

Dokumentacja końcowa

System rekomendacji filmów

Michał Padzik

Albert Wolant

31 maja 2015

1. Ogólny opis projektu

Celem projektu było zaimplementowanie i porównanie metod rekomendacji. Zadanie rekomendacji polega wybraniu dla pewnego użytkownika najbardziej odpowiednich dla niego elementów ze zbioru ocenianych obiektów. Typowo przedmiotami oceny są filmy, książki czy strony internetowe. Istotny jest fakt, że zbiór ocen zawiera wiele elementów – oceny wielu użytkowników.

Przy rozwiązywaniu zadania rekomendacji dostępne metody można podzielić na dwie zasadnicze grupy: metody bazujące bezpośrednio na danych i metody z modelem. Metody z modelem, na podstawie dostępnej bazy ocen, budują pewną strukturę, inaczej model. Na podstawie analizy tej struktury realizują rekomendację. Metody bazujące na danych, inaczej niż metody z modelem, wyznaczają rekomendacje bezpośrednio z bazy ocen, poprzez obliczanie pewnych statystyk.

W projekcie skupiono się na metodach bazujących na danych. Metody te mają pewną istotną z praktycznego punktu widzenia zaletę. W środowisku, w którym baza ocen jest intensywnie modyfikowana, przy zastosowaniu metody z modelem, model musi być nieustannie przebudowywany. W takim przypadku metody bazujące na danych mogą wyznaczać rekomendację na podstawie najbardziej aktualnych danych.

Zadanie rekomendacji może mieć dwa podstawowe warianty. Pierwszy z nich to obliczanie wartości rekomendacji dla podanego użytkownika i podanego przedmiotu oceny. Drugi to wyznaczenie listy najlepszych elementów dla podanego użytkownika. W projekcie zaimplementowano oba warianty.

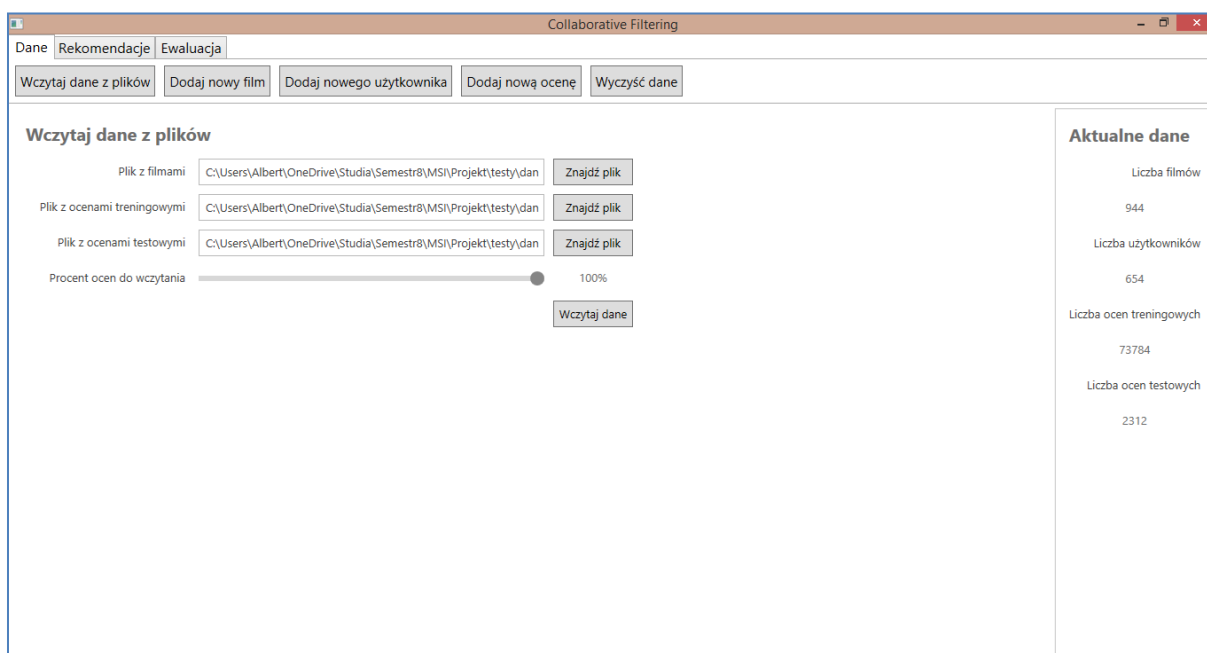
Zaimplementowane algorytmy rekomendacji zostały w fazie eksperymentów ocenione różnymi metodami. Każda z nich daje pewne informacje na temat jakości wyznaczonych rekomendacji w jednym z dwóch przedstawionych powyżej wariantów.

