



POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

ポ
ー
ラ
ン
ド
日
本
情
報
工
科
大
学

Warszawa, 03.07.2023

prof. dr hab. Adam Wierzbicki
Polsko-Japońska Akademia
Technik Komputerowych

OPINIA O ROZPRAWIE DOKTORSKIEJ

MGR INŻ. BARBARY RYCHALSKIEJ

“ A MULTIMODAL AND INTERPRETABLE RECOMMENDER SYSTEM FOR
MASSIVE DATASETS ”

1. Jakie zagadnienie naukowe jest rozpatrzone w pracy i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Mgr inż. Barbara Rychalska sformułowała cztery cele badawcze swojej rozprawy doktorskiej:

1. Poprawa wydajności i skalowalności grafowych sieci neuronowych poprzez opracowanie nowych algorytmów odwzorowujących grafy na reprezentację w przestrzeni wielowymiarowej
2. Zaprojektowanie nowego algorytmu rekomendacyjnego, wykorzystującego dane wielomodalne i sieci neuronowe, który będzie miał lepszą wydajność oraz porównywalną dokładność w porównaniu z algorytmami używającymi głębokich sieci neuronowych
3. Zaprojektowanie nowego, wydajnego algorytmu rekomendacyjnego typu „content-based”
4. Zaprojektowanie i ewaluacja nowych metod zwiększania odporności systemów rekomendacyjnych na błędy.

Doktorantka sformułowała także pięć szczegółowych hipotez, które precyzują cele badawcze rozprawy, i które są powiązane z pięcioma głównymi rozdziałami rozprawy.



POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

ポ ー ラ ン ド 日 本 情 報 工 科 大 学

Rozprawa ma charakter badań stosowanych i jest wynikiem doktoratu wdrożeniowego realizowanego przez doktorantkę we współpracy z firmą Synerise. Firma Synerise wdrożyła wiele wyników badań doktorantki w swojej działalności.

Należy podkreślić, że choć rozprawa dotyczy badań stosowanych, to wyniki opisane w rozprawie mają istotne konsekwencje teoretyczne. Doktorantka wykazała, że w wielu przypadkach, stosowane obecnie powszechnie głębokie sieci neuronowe z mechanizmem „attention” lub „self-attention” można zastąpić metodami prostszymi, o złożoności liniowej (wspomniane sieci neuronowe mają złożoność kwadratową).

Cele badawcze i hipotezy rozprawy zostały sformułowane bardzo klarownie i precyzyjnie. Doktorantka podjęła także wysiłek umieszczenia swoich badań w szerszym kontekście, zarówno względem aktualnego stanu badań naukowych, jak i względem zastosowań praktycznych.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczący o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Rozprawa zawiera dwa rozdziały, które dotyczą analizy źródeł i aktualnego stanu badań w jej dziedzinie. Rozdział drugi rozprawy pod tytułem „Introduction to Neural Networks” podejmuje próbę wprowadzenia czytelnika w tematykę sieci neuronowych od podstaw aż do współcześnie stosowanych głębokich, rekursywnych sieci neuronowych z mechanizmem „attention”. Próba ta nie jest całkiem udana, głównie ze względu na brak miejsca – tematyka jest godna obszernego podręcznika. Mimo to doktorantce udaje się nie tylko dobrze opisać aktualny stan wiedzy, ale przedstawić historię rozwoju sieci neuronowych, wyjaśniając aktualnie używane metody.

Rozdział trzeci rozprawy pod tytułem „Recommender Systems” to systematyczny opis aktualnego stanu badań w tej dziedzinie, zaczynając od podstaw, aż po współczesne systemy rekomendacyjne wykorzystujące głębokie sieci neuronowe. Rozdział zawiera szczegółowy opis metod i miar oceny dokładności systemów rekomendacyjnych (sekcja „Recommender Evaluation Metrics”). Rozdział trzeci zawiera także omówienie interpretowalnych i wielomodalnych systemów rekomendacyjnych. Rozdział ten kończy się omówieniem zbiorów danych stosowanych we współczesnych badaniach nad systemami rekomendacyjnymi oraz stosowanych przez doktorantkę do uczenia i oceny jej algorytmów oraz metod.

W rozdziałach dotyczących przeglądu literatury formułowane są wnioski dotyczące wad obecnie stosowanych systemów rekomendacyjnych i sieci neuronowych. Na przykład, grafowe sieci



POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

ポ
ー
ラ
ン
ド
日
本
情
報
工
科
大
学

neuronowe z mechanizmem „attention”, wykorzystywane w systemach rekomendacyjnych typu „collaborative”, mają zbyt wysoką (kwadratową) złożoność obliczeniową, co oznacza, że mogą być uczone i testowane jedynie na małych zbiorach danych, lub na małych częściach dużych zbiorów.

