

Gliwice, 4.10.2023

Prof. dr hab. inż. Marcin Woźniak
Wydział Matematyki Stosowanej
Politechnika Śląska

Sz. P.
Prof. dr hab. inż. Robert Nowicki
Przewodniczący
Rady Dyscypliny Naukowej
Informatyka Techniczna i Telekomunikacja
Politechnika Częstochowska

Recenzja rozprawy doktorskiej mgr inż. Piotra Woldana pt.: „Hybrid deep learning structures for recommender systems” opracowanej pod kierunkiem dr hab. Piotra Dudy, Prof. PCz.

W odpowiedzi na pismo R-WIMil-510-8/18 z dnia 14.07.2023 przedkładam niniejszą recenzję.

Przedstawiona do recenzji praca została napisana w języku angielskim i składa się z 8 głównych rozdziałów obejmujących w sumie 85 stron. Praca przedstawia badany temat oraz osiągnięcia Doktoranta, streszczenia w języku polskim, spisu rysunków, spisu tabel oraz bibliografii. Bibliografia składa się 87 pozycji, wśród których niestety nie ma wymienionej żadnej pracy Doktoranta. Rozprawa zawiera 36 rysunków oraz 6 tabel.

W pierwszym rozdziale Doktorant opisuje podstawowe pojęcia związane z systemami rekomendacyjnymi. Rozdział został podzielony na trzy podrozdziały. W pierwszym z nich wyróżnione zostały dwa typy systemów, to jest systemy oparte na treści (ang. Content-Based Filtering) oraz systemy podejmujące decyzję na podstawie opinii użytkowników (ang. Collaborative Filtering). Rozdział zawiera charakterystykę systemów oraz wspomniane zostały w nim narzędzia pozwalające je zrealizować. Przedstawiono przykłady unikając jednak opisu szczegółowego, co jest nietypowe w pracy naukowej. Inne typy systemów rekomendacyjnych, jak systemy hybrydowe, systemy oparte o wiedzę oraz systemy wykorzystujące modele uczenia maszynowego zostały jedynie wspomniane. Kolejny podrozdział dotyczy trudności związanych z oceną działania systemów rekomendacyjnych. Pokazane zostały najpopularniejsze kryteria oceny systemów uczących się, wraz ze wzorami je opisującymi. Zostały również przedstawione miary dla systemów rekomendacyjnych, w tym przypadku jednak zabrakło wzorów. W dalszej części znajduje się dyskusja nad problemem tzw. zimnego startu. Rozdział pierwszy kończy część w której wskazane są obszary zastosowań systemów rekomendacyjnych.

W rozdziale drugim Doktorant przedstawia motywację do podjęcia problemu badawczego, co finalizuje sformułowaniem hipotezy badawczej. Jako cel pracy przedstawiono analizę nowoczesnych technik przetwarzania danych dla systemów rekomendacyjnych oraz umożliwiających interpretację przesłanek. Brakuje tu jednak jednoznacznego zdefiniowania co Doktorant definiuje jako interpretowalna rekomendacja.

Rozdział trzeci to spis podstawowych informacji dotyczących pięciu zbiorów danych wybranych przez Doktoranta do przeprowadzenia eksperymentów numerycznych. Możemy się z niego dowiedzieć niezbędnych informacji, takich jak liczba danych, liczba klas abstrakcji, liczba atrybutów czy rozdzielczość obrazów.

W rozdziale czwartym opisane są podstawowe modele sieci neuronowych powszechnie stosowane w analizie danych, np. autoencodery, ograniczona maszyna Boltzmanna, sieci konwolucyjne oraz algorytm tSNE. Doktorant skupia się na przekazaniu idei stojącej za każdym z wymienionych modeli. Dobór opisanych narzędzi jest rozsądny. Jednak brakuje opisu mechanizmu uwagi, będącego jednym z kluczowych elementów takich systemów.

Rozdział piąty szczegółowo opisuje proponowane rozwiązanie. Pierwszym krokiem w proponowanym systemie jest wstępne przygotowanie danych. Różne procedury zostały zaproponowane dla atrybutów nominalnych oraz, jak pisze Doktorant, nie-nominalnych.

