



Politechnika Wrocławska

Wydział Informatyki i Zarządzania

kierunek studiów: Informatyka

specjalność: Projektowanie Systemów Informatycznych

Praca dyplomowa - magisterska

HYBRYDOWY ALGORYTM REKOMENDACJI
WYKORZYSTUJĄCY METODY FILTROWANIA Z
ANALIZĄ ZAWARTOŚCI I FILTROWANIA
KOLABORATYWNEGO

HYBRID RECOMMENDER ALGORITHM BASED ON CONTENT-BASED FILTERING
AND COLLABORATIVE FILTERING

Katarzyna Biernat

słowa kluczowe:
KEYWORDS

krótkie streszczenie:
SHORT ABSTRACT

Promotor:	dr inż. Bernadetta Maleszka
	<i>imię i nazwisko</i>	<i>ocena</i>	<i>podpis</i>

Do celów archiwalnych pracę dyplomową zakwalifikowano do:*

a) kategorii A (akta wieczyste)

b) kategorii BE 50 (po 50 latach podlegające ekspertyzie)

* niepotrzebne skreślić

pieczęć wydziałowa

Wrocław 2016

Niniejszy dokument został złożony w systemie L^AT_EX.

Spis treści

Todo list	1
Rozdział 1. Wstęp	3
Rozdział 2. Przegląd istniejących rozwiązań	5
2.1. Filtrowanie w oparciu o zawartość	5
2.1.1. Metody tworzenia profilu użytkownika	6
2.1.2. Zalety filtrowania w oparciu o zawartość	6
2.1.3. Wady filtrowania w oparciu o zawartość	6
2.2. Filtrowanie kolaboratywne	6
2.2.1. Zalety filtrowania kolaboratywnego	8
2.2.2. Wady filtrowania kolaboratywnego	8
2.2.3. Algorytmy hybrydowe	9
2.3. Popularne serwisy wykorzystujące algorytmy rekomendacji	9
2.3.1. Rekomendacja muzyki	9
2.3.2. Rekomendacja filmów	9
2.3.3. Platformy typu e-commerce	9
Rozdział 3. Model systemu	11
Rozdział 4. Algorytmy	13
4.1. Filtrowanie kolaboratywne	13
4.1.1. Matrix Factorization	13
4.1.2. Biased Matrix Factorization	17
4.1.3. SVD++	18
4.2. Filtrowanie w oparciu o zawartość	19
4.2.1. Konstrukcja sieci neuronowej	19
4.2.2. Uczenie sieci neuronowej	21
4.2.3. Propagacja wsteczna	22
4.2.4. Algorytm RPROP	24
4.2.5. Algorytm genetyczny	25
4.3. Algorytmy hybrydowe	26
4.3.1. Zalety zaproponowanej metody	26
4.3.2. Wady zaproponowanej metody	27
4.4. Analiza złożoności i poprawności	27
Rozdział 5. Ocena eksperymentalna	29
5.1. Opis metody badawczej	29

5.1.1.	Miara oceny	29
5.1.2.	Zbiory danych	30
5.2.	Środowisko symulacyjne	31
5.2.1.	Oprogramowanie	31
5.2.2.	Sprzęt	32
5.3.	Przeprowadzone eksperymenty	32
5.3.1.	Dopasowanie parametrów sieci neuronowej	33
5.3.2.	Porównanie algorytmu hybrydowego z filtrowaniem w oparciu o zawartość i filtrowaniem kolaboratywnym	37
5.3.3.	Analiza statystyczna wyników	37
Rozdział 6. Wnioski		39
Rozdział 7. CHAPTER 1		41
7.1.	SECTION	41
7.2.	Section 2	41
7.2.1.	Subsection 1	41
Dodatek A. Appendix 1		43
Bibliografia		45

ABSTRACT PL

Streszczenie

ABSTRACT EN

Abstract

Todo list

■ CEL PRACY: trzeba rozszerzyć opis, w szczególności o to, co jest nowego i czym różni się od innych już istniejących algorytmów	4
■ streszczenie spisu treści	4
■ krótka wzmianka o tym, że w kolejnych podrozdziałach będą opisane wybrane techniki - te, które będą wykorzystywane w Pani systemie	5
■ ktoś może Pani zarzucić, że to wada, bo nie korzystamy z informacji, które są dostępne	6
■ Zalety filtrowania kolaboratywnego	8
■ opisać istniejące algorytmy hybrydowe i na czym polegają etc	9
■ Napisać wstęp. Co powinien zawierać?	13
■ czyby będzie lepiej jednak przenieść te algorytmy do rozdziału z przeglądem, bo tutaj raczej czytelnik oczekuje już konkretnego algorytmu— omówić reorganizację	13
■ OMÓWIĆ dwa ostatnie zdania brzmią razem tak, jakby wynik był losowy .	16
■ to będzie początek rozdziału	19
■ NAPISAĆ TUTAJ WIĘCEJ o tych wszystkich wariantach, dodać analizę własności	26
■ Analiza złożoności i poprawności	27
■ Opisać eksperyment w punktach - co się po kolei dzieje	29
■ Dodać pseudokod	29
■ Przenieść niżej w rozdziale	29
■ i od razu podkłada się Pani recenzentowi... Dwa zdania temu pisała Pani o tym, że MAE i RMSE są dobre dla tych samych danych, a teraz że będzie Pani porównywać trzy różne zbiory	30
■ Opisać Yahoo Music	31
■ Opisać Amazon Meta	31
■ Do parametrów dodać symbole takie jak były wcześniej wykorzystane w modelu	31
■ Napisać tu ładny wstęp	32
■ Opis i schemat eksperymentów	32
■ ?? Jak bardzo wchodzić w szczegóły?	32

■ Zweryfikować, czy nie kłamię	33
■ Koniecznie TODO	37
■ Czy trzeba?	37
■ Czy trzeba	37
■ Porównanie algorytmu hybrydowego z content-based i collaborative	37
■ Analiza statystyczna wyników, PQStat	37
■ Wnioski	39

Rozdział 1

Wstęp

Wraz z rozwojem Internetu zmienił się sposób dostępu do informacji. Kiedyś to użytkownik musiał walczyć pozyskanie wiedzy; dzisiaj to informacje walczą u uwagę użytkowników. W świecie zalanym wiadomościami koniecznym wydaje się być zastosowanie filtra, który odsieje interesującą i wartościową zawartość od tej niechcianej. Pomocne okazują się zautomatyzowane mechanizmy rekomendacji.

Jednakże sama idea rekomendacji nie jest niczym nowym. Co więcej, zjawisko to możemy zaobserwować w naturze – na przykład wśród mrówek, które podążają wyznaczoną (rekomendowaną) ścieżką feromonową w poszukiwaniu pożywienia.

Ludzie od niepamiętnych czasów posiłkowali się opiniami innych, aby ułatwić sobie dokonanie wyboru, od najbliższego grona znajomych do ekspertów i autorytetów.

Wraz z rozwojem nauk informatycznych problem rekomendacji stał się problemem interesującym badaczy. Za pierwszy system rekomendacji uznaje się *Tapestry* stworzony w laboratoriach Xerox Palo Alto Research Center w 1992 roku. Motywacją było odfiltrowanie rosnącej liczby niechcianej poczty elektronicznej [16].

Wkrótce później idea ta została rozszerzona przez takich graczy jak Amazon, Google, Pandora, Netflix, Youtube, Yahoo etc. aż do formy, jaką znamy dzisiaj: systemu, który sugeruje użytkownikom produkty, filmy, muzykę, strony internetowe na podstawie ich aktywności w sieci [47].

Wielkie koncerny internetowe stale poprawiają jakość swoich algorytmów rekomendacji. Najlepszym przykładem jest tutaj Netflix, który w październiku 2006 zorganizował ogólnodostępny konkurs na najlepszy algorytm. Zadaniem uczestników było ulepszenie algorytmu Cinematch. Już po siedmiu dniach od ogłoszenia konkursu trzy zespoły zdołały poprawić wynik Cinematch o 1.06% [33, 35]. 18 września 2009 Netflix ogłosił, że zespół BellKor's Pragmatic Chaos poprawił Cinematch o 10,06% osiągając wynik $RMSE = 0.8567$. Tym samym wygrał nagrodę w wysokości \$1,000,000 i zakończył konkurs [34, 36].

Systemy rekomendacji ulepszone są nieustannie, o czym świadczy chociażby organizowana rokrocznie konferencja *ACM International Conference on Recommender Systems*. Tematyka ta poruszana jest także na konferencjach *European Conference on Information Retrieval*, *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* i wielu innych. Mimo dużego stopnia

zaawansowania wciąż istnieje pole manewru do ulepszania algorytmów rekomendacji i co za tym idzie zwiększanie zadowolenia użytkowników, które z kolei prowadzi do osiągania korzyści biznesowych.

Celem pracy jest opracowanie i zbudowanie hybrydowego algorytmu rekomendacji. Składowymi docelowego algorytmu są metody kolaboratywnego filtrowania oraz metody filtrowania z analizą treści.

W kolejnych rozdziałach przedstawione są istniejące oraz proponowane ulepszenia, które pozwalają na jeszcze lepsze dopasowanie wyników rekomendacji do oczekiwań użytkowników.

CEL PRACY: trzeba rozszerzyć opis, w szczególności o to, co jest nowego i czym różni się od innych już istniejących algorytmów

streszczenie spisu treści

Rozdział 2

Przegląd istniejących rozwiązań

Tradycyjnie wyróżniamy następujące techniki rekomendacji:

- o **filtrowanie w oparciu o zawartość** (ang. *content-based filtering*), technika koncentrująca się na atrybutach elementów. Użytkownikowi rekomendowane są elementy, które podobne są do tych wybieranych przez niego w przeszłości;
- o **filtrowanie kolaboratywne** (ang. *collaborative filtering*), technika polegająca na odnajdywaniu użytkowników o podobnych gustach i sugerowaniu lubianych przez nich elementów aktualnie aktywnemu użytkownikowi;
- o **filtrowanie demograficzne** (ang. *demographic filtering*), technika koncentrująca się na sugerowaniu aktywnemu użytkownikowi elementów popularnych wśród użytkowników z tej samej okolicy bądź w podobnym przedziale wiekowym;
- o **filtrowanie z analizą domeny wiedzy** (ang. *knowledge-based filtering*), technika dobierająca kolejne elementy na podstawie określonej domeny wiedzy na temat tego, jak dany element spełnia potrzeby i preferencje użytkownika;
- o **filtrowanie z analizą społecznościową** (ang. *community-based filtering*), technika dobierająca rekomendacje dla użytkownika w zależności od preferencji innych użytkowników z jego sieci społecznościowej. W myśl zasady „powiedz mi kim są twoi przyjaciele a powiem ci kim jesteś”;
- o **hybrydowe systemy rekomendacji** to kombinacja dowolnych powyższych technik.

Każda z tych technik ma swoje wady i zalety w zależności od kontekstu, w którym ma być stosowana [40].

2.1. Filtrowanie w oparciu o zawartość

Filtrowanie w oparciu o zawartość opiera się na cechach elementów w systemie. Rekomendowane są obiekty, które podobne są do tych pozytywnie ocenionych wcześniej przez użytkownika [17]. W zależności od domeny pod uwagę mogą być brane słowa kluczowe, cechy takie jak rok wydania, reżyser, autor, kompozytor, gatunek itp.

krótka
wzmianka
o tym, że w
kolejnych
podrozdzia-
łach będą
opisane wy-
brane tech-
niki - te,
które będą
wykorzysty-
wane w Pani
systemie

2.1.1. Metody tworzenia profilu użytkownika

Profil użytkownika może być tworzony na dwa sposoby. Jeżeli użytkownik jawnie pozostawia informacje można mówić bezpośredniej informacji zwrotnej (ang. *explicit feedback*). Do takich informacji należą: ocena konkretnych elementów, przycisk lubienia lub nie lubienia (tzw. „łapka” w górę lub w dół), komentarz itp.

Jednakże nawet jeżeli użytkownik nie jest skory do zostawiania tego typu śladów, to i tak można wiele na jego temat wywnioskować korzystając z pośredniej informacji zwrotnej (ang. *implicit feedback*). System bierze wówczas pod uwagę aktywność użytkownika taką jak: historia zakupów, historia przeglądarki a nawet ruchy myszką. W przypadku serwisu z muzyką czy filmem cenną informacją będzie fakt, czy użytkownik wysłuchał lub obejrzał dany materiał do końca czy też wyłączył go po paru sekundach [24, 29].

2.1.2. Zalety filtrowania w oparciu o zawartość

Do zalet filtrowania w oparciu o zawartość należy niezależność użytkownika. Podczas budowania rekomendacji brany pod uwagę jest tylko jego profil; aktywność innych aktorów w systemie nie wpływa na wynik końcowy. Inną przewagą jest przejrzystość – każda propozycja jest w pełni uzasadniona, gdyż opiera się na działaniach użytkownika w przeszłości. Ponadto, tego typu algorytm ma możliwość zaproponowania elementu, który nie był nigdy wcześniej oceniany przez nikogo. Zapobiega to zjawisku długiego ogona [29].

ktoś może
Pani zarzu-
cić, że to
wada, bo nie
korzystamy
z informa-
cji, które są
dostępne

2.1.3. Wady filtrowania w oparciu o zawartość

Aby rekomendacja była skuteczna użytkownik powinien ocenić jak najwięcej elementów. Problematiczni są zatem użytkownicy, którzy dopiero co dołączyli do serwisu oraz tacy, którzy nie są aktywni i rzadko zostawiają po sobie ślad [30].

