



Prof. dr hab. Marcin Szpyrka  
Akademia Górniczo-Hutnicza  
Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej  
Katedra Informatyki Stosowanej  
Al. Mickiewicza 30, 30-059 Kraków  
tel.: 012 617 51 94  
e-mail: mszpyrka@agh.edu.pl

Kraków, 12 czerwca 2023 r.

## **Recenzja rozprawy doktorskiej mgr Barbary Rychalskiej pt. *A Multimodal and Interpretable Recommender System for Massive Datasets***

Rozprawa doktorska została opracowana na Politechnice Warszawskiej w dyscyplinie Informatyka techniczna i telekomunikacja. Rozprawa została przygotowana w ramach ministerialnego programu „Doktorat wdrożeniowy”.

### **1. Cel i zakres rozprawy**

Rozprawa doktorska mgr Barbary Rychalskiej dotyczy algorytmów wykorzystywanych przy konstruowaniu multimodalnych systemów rekomendacyjnych, które pracują na bardzo dużych zbiorach danych. We współczesnym świecie obserwujemy niezwykle dynamiczny wzrost zastosowań uczenia maszynowego. Przykładem takiego obszaru zastosowań są systemy rekomendacyjne, których podstawową funkcją jest analiza danych dotyczących użytkownika i ekstrakcja przydatnych informacji do użycia ich w celu rekomendacji. Są one często projektowane w celu umożliwienia użytkownikom szybkiego zlokalizowania poszukiwanych przedmiotów, przy jednoczesnym ograniczeniu ilości dostarczanych informacji. Ich działanie ma kluczowe znaczenie dla biznesu. Skutecznie przewidując preferencje użytkownika możemy zwiększyć sprzedaż, a co za tym idzie zwiększyć zyski firmy.

Podstawowym problemem współczesnego świata wydaje się ilość i różnorodność informacji, którą musimy przetwarzać. Wymaga to poszukiwania nowych, bardziej wydajnych algorytmów, które radzą sobie z tymi problemami. W tym kontekście rozprawa doktorska Pani Barbary Rychalskiej doskonale wpisuje się w te potrzeby.

Podstawowym celem rozprawy jest analiza zdolności i ograniczeń istniejących systemów rekomendacji, oraz zaproponowanie nowych metod dla systemów multimodalnych pracujących na dużych zbiorach danych.

W podrozdziale 1.2 doktorantka zdefiniowana 5 tez, które stara się udowodnić w swojej rozprawie doktorskiej:

1. Niektóre formy grafowych sieci neuronowych (GNN) spotykane w systemach rekomendacji, można uprościć, usuwając optymalizację wag i nieliniowości, aby uzyskać czysto nienadzorowany algorytm oparty na ważonym uśrednianiu zagnieżdżeń węzłów. Taki uproszczony algorytm może zapewnić wysoką jakość działania. Jednocześnie może on być znacznie szybszy w uczeniu niż pełna grafowa sieć neuronowa.

2. Modelowanie przestrzeni elementów użytkownika oparte na estymacji gęstości pozwala zwiększyć jakość rekomendacji top- $k$ .
3. Wyszukiwanie obrazów oparte na centroidach w połączeniu z modelem neuronowym pozwala zwiększyć jakość i szybkość jednomodalnych rekomendacji treści opartych na obrazach.
4. Transformacja osadzeń z niepasujących przestrzeni cech do ujednocionej przestrzeni cech pozwala zwiększyć jakość i szybkość rekomendacji treści multimodalnych (cross-modal).
5. Zwiększenie odporności modeli na niektóre przykłady mylące pozwala wzmocnić model przed powiązanymi przykładami myłymi.

Rozprawa została przygotowana w ramach programu „Doktorat wdrożeniowy” we współpracy z firmą Synerise. Opracowane rozwiązania zostały wdrożone w ramach dwóch platform komercyjnych, które są dostępne na rynku i osiągają sukces biznesowy.

## 2. Struktura i zawartość rozprawy

Opiniowana rozprawa doktorska została przygotowana w języku angielskim i wydana w formie monografii w 2023 r. Praca składa się z pięciu rozdziałów wliczając w to wstęp i podsumowanie. Zawiera ponadto streszczenie w języku polskim i angielskim, spis rysunków, spis tabel oraz wykaz literatury. Praca liczy łącznie 124 strony. Wspomniany wykaz literatury zawiera łącznie 141 nienumerowanych pozycji.