W przypadku atrybutów nominalnych zaproponowano zastosowanie binaryzacji z wieloma jedynekami (ang. Multi-Hot-Encoding). W przypadku atrybutów numerycznych zalecono przeskalowanie ich do przedziału (0, 1]. Doktorant nie wyjaśnił dlaczego w procedurze na rysunku 5.2 wartości bardzo bliskie wartości minimalnej są przekształcane w wartość bliską zeru, a wartość dokładnie równa wartości minimalnej na wartość 0.001. Następnie dane przetwarzane są przez sieci neuronowe. W przypadku danych numerycznych jest to dwuwarstwowa sieć typu fully-connected zakończona jednym neuronem wyjściowym. W przypadku danych nominalnych neuron wyjściowy poprzedzony jest warstwą konwolucyjną. Ostatecznie, tak przetworzone atrybuty stanowią wejście dla sieci regresyjnej posiadającej jedną warstwę ukrytą. Brak jest informacji o testowaniu przez Doktoranta innych struktur sieci. W rozdziale 5.2.3 pojawia się opis uczenia sieci typu deep belief network, brak jednak wskazania jak odnosi się on do prezentowanego wcześniej rozwiązania. Wyniki rekomendacji zostały zilustrowane na danych pochodzących ze zbioru MovieLens 20M, będącego jednym z najpopularniejszych benchmarków dla systemów rekomendacyjnych typu collaborative filtering. W ostatnim podrozdziale opisana jest kwestia interpretowalności uzyskanych rekomendacji. Proponowane jest rzutowanie danych obiektów (tu filmów) na przestrzeń trójwymiarową, a następnie znalezienie grupy najbliższych sąsiadów. Podkreślono, że przy użyciu tej metody możliwe jest wyszukanie nie tylko obiektów najbardziej podobnych, które należy zarekomendować, ale również tych które nie powinny być rekomendowane. W tekście autor rozróżnia pojęcia grupowania danych oraz ich klastrowania. Podsumowując rozdział ten prezentuje proponowaną przez Doktoranta procedurę wyszukiwania rekomendowanych w podejściu collaborative filtering. Interpretowalność została tu przedstawiona w formie wizualizacji najbliższych sąsiadów dla danego obiektu według preferencji użytkownika.

W rozdziale szóstym podejmowany jest problem stworzenia hybrydowego systemu rekomendacyjnego opartego o treść. System ten, przeznaczony do rekomendowania obszarów na zdjęciach, które dla użytkownika mogą być szczególnie istotne,

inspirowany jest mechanizmem uwagi Bahdanau. Metoda została przedstawiona na tle zbioru danych Zappo50K, zawierającego zdjęcia butów podzielone na 5 zasadniczych kategorii, które następnie są rozdzielane na 21 klas. Liczba elementów w klasach jest wysoko niezbalansowana. W pierwszym kroku Doktorant zaproponował stworzenie klasyfikatora w postaci konwolucyjnej sieci neuronowej złożonej z 13 warstw, gdzie 3 ostatnie warstwy są typu full-connected. Skuteczność sieci została zbadana pod kilkoma kryteriami charakterystycznymi dla danych niezbalansowanych. W kolejnym kroku nazwa kategorii głównej i klasy abstrakcji zostały tokenizowane, by służyły jako obiekty poddane mechanizmowi uwagi. Ostatecznie tworzona jest nowa sieć konwolucyjna do klasyfikacji wykorzystująca istotność informacji wyznaczoną przez mechanizm uwagi. Przedstawiony mechanizm, poprzez połączenie informacji z ostatniej warstwy konwolucyjnej i informacji z mechanizmu uwagi pozwala wyświetlić najistotniejsze obszary na obrazie. Wektor wyliczony przez ostatnią warstwę ukrytą klasyfikatora w trzecim kroku zostaje użyty do wyznaczenia najbardziej podobnych obrazów.

W rozdziale siódmym Doktorant opisuje algorytm Occlusion Sensivity. Autor proponuje rekurencyjną wersję tej metody mającą umożliwić wyznaczanie obszarów o różnym kształcie. W podrozdziale 7.3 przedstawiony jest pseudokod proponowanego rozwiązania. Podejście używane do klasyfikacji zdjęć zostało przebadane z wykorzystaniem zbioru ImageNet, natomiast do segmentacji z wykorzystaniem zbioru COCO2017. W obu przypadkach skorzystano z pre-trenowanej sieci EfficientNetB7. W podrozdziale 7.6 przedstawiona została modyfikacja popularnego algorytmu GradCam. Działanie tego algorytmu polega na przesłaniu informacji z warstwy wyjściowej w głąb sieci w celu odczytania najistotniejszych regionów dla zadania klasyfikacji. Algorytm umożliwia wykorzystanie różnych kryteriów przesłania informacji w głąb sieci. Doktorant proponuje własną modyfikację polegającą na przesyłaniu również negatywnych gradientów.

W ostatnim rozdziale Doktorant podsumowuje zaproponowane przez siebie podejścia oraz wskazuje konieczność dalszego rozwijania metod służących rekomendacji, w szczególności dostosowywania do tego celu dynamicznie rozwijających się modeli głębokich sieci neuronowych.

Ocena merytoryczna przedstawionej rozprawy doktorskiej:

Przedłożona rozprawa doktorska mgr inż. Piotra Woldana dotyczy metod uczenia maszynowego dedykowanych analizie obrazów i detekcji cech charakterystycznych. Badania pokazują, że najbardziej skutecznymi w tej dziedzinie są metody oparte o głębokie sieci neuronowe. Doktorant skutecznie modyfikuje nowoczesne podejścia do wyznaczania istotnych fragmentów zdjęć. Praca pokazuje zastosowanie głębokiego uczenia w systemach rekomendacyjnych.

