzystywane. Propozycja algorytmów i analiza ich skuteczności są też wartościowym materiałem dla innych firm pracujących nad systemami rekomendacyjnymi.

O ile potencjał i wartość wdrożeniowa zaproponowanych nowych metod nie budzi żadnych wątpliwości, pewien niedosyt pozostawia analiza czysto aka-



demickich aspektów doktoratu. Najbardziej zaawansowane z proponowanych metod (RP3LRP i RP3GNN) mogłyby być zaprezentowane na konferencjach lub czasopismach skupionych na systemach rekomendacyjnych. Również analiza porównawcza proponowanych metod mogłaby być zestawiona z rozwiązaniami stosowanymi w innych firmach/organizacjach rozwijających systemy rekomendacyjne. Pod tym względem „analiza konkurencji” badawczej jest w rozprawie doktorskiej raczej skąpa.

Niedosyt pozostawia też część rozprawy z prezentacją wyników. Wykresy są mało czytelne, z trudnymi do odczytania napisami na osiach i skąpymi podpisami pod rysunkami. Podobnie zredukowane i lakoniczne są podpisy pod tabelami. Przykładowo tabela 4.3 przedstawia wyniki dla kilku metod, ale pogrubione nie są najlepsze wyniki dla każdego z kryteriów (co jest standardem w tego typu artykułach) ale metoda proponowana przez autora. W tabeli 4.4 przedstawione są p-wartości, z których większość jest równa 0, choć biorąc pod uwagę matematyczną interpretację p-wartości nie powinno się nigdy podawać takiej wartości. Oczywiście są to względnie niewielkie trudności, które nie uniemożliwiają zrozumienia i docenienia głównych wyników rozprawy.

Największym brakiem, co do którego mam nadzieję, że doktorant ustosunkuje się podczas obrony, jest kwestia walidacji proponowanych modyfikacji systemu rekomendacyjnego. Autor skupia się na podstawowych klasycznych metodach liczenia skuteczności dla rekomendacji, takich jak Recall, Precision, MAP itp, ignorując współczesne wątki poruszane np. przez ustawę o Sztucznej Inteligencji (AI act) takie jak możliwość dyskryminacji przez systemy SI, kwestie transparentności i wyjaśnialności automatyzowanych decyzji, reprezentatywności danych treningowych i innych tematów które należy badać i monitorować by opracowana SI była godna zaufania (ang. Trustworthy). Biorąc pod uwagę niedawne zeznania Marka Zuckerberga przed Senatem USA w sprawie potencjalnych negatywnych wpływów systemów rekomendacyjnych Facebooka na nastolatki w USA warto ten temat potrak-



tować z odpowiednią uwagą. Biorąc pod uwagę, wdrożeniowy charakter pracy, taka pogłębiona krytyczna analiza nowych algorytmów i ich potencjalnego wpływu na społeczeństwo jest niezbędna i liczę, że zostanie przedyskutowana szeroko podczas publicznej obrony.

Pomimo powyższych krytycznych uwag, chcę wyraźnie podkreślić, że wyniki zaprezentowane w rozprawie są ciekawe i dobrze umotywowane oraz dobrze opisane. Udostępnienie zbioru danych, sekwencja trzech kolejnych usprawnień metody RP3Beta weryfikowanych na tym zbiorze danych tworzą spójną ścieżkę rozwiązywania istotnego problemu badawczego. Opracowane metody dowodzą wiedzy, intuicji badawczej autora i umiejętności samodzielnego prowadzenia prac naukowych. Zostały zaprezentowane w postaci jednego artykułu konferencyjnego i jednego artykułu w czasopiśmie, spełniają więc warunki ustawowe stawiane rozprawom doktorskim. Dlatego oceniam tę rozprawę pozytywnie i wnioskuję o dopuszczenie rozprawy doktorskiej pana mgr Roberta Kwiecińskiego do publicznej obrony.

Przemysław Biecek