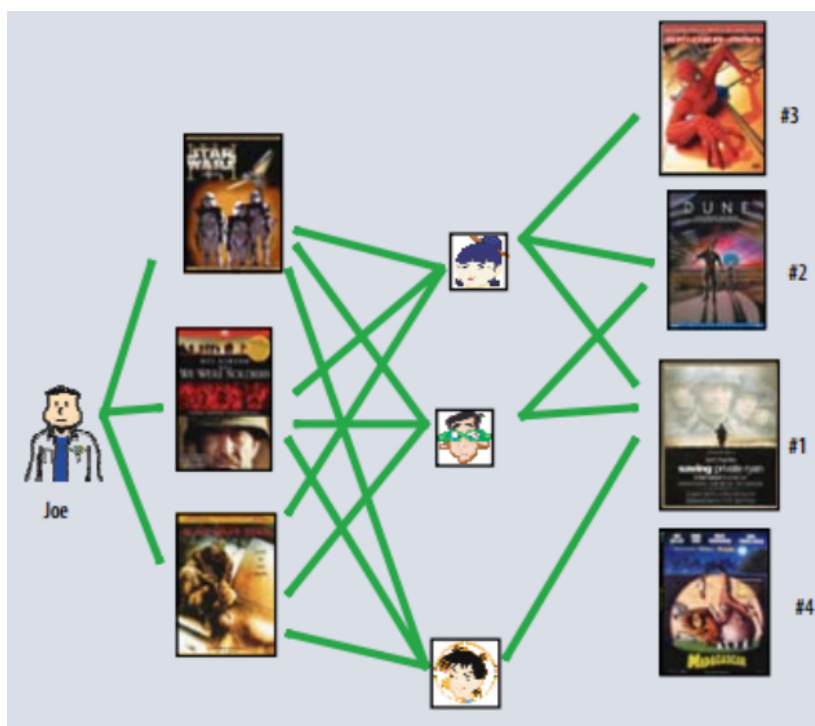
Filtrowanie w oparciu o zawartość jest podatne na pułapkę tzw. bańki informacyjnej. Jeżeli w systemie rekomendującym produkcje kinowe użytkownik do tej pory oceniał jedynie filmy akcji, to mało prawdopodobne jest, że algorytm zaproponuje mu ciekawy dramat obyczajowy. Nowe propozycje nie są zaskakujące [29].

2.2. Filtrowanie kolaboratywne

Filtrowanie kolaboratywne opiera się na założeniu, że ludzie o zbliżonym guście dokonują podobnych wyborów. Użytkownicy o zbliżonym guście to osoby, które oceniły konkretne elementy w podobny sposób [40, 46, 17].

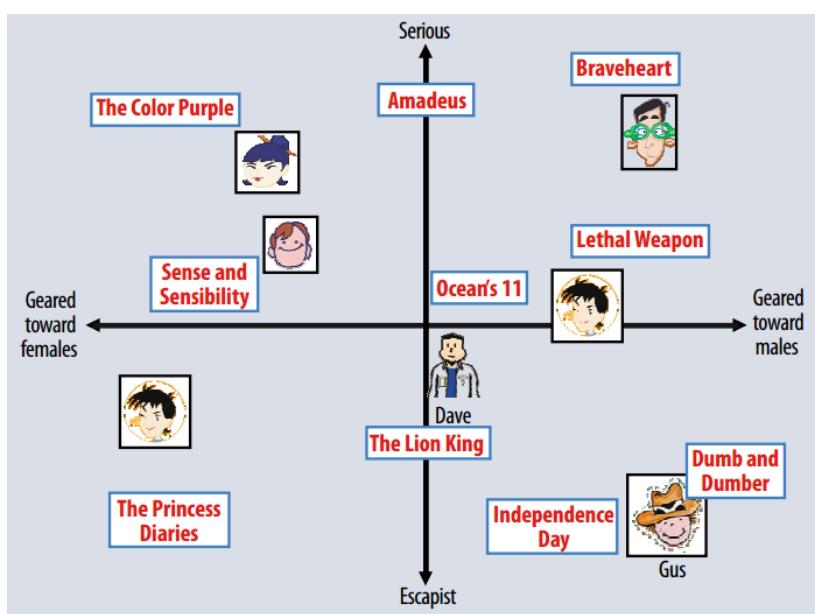
W przypadku filtrowania kolaboratywnego można wyróżnić dwa główne podejścia: oparte o regułę sąsiedztwa (ang. *neighborhood*) oraz oparte o model (ang. *model-based*), wykorzystujące modele ukrytych parametrów [24, 25].

Filtrowanie w oparciu o regułę sąsiedztwa koncentruje się na związkach element–element bądź użytkownik–użytkownik [24]. Rysunek 2.1 pokazuje regułę sąsiedztwa skoncentrowaną na relacji użytkownik–użytkownik. Joe ocenił trzy filmy. System odnajduje innych użytkowników, którzy ocenili te trzy pozycje podobnie jak Joe. Każdy z nich



Rys. 2.1: Filtrowanie kolaboratywne metodą sąsiedztwa, zorientowane na użytkownika [25].

pozytywnie ocenił film „Saving Private Ryan”, zatem jest to pierwsza rekomendacja dla Joe.



Rys. 2.2: Filtrowanie kolaboratywne z wykorzystaniem modelu ukrytych parametrów [25].

Ideą podejścia w oparciu o model jest zbadanie i modelowanie zależności element–użytkownik

wraz z czynnikami reprezentującymi ukryte własności elementów i użytkowników. Taki model jest następnie uczony przy użyciu dostępnych danych. W rezultacie można z niego odczytać przewidywaną ocenę elementu dla konkretnego użytkownika [8, 24].

Rysunek 2.2 pokazuje w sposób uproszczony podejście oparte o model. W układzie współrzędnym oznaczeni są użytkownicy wedle swoich preferencji oraz konkretnych cech (np. płeć) a także filmy, które stanowią odpowiedź na dany zestaw preferencji/cech [25].

2.2.1. Zalety filtrowania kolaboratywnego

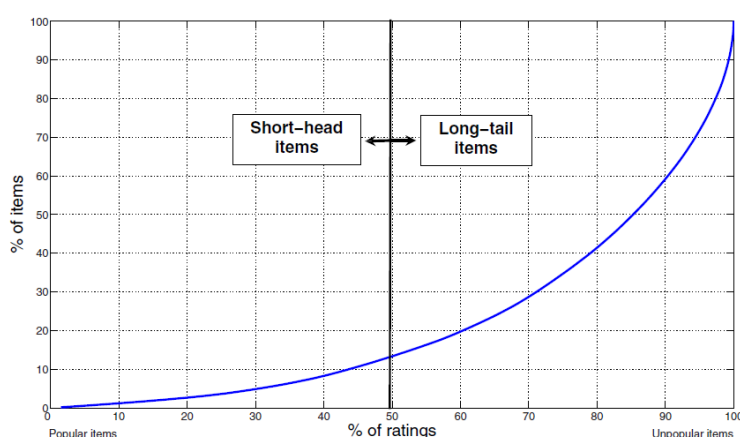
Zalety filtrowania kolaboratywnego

2.2.2. Wady filtrowania kolaboratywnego

Jednym z problemów klasycznego podejścia do kolaboratywnego filtrowania jest brak uwzględnienia dynamiki zmian w gustach użytkowników. Ten sam użytkownik na przestrzeni kilku lat lub miesięcy może zupełnie inaczej ocenić ten sam film bądź piosenkę. Rozwiązaniem jest dodanie czynnika czasu podczas obliczania wag kolejnych ocen. [6, 20, 25].

Innym problemem jest tzw. zimny start (ang. *cold start*). Polega on na tym, że nowi użytkownicy w systemie ocenili zbyt mało elementów, aby można było zbudować dla nich dobre rekomendacje [44, 50].

Powszechnym zjawiskiem jest tzw. efekt długiego ogona. Rysunek 2.3 przedstawia jak rozkłada się procentowa ilość ocen danych elementów w zależności od ich popularności. Jeżeli algorytm rekomendacji nie wspiera mniej popularnych elementów, to istnieje ryzyko, że użytkownicy nie otrzymają możliwości eksplorowania nowych, niszowych materiałów [5, 44].



Rys. 2.3: Problem długiego ogona: 50% ocen dotyczy 10-12% najpopularniejszych elementów w systemie [44].

Systemy rekomendacji wykorzystujące filtrowanie kolaboratywne nie są skalowalne. Złożoność rośnie proporcjonalnie do ilości użytkowników i elementów. Wielkie koncerny

internetowe takie jak Twitter wykorzystają klastry i maszyny z bardzo dużą pamięcią aby zachować płynność działania serwisu [12].

2.2.3. Algorytmy hybrydowe

opisać ist-
niejące al-
gorytmy
hybrydowe
i na czym
polegają etc

2.3. Popularne serwisy wykorzystujące algorytmy rekomendacji

W przeciągu ostatnich lat algorytmy rekomendacji stały się nieodłącznym elementem wielu popularnych serwisów internetowych z różnych domen. Poniższa lista prezentuje wybrane produkty.

2.3.1. Rekomendacja muzyki

- o **YouTube** – serwis powstały w 2005 roku, pozwalający na bezpłatne umieszczanie, odtwarzanie, ocenianie i komentowanie filmów. Od 2006 roku przejęty przez Google. YouTube buduje profil użytkownika w oparciu o jego aktywność w serwisie. Brane pod uwagę są polubienia, subskrypcje, udostępnianie a także informacje czy użytkownik obejrzał film do końca czy tylko pewien jego procent. Techniki rekomendacji stosowane przez serwis to przede wszystkim wydobywanie reguł asocjacyjnych i licznik wspólnych odwiedzin danego wideo w czasie trwania pojedynczej sesji [7].
- o **LastFM** – internetowa radiostacja oferująca rozbudowany mechanizm rekomendacji piosenek "Audioscrobber".
- o **Pandora** – spersonalizowane radio internetowe wykorzystujące projekt Music Genome Project. Każda piosenka przeanalizowana jest pod kątem maksymalnie 450 cech; na tej podstawie budowane są rekomendacje [1].

2.3.2. Rekomendacja filmów

- o **Netflix** – amerykańska platforma oferująca strumieniowanie filmów i seriali. Działający od 2007 roku gigant oferuje rozbudowany system rekomendacji Cinematch [39].
- o **Filmweb** – polski serwis poświęcony filmom i jednocześnie druga największa baza filmowa na świecie. Oferuje system rekomendacji Gustomierz, który umożliwia poznawanie nowych filmów w guście użytkownika [9].
- o **Internet Movie Database (IMDb)** – największa internetowa baza filmów. Baza zawiera 3,837,014 pozycji, które są oceniane w skali od 1 do 10 przez użytkowników [19].

2.3.3. Platformy typu e-commerce

- o **Allegro** – polski portal aukcyjny. Swoim użytkownikom oferuje panel rekomendacji. Prezentowane produkty wybierane są w oparciu o to, co dotychczas kupował i oglądał użytkownik [2].

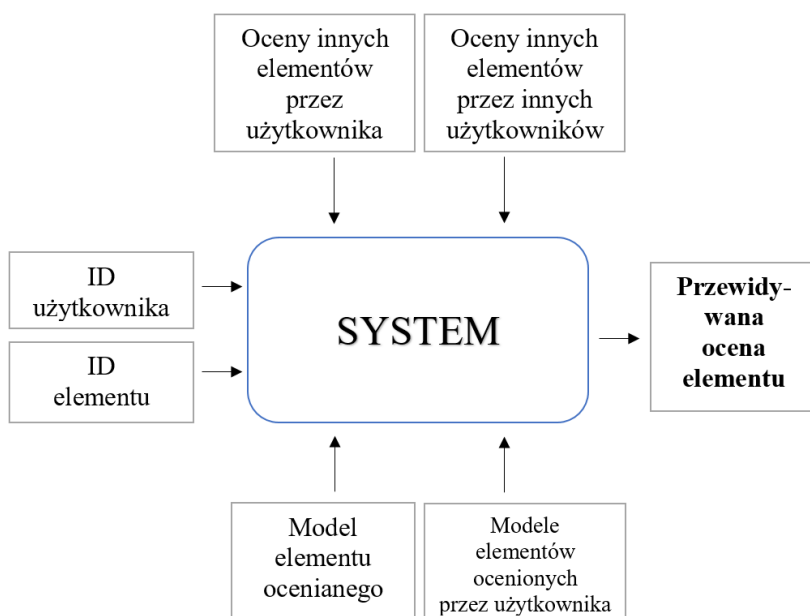
- **Amazon** – największy na świecie sklep internetowy typu B2C. Amazon w swoich mechanizmach rekomendacji wykorzystuje algorytmy filtrowania kolaboratywnego typu element–element [28].

Rozdział 3

Model systemu

Głównym założeniem systemu zaproponowanego przez autorkę jest połączenie zalet kolaboratywnego filtrowania i filtrowania w oparciu zawartość minimalizując jednocześnie wady obu podejść.