2. Interfejs użytkownika

Poniższy rozdział zawiera opis interfejsu i instrukcję użytkownika.

2.1. Wczytywanie danych

Aplikacja umożliwia wczytywanie danych z plików w formacie zgodnym z obowiązującym w konkursie Netflix Prize. Aby wczytać dane z plików należy przejść do ekranu „Wczytaj dane z plików” w menu „Dane”.



Rysunek 1 Wczytywanie danych z plików

Do poprawnego wczytania danych należy podać ścieżki do plików:

- Pliku z nazwami filmów
- Pliku z ocenami treningowymi
- Pliku z ocenami testowymi

Suwak „Procent ocen do wczytania” służy do ograniczenia ilości wczytywanych danych. Po podaniu wszystkich niezbędnych wartości należy użyć przycisku „Wczytaj dane”. Po wczytaniu, informacje o aktualnych danych znajdą się w panelu po prawej stronie okna.

Ponadto, używając przycisków „Dodaj nowy film”, „Dodaj nowego użytkownika” i „Dodaj nową ocenę” można przejść do ekranów umożliwiających dodanie pojedynczych rekordów do aktualnych danych.

Użycie przycisku „Wyczyść dane” usunie wszystkie rekordy z aktualnych danych. Pozwoli to wczytać nowy zbiór danych.

2.2. Wyznaczanie rekomendacji

W menu „Rekomendacje” dostępne są ekrany umożliwiające wyznaczenie wartości rekomendacji i zarekomendowanie filmów dla wybranego użytkownika.

Aby wyznaczyć wartość rekomendacji dla podanego użytkownika i filmu należy z menu „Rekomendacje” wybrać opcję „Sprawdź wartość rekomendacji”.

The screenshot shows a web application titled "Collaborative Filtering". It has three tabs: "Dane", "Rekomendacje", and "Ewaluacja". The "Rekomendacje" tab is active, and the "Sprawdź wartość rekomendacji" button is selected. The interface is divided into several sections:

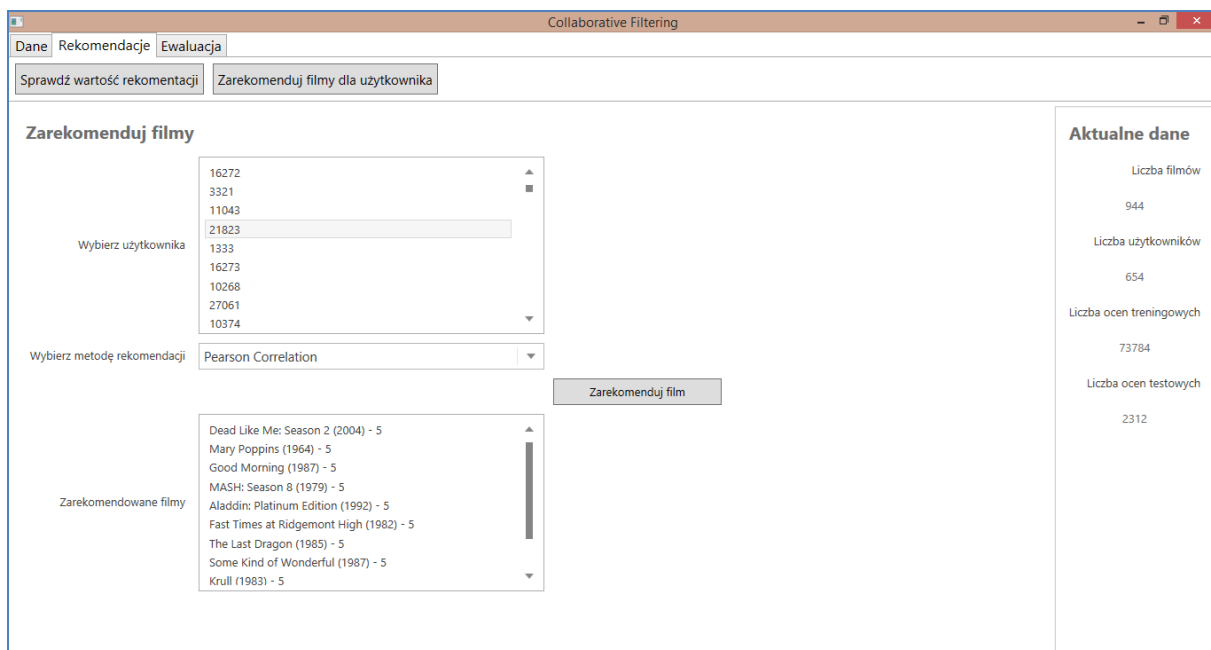
- Wybierz oceniony film:** A list of movie IDs (16273, 10268, 27061, 10374) and a list of movie titles including "Mad About You: Season 1 (1992)", "Mermaids (1990)", "Charmed: Season 1 (1998)", "Seven (1995)", "The World Is Not Enough (1999)", "The Rose (1979)", "3000 Miles to Graceland (2001)", "Predator: Collector's Edition (1987)", and "The Munsters: Season 1 (1964)".
- Wybierz nie oceniony film:** A list of movie titles including "DIGI (2004)", "The Day of the Triffids (1963)", "This Island Earth (1954)", "Tabu: A Story of the South Seas (1931)", "South Pacific (2001)", "Fire Birds (1990)", "Notorious (1946)", "Sliding Doors (1998)", and "Chain Reaction (1996)".
- Wybierz metodę rekomendacji:** A dropdown menu currently showing "Pearson Correlation".
- Wartość rekomendacji:** A text field displaying "This Island Earth (1954) - 4,6854".
- Wyznacz wartość rekomendacji:** A button at the bottom right.
- Aktualne dane:** A sidebar on the right showing statistics: "Liczba filmów: 944", "Liczba użytkowników: 654", "Liczba ocen treningowych: 73784", and "Liczba ocen testowych: 2312".

Rysunek 2 Wyznaczenie wartości rekomendacji

Z list należy wybrać użytkownika oraz oceniony lub nie oceniony film. Następnie jedną z dostępnych metod wyznaczania rekomendacji. Po wybraniu wszystkich parametrów i wciśnięciu przycisku „Wyznacz wartość rekomendacji” zostanie wyświetlona wartość rekomendacji dla wybranego użytkownika i filmu.

Aby zarekomendować filmy dla użytkownika należy z menu „Rekomendacje” wybrać opcję „Zarekomenduj filmy dla użytkownika”.