Podsumowując – przegląd źródeł oraz aktualnego stanu wiedzy oraz wnioski sformułowane na jego podstawie przez doktorantkę oceniam bardzo wysoko.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Doktorantka używa metodologii empirycznej, będącą podstawą uczenia maszynowego. Wszystkie jej algorytmy są oceniane poprzez testowanie ich na wielu publicznie dostępnych zbiorach danych, oraz poprzez porównanie z licznymi konkurencyjnymi algorytmami, które są aktualnym stanem wiedzy i techniki w rozpatrywanym zagadnieniu badawczym. Algorytmy i metody doktorantki były także testowane poprzez wzięcie udział przez doktorantkę (wraz z zespołem firmy Synerise) w wielu konkursach z dziedziny systemów rekomendacyjnych oraz „data science”.

Algorytmy i metody opracowane przez doktorantkę są dostępne na otwartej licencji (MIT) w serwisie GitHub, co umożliwia replikację badań doktorantki:

- Cleora: <https://github.com/Synerise/cleora>
- System rekomendacyjny wykorzystujący EMDE: <https://github.com/Synerise/booking-challenge>
- Metoda wykorzystująca centroidy: <https://github.com/mikwieczorek/centroids-reid>
- WildNLP: <https://github.com/MI2DataLab/WildNLP>

Doktorantka weryfikuje swoje hipotezy empirycznie, poprzez opisane powyżej testy i porównania z konkurencyjnymi lub alternatywnymi algorytmami i metodami. Opisane przez doktorantkę badania naukowe mogą zostać zreplikowane dzięki dostępności algorytmów. Stwierdzam, że przyjęta przez doktorantkę metodologia badawcza jest właściwa i nie budzi zastrzeżeń.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Doktoranta wymieniła w sekcji 1.3 („Original Elements of the Thesis”) osiem oryginalnych, własnych osiągnięć. Osiągnięcie numer 8 jest ogólne i dotyczy jej znacznego udziału w testowaniu wszystkich hipotez i implementacji wszystkich metod opisanych w pracy. Osiągnięcie numer 7 dotyczy



POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

ポ
ー
ラ
ン
ド
日
本
情
報
工
科
大
学

tworzenia testów skalowalności metod reprezentacji grafów w przestrzeniach wielowymiarowych. Pozostałe osiągnięcia dotyczą oryginalnych metod zaproponowanych przez autorkę i opisanych w rozprawie:

1. WildNLP, metody zwiększania odporności algorytmów rekomendacyjnych działających w oparciu o dane tekstowe na błędy
2. Zastosowania algorytmu EMDE w połączeniu z siecią neuronową jako algorytmu rekomendacyjnego
3. Algorytmu Cleora, który odwzorowuje węzły grafu w przestrzeń wielowymiarową, tworząc reprezentacje grafów („graph embeddings”) jako dane wejściowe do sieci neuronowych
4. Algorytmu T-EMDE, który służy do estymacji gęstości występowania danych w przestrzeniach wielowymiarowych na podstawie podziału tych przestrzeni w oparciu o centroidy
5. Metody użycia centroidów do budowy jednomodalnych algorytmów rekomendacyjnych

Osiągnięcia opisane przez doktorantkę są odzwierciedlone w publikacjach, na których opiera się rozprawa. Rozprawa nie jest pakietem publikacji, ale jej pięć głównych rozdziałów jest powiązane z publikacjami opisującymi omawiane metody lub algorytmy. W związku z tym, oceniając osiągnięcia doktorantki, recenzent powinien wziąć pod uwagę kolejność autorów wymienionych publikacji, lub analogiczną informację wskazującą na wkład doktorantki w opisane w rozprawie metody.

Mgr inż. Barbara Rychalska jest pierwszą i główną autorką publikacji:

- A. Barbara Rychalska, Piotr Bąbel, Konrad Gołuchowski, Andrzej Michałowski, Jacek Dąbrowski and Przemysław Biecek. "Cleora: a Simple, Strong and Scalable Graph Embedding Scheme" accepted for publication in the proceedings of the International Conference on Neural Information Processing (ICONIP 2021)

Publikacje dotyczące pozostałych wymienionych przez autorkę osiągnięć wszystkie mają dwóch głównych autorów o równorzędnych wkładach. Jest to sytuacja często spotykana w naukach przyrodniczych, ale rzadziej w informatyce. Mimo to jest to zrozumiałe wobec szerokich zainteresowań i obszernego zakresu badań doktorantki, które realizowała pracując zespołowo. Jest także dowodem jej zdolności do zespołowej pracy badawczej. Jednakże w mojej ocenie, uwzględniając wkład doktorantki w publikacje metod, uznaję algorytm Cleora za jej najważniejsze, oryginalne osiągnięcie.



POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

ポ
ー
ラ
ン
ド
日
本
情
報
工
科
大
学

Doktorantka jest również pierwszą autorką publikacji:

B. Barbara Rychalska, Szymon Łukasik and Jacek Dąbrowski. "Synerise Monad: a Foundation Model for Behavioral Event Data" accepted for publication in Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'23).

Publikacja ta została przyjęta na konferencję SIGIR (200 punktów MNiSW) w ramach „Industrial Track”.

W chwili pisania tej recenzji publikacja nie była jeszcze publicznie dostępna, jednak recenzent uzyskał ją od doktorantki. Publikacja opisuje model bazowy nauczony na danych behawioralnych użytkowników (kliknięcia, oceny, zakupy) w oparciu o głębokie sieci neuronowe, który wykorzystuje odwzorowania użytkowników uzyskane z algorytmów Cleora i EMDE. Model bazowy może zostać douczony na zasadzie transfer learning do przewidywania np. decyzji zakupowych użytkowników, i uzyskuje wyższą dokładność w porównaniu z konkurencyjnymi metodami. Artykuł ma charakter praktyczny i jest dość krótki.

Pozostałe publikacje doktorantki ukazały się na konferencji ICONIP. Jedna praca dotycząca algorytmu T-EMDE nie została jeszcze opublikowana poza wersją preprint w systemie Arxiv. Uwzględniając prace, w których doktorantka jednym z dwojga równorzędnych współautorów, z wagą 50%, publikacje doktorantki, w których jest główną autorką (lub równorzędną główną autorką), uzyskały 550 punktów MNiSW, co jest w mojej ocenie bardzo dobrym dorobkiem publikacyjnym.

Pozycja wyników uzyskanych przez doktorantkę w stosunku do aktualnego stanu wiedzy i techniki jest potwierdzona licznymi sukcesami w konkursach dotyczących algorytmów rekomendacyjnych i Data Science:

- SIGIR eCom Rakuten Challenge 2020, 1 miejsce przed zespołem Rakuten (organizatora konkursu)
- Twitter RecSys Challenge, 2021, 2 miejsce (przed zespołami IBM i Amazon)
- KDD Cup, 2 miejsce (przed zespołami Intel Research, czołowych uczelni chińskich)
- WSDM Booking.com Challenge, 2 miejsce (przed zespołem Amazon)

We wszystkich wymienionych konkursach wykorzystywany była algorytm Cleora.

Oryginalność wyników uzyskanych przez doktorantkę jest tym wyższa, że proponowane przez nią algorytmy i metody nie są zgodne z powszechnym trendem badań. Obecnie najczęściej podejmowane są próby zwiększania złożoności głębokich sieci neuronowych w celu poprawy dokładności klasyfikacji lub predykcji. Tymczasem prace autorki mają przede wszystkim na celu poprawę złożoności obliczeniowej bez redukcji dokładności poprzez upraszczanie, a nie komplikowanie algorytmów i sieci.

Znakomitym przykładem jest algorytm Cleora, który pozwala zastąpić złożone grafowe sieci neuronowe z mechanizmem „attention” znacznie prostszą siecią neuronową, która jako dane

ポ wejściowe otrzymuje odwzorowania węzłów grafu w wielowymiarowej przestrzeni. Odwzorowanie jest oparte na prostej koncepcji iteracyjnego mnożenia macierzy odwzorowań węzłów w przestrzeń d-wymiarową przez macierz tranzycji, zawierającej prawdopodobieństwa wybrania dowolnej krawędzi w grafie przez spacer losowy. Algorytm Cleora nie używa wag stosowanych przez grafowe sieci neuronowe i nie jest w ogóle siecią neuronową. Wyjście algorytmu Cleora może być wejściem do prostej sieci neuronowej typu feed-forward. Skuteczność tak prostego algorytmu w porównaniu z grafowymi sieciami konwolucyjnymi pokazuje, że znaczna część operacji w tych sieciach jest prawdopodobnie zbędna, co jest bardzo istotnym wynikiem naukowym w dziedzinie uczenia maszynowego.