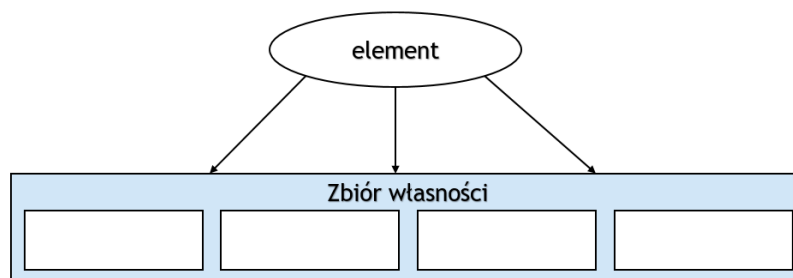
Rys. 3.1 przedstawia czarnoskrzynkowy model systemu. Danymi wejściowymi są numery identyfikacyjne użytkownika, dla którego ma być zbudowana rekomendacja oraz elementu, dla którego ma być przewidziana ocena. System pobiera model elementu a także modele wszystkich innych elementów, które użytkownik ocenił w przyszłości. Jednocześnie pobierane są informacje o tym, jak użytkownicy systemu ocenili inne elementy. Wynikiem wyjściowym jest predykcja – jak aktywny użytkownik oceni element.



Rys. 3.1: Model czarnoskrzynkowy

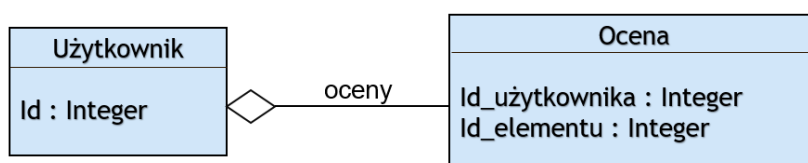
Założeniem systemu jest uniwersalność, zatem model elementu jest uogólniony i dostosowuje się w zależności do domeny, w której system jest wykorzystywany. Rys.

3.2 przedstawia reprezentację elementu w systemie. W zbiorze wartości mogą znaleźć się takie pozycje jak lista aktorów, reżyser (w przypadku filmów), gatunek, wykonawca (w przypadku muzyki), typ produktu lub cena (w przypadku systemów typu e-commerce).



Rys. 3.2: Uogólniony model elementu

Każdy użytkownik systemu jest anonimowy. Nie jest znana jego płeć, wiek, pochodzenie itp. System nie przechowuje także informacji właściwych mediom społecznościowym, takich jak relacje między użytkownikami (przyjaźnie, śledzenie). Wiadomym jest jedynie, jakie elementy zostały ocenione i jak zostały ocenione. Rys. 3.3 przedstawia reprezentację użytkownika w systemie.



Rys. 3.3: Uogólniony model elementu

Rozdział 4

Algorytmy

Napisać
wstęp. Co
powinien
zawierać?

4.1. Filtrowanie kolaboratywne

Implementacja algorytmów filtrowania kolaboratywnego, które wykorzystane zostały w projekcie pochodzą biblioteki MyMediaLite [10, 11].

Przed podjęciem decyzji dotyczącej wyboru algorytmu kolaboratywnego filtrowania zostały przeanalizowane testy na bazie MovieLens M1 [13]. Testy przeprowadzone zostały z pięciokrotną walidacją krzyżową. Rys. 4.1 przedstawia wyniki (im mniejsza wartość RMSE i MAE tym lepiej).

Najefektywniejsze okazały się algorytmy SVD++, Biased Matrix Factorization i Matrix Factorization bazujące na podejściu w oparciu o model.

chyba będzie
lepiej jednak
przenieść te
algorytmy
do rozdziału
z przeglą-
dem, bo tu-
taj raczej
czytelnik
oczekuje już
konkretnego
algorytmu—
omówić reor-
ganizację

4.1.1. Matrix Factorization

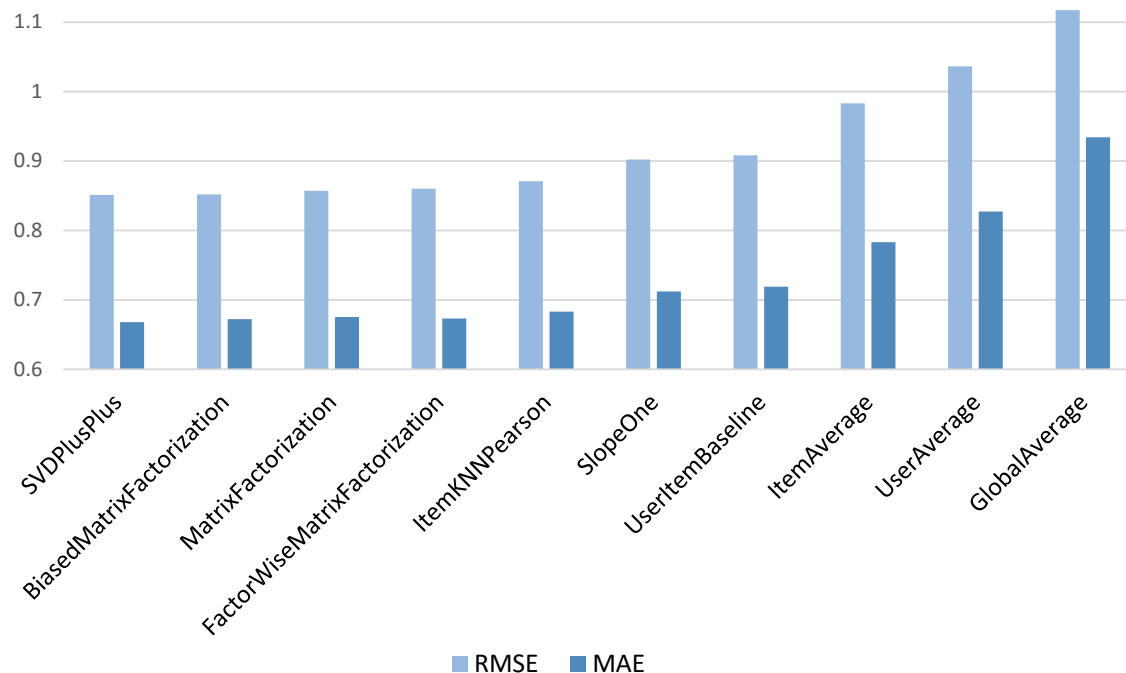
Metoda matematyczna

Algorytm *Matrix Factorization* bazuje na matematycznej metodzie rozkładu macierzy na czynniki.

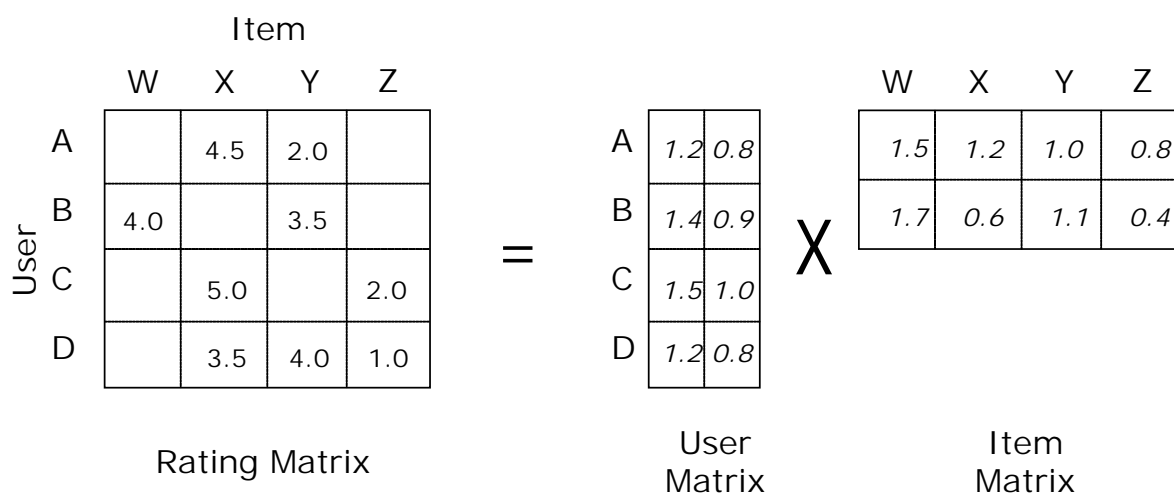
Początkowo dana jest niekompletna macierz zawierająca oceny, jakie użytkownicy wystawili konkretnym elementom (*Rating Matrix*). Celem metody jest odnalezienie wartości, jakie można wstawić w puste miejsca, czyli przewidzenie jaką ocenę dany użytkownik wystawi nieocenionemu jeszcze elementowi.

W tym celu tworzone są odrębne macierze dla użytkowników i elementów zawierające ukryte własności. Każdy element powiązany jest z wektorem $q_i \in \mathbb{R}^f$ (wektory W, X, Y, Z na rys. 4.2) a każdy użytkownik z wektorem $p_u \in \mathbb{R}^f$ (wektory A, B, C, D na rys. 4.2). Wartości czynników ukrytych determinują stopień zainteresowania daną cechą (w przypadku macierzy użytkowników) bądź stopień, w jakim dany element posiada tę cechę (w przypadku macierzy elementów).

Iloczyn skalarny $q_i^T p_u$ przedstawia relację pomiędzy użytkownikiem a elementem. Na tej podstawie można wnioskować ocenę r_{ui} , jaką użytkownik może wystawić ele-



Rys. 4.1: Test algorytmów filtrowania kolaboratywnego [32]



Rys. 4.2: Faktoryzacja macierzy [43]

mentowi: $r_{ui} = q_i^T p_u$ i w konsekwencji estymować zainteresowanie użytkownika danym elementem [25].

Głównym wyzwaniem jest odnalezienie wartości macierzy użytkownika i elementu, które po przemnożeniu przez siebie dadzą kompletną macierz ocen. W przypadku omawianych algorytmów stosowana jest metoda stochastycznego gradientu prostego.

Stochastyczny gradient prosty

Stochastyczny gradient prosty (ang. *stochastic gradient descent*, *SGD*) jest iteracyjnym algorytmem optymalizacyjnym mającym za zadanie odnalezienie minimum bądź

maksimum zadanej funkcji. Jest on uproszczeniem popularnej metody gradientu prostego [4].

Dana jest funkcja celu w postaci

$$Q(w) = \sum_{i=1}^n Q_i(w). \quad (4.1)$$

Zadaniem algorytmu jest odnalezienie takiej wartości parametru w , dla którego $Q(w)$ będzie minimalne. Wykorzystując klasyczną metodę gradientu prostego poszukiwania można zapisać następującym wzorem:

$$w := w - \eta \sum_{i=1}^n \nabla Q_i(w), \quad (4.2)$$

gdzie η symbolizuje współczynnik uczenia (ang. *learning rate*). W przypadku algorytmu stochastycznego następuje uproszczenie. Zamiast obliczać dokładny gradient $Q(w)$, w każdej iteracji jest on aproksymowany na podstawie pojedynczego, losowo wybranego przypadku:

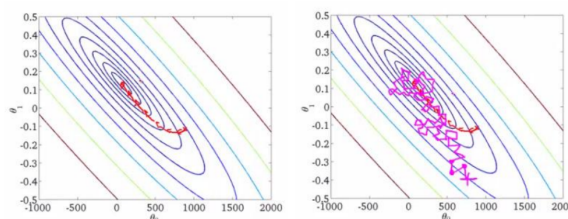
$$w := w - \eta \nabla Q_i(w). \quad (4.3)$$

Algorytm przetwarza wszystkie elementy zbioru treningowego i dla każdego przypadku wykonuje aktualizację wartości w . Przebieg algorytmu przedstawiony jest poniżej (algorytm 1).

Algorytm 1 Stochastyczny gradient prosty

```
 $w \leftarrow$  początkowy wektor  $w$ 
 $\eta \leftarrow$  współczynnik uczenia
while nie odnaleziono minimum do
  losowo mieszkaj przykłady ze zbioru testowego
  for  $i \in \{1, 2, \dots, n\}$  do
     $w := w - \eta \nabla Q_i(w)$ 
  end for
end while
```

Stochastyczny gradient prosty wykonuje dużo więcej kroków niż jego klasyczny odpowiednik aby odnaleźć rozwiązanie optymalne. Mimo tego jest on szybszy, gdyż każdy pojedynczy krok jest mniej kosztowny niż w oryginale. Różnicę w działaniu przedstawia rys. 4.3.



Rys. 4.3: Różnica pomiędzy klasycznym gradientem prostym (po lewej) a jego stochastyczną wersją (po prawej) [3]

Przebieg algorytmu Matrix Factorization

Przebieg algorytmu składa się z dwóch faz. W pierwszej fazie inicjowany jest model (zob. algorytm 2). Daną wejściową jest macierz zawierająca dotychczasowe oceny elementów przez użytkowników w systemie. Na wyjściu otrzymywane są dwie nowe macierze reprezentujące ukryte własności użytkowników i elementów. Na tym etapie są one wypełnione wartościami losowymi.