- Rozdział 1 zawiera wprowadzenie do problematyki pracy, w szczególności przedstawiono cele badań i zdefiniowano 5 wcześniej przytoczonych tez. Rozdział zawiera również zestawienie najważniejszych wyników naukowych doktorantki, wykaz wybranych publikacji doktorantki, w tym prace bezpośrednio powiązane z rozprawą doktorską. Opisano również znaczenie wypracowanych wyników oraz w skrócie przedstawiono strukturę rozprawy.

Warto w tym miejscu dodać, że w systemie Google Scholar widocznych jest 19 publikacji doktorantki (pominięto prace zamieszczone w arXiv), które mają łącznie 230 cytowań (h-index = 7). 5 prac widocznych jest w systemie Web of Science. Mają one łącznie 13 cytowań (h-index = 2). Jest to bardzo dobry wynik w odniesieniu do doktoranta.

- Rozdział 2 zawiera krótkie wprowadzenie do teorii sieci neuronowych. Autorka skupia się głównie na bardzo ogólnym opisanu podstawowych typów sieci neuronowych, które mogą być wykorzystywane przy budowaniu systemów rekomendacyjnych.
- Rozdział 3 stanowi wprowadzenie do tematyki systemów rekomendacyjnych. Omówiono w nim typy systemów rekomendacyjnych oraz przedstawiono metryki, które są wykorzystywane do oceny ich działania. W rozdziale znalazły się również podstawowe informacje na temat interpretowalności i niezawodności modeli rekomendacji oraz multimodalności zbiorów danych wejściowych, na których te systemy pracują. Rozdział kończy przegląd ogólnodostępnych zestawów danych, które doktorantka wykorzystywała do testowania rozwiązań zaproponowanych w rozprawie.
- Rozdział 4 stanowi główną część ocenianej rozprawy doktorskiej. Zawarto w nim opis autorских wyników badań prowadzonych przez autorkę rozprawy.

Pierwszy omawiany problem dotyczy kwestii osadzania różnego typu danych wejściowych (tekst, grafika, struktury grafowe) tak, aby ich reprezentacja pozwoliła na wykorzystanie tych danych przez system rekomendacyjny. Pierwszy zaproponowany przez doktorantkę algorytm

dotyczy osadzania danych o strukturze grafowej. Wychodząc od problemu z wykorzystaniem grafowych konwolucyjnych sieci neuronowych do przetwarzania bardzo dużych grafów (miliony węzłów i miliardy krawędzi) doktorantka pokazała, że całkowite usunięcie wag  $W$  (trenowalna macierz wag z  $i$ -tej warstwy) i nieliniowości  $\sigma$  (wzór 26, str. 59) prowadzi do wykonywalnego, pozbawionego większości problemów skalowalności, choć nieneuronowego algorytmu (algorytm Cleora). Przedstawione w pracy wyniki eksperymentów potwierdziły wysoką jakość osadzania węzłów z użyciem tej metody.

Drugi z omawianych algorytmów dotyczy rekomendacji tok- $k$  bazującej na wielu modalnościach. Algorytm ten wykorzystuje wcześniej wspomniane osadzanie węzłów Cleora, wraz z osadzeniami dla innych modalności (tekst, obrazy). Główny rozważany problem dotyczy kwestii, jak połączyć różne historyczne zdarzenia danego użytkownika w jedną reprezentację, która mogłaby stworzyć zagregowany profil behawioralny. Jako podstawę systemu rekomendacji zaproponowano wielowymiarowy estymator gęstości. Jako metodę estymacji gęstości wykorzystano Efficient Manifold Density Estimator (EMDE) opracowany w firmie Synerise. Zaproponowany został model bazujący na rzadkiej sieci neuronowej, która na wejściu otrzymuje zagregowane szkice EMDE reprezentujące część historii zakupów użytkownika, a na wyjściu dostarcza szkic EMDE brakującej części historii zakupów użytkownika, który należy przewidzieć. Eksperymenty przeprowadzone przez doktorantkę potwierdziły, że podejście takie zwiększa jakość rekomendacji dla rekomendacji top- $k$ . Jednocześnie metoda pozwala na łatwe włączanie dodatkowych modalności.