ド **5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienie uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?**

日 Rozprawa jest napisana bardzo dobrym językiem angielskim. Ma także przejrzystą, czytelną strukturę. Wprowadzenie jest napisane bardzo dobrze, przedstawia jasno problem badawczy, jego znaczenie, oraz proponowane metody rozwiązania (w formie hipotez). Po nim następują dwa omówione powyżej rozdziały z przeglądem literatury. Rozdział czwarty zawiera zasadnicze wyniki rozprawy i jest podzielony na sześć części. Pięć pierwszych sekcji rozdziału czwartego odpowiada pięciu hipotezom rozprawy, a ostatni rozdział omawia przemysłowe zastosowania proponowanych metod w formie jednego systemu rekomendacyjnego. Ostatni (piąty) rozdział podsumowuje rozprawę.

本 Struktura rozprawy ma pewne drobne wady. Sekcje 4.1-4.4 zbudowane są według podobnego wzorca. Na początku jest podsekcja „Introduction”, a po niej podsekcja „Background”. Nazwy tych sekcji są mylące. W sekcji „Background” przedstawiana jest ogólna koncepcja omawianej metody lub algorytmu. Zawiera ona często przedstawienie nowych, istotnych problemów badawczych, które nie były opisywane we wprowadzeniu do rozprawy (rozdziale pierwszym). Podsekcja „Introduction” na początku każdego z podrozdziałów rozdziału czwartego podejmuje próby umieszczenia rozważanego problemu w szerszym kontekście i powiązania go z innymi elementami systemu. Jest to często dość mylące. Moim zdaniem, cały rozdział czwarty powinien mieć obszerne wprowadzenie, w którym przedstawiona jest architektura całego systemu rekomendacyjnego (a może także procesu jego uczenia i testowania) i miejsce, jakie zajmują w tym systemie kolejno omawiane później algorytmy i

情

報

工

科

大

学

metody. Można by wtedy połączyć wielokrotnie powtarzane podsekcje „Introduction” z podsekcjami „Background”.

Rozdział piąty („Summary”) ma zaledwie jedną stronę i zawiera właściwie jedynie powtórzenie osiągnięć wymienionych we wprowadzeniu. Brakuje sekcji „Limitations”, która omawiałaby wady lub ograniczenia proponowanych rozwiązań. Brakuje także jakichkolwiek propozycji dalszych prac badawczych (sekcji „Future work”).

6. Jakie są słabe strony rozprawy i jej główne wady?

Rozprawa ma kilka drobnych wad. Na stronie 64, w podsekcji 4.1.5, wprowadzany jest nowy problem „Node classification”, który nie był wcześniej omawiany w rozdziale opisującym miary jakości systemów rekomendacyjnych, ani w rozdziale o zbiorach danych. Następnie dokładność klasyfikacji węzłów jest używana jako miara do porównania algorytmu Cleora z konkurencyjnymi algorytmami używającymi grafowych, konwolucyjnych sieci neuronowych. Nie jest przy tym wyjaśnione, jaki jest związek dokładności klasyfikacji węzłów z jakością działania systemów rekomendacyjnych. Do klasyfikacji węzłów za pomocą algorytmu Cleora użyty jest klasyfikator oparty na regresji logistycznej. Czy ten sam klasyfikator jest używany potem w systemach rekomendacyjnych? Jeśli nie, to jaki jest związek tych wyników z działaniem systemu rekomendacyjnego opartego na Cleorze? Do porównania używane są wyłącznie małe zbiory danych, co powinno być skomentowane jako ograniczenie.

Rozdział poświęcony Cleorze pozostawia pewien niedosyt. Grafy, które są wejściem do Cleory, to zwykle różne formy sieci społecznych (relacje pomiędzy użytkownikami systemów rekomendacyjnych lub grafy dwudzielne odwzorowujące użytkowników na ocenione przez nich produkty). Czy struktura tych sieci nie ma wpływu na działanie Cleory?

Opis algorytm EMDE w rozdziale 4.2 jest miejscami niezrozumiały. Szczególnie dotyczy to opisu agregacji „sketches” na stronie 73. Opis ma błędy językowe, a wagi W pojawiają się „znikąd” – nie zostały wcześniej zdefiniowane.

W niektórych miejscach, ale dość rzadko, używane są testy statystyczne. Moim zdaniem, należałoby robić to częściej, szczególnie, jeśli porównywane są średnie. Na przykład, w rozdziale 4.4 (opis algorytmu T-EMDE), na stronie 92 (Tabela 16), porównywane są średnie z pięciu uruchomień. Dlaczego nie zrobić więcej uruchomień i nie przeprowadzić testu statystycznego?