Algorytm 2 Matrix Factorization – Inicjacja modelu

$N \leftarrow$ liczba użytkowników
 $M \leftarrow$ liczba elementów
 $F \leftarrow$ liczba ukrytych własności
 $ratings \leftarrow$ Macierz $N \times M$ zawierająca dotychczasowe oceny wszystkich elementów przez wszystkich użytkowników
 $user_factors \leftarrow$ Macierz $N \times F$ reprezentująca ukryte własności użytkowników
 $item_factors \leftarrow$ Macierz $M \times F$ reprezentująca ukryte własności elementów
for each $uf \in user_factors, if \in item_factors$ **do**
 wstaw losową wartość za pomocą transformacji Boxa–Mullera
end for
for each $user, item \in ratings$ **do**
 if $ratings_{user,item} = NULL$ **then**
 wstaw 0 do wiersza $user_factors_{user}$ i $item_factors_{item}$
 end if
end for
return $user_factors, item_factors$

W fazie drugiej następuje uczenie metodą stochastycznego gradientu prostego (zob. algorytm 3). Wynikiem tej fazy są macierze reprezentujące ukryte własności użytkowników i elementów. Mnożąc je ze sobą uzyskiwana jest przewidywana ocena każdego z elementów przez użytkowników.

Przed rozpoczęciem uczenia ustalane są parametry: parametr regulujący (ang. *regularization*), współczynnik uczenia (ang. *learning rate*), parametr zanikania (ang. *decay*) i liczba iteracji. W trakcie trwania głównej pętli parametr regulujący pozostaje niezmienny. Służy on zapobieganiu zjawisku nadmiernego dopasowania (ang. *overfitting*). Tempo uczenia jest przy każdym przebiegu pętli mnożone przez parametr zanikania,

OMÓWIĆ
dwa ostat-
nie zdania
brzmia ra-
zem tak,
jakby wynik
był losowy

dzięki czemu można kontrolować w jakim stopniu kolejne przebiegi pętli wpływają na finalny wynik.

Ostatnim parametrem ustalonym przed główną pętlą jest skośność globalna (ang. *global bias*), która jest średnią wszystkich znanych ocen.

W pętli uczenia wykonywane są następujące operacje: dla każdej pary użytkownik – element budowana jest przewidywana ocena poprzez obliczenie iloczynu skalarnego odpowiednich wartości z macierzy wartości ukrytych. Ocena ta jest modyfikowana poprzez dodanie globalnej skośności a następnie porównywana z faktyczną oceną elementu przez użytkownika. Tak uzyskany błąd służy do wyliczenia delty. Macierze wartości ukrytych uaktualniane są o wyliczoną deltę.

Pod koniec każdej iteracji aktualizowany jest współczynnik uczenia.

Algorytm 3 Matrix Factorization – Faza uczenia

```

global_bias ← średnia wszystkich ocen
X ← liczba iteracji
regularization ← parametr regulujący
current_learnrate ← współczynnik uczenia
decay ← paramert zanikania
for each x ∈ X do
  for each user, item ∈ ratings do
    prediccion = global_bias + IloczynSkalarny(user_factorsuser, item_factorsitem);
    error = ratingsuser, item − prediccion
    //dopasowanie własności ukrytych:
    for each f ∈ F do
      deltau = error * item_factorsitem, f − regularization * user_factorsuser, f
      deltai = error * user_factorsuser, f − regularization * item_factorsitem, f
      user_factorsuser, f += current_learnrate * deltau
      item_factorsitem, f += current_learnrate * deltai
    end for
  end for
  current_learnrate *= decay
end for
return user_factors, item_factors

```

4.1.2. Biased Matrix Factorization

Algorytm *Biased Matrix Factorization* jest modyfikacją wyżej opisanego algorytmu *Matrix Factorization*. Podobnie jak jego pierwowzór składa się z dwóch faz. W pierwszej fazie dodatkowo inicjowane są dwa dodatkowe wektory: skośność użytkowników (ang. *user bias*) i skośność elementów (ang. *item bias*).

Inaczej jest też obliczana skośność globalna:

$$global_bias = \frac{\frac{a - r_{min}}{r_{max} - r_{min}}}{1 - \frac{a - r_{min}}{r_{max} - r_{min}}}, \quad (4.4)$$

gdzie

a to średnia wszystkich ocen
 r_{min} to minimalna ocena w systemie
 r_{max} to maksymalna ocena w systemie

Faza druga wygląda podobnie jak w przypadku algorytmu *Matrix Factorization*, jednak są uwzględniane dodatkowe parametry i wykonywane dodatkowe kroki.

Algorytm 4 Biased Matrix Factorization – Faza uczenia

```

global_bias ← średnia wszystkich ocen
X ← liczba iteracji
regU, regI, BiasReg ← parametry regulujące dla użytkownika, elementu i ogólny
current_learnrate ← współczynnik uczenia
BiasLearnRate ← współczynnik uczenia skośności
decay ← paramert zanikania
for each  $x \in X$  do
  for each  $user, item \in ratings$  do
    score = global_bias + user_biasuser + item_biasitem +
      IloczynSkalarny(user_factorsuser, item_factorsitem)
    sig_score =  $\frac{1}{1+\exp(-score)}$ 
    predicion = ratingmin + sig_score + (ratingmax - ratingmin)
    error = ratingsuser,item - prediction
    gradient_common = err * sig_score * (1 - sig_score) * (ratingmax - ratingmin)
    //dopasowanie skośności:
    user_biasuser += BiasLearnRate * current_learnrate * (gradient_common -
      BiasReg * RegU * user_biasuser)
    item_biasitem += BiasLearnRate * current_learnrate * (gradient_common -
      BiasReg * RegI * item_biasitem)
    //dopasowanie własności ukrytych:
    for each  $f \in F$  do
      deltau = gradient_common * item_factorsitem,f - RegU * user_factorsuser,f
      deltai = gradient_common * user_factorsuser,f - RegI * item_factorsitem,f
      user_factorsuser,f += current_learnrate * deltau
      item_factorsitem,f += current_learnrate * deltai
    end for
  end for
  current_learnrate *= decay
end for
return user_factors, item_factors
  
```

4.1.3. SVD++

SVD++ jest rozszerzeniem metody SVD (dekompozycja głównych składowych, ang. *singular value decomposition*). Od poprzednich omawianych algorytmów różni go przede wszystkim to, że korzysta nie tylko z bezpośredniej informacji zwrotnej do tworzenia profilu użytkownika ale także z pośredniej (zob. 2.1.1).

Model SVD++ opisywany jest równaniem:

$$r_{ui} = \mu + b_u + b_i + q_i^T(p_u + \frac{1}{\sqrt{|N(u)|}} \sum_{j \in N(u)} y_j) \quad (4.5)$$

Użytkownik powiązany jest z wektorem $p_u \in \mathbb{R}^f$ reprezentującym zainteresowanie konkretnymi cechami. Taki model uzupełniany jest sumą $\frac{1}{\sqrt{|N(u)|}} \sum_{j \in N(u)} y_j$, która reprezentuje pośrednie informacje zwrotne. Przykładem informacji niejawnej jest fakt, że użytkownik w ogóle zareagował na dany element (np. ocenił go), bez względu na wynik tej interakcji.

Zmienne b_u i b_i reprezentują obserwowane odchylenie od średniej dla użytkowników i elementów, natomiast μ to średnia wszystkich ocen [23].

4.2. Filtrowanie w oparciu o zawartość

Algorytmny filtrowania w oparciu o zawartość budują rekomendację na podstawie ocen, jakie zostały dotychczas wystawione przez użytkownika. Analizowane są cechy elementów i ich wartości oraz określana jest ich siła wpływu na finalną ocenę.

W tym celu dla każdego użytkownika tworzona jest sieć neuronowa, która uczy się jego preferencji.

W projekcie wykorzystana została implementacja sieci neuronowych z biblioteki AForge.NET Framework [21].

4.2.1. Konstrukcja sieci neuronowej

Struktura perceptronów

Sieć neuronowa składa się z trzech warstw neuronów (perceptronów). W każdej warstwie wszystkie neurony mają konstrukcję na jak rys. 4.4

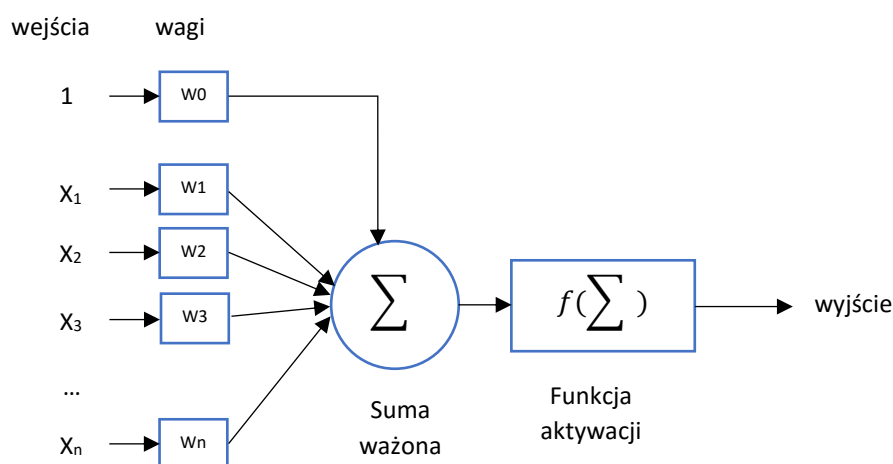
Do neuronu przekazywany jest zestaw wartości w postaci wektora x . Następnie obliczana jest suma ważona tych wartości w zależności od nadanych wag w . Proces dobierania odpowiednich wag jest nazywany uczeniem (zob. 4.2.2). W następnym kroku suma ważona przekazywana jest do funkcji aktywacji neuronu. Jeżeli funkcja przyjmie wartość wyższą lub równą niż określony próg aktywacji, to perceptron zostanie pobudzony (zwróci wartość 1). Proces ten obrazuje równanie 4.6.

$$N(x_1, x_2, x_3, \dots, x_n) = \begin{cases} 1 & \text{jeśli } f(w_0 + \sum_{i=1}^n w_i x_i) \geq \eta \\ 0 & \text{jeśli } f(w_0 + \sum_{i=1}^n w_i x_i) < \eta \end{cases}, \quad (4.6)$$

gdzie

- w to wagi kolejnych wejść
- x to wartości przekazywane do wejść
- $f(u)$ to funkcja aktywacji neuronu
- η to próg aktywacji neuronu

to będzie
początek
rozdziału

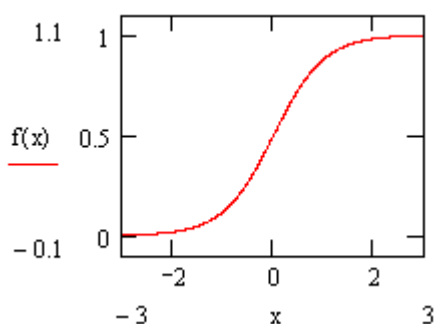


Rys. 4.4: Schemat perceptronu

Na potrzeby algorytmu rekomendacji zdecydowano się przyjąć sigmoidalną unipolarną funkcję aktywacji neuronu (równanie 4.7). Funkcja przyjmuje wartości z zakresu $[0, 1]$.

$$f(x) = \frac{1}{1 + \exp(-\alpha x)} \quad (4.7)$$

Wykres funkcji wygląda jak na rys. 4.5.

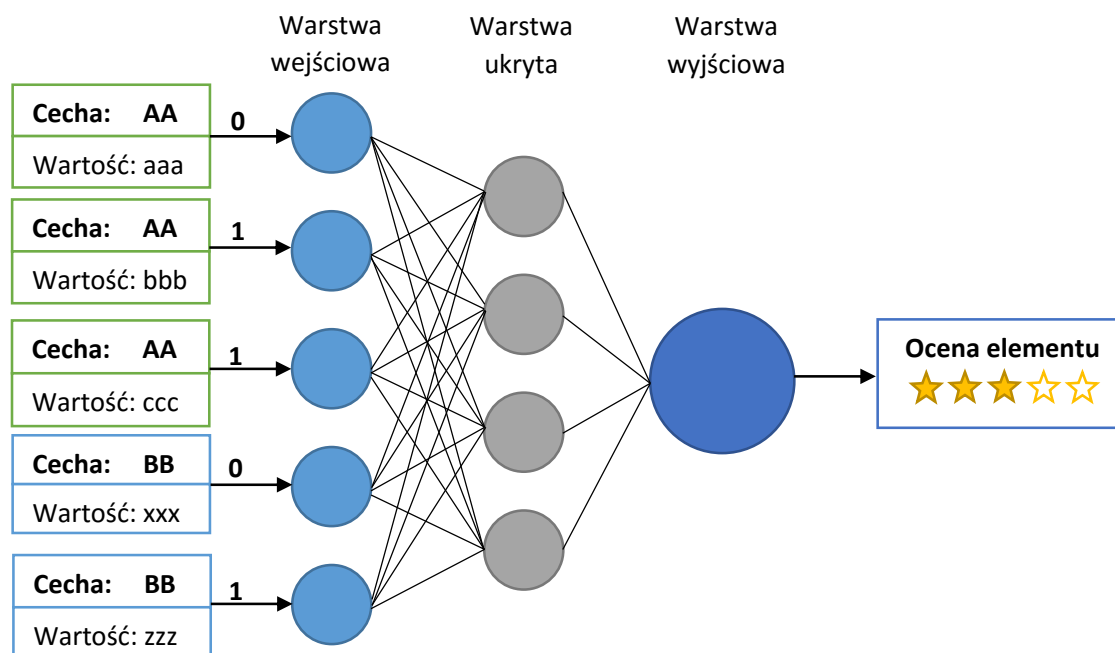


Rys. 4.5: Wykres sigmoidalnej funkcji aktywacji perceptronu [21]

Struktura sieci i przebieg algorytmu

Pierwszym etapem algorytmu jest analiza cech elementów ocenionych przez użytkownika. Tworzona jest lista wszystkich występujących cech które powtarzają się minimum tyle razy, ile wynosi wartość parametru *minimumRepeatingFeatures*.