Kolejny z przedstawionych algorytmów dotyczy jednomodalnych rekomendacji treści opartych na obrazach. Dotyczy on sytuacji, gdy dla danego elementu nie istnieją jeszcze żadne interakcje, a jedynym sposobem na zarekomendowanie elementu mogą być jego cechy. Doktorantka zaproponowała podejście, w którym zamiast obliczania podobieństw z poszczególnymi elementami w galerii, wyznaczone są podobieństwa między zapytaniem, a reprezentacjami całych galerii, zwanych centroidami (np. dla wielu fotografii tego samego obiektu wyznacza się jeden centroid). Algorytm wykorzystuje zmienioną metodę Triplet Loss nazywaną w pracy Centroid Triplet Loss (CTL). Eksperymenty z danymi pokazały, że proponowany model zwiększa jakość i szybkość wyszukiwania unimodalnego. Dodatkowo metoda może być stosowana dla dowolnej modalności (innej niż grafika).

Czwarty z opisanych algorytmów dotyczy wyszukiwania między modalnego, tj. dopasowywania elementów w oparciu o dwie różne modalności, np. wyszukanie obrazu na podstawie opisu tekstowego. Kluczowy problem w porównywaniu elementów dotyczy ich reprezentacji numerycznej. Doktorantka wyeliminowała ten problem przekształcając najpierw reprezentacje obrazów i tekstu we wspólną abstrakcyjną przestrzeń reprezentacji (ponownie z wykorzystaniem koncepcji centroidów). W dalszym etapie porównywane są już takie przekształcone reprezentacje. Doktorantka zaproponowała algorytm określony jako T-EMDE (wytrenowana wersja EMDE, która może kodować dowolną modalność) o złożoności liniowej (konkurencyjna metoda ma złożoność kwadratową).

Ostatnia z omawianych metod dotyczy analizy odporności modeli na przykłady mylące. Doktorantka skupiła się na domenie tekstowej. Podrozdział 4.5 rozpoczyna się od badania, które dane wejściowe (części zdania) są najbardziej newralgiczne z perspektywy przykładów mylących, tj. zmiana/błąd w której części zdania może mieć najbardziej negatywny wpływ na funkcjonowanie systemu rekomendacji. W dalszej części podrozdziału badane są przykłady mylące i sposoby ich zwalczania. Analizy były prowadzone użyciem benchmarku WildNLP, który symuluje błędy w tekście, np. zmiana partykuł, wprowadzanie typowych błędów ortograficznych, zamiana sąsiadujących na klawiaturze liter itp. Ponieważ podejście to pozwala generować duże ilości uszkodzonych danych, doktorantka zaproponowała wykorzystanie takich danych przy tre-

nowaniu modelu (adversarial training). W założeniu miało to wprowadzić odporność modelu na przykłady mylące. Wyniki eksperymentów potwierdziły lepsze wyniki po zastosowaniu proponowanego podejścia.

- Ostatni rozdział zawiera krótkie jednostronicowe podsumowanie rozprawy.

Prezentacja materiału przedstawionego w pracy dokonana jest w sposób relatywnie czytelny i zrozumiały. Stosowana terminologia i symbole nie budzą w zasadzie zastrzeżeń. Drobne uwagi dotyczące składu i języka prezentacji zawarto w dalszej części recenzji.

### 3. Ocena rozprawy

Główne wyniki naukowe opisane w rozprawie doktorskiej dotyczą modyfikacji wybranych elementów systemów rekomendacyjnych w celu dopasowania ich do wydajnego radzenia sobie z dużymi zbiorami danych. Biorąc pod uwagę wdrozeniowy charakter rozprawy doktorskiej i firmę, w której pracuje doktorantka, przyjęte cele są w pełni zrozumiałe. Rozprawa jako całość nie stanowi rozwiązania jednego problemu naukowego. Są to raczej efekty poszukiwania tych elementów funkcjonowania systemów rekomendacyjnych, w których można było uzyskać redukcję kosztów obliczeniowych. Duże znaczenie dla przedstawionych wyników badań mają eksperymenty obliczeniowe zrealizowane przez doktorantkę, w przypadku niektórych zadań również opracowane metody testowania/porównywania rozwiązań. Pokazują one, że uzyskane wyniki wypadają lepiej lub co najmniej nie gorzej od wyników uzyskanych przez współczesne wiodące rozwiązania.