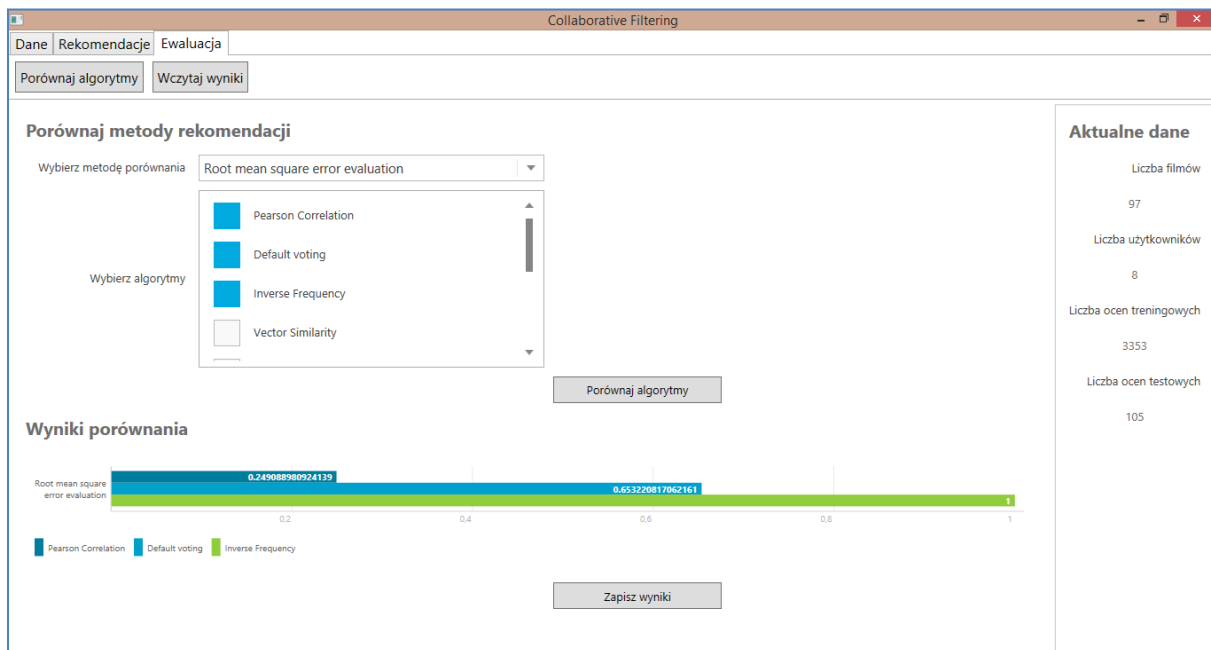
Po wybraniu użytkownika i metody rekomendacji oraz przyciśnięciu przycisku „Zarekomenduj film” na liście zostaną wyświetlone zarekomendowane filmy posortowane malejąco po wartości rekomendacji.



Rysunek 3 Rekomendacja filmów dla użytkownika

2.3. Ewaluacja metod rekomendacji

W celu oceny metod rekomendacji należy przejść do ekranu „Porównaj algorytmy” w menu „Ewaluacja”. Przed przeprowadzeniem ewaluacji należy wybrać metodę oceny algorytmów z listy oraz wskazać, które metody mają być ocenione.



Rysunek 4 Ekran oceny algorytmów

Po przeprowadzeniu oceny algorytmów, uruchamianej za pomocą przycisku „Porównaj algorytmy”, wyświetlony zostanie wykres wyników ewaluacji. Utworzony wykres może być zapisany do pliku przy użyciu przycisku „Zapisz wyniki” dla celu późniejszej analizy. Tak zapisane wyniki można później wczytać w ekranie „Wczytaj wyniki” z menu „Ewaluacja”.

3. Dane wejściowe

Użyte w projekcie dane wejściowe są podzbiorem danych udostępnionych uczestnikom konkursu Netflix Prize. Dane zawierają oceny filmów zebrane od użytkowników serwisu Netflix przez okres około 3 lat. Filmy oceniane były w skali od 1 do 5. Zbiór danych treningowych zawiera około 3,25 miliona ocen. Zbiór danych testowych składa się z około 100 000 rekordów. Dane obejmują 17 000 filmów ocenianych przez 30 000 użytkowników serwisu.

Ze względów praktycznych, w przeprowadzonych eksperymentach, zbiór danych został ograniczony. Zmniejszenie ilości danych zostało wykonane w następujący sposób:

- Posortowano dane w zbiorze testowym według identyfikatorów użytkowników,
- Wybrano M pierwszych rekordów, dotyczących pewnych N użytkowników,
- Ze zbioru treningowego wybrano wszystkie rekordy reprezentujące oceny jednego z N wcześniej wybranych użytkowników,
- Do zbioru ograniczonego dodano opisy wszystkich filmów.