Następnie inicjowana jest sieć neuronowa. Ilość neuronów warstwy wejściowej jest równa ilości wyodrębnionych cech. Warstwa ukryta zawiera tyle neuronów ile jest to



Rys. 4.6: Schemat sieci neuronowej

określone parametrem *hiddenLayerNeurons*. Warstwa wyjściowa składa się z tylko jednego neuronu (rys. 4.6).

W kolejnym etapie dla każdego elementu tworzona jest mapa cech. Jeżeli element zawiera daną cechę o danej wartości przypisywana jest wartość 1. W przeciwnym razie wstawiane jest 0. Rys. 4.7 przedstawia przykładową mapę cech. Tak przygotowana lista przekazywana jest do sieci neuronowej.

Na wyjściu sieć zwraca przewidywaną ocenę elementu.

4.2.2. Uczenie sieci neuronowej

Aby sieć zwracała jak najlepsze wyniki musi wcześniej zostać nauczona preferencji użytkownika. Uczenie sieci polega na odnalezieniu odpowiednich wag dla każdego wejścia każdego perceptronu sieci (budowa perceptronu zob. 4.2.1). Proces ten można zapisać w postaci

$$W_{ij}(n+1) = W_{ij}(n) + \Delta W_{ij}(n), \quad (4.8)$$

gdzie

$W_{ij}(n)$ to poprzednie wagi wejść

$W_{ij}(n+1)$ to nowe wagi wejść

n numer cyklu uczącego

Dostrajanie wartości wag można wykonać na wiele sposobów. Można wyróżnić uczenie nadzorowane (z nauczycielem), uczenie z krytykiem (ang. *reinforcement learning*) i uczenie samoorganizujące się (bez nadzoru) [37].

Cecha	Wartość	Czy zawiera?
Aktor	Julia Roberts	1
Aktor	Al Pacino	1
Aktor	Brad Pitt	0
...		
Reżyser	Francis Ford Coppola	1
Reżyser	Darren Aronofsky	0
...		

Rys. 4.7: Mapa cech elementu. Wiadomo, że element X zawiera cechę „Aktor” o wartościach „Julia Roberts, Al Pacino” oraz cechę „Reżyser” o wartości „Francis Ford Coppola”. Element nie zawiera cechy „Aktor” o wartości „Brad Pitt” ani cechy „Reżyser” o wartości „Darren Aronofsky” więc w te miejsca wstawiane jest 0.

Na potrzeby systemu rekomendacji w oparciu o zawartość opisywanego w tej pracy wykorzystane zostały trzy metody uczenia: algorytm propagacji wstecznej, algorytm RPROP i algorytm genetyczny. Wymienione metody należą do kategorii uczenia nadzorowanego.

4.2.3. Propagacja wsteczna

Algorytm propagacji wstecznej jest jedną z popularniejszych metod uczenia nadzorowanego jednokierunkowych sieci neuronowych. Zbiór uczący składa się z danych wejściowych (wektory cech elementu) i z informacji o oczekiwanym wyniku (ocena elementu). Różnica między wynikiem zwróconym przez sieć a wartością oczekiwaną stanowi miarę błędu sieci neuronowej.

Wiadomo, że funkcja celu jest funkcją ciągłą. Bazując na gradientowych metodach optymalizacji, wagi w sieci neuronowej aktualizowane są w następujący sposób:

$$W_{ij}(n+1) = W_{ij}(n) + \Delta W_{ij}(n), \quad (4.9)$$

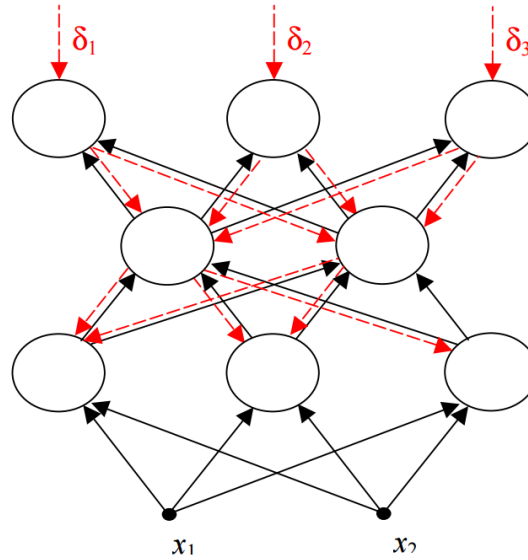
$$\Delta W_{ij}(n) = \eta p(W), \quad (4.10)$$

gdzie

η to współczynnik uczenia (ang. *learning rate*)

$p(W)$ to kierunek w przestrzeni wielowymiarowej W

Aby wyznaczyć kierunek $p(W)$ dla wszystkich warstw sieci należy przejść przez kolejne etapy uczenia [14, 15, 26, 37, 48].



Rys. 4.8: Algorytm propagacji wstecznej w sieci trójwarstwowej – idea działania [26]

1. W kroku pierwszym sieć neuronowa poddawana jest analizie o zwykłym kierunku przepływu sygnałów. Wynikiem są wartości sygnałów wychodzących z neuronów warstw wyjściowej i ukrytych oraz pochodne funkcji aktywacji w kolejnych warstwach.
2. W kroku drugim kierunek przepływu sygnałów zostaje odwrócony (stąd nazwa propagacja wsteczna). Funkcje aktywacji zostają zastąpione przez swoje pochodne. Na oryginalne wyjście sieci (aktualnie wejście) podana zostaje wartość równa różnicy pomiędzy wynikiem oczekiwanym a wynikiem zwróconym przez sieć.
3. Obliczana zostaje wartość różnic wstecznych.
4. W kolejnym kroku następuje wreszcie proces adaptacji wag. Odbyna się on zarówno dla sieci zwykłej jak i dla sieci o propagacji wstecznej. Reguła modyfikacji ma postać:

$$\Delta W_{ij} = \eta \sum_{\text{wzroce}} \delta_{\text{wyjście}} \cdot x_{\text{wejście}}, \quad (4.11)$$

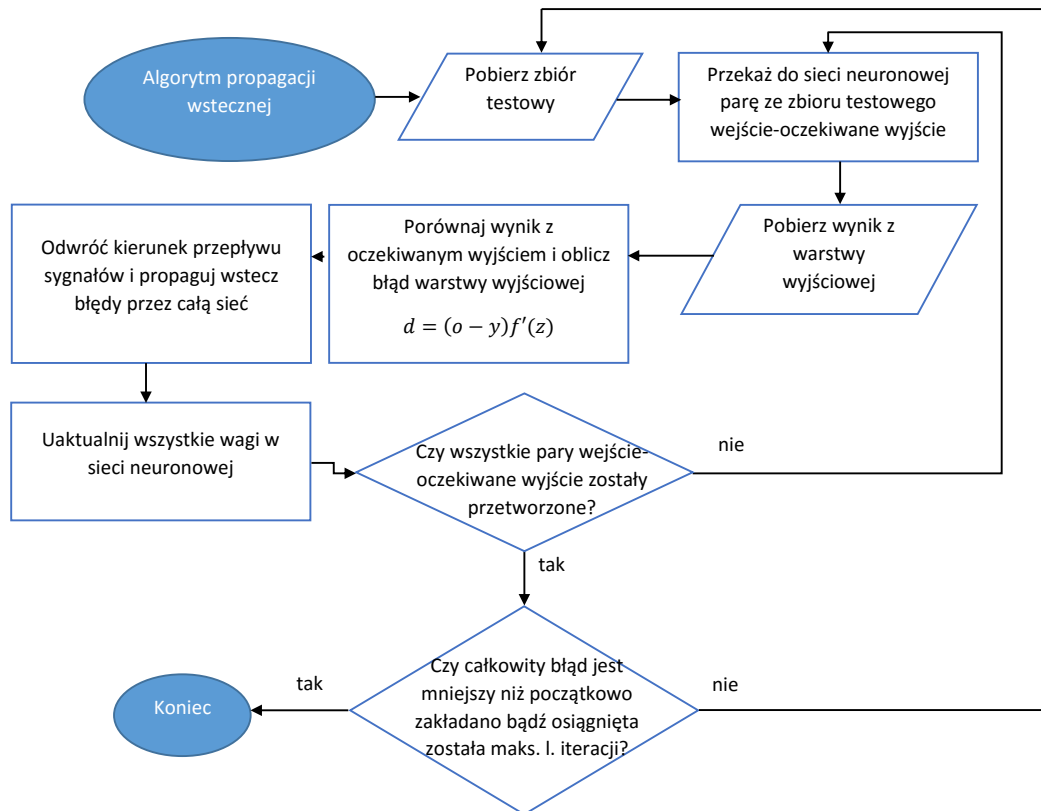
5. Powyższe kroki potarzane są dla każdej pary dane wejściowe – oczekiwany wynik tak długo, aż poziom błędu spadnie poniżej akceptowalnej wartości bądź osiągnięta zostanie maksymalna liczba iteracji.

Algorytm propagacji wstecznej obrazuje schemat blokowy 4.9.

Metoda momentum

W celu zwiększenia efektywności działania algorytmu została zastosowana metoda momentum. I tak też standardowy sposób aktualizacji wag sieci (równanie 4.9) został zmodyfikowany w następujący sposób:

$$W_{ij}(n+1) = W_{ij}(n) + \Delta W_{ij}(n), \quad (4.12)$$



Rys. 4.9: Algorytm propagacji wstecznej

$$\Delta W_{ij}(n) = \eta(n) \cdot p(n) + \alpha(W_{ij}(n) - W_{ij}(n-1)), \quad (4.13)$$

gdzie α jest współczynnikiem momentum z zakresu $[0, 1]$. Im wyższa wartość współczynnika tym większy jego wpływ na ostateczny kształt kolejnych wag. W dalszej części pracy przeprowadzono eksperymenty mające na celu dobranie optymalnej wartości współczynnika momentum.

4.2.4. Algorytm RPROP

Pomimo dużej popularności algorytmu propagacji wstecznej nie jest on pozbawiony wad. Sporym problemem jest jego relatywnie niska szybkość działania, szczególnie w przypadku bardzo dużych sieci neuronowych – a więc w przypadku odpowiadającemu potrzebom systemów rekomendacji. W odpowiedzi na te problemy Martin Riedmiller i Heinrich Braun zaproponowali w 1992 roku alternatywny algorytm – Resilient Backpropagation (RPROP).

Główną różnicą jest wykorzystywanie jedynie informacji o dodatniości lub ujemności każdej składowej gradientu zamiast o ich wartości (jak to się odbywa w oryginalnym algorytmie). Ponadto, współczynnik uczenia modyfikowany jest w każdym kolejnym kroku.

Modyfikacja współczynników odbywa się zgodnie ze wzorem:

$$\eta(t) = \begin{cases} \min\{a\eta(t-1), \eta_{\max}\}, & \text{gdy } \frac{\partial E^2(t)}{\partial v(t)} \frac{\partial E^2(t-1)}{\partial v(t-1)} > 0 \\ \max\{b\eta(t-1), \eta_{\min}\}, & \text{gdy } \frac{\partial E^2(t)}{\partial v(t)} \frac{\partial E^2(t-1)}{\partial v(t-1)} < 0 \\ \eta(n-1), & \text{w każdym innym przypadku} \end{cases}, \quad (4.14)$$

gdzie $\frac{\partial E^2(t)}{\partial v(t)}$ jest dokładną wartością składowej v gradientu. Ponadto stałe a , b , η_{\min} i η_{\max} są zdefiniowane następująco:

$$a = 1.2$$

$$b = 0.5$$

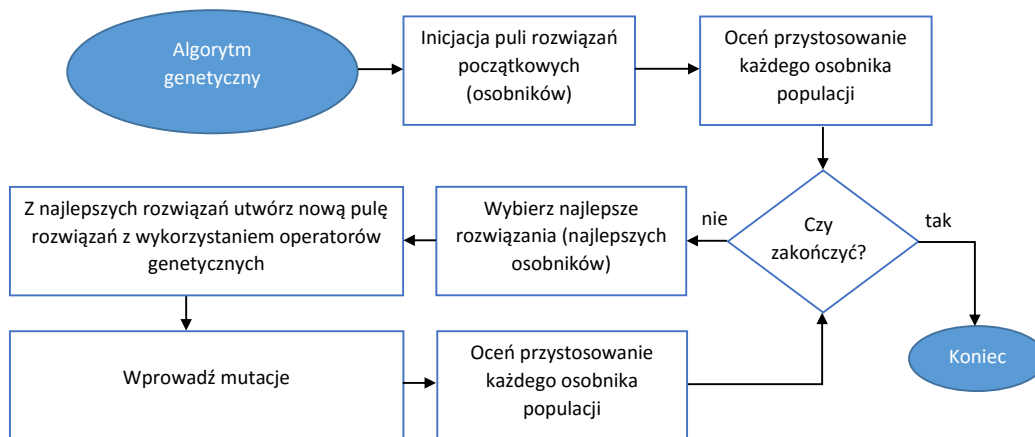
$$\eta_{\min} = 10^{-6}$$

$$\eta_{\max} = 50$$

Algorytm RPROP jest efektywniejszy pod kątem prędkości działania względem tradycyjnej propagacji wstecznej, więc dobrze sprawdza się tam, gdzie szybkość jest ważniejsza od wysokiej poprawności wyników [41, 42].