Do najważniejszych osiągnięć doktorantki przedstawionych w ocenianej rozprawie doktorskiej należy zaliczyć:

1. Wykazanie, że grafowe konwolucyjne sieci neuronowe, używane do obliczania osadzenia węzłów, mogą być uproszczone do ważonego mnożenia macierzy, które zachowuje jakość osadzenia i oferuje znaczny wzrost szybkości i skalowalności.
2. Wykazanie, że metoda estymacji gęstości dla danych multimodalnych (EMDE) odpowiednio połączona z siecią neuronową może służyć jako wysokowydajny rekomendator top- $k$ .
3. Wykazanie, że przy wyszukiwaniu jednomodalnym można skutecznie wykorzystywać wyszukiwanie oparte na centroidach.
4. Zaproponowanie algorytmu T-EMDE wykorzystywanego przy wyszukiwaniu międzymodalnym.
5. Zaproponowanie metody zwiększania odporności na przykłady mylące dla danych tekstowych poprzez wykorzystanie na etapie uczenia modelu danych wygenerowanych za pomocą WildNLP.
6. Praktyczna weryfikacja zaproponowanych rozwiązań na danych testowych i wykazanie, że spełniają one założone wymagania dotyczące wydajności przy pracy na dużych zbiorach danych.

#### 3.1. Uwagi

1. Str. 37 – „Formally, a graph  $G$  can be defined as  $G = (V, E)$ , where  $V$  is the set of nodes, and  $E$  are the edges between them.” – Trudno ten zapis uznać za definicję grafu. Jeżeli decydujemy się użyć słowa „formally”, to należałoby podać kompletną definicję (uzupełnić opis  $E$ ).

2. W sekcji 3.4.1 (str. 46) przedstawiona została algorytm rozłożenia macierzy Opis w tym rozdziale sugeruje, że operator  $\times$  użyty na rys. 9 oznacza mnożenie macierzy. Jeżeli rzeczywiście tak jest, to trudno macierz *Item Rating Matrix* uznać za poprawny wynik mnożenia dwóch macierzy wskazanych po prawej stronie.
3. Metoda radzenia sobie z przykładami mylącymi opisana w podrozdziale 4.5 została opracowana kilka lat temu (publikacje z roku 2018 i 2019) i dotyczy wyłącznie analizy danych tekstowych. Czy były prowadzone jakiegokolwiek próby zastosowania podobnego podejścia dla innych modalności (grafika, grafy)?
4. Bibliografia nie jest numerowana. Trudno też określić jaki jest porządek sortowania pozycji bibliograficznych. W rezultacie poszukiwanie jakiegokolwiek pozycji w bibliografii wymaga przejrzania całości.
5. Żadna z publikacji bezpośrednio powiązanych z rozprawą doktorską nie została opublikowana w dobrym czasopiśmie. Są to praktycznie wyłącznie referaty konferencyjne. Dlaczego, skoro eksperymenty obliczeniowe przedstawione w rozprawie wskazują na wysoką jakość uzyskanych wyników naukowych?

### 3.2. Uwagi techniczno-redakcyjne

Pod względem edytorskim oceniana rozprawa doktorska lokuje się zdecydowanie powyżej przeciętnej. Została dość starannie złożona z użyciem systemu składu L<sup>A</sup>T<sub>E</sub>X. Tym niemniej autorka nie ustrzegła się kilku „typowych” błędów, np.: brak stosowania polecenia `\mathit` dla wieloliterowych nazw w trybie matematycznym, zbędne wcięcia dla kontynuacji akapitu po wyśrodkowanych wzorach matematycznych.

## 4. Wniosek końcowy

Przedstawiona w pracy problematyka dotyczy aktualnych i interesujących zagadnień naukowych związanych z problemem multimodalnych systemów rekomendacyjnych pracujących na dużych zbiorach danych. Rozprawa doktorska zawiera szereg interesujących oryginalnych wyników pracy badawczej autorki, które zostały z sukcesem wdrożone w dwóch komercyjnych systemach. Nieliczne drobne uwagi nie wpływają na pozytywną ocenę rozprawy.

Stwierdzam, że autorka wykazała się ogólną wiedzą teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja, trafnie sformułowała problem naukowy i następnie przedstawił jego oryginalne rozwiązanie. Podsumowując, stwierdzam, że recenzowana rozprawa doktorska pt. *A Multi-modal and Interpretable Recommender System for Massive Datasets* spełnia wymagania stawiane rozprawom doktorskim przez ustawę *Prawo o Szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 r. (z późniejszymi zmianami) i wnoszę o dopuszczenie mgr Barbary Rychalskiej do publicznej obrony.

Biorąc pod uwagę dorobek publikacyjny doktorantki bezpośrednio powiązany z rozprawą doktorską (tabela 1, str. 23), w szczególności pozycje bibliograficzne, w których doktorantka jest pierwszym autorem, wnoszę o wyróżnienie rozprawy doktorskiej.