Tabele 3-7 mają tytuł nad tabelą, co nie jest konsekwentne – dalsze tabele mają tytuły pod tabelami.

Do największych słabości rozprawy należy brak omówienia słabych stron i ograniczeń proponowanych metod. Trzeba przyznać, że empiryczna ocena proponowanych metod wykonana jest bardzo starannie, proponowane algorytmy najczęściej przewyższają konkurencję. Nie dzieje się tak jednak zawsze, a tym sytuacjom nie poświęcono w rozprawie komentarza. Każdy algorytm ma przy tym ograniczenia; na przykład, o ograniczeniach algorytmu EMDE dowiadujemy się dopiero w momencie, gdy wprowadzony jest algorytm T-EMDE, który nie ma tych ograniczeń (chodzi o statyczny podział przestrzeni). Brakuje również porównania między algorytmami EMDE i T-EMDE, a także dokładniejszego omówienia ich wad oraz zalet.

Przedstawiona w rozdziale 4.5 metoda WildNLP, służąca do testowania i poprawy odporności systemów rekomendacyjnych działających w oparciu o dane tekstowe, budzi pewne wątpliwości.

Przede wszystkim, doktorantka posłużyła się w tej części rozprawy terminologią zapożyczoną z badań nad ochroną informacji – „adwersarzy”, „podatności” i „zagrożeń”. Nie jest dla mnie jasne, czy jest to terminologia adekwatna do problemu. Przede wszystkim, jeśli chcemy rozważyć rolę adwersarzy, potrzebne jest opisanie tzw. „modelu adwersarza”. Taki model określa, jakie adwersarz ma cele i środki, i co z tego wynika. Doktorantka co pewien czas pisze o „złośliwych adwersarzach”, ale kiedy przechodzi do opisu metody testowania, skupia się na „naturalnie występujących błędach” (strona 99). Nie chodzi zatem o złośliwych adwersarzy, a jedynie o użytkowników, którzy mogą popełniać błędy we wprowadzaniu danych tekstowych.

Nie jest też jasne, czy trenowanie sieci na przykładach danych, w których występują błędy QWERTY lub przestawione litery, ma sens. Czemu system rekomendacyjny nie może być wyposażony w prosty moduł, który za pomocą słowników poprawia takie błędy?

Pomimo dużej liczby publikacji doktorantki, pewien niedosyt budzi to, że prawie wszystkie publikacje ukazywały się na konferencji ICONIP. Nie chcę krytykować tej konferencji (choć w nowym rankingu CORE uzyskała rangę B). Uważam jednak, że artykuł dotyczący T-EMDE należałoby opublikować na innej konferencji lub w czasopiśmie.

7. Jaka jest przydatność rozprawy dla Informatyki?

Informatyka techniczna jest nauką stosowaną. Rola badań stosowanych jest dlatego kluczowa dla tej dyscypliny, ponieważ pozwalają one nie tylko na empiryczną walidację metod teoretycznych, ale także wskazują na istotne, nowe problemy badawcze, inspirujące kolejne badania teoretyczne. Recenzowana rozprawa jest doktoratem wdrożeniowym, ale nie ogranicza się do rozwiązywania wąskich,



POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

ポーランド 実践的 problemów firmy. Zamiast tego, opisane wyniki badań rozpatrują bardzo szerokie, różnorodne aspekty współczesnych systemów rekomendacyjnych oraz głębokich sieci neuronowych. Algorytmy i metody proponowane przez doktorantkę są często proste, ale bardzo dobrze przemyślane, skuteczne i inspirujące. Sądzę, że proponowane metody będą miały duży wpływ na przyszłe badania w zakresie systemów rekomendacyjnych, grafowych sieci neuronowych i sieci z mechanizmem „attention” lub „self-attention”.

WNIOSKI

ド Rozprawa przedłożona przez mgr inż. Barbarę Rychalską jest, w moim odczuciu, bardzo wartościowa. Rozprawa jest interesująca, jasno i klarownie zredagowana, i mimo pewnych mankamentów opisanych w recenzji, ma bardzo dużą wartość naukową.

日 Podsumowując stwierdzam, że **rozprawa spełnia wymagania stawiające rozprawom doktorskim przez obowiązujące przepisy z wyraźnym nadmiarem.**

本 Wniosuję o wyróżnienie rozprawy na podstawie jej oryginalności oraz wysokiej praktycznej skuteczności zaproponowanych algorytmów.

情報
報工
科大
学

prof. dr hab. Adam Wierzbicki