4.2.5. Algorytm genetyczny

Koncepcja algorytmów genetycznych została zaproponowana już 1960 roku przez Johna Hollanda. Należą one do klasy algorytmów ewolucyjnych, inspirowanych zasadą doboru naturalnego Darwina. Analogią do środowiska naturalnego jest pewien problem, dla którego poszukiwane jest optymalne rozwiązanie. Populacja składa się z potencjalnych rozwiązań, które są oceniane funkcją przystosowania. Zgodnie z zasadą Darwina zwycięża najsilniejszy, czyli najlepsze (choć nie zawsze optymalne) rozwiązanie [38]. Schemat działania algorytmów genetycznych przedstawia rys. 4.10.



Rys. 4.10: Ogólny schemat algorytmu genetycznego

Jako, że algorytmy genetyczne należą do grupy algorytmów optymalizacyjnych mogą zostać wykorzystane do uczenia sieci neuronowych. W przypadku tej implementacji populację stanowią propozycje wag dla każdego perceptronu. Preferowane są

takie zestawy wag, dla których sieć zwraca rezultaty obciążone najmniejszym błędem. Najlepsze zestawy są ze sobą krzyżowane oraz podlegają mutacji. W wyniku takich mechanizmów ewolucyjnych powstaje rozwiązanie, dla których sieć neuronowa zwraca oczekiwane rezultaty [22, 31].

4.3. Algorytmy hybrydowe

Zaproponowany nowy algorytm jest hybrydą filtrowania kolaboratywnego i filtrowania w oparciu o zawartość. Zaimplementowanych zostało 9 wariantów:

1. Matrix Factorization + sieć neuronowa uczona metodą propagacji wstecznej;
2. Matrix Factorization + sieć neuronowa uczona metodą RPROP;
3. Matrix Factorization + sieć neuronowa uczona algorytmem genetycznym;
4. Biased Matrix Factorization + sieć neuronowa uczona metodą propagacji wstecznej;
5. Biased Matrix Factorization + sieć neuronowa uczona metodą RPROP;
6. Biased Matrix Factorization + sieć neuronowa uczona algorytmem genetycznym;
7. SVD++ + sieć neuronowa uczona metodą propagacji wstecznej;
8. SVD++ + sieć neuronowa uczona metodą RPROP;
9. SVD++ + sieć neuronowa uczona algorytmem genetycznym.

NAPISAĆ
TUTAJ
WIĘCEJ o
tych wszyst-
kich wariant-
tach, dodać
analizę wła-
sności

Algorytm hybrydowy uruchamia osobno metodę filtrowania kolaboratywnego i filtrowania z analizą zawartości. W efekcie dla każdej pary użytkownik–element uzyskiwane są dwie predykowane oceny z obu metod: $r_{content-based_{ui}}$ i $r_{collaborative_{ui}}$. Ponadto, znany jest błąd $e \geq 0$, z jakim zakończyło się uczenie sieci neuronowej dla użytkownika u . Przewidywana ocena obliczana jest wzorem:

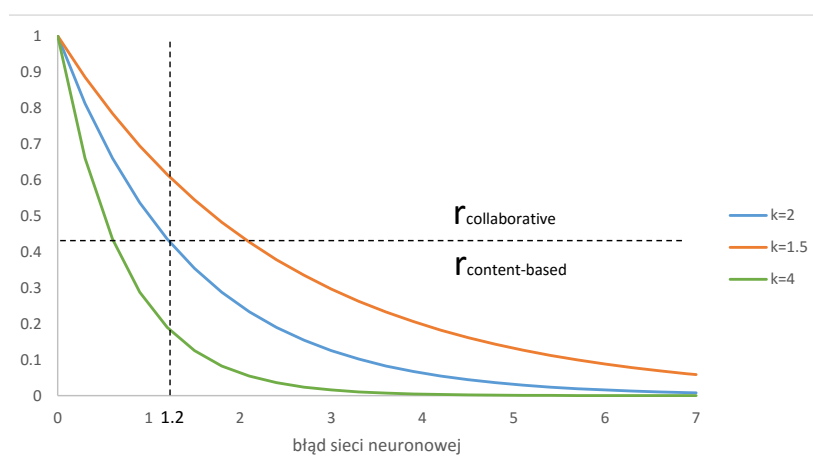
$$r_{ui} = \left(\frac{1}{k}\right)^e \cdot r_{content-based_{ui}} + \left(1 - \left(\frac{1}{k}\right)^e\right) \cdot r_{collaborative_{ui}}, \quad (4.15)$$

gdzie k jest parametrem konfigurowalnym (domyślnie $k = 2$), który determinuje stopień nachylenia funkcji wykładniczej. Parametr musi spełniać warunek $k > 1$. Im wyższa wartość k , tym większy wpływ na rezultat ma wynik algorytmu kolaboratywnego. Zależność tę obrazuje rys. 4.11.

4.3.1. Zalety zaproponowanej metody

Największą przewagą zaproponowanego algorytmu jest to, że łączy on zalety filtrowania z analizą zawartości i filtrowania kolaboratywnego minimalizując jednocześnie ich wady poprzez dynamiczne dobieranie proporcji ich wpływu na końcowy wynik. Nowa metoda cechuje się uniwersalnością, gdyż model elementu zawierający jego własności budowany jest automatycznie. Jednocześnie nie są wymagane informacje o użytkowniku inne, niż to jakie elementy ocenił i jak je ocenił.

Im więcej zostanie dostarczonych informacji na temat elementów, tym lepsze będą wyniki. Jednakże w przypadku braku takowych informacji metoda ciągle będzie działać, wówczas wynik będzie opierał się głównie na rezultatach filtrowania kolaboratywnego.



Rys. 4.11: Wpływ parametru k na wynik algorytmu. Jeżeli parametr k wynosi 2 a uczenie sieci neuronowej zakończyło się z błędem równym 1,2, to końcowy stosunek wag wyniku filtrowania z analizą zawartości do wyniku filtrowania kolaboratywnego wyniesie ok. 43,5% do 56,5%.

Analogicznie, w przypadku małej liczby powiązań z innymi użytkownikami, zaproponowany algorytm hybrydowy ulepszy wynik filtrowania kolaboratywnego dzięki czynnikowi filtrowania w oparciu o zawartość.

4.3.2. Wady zaproponowanej metody

Największą wadą nowej metody jest wyższa złożoność. Czas budowania modelu jest sumą czasów budowania modelu kolaboratywnego i modelu filtrowania w oparciu o zawartość.

Istnieje też ryzyko, że w momencie gdy wynik filtrowania kolaboratywnego będzie bardzo zły, to rezultat osiągnięty przez nową metodę będzie gorszy niż z samego filtrowania z analizą zawartości.

4.4. Analiza złożoności i poprawności

Rozdział 5

Ocena eksperymentalna

5.1. Opis metody badawczej

5.1.1. Miara oceny

W celu zbadania jakości algorytmów zostały zastosowane miary oceny: średnia kwadratowa błędów (RMSE) i średni błąd bezwzględny (MAE).

Średnia kwadratowa błędów

Średnia kwadratowa błędów (ang. RMSE – *root mean square error*) jest często wykorzystywaną miarą służącą zmierzeniu różnicy pomiędzy wartościami przewidywanymi a rzeczywistymi (obserwowanymi).

RMSE jest stosunkowo dobrą miarą dokładności ale tylko w celu porównania różnych modeli dla tego samego zestawu danych. RMSE jest zależne od skali, zatem nie sprawdza się najlepiej w przypadku porównywania ze sobą różnych zmiennych [18].

Średnią kwadratową błędów wylicza się ze wzoru:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (5.1)$$

gdzie

\hat{y}_i to wartość przewidywana

y_i to wartość rzeczywista

Im niższa wartość RMSE tym bardziej zbliżone są wartości przewidywane do rzeczywistych, zatem tym lepszy jakościowo jest model.

Opisać eksperyment w punktach - co się po kolei dzieje

Dodać pseudokod

Przenieść niżej w rozdziale

Średni błąd bezwzględny

Inną miarą mierzenia jakości modeli predykcyjnych jest MAE (ang. *mean absolute error*). Podobnie jak RMSE miara ta jest zależna od skali, zatem najlepiej sprawdza się w działaniu na tym samym zestawie danych [18].

Średni błąd bezwzględny wylicza się ze wzoru:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|, \quad (5.2)$$

gdzie

\hat{y}_i to wartość przewidywana

y_i to wartość rzeczywista

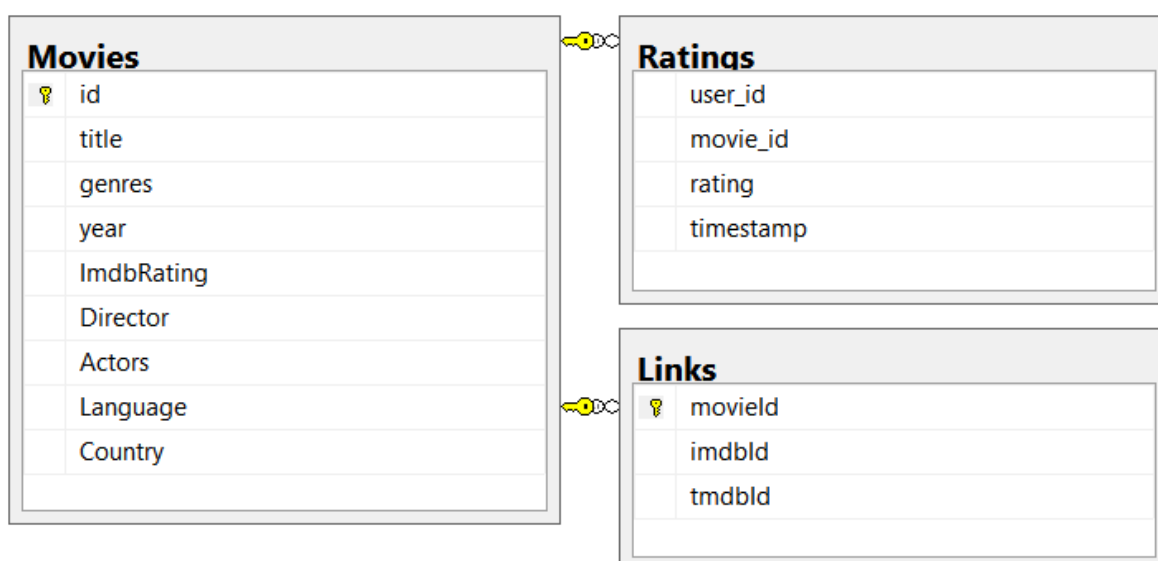
5.1.2. Zbiory danych

W celu uzyskania jak najbardziej miarodajnych wyników, badania zostały przeprowadzone na trzech różnych bazach danych z trzech różnych domen.

MovieLens

MovieLens [13] to baza zawierająca oceny filmów przez użytkowników portalu movielens.org. Baza zawiera 3706 filmów i 1000209 ocen wystawionych przez 6040 unikalnych użytkowników pomiędzy 25 kwietnia 2000 a 28 lutym 2003. Filmy oceniane są w skali od 1 do 5, gdzie 1 jest oceną najgorszą a 5 najlepszą.

Baza zawiera tabelę łączącą numery identyfikacyjne filmów z bazą IMDB.com. Wykorzystując te informacje baza filmów została rozszerzona o informacje pobrane z IMDB.com. Ostateczny kształt bazy widoczny jest na rys. 5.1.



Rys. 5.1: Schemat bazy MovieLens

i od razu podkłada się Pani recenzentowi... Dwa zdania temu pisała Pani o tym, że MAE i RMSE są dobre dla tych samych danych, a teraz że będzie Pani porównywać trzy różne zbiory

Yahoo Music

Opisać
Yahoo Music

Amazon Meta

Opisać Ama-
zon Meta

[27]

5.2. Środowisko symulacyjne

5.2.1. Oprogramowanie

Rys. 5.2 przedstawia zrzut ekranu środowiska symulacyjnego opracowanego przez autorkę pracy na potrzeby przeprowadzenia badań algorytmów. Główny interfejs programu składa się z trzech części: ustawienia podstawowe, okno z wynikiem oraz panel sterujący ustawieniami zaawansowanymi.

W sekcji z ustawieniami podstawowymi możliwy jest wybór:

- o bazy danych, która zostanie wykorzystana do pomiarów;
- o typu algorytmu rekomendującego (filtrowanie w oparciu o zawartość, filtrowanie kolaboratywne bądź algorytm hybrydowy)
- o w przypadku wyboru kolaboratywnego filtrowania lub filtrowania hybrydowego możliwy jest wybór typu algorytmu: Matrix Factorization, Biased Matrix Factorization lub SVD++.
- o w przypadku wyboru filtrowania z analizą zawartości lub filtrowania hybrydowego automatycznie wybierany jest algorytm oparty na sieci neuronowej.