Atutem pracy jest udostępnienie przez Doktoranta kodów źródłowych. Praca napisana jest poprawnie i zwięźle, choć w kilku miejscach brakuje należytej staranności w naukowym opisie badań. Praca została napisana w języku angielskim. Dzięki temu uniknięto konieczności tłumaczenia nazw i terminów z dziedziny uczenia maszynowego, które zdecydowanie zdominowane jest przez literaturę angielskojęzyczną.

Po przeczytaniu przedstawionej rozprawy doktorskiej nasuwają się następujące pytania i sugestie:

- i. Jak Doktorant definiuje interpretowalność w przypadku rekomendacji wskazanych przez system? Jakie systemy pozwalają na tę interpretowalność, a jakie nie? Czy wszystkie systemy zaproponowane w pracy spełniają tę definicję?
- ii. Dla procedury przedstawionej na rysunkach 5.3 do 5.5 należy wyjaśnić jakie znaczenie ma zastosowana struktura sieci neuronowej. Z czego wynika dobór liczby warstw, liczby neuronów, czy rozmiaru kernela w sieci konwolucyjnej?
- iii. W jaki sposób sieć StackedRBM prezentowana w sekcji 5.2.3 odnosi się do poprzedniej części rozdziału? Jaki wpływ na wykorzystanie metody contrastive divergence na zastąpienie brakujących wartości w danych uczących?
- iv. W metodzie wizualizacji danych przedstawionej w podrozdziale 5.4 zastosowano algorytm t-SNE. Jakie znaczenie ma użycie właśnie tego algorytmu? Czy można użyć innego algorytmu rzutowania danych na mniejszą podprzestrzeń?
- v. W jaki sposób została przeprowadzona tokenizacja? Dlaczego autoencoder był uczony tylko przez 10 epok? Jaki był przebieg jego uczenia? W pracy nie została zamieszczona należyta informacja na ten temat.

- vi. Należy doprecyzować, jak się ma wymiar ostatniej warstwy konwolucyjnej przedstawiony na początku rozdziału 6.3 do struktury klasyfikatora z rysunku 6.2.
- vii. Czy do mechanizmu przedstawionego w rozdziale szóstym można w jakiś sposób dodać informację o preferencjach indywidualnego użytkownika?
- viii. Jak Doktorant rozumie hybrydowość systemu zaproponowanego w rozdziale 6?

Ocena formy, taksonomii, języka i edycji pracy:

Przedstawiona rozprawa doktorska jest napisana poprawnym językiem, choć w kilku miejscach można dostrzec różnice w czcionce i numeracji rysunków oraz przedstawianych opisach. Edycja i forma dokumentu są poprawne. Rozprawę napisano w języku angielskim, a na jej końcu dołączono streszczenie w języku polskim. Przedstawione rysunki, tabele i wzory zostały ponumerowane stosownie do przynależności do rozdziałów opracowania.

Ocena przedstawionego dorobku naukowego:

Doktorant jest współautorem dziesięciu artykułów opublikowanych zarówno w czasopiśmie, jak w wolumenach dla międzynarodowych konferencjach. Większość jego prac skupia się na systemach rekomendacyjnych, natomiast kilka z nich wykracza poza ten obszar. Bezpośrednio do tematu odnoszą się dwa artykuły.

Przedstawiony dorobek naukowy Doktoranta pokazuje zaangażowanie w prace naukowe, które zostały uznane w międzynarodowym środowisku naukowym. Moim zdaniem zaprezentowany dorobek pozwala stwierdzić opanowanie Doktoranta z dyscypliną prowadzenia badań naukowych, analizą wyników i podejmowaniem wniosków z prowadzonych badań naukowych.

Podsumowanie:

Przedstawiona do recenzji rozprawa doktorska porusza temat zastosowania modeli uczenia maszynowego do analizy obrazów. Uczenie maszynowe za pomocą modeli głębokich sieci neuronowych jest tematem wielu aktualnych opracowań naukowych.

Jest to temat ważny z punktu widzenia rozwoju informatyki. Doktorant rozwiązał problem naukowy dotyczący opisu danych wejściowych przy pomocy autorskiego modelu i wykonał testy dla podstawowych architektur głębokich sieci neuronowych. Wyniki pokazały, że zastosowanie proponowanego modelu przetwarzania pozwala na zwiększenie wydajności.

Na podstawie przedstawionej powyżej oceny stwierdzam, iż rozprawa doktorska mgr. inż. Piotra Woldana pt.: „Hybrid deep learning structures for recommender systems” spełnia warunki określone w art. 13 ust 1 ustawy z dnia 14 marca 2003r. o stopniach naukowych i tytule naukowym (Dz. U. z 2017 r. poz.1789) stosowane w postępowaniach o nadanie stopnia doktora. **Wnoszę o przyjęcie w/w rozprawy doktorskiej i dopuszczenie jej do publicznej obrony w dyscyplinie informatyka.**