Dzięki takiemu sposobowi ograniczenia ilości danych, udało się uzyskać praktyczny czas przeprowadzania eksperymentów. Ponadto zaprezentowana powyżej metoda pozwala zachować pełną informację o wszystkich wybranych użytkownikach. W eksperymentach dobrano M tak, aby przeprowadzić ewaluację na około 1% danych. Pozwoliło to wykonać pełen zestaw eksperymentów, opisanych w sekcji 5 w około 6 godzin.

4. Metody rekomendacji

Poniższy rozdział zawiera opisu zaimplementowanych algorytmów rekomendacji.

4.1. Pearson Correlation

W tej metodzie używany jest współczynnik liniowej korelacji Pearsona. Na podstawie ocenionych przez dwóch użytkowników filmów, pozwala on określić jak bardzo skorelowane są oceny tych dwóch użytkowników, czyli patrząc praktycznie, czy Ci użytkownicy mają podobny gust. Wówczas można na podstawie ocen użytkowników o podobnych gustach wyznaczyć wartość rekomendacji.

4.2. Vector similarity

W tej metodzie, dla każdego użytkownika tworzony jest wektor. Jego komórki reprezentują filmy, a wartości w komórkach oceny filmów danego użytkownika. Następnie liczone jest podobieństwo wektorów rozumiane jako cosinus kąta tych wektorów w przestrzeni wielowymiarowej. Wartości rekomendacji są wyznaczone na podstawie wartości komórek wektorów podobnych.

4.3. Default voting

Metoda głosów domyślnych polega na uzupełnieniu domeny problemu o głosy o pewnej z góry ustalonej wartości. Ma to szczególne znaczenie, jeśli w dostępnym zbiorze danych jest stosunkowo mało ocen. Następnie obliczenia wykonywane są algorytmem współczynnika korelacji.

4.4. Inverse frequency

W metodzie Vector similarity, stosowanej do analizy tekstów, występuje problem słów nie znaczących. Polega on na tym, że pewne słowa, na przykład spójniki, często występują w tekstach. Nie niosą one dodatkowego znaczenia, a w pewnym sensie sztucznie, podbijają podobieństwo tekstów. Odwrócenie częstości jest próbą zniwelowania tego efektu. W prezentowanym rozwiązaniu polega na zmniejszeniu znaczenia filmów oglądanych przez prawie wszystkich użytkowników.

4.5. Technika Case Amplification

Zadaniem techniki Case Amplification w systemach rekomendacyjnych jest redukcja szumu w danych. Polega ona na zwiększaniu wysokich podobieństw i zmniejszaniu małych. Dzięki temu, przypadkowe podobieństwa gustów użytkowników lub podobieństwa wynikające z bardzo dużej popularności pewnych filmów będą ograniczone. Praktyczna implementacja polega na podniesieniu wagi, czy współczynnika podobieństwa, do pewnej potęgi większej od 1. Wówczas wysokie podobieństwa pozostają wysokie, a małe stają się zaniedbywalne.

4.6. Uwagi dotyczące implementacji

Jako, że w przedstawionych powyżej metodach trzeba obliczać statystyki na dużym zbiorze treningowym w implementacji zastosowano programowanie równoległe. Dało to bardzo pozytywne wyniki, jeśli chodzi o czas uzyskania wartości rekomendacji. Dla zbioru danych używanego w eksperymentach, na średniej klasy komputerze PC z procesorem dysponującym ośmioma wątkami, czas uzyskania listy rekomendacji dla użytkownika wynosi około kilku sekund.

5. Opis i wyniki eksperymentów

W poniższej sekcji zostały przedstawione użyte w eksperymentach metody oceny algorytmów rekomendacji oraz wyniki eksperymentów. Eksperymenty zostały przeprowadzone na komputerze klasy PC wyposażonym w procesor Intel Core i7-2630QM z czterema rdzeniami i ośmioma wątkami oraz 6Gb pamięci operacyjnej. Wyniki uzyskano w około 6 godzin.