Widok ustawień zaawansowanych zmienia się w zależności od wybranego filtrowania. W przypadku wyboru filtrowania w oparciu o zawartość istnieje możliwość regulowania parametrów sieci neuronowej. W każdym wariancie można regulować kryteria doboru zestawu danych.

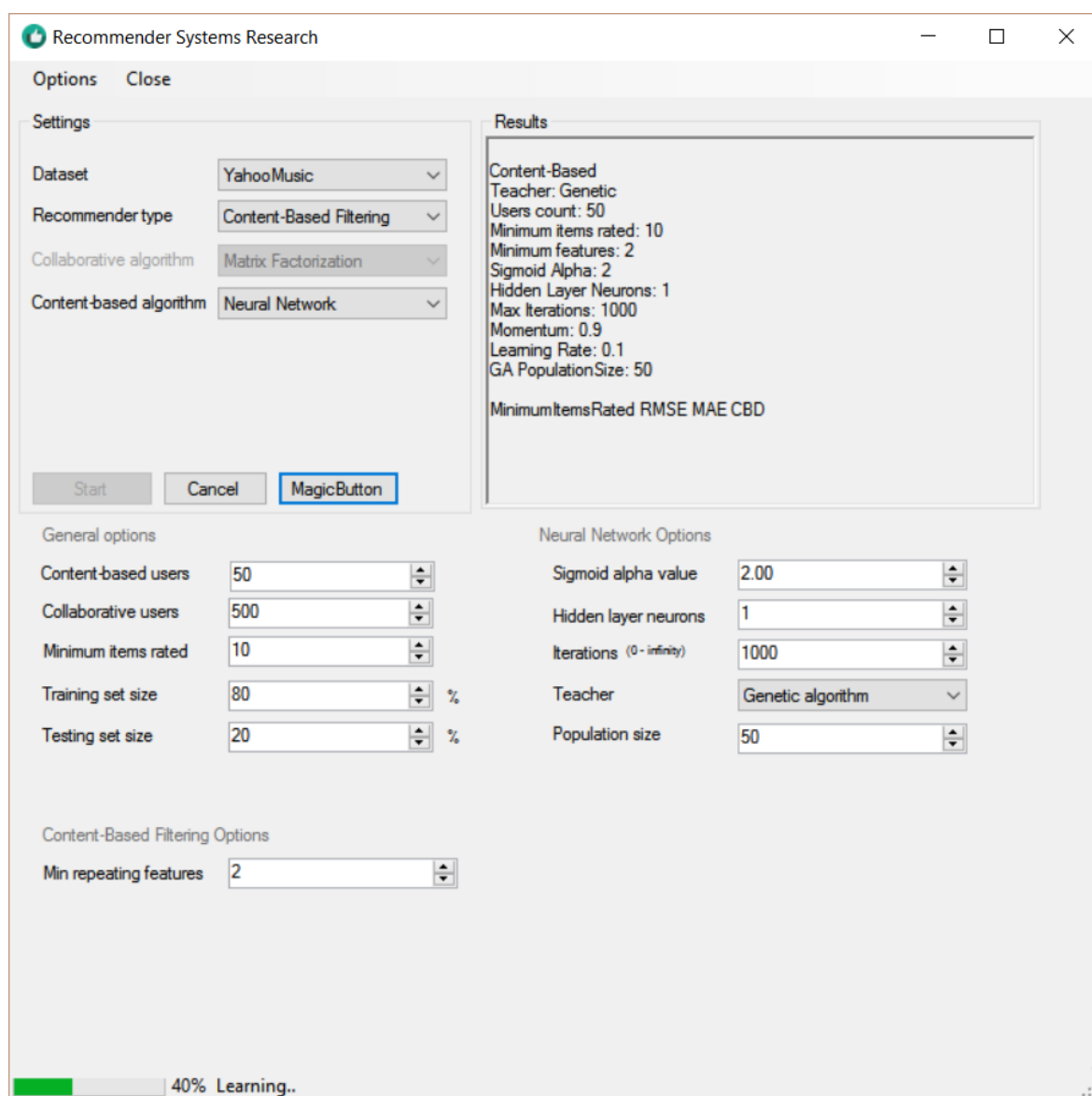
Kryteria doboru zestawu danych są następujące:

- o liczba użytkowników do pobrania do filtrowania w oparciu o zawartość;
- o liczba użytkowników do pobrania do filtrowania kolaboratywnego;
- o minimum elementów, jakie zostały ocenione przez każdego pobranego użytkownika;
- o stosunek wielkości zbioru treningowego do zbioru testowego (domyślnie 80%-20%).

Parametry sieci neuronowej podlegające strojeniu to:

- o Parametr funkcji aktywacji α ;
- o liczba neuronów w warstwie ukrytej;
- o maksymalna liczba iteracji uczenia sieci neuronowej;
- o algorytm uczący sieć neuronową: propagacja wsteczna, rprop (resilient backpropagation) lub algorytm genetyczny;
- o w przypadku wyboru algorytmu genetycznego – rozmiar każdej kolejnej populacji;
- o minimalna wymagana ilość powtórzeń danej cechy, aby była brana pod uwagę w trakcie budowania rekomendacji.

Do parametrów dodać symbole takie jak były wcześniej wykorzystane w modelu



Rys. 5.2: Zrzut ekranu środowiska symulacyjnego

5.2.2. Sprzęt

Wszystkie badania zostały przeprowadzone na komputerze klasy PC działającym pod kontrolą 64-bitowego systemu Windows 10 Pro. Konfiguracja sprzętowa wygląda następująco:

Procesor: Intel Core i5-6200U 2.30GHz

RAM: 8 GB

Typ dysku twardego: SSD

5.3. Przeprowadzone eksperymenty

Napisać tu ładny wstęp

Opis i schemat eksperymentów

?? Jak bardzo wchodzić w szczegóły

Dla wszystkich przeprowadzonych eksperymentów stosunek zbioru treningowego do zbioru testowego wynosił 80%/20%.

5.3.1. Dopasowanie parametrów sieci neuronowej

Pierwszy przeprowadzony eksperyment miał na celu dopasowanie optymalnych parametrów sieci neuronowej dla filtrowania w oparciu o zawartość. Badania przeprowadzone zostały w oparciu o bazę MovieLense.

Dopasowanie maksymalnej liczby iteracji

Uczenie sieci neuronowej odbywa się w pętli tak długo, jak wartość błędu na wyjściu spadnie poniżej określonej wartości bądź przekroczona zostanie maksymalna ilość iteracji. Eksperyment ma na celu zbadanie wartości błędów w zależności od liczby iteracji.

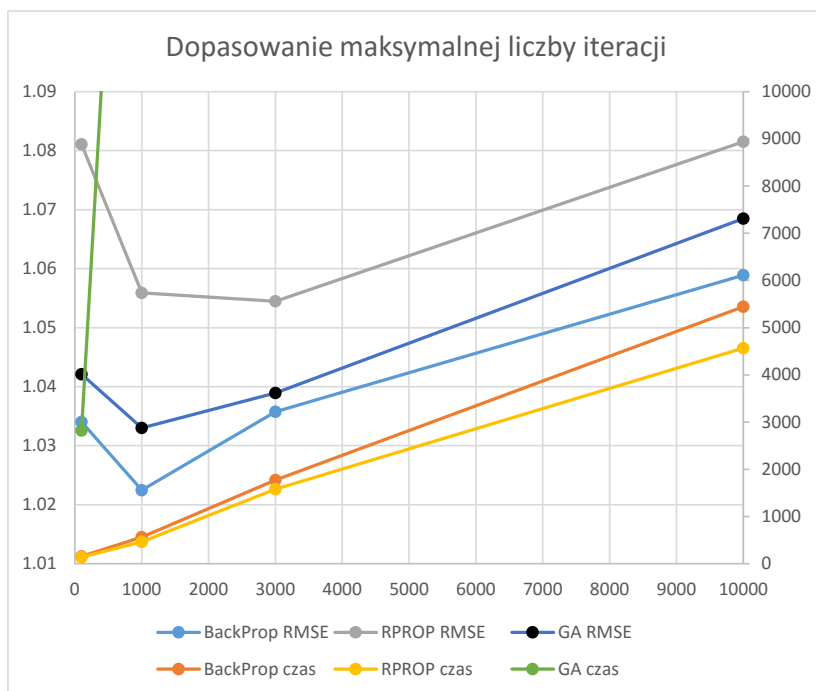
Konfiguracja parametrów podstawowych	
Baza danych:	MovieLens
Testowana metoda:	Filtrowanie w oparciu o zawartość
Konfiguracja filtrowania w oparciu o zawartość	
Algorytm:	Sieć neuronowa uczona propagacją wsteczną, RPROP i algorytmem genetycznym
Maksymalna ilość iteracji nauki:	?
Ilość neuronów w warstwie ukrytej:	1
Współczynnik momentum:	0.9
Współczynnik uczenia:	0.1
Parametr funkcji aktywacji α :	2.0
Minimum powtarzających się cech:	10
Minimum ocenionych elementów przez użytkownika:	10
Ilość testowanych użytkowników:	10

Tabela 5.1: Konfiguracja dla eksperymentu maksymalnej liczby iteracji

Wyniki eksperymentu przedstawia wykres 5.3 i tabela 5.2. Wraz ze wzrostem liczby iteracji wartość błędu maleje, jednakże po przekroczeniu pewnej wartości następuje ponowny wzrost. Wynika to z występowania zjawiska nadmiernego dopasowania (ang. *overfitting*).

Zgodnie z oczekiwaniami, czas wykonania jest wprost proporcjonalny do liczby iteracji. W trakcie tego eksperymentu potwierdzone zostały również przypuszczenia dotyczące prędkości działania algorytmów. Najszybszy okazał się RPROP, jednakże kosztem nieco większych błędów. Najgorzej wypadł algorytm genetyczny, który cha-

rakteryzował się bardzo szybko rosnącym czasem wykonania przy jednoczesnym nie najlepszym wynikiem RMSE.



Rys. 5.3: Wyniki eksperymentu maksymalnej liczby iteracji

L. iter.	Propagacja wsteczna		RPROP		Algorytm genetyczny	
	RMSE	czas	RMSE	czas	RMSE	czas
100	1.03402	155	1.081065	139	1.042103	2824
1000	1.022469	564	1.055915	463	1.033005	24831
3000	1.035766	1770	1.054472	1581	1.038922	70579
10000	1.058892	5444	1.081511	4567	1.068489	239775

Tabela 5.2: Wyniki eksperymentu maksymalnej liczby iteracji

Dopasowanie parametru momentum

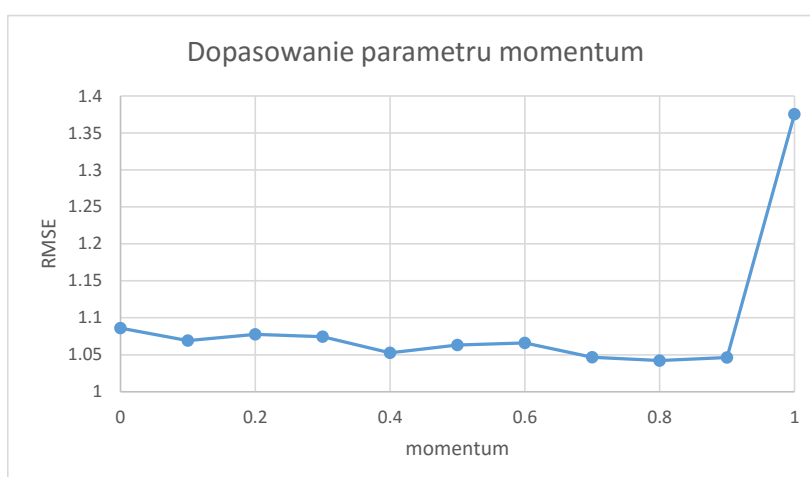
Parametr momentum jest dodatkowym współczynnikiem mającym na celu ulepszenie działania algorytmu propagacji wstecznej (zob. 4.2.3). Eksperyment ma na celu zbadanie dla jakich wartości osiągany jest najniższy błąd.

Konfiguracja parametrów podstawowych
Baza danych: MovieLens
Testowana metoda: Filtrowanie w oparciu o zawartość
Konfiguracja filtrowania w oparciu o zawartość

Algorytm:	Sieć neuronowa z propagacją wsteczną
Maksymalna ilość iteracji nauki:	1000
Ilość neuronów w warstwie ukrytej:	2
Współczynnik momentum:	?
Współczynnik uczenia:	0.1
Parametr funkcji aktywacji α :	2.0
Minimum powtarzających się cech:	10
Minimum ocenionych elementów przez użytkownika:	100
Ilość testowanych użytkowników:	5

Tabela 5.3: Konfiguracja dla eksperymentu dopasowania wartości momentum

Wyniki eksperymentu przedstawia wykres 5.4 i tabela 5.4. W zakresie wartości $[0, 0.9]$ błąd oscyluje w granicach podobnych wartości z nieznaczną tendencją malejącą. Wyraźny skok na niekorzyść odnotowany jest dopiero przy momentum = 1.



Rys. 5.4: Wyniki eksperymentu momentum

Momentum	RMSE	MAE
0	1.086007	0.855964
0.1	1.069125	0.851345
0.2	1.077596	0.860146
0.3	1.074588	0.851467
0.4	1.052529	0.8424
0.5	1.063205	0.846198

0.6	1.065994	0.856964
0.7	1.046752	0.83462
0.8	1.042235	0.842353
0.9	1.046243	0.840319
1	1.375336	1.071652

Tabela 5.4: Wyniki eksperymentu momentum

Dopasowanie rozmiaru populacji

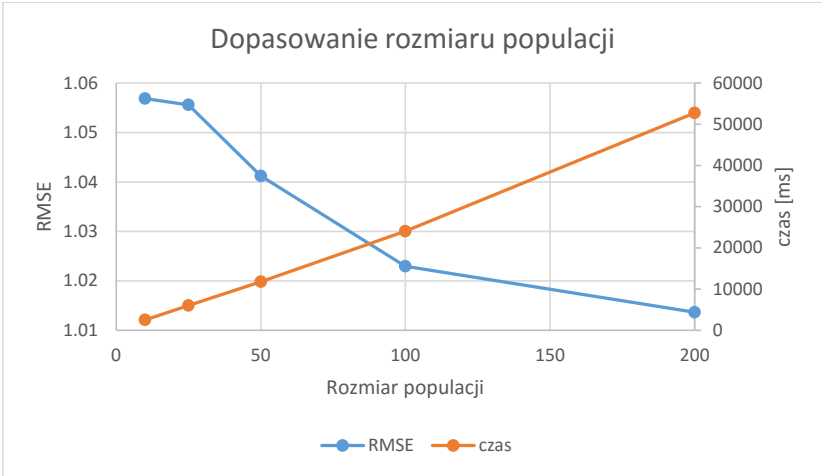
Parametr rozmiaru populacji determinuje ilość osobników, jaka zostanie wygenerowana w każdej epoce przebiegu algorytmu genetycznego (zob 4.2.5). Eksperyment ma na celu zbadanie korelacji pomiędzy rozmiarem populacji a błędem algorytmu rekomendacji. Mierzony jest również czas uczenia sieci neuronowej.

Konfiguracja parametrów podstawowych	
Baza danych:	MovieLens
Testowana metoda:	Filtrowanie w oparciu o zawartość
Konfiguracja filtrowania w oparciu o zawartość	
Algorytm:	Sieć neuronowa uczona algorytmem genetycznym
Maksymalna ilość iteracji nauki:	1000
Ilość neuronów w warstwie ukrytej:	1
Współczynnik momentum:	0.9
Współczynnik uczenia:	0.1
Parametr funkcji aktywacji α :	2.0
Minimum powtarzających się cech:	10
Minimum ocenionych elementów przez użytkownika:	10
Ilość testowanych użytkowników:	10

Tabela 5.5: Konfiguracja dla eksperymentu dopasowania wielkości populacji

Wyniki eksperymentu przedstawia wykres 5.5 i tabela 5.6. Zgodnie z oczekiwaniami występuje dodatnia korelacja pomiędzy rozmiarem populacji a czasem wykonania i ujemna pomiędzy rozmiarem populacji a błędem algorytmu rekomendacji.

Zależność pomiędzy rozmiarem populacji a czasem wykonania jest liniowa, natomiast błąd maleje wykładniczo gdy populacja osiąga większe rozmiary. Optymalnym zatem jest przyjęcie rozmiaru populacji równej 100.



Rys. 5.5: Wyniki eksperymentu rozmiaru populacji

Populacja	RMSE	MAE	czas
10	1.056848	0.861325	2519
25	1.055562	0.859391	6015
50	1.041198	0.850866	11804
100	1.022943	0.825477	24062
200	1.013656	0.816789	52751

Tabela 5.6: Wyniki eksperymentu rozmiaru populacji

Dopasowanie ilości neuronów w warstwie ukrytej

Konieczne
TODO

Dopasowanie współczynnika uczenia

Czy trzeba?

Dopasowanie wartości alfa

Czy trzeba

5.3.2. Porównanie algorytmu hybrydowego z filtrowaniem w oparciu o zawartość i filtrowaniem kolaboratywnym

Porównanie
algorytmu
hybrydo-
wego z
content-based
i
collaborative

5.3.3. Analiza statystyczna wyników

Analiza sta-
tystyczna
wyników,
PQStat

Rozdział 6

Wnioski

Rozdział 7

CHAPTER 1

7.1. SECTION

Algorytm 5 Alghoritm 4

```
T ← text under analysis
for each word w ∈ T do
  Sw ← FIND_SENTIMENT(w)
  if Sw = POSITIVE then
    Sentiment[POSITIVE] ++
  else if Sw = NEGATIVE then
    Sentiment[NEGATIVE] ++
  else
    Sentiment[NEUTRAL] ++
  end if
end for
return arg maxx Sentiment[x]
```

Rys. 7.1: Schema 1

¡GRAPHIC¿

7.2. Section 2

7.2.1. Subsection 1

Subsubsection 1
Definicja 1
Definicja - pierwsza

Dodatek A

Appendix 1

Spis rysunków

7.1 Schema 1	41
------------------------	----

Spis wzorów

Spis algorytmów

1 Stochastyczny gradient prosty	15
2 Matrix Factorization – Inicjacja modelu	16
3 Matrix Factorization – Faza uczenia	17
4 Biased Matrix Factorization – Faza uczenia	18
5 Alghoritm 4	41

Bibliografia

- [1] About the Music Genome Project. <http://www.pandora.com/about/mgp>. Data dostępu: 2016-07-19.
- [2] Allegro – korzystanie z systemu rekomendacji. <http://faq.allegro.pl/artikul/27613/korzystanie-z-systemu-rekomendacji>. Data dostępu: 2016-07-19.
- [3] Avron H., Kale S., Kasiviswanathan S., Sindhvani V. Efficient and practical stochastic subgradient descent for nuclear norm regularization. <https://www.cs.cmu.edu/~yuxiangw/docs/SSGD.pdf>. Data dostępu: 2016-09-07.
- [4] Bottou L. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade*, pages 421–436. Springer, 2012.
- [5] Celma O. *The Long Tail in Recommender Systems*, pages 87–107. Springer-Verlag Berlin Heidelberg, 2010.
- [6] Cheng J., Liu Y., Zhang H., Wu X., Chen F. A new recommendation algorithm based on user’s dynamic information in complex social network. *Mathematical Problems in Engineering*, 2015, 2015.
- [7] Davidson J., Liebal B., Liu J., Nandy P., Van Vleet T., Gargi U., Gupta S., He Y., Lambert M., Livingston B. et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.
- [8] Desrosiers C., Karypis G. *Recommender Systems Handbook*, chapter A Comprehensive Survey of Neighborhood-based Recommendation Methods, pages 107–144. Springer, New York Dordrecht Heidelberg London, 2010.
- [9] Filmweb – najczęściej zadawane pytania. <http://www.filmweb.pl/help>. Data dostępu: 2016-07-19.
- [10] Gantner Z., Rendle S., Drumond L., Freudenthaler C. Mymedialite recommender system library. <http://www.mymedialite.net/>. Data dostępu: 2016-09-05.
- [11] Gantner Z., Rendle S., Freudenthaler C., Schmidt-Thieme L. Mymedialite: a free recommender system library. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 305–308. ACM, 2011.

- [12] Gupta P., Goel A., Lin J., Sharma A., Wang D., Zadeh R. Wtf: The who to follow service at twitter. In *Proceedings of the 22nd international conference on World Wide Web*, pages 505–514. ACM, 2013.
- [13] Harper F. M., Konstan J. A. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):19, 2016.
- [14] Haykin S. Neural networks: A comprehensive foundation: Macmillan college publishing company. *New York*, 1994.
- [15] Hertz J. A., Krogh A. S., Palmer R. G., Jankowski S. *Wstęp do teorii obliczeń neuro-nowych*. Wydawnictwa Naukowo-Techniczne, 1993.
- [16] Huttner J. From Tapestry to SVD: A survey of the algorithms that power recommender system. Master’s thesis, Haverford College Department of Computer Science, 05 2009.
- [17] Huynh T., Hoang K. Modeling collaborative knowledge of publishing activities for research recommendation. In *International Conference on Computational Collective Intelligence*, pages 41–50. Springer, 2012.
- [18] Hyndman R. J., Koehler A. B. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.
- [19] IMDb database statistics. <http://www.imdb.com/stats>. Data dostępu: 2016-07-19.
- [20] Ji K., Sun R., Shu W., Li X. Next-song recommendation with temporal dynamics. *Knowledge-Based Systems*, 88:134–143, 2015.
- [21] Kirillov A. AForge.NET framework. <http://www.aforgenet.com/framework/>. Data dostępu: 2016-09-05.
- [22] Kirillov A. AForge.NET framework – evolutionary learning class. <http://www.aforgenet.com/framework/docs/html/cc8bebc5-da54-5c56-6ddf-6a93aec7b9cd.htm>. Data dostępu: 2016-09-06.
- [23] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.
- [24] Koren Y., Bell R. *Recommender Systems Handbook*, chapter Advances in Collaborative Filtering, pages 145–186. Springer, New York Dordrecht Heidelberg London, 2010.
- [25] Koren Y., Bell R., Volinsky C. et al. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
- [26] Kwater T. Algorytmy uczenia sieci neuronowych. http://www.neurosoft.edu.pl/media/pdf/tkwater/sztuczna_inteligencja/2_alg_ucz_ssn.pdf. Data dostępu: 2016-09-06.
- [27] Leskovec J., Adamic L. A., Huberman B. A. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5, 2007.
- [28] Linden G., Smith B., York J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.
- [29] Lops P., de Gemmis M., Semeraro G. *Recommender Systems Handbook*, chapter Content-based Recommender Systems: State of the Art and Trends, pages 73–100. Springer, New York Dordrecht Heidelberg London, 2010.

- [30] Maleszka M., Mianowska B., Nguyen N. T. A method for collaborative recommendation using knowledge integration tools and hierarchical structure of user profiles. *Knowledge-Based Systems*, 47:1–13, 2013.
- [31] Montana D. J., Davis L. Training feedforward neural networks using genetic algorithms. In *IJCAI*, volume 89, pages 762–767, 1989.
- [32] Mymedialite: Example experiments. <http://www.mymedialite.net/examples/datasets.html>. Data dostępu: 2016-07-24.
- [33] Netflix Prize (I tried to resist, but...). <https://www.snellman.net/blog/archive/2006-10-15-netflix-prize.html>. Data dostępu: 2016-07-08.
- [34] Netflix Prize: forum. <http://www.netflixprize.com/community/viewtopic.php?id=1537>. Data dostępu: 2016-07-08.
- [35] Netflix Prize Rankings. http://www.hackingnetflix.com/2006/10/netflix_prize_r.html. Data dostępu: 2016-07-08.
- [36] Netflix Prize Rules. <http://www.netflixprize.com/rules>. Data dostępu: 2016-07-08.
- [37] Osowski S. *Sieci neuronowe w ujęciu algorytmicznym*. Wydawnictwa Naukowo-Techniczne, 1996.
- [38] Pena-Reyes C. A., Sipper M. Evolutionary computation in medicine: an overview. *Artificial Intelligence in Medicine*, 19(1):1–23, 2000.
- [39] Pogue D. A Stream of Movies, Sort of Free. *The New York Times*, 2007.
- [40] Ricci F., Rokach L., Shapira B. *Recommender Systems Handbook*, chapter Introduction to Recommender Systems Handbook, pages 1–35. Springer, New York Dordrecht Heidelberg London, 2010.
- [41] Riedmiller M., Braun H. A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *Neural Networks, 1993., IEEE International Conference On*, pages 586–591. IEEE, 1993.
- [42] Riedmiller M., Rprop I. Rprop-description and implementation details. 1994.
- [43] Rohrmann T. Computing recommendations at extreme scale with apache flink. <http://data-artisans.com/computing-recommendations-at-extreme-scale-with-apache-flink>, 2015.
- [44] Rubens N., Kaplan D., Sugiyama M. Active learning in recommender systems. In Kantor P., Ricci F., Rokach L., Shapira B., editors, *Recommender Systems Handbook*, pages 735–767. Springer, 2011.
- [45] Sarwar B., Karypis G., Konstan J., Riedl J. Application of dimensionality reduction in recommender system-a case study. Technical report, DTIC Document, 2000.
- [46] Schafer J., Frankowski D., Herlocker J., Sen S. *The Adaptive Web*, chapter Collaborative filtering recommender systems, page 291–324. Springer Berlin / Heidelberg, 2007.
- [47] Sharma R., Singh R. Evolution of Recommender Systems from Ancient Times to Modern Era: A Survey. *Indian Journal of Science and Technology*, 9(20), 2016.
- [48] Timothy M. Sieci neuronowe w praktyce. *WNT, Warszawa*, 1996.

- [49] Willmott C. J., Matsuura K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.
- [50] Zhang H.-R., Min F., He X., Xu Y.-Y. A hybrid recommender system based on user-recommender interaction. *Mathematical Problems in Engineering*, 2015, 2015.