5.1. Metody ewaluacji

5.1.1. Mean error

Porównywany jest średni błąd dla każdej oceny ze zbioru ocen testowych. Każda z porównywanych metod, na podstawie danych treningowych, wyznacza wartość rekomendacji dla każdej oceny ze zbioru ocen testowych. Następnie porównuje wartość rekomendacji z faktyczną wartością oceny.

5.1.2. Root mean error

Działa podobnie do metody Mean error, ale liczy błąd średniokwadratowy.

5.1.3. Average rater mean error

Liczy błąd podobnie do metody Mean error, ale średni błąd jest liczony z podziałem na użytkowników.

5.1.4. Top list relevancy

Metoda sprawdza, ile spośród ocen w zbiorze testowym jest zgodnych z rekomendacjami wyznaczonymi ocenianym algorytmem rekomendacji.

5.1.5. Top list significancy

Metoda sprawdza, ile spośród ocen na liście najbardziej odpowiednich filmów dla użytkownika jest zgodnych z ocenami w zbiorze testowym.

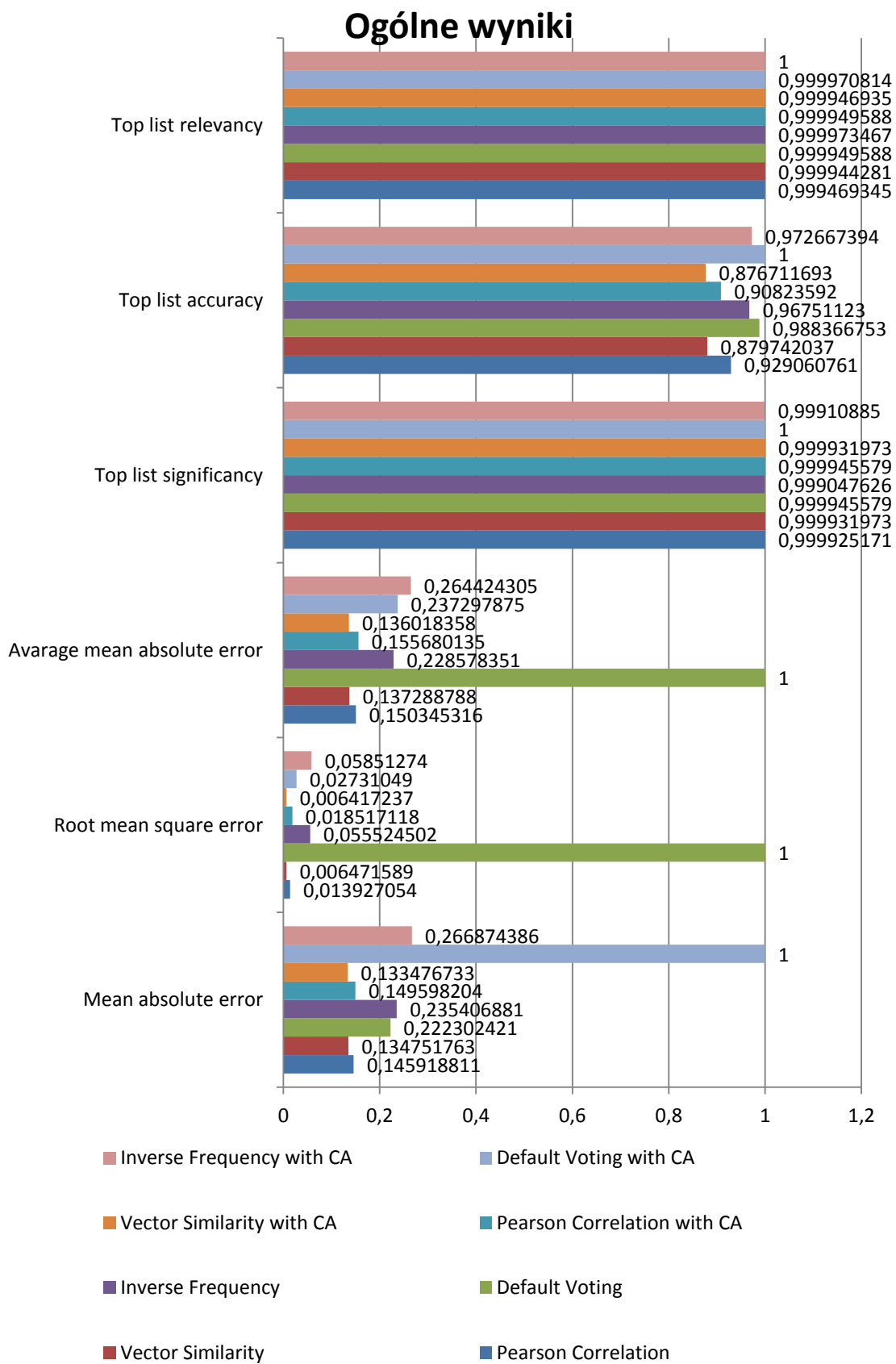
5.1.6. Top list accuracy

Metoda, dla każdego algorytmu rekomendacji, wyznacza listę najbardziej odpowiednich filmów dla każdego użytkownika. Następnie każdą rekomendację z wyznaczonych list porównuje z zawartością zbioru testowego. Jeśli w zbiorze testowym istnieje ocena aktualnego użytkownika dla aktualnego filmu, jako błąd przyjmuje się wartość bezwzględną różnicy faktycznej oceny i wartości rekomendacji. Jeśli nie, jako błąd przyjmuje się wartość rekomendacji.

5.2. Wyniki eksperymentów

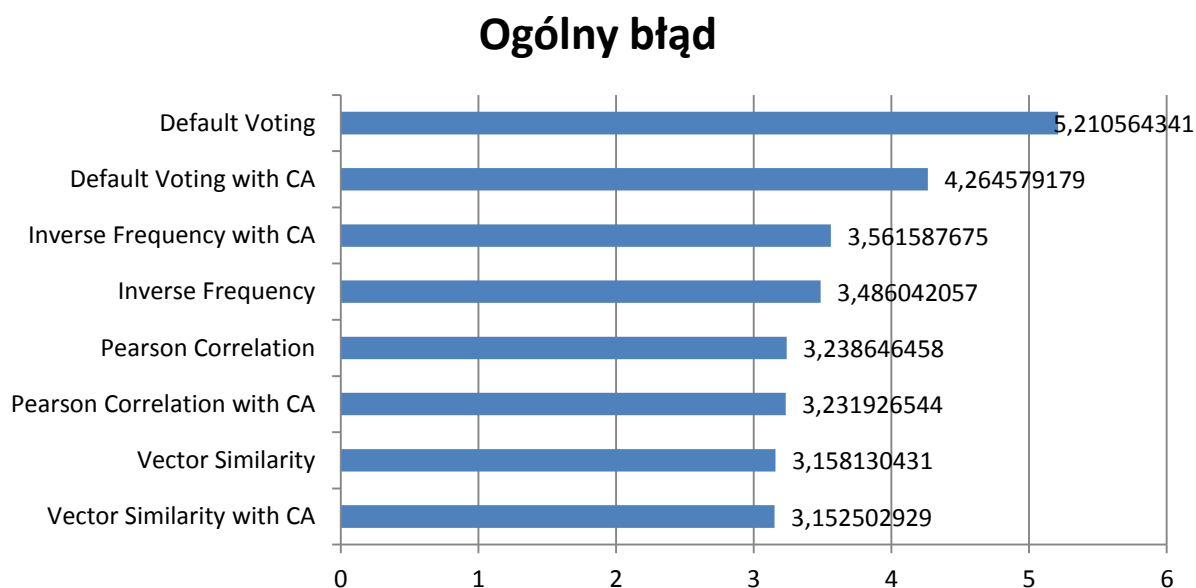
W trakcie eksperymentów wykonano ewaluację wszystkich metod wyznaczania rekomendacji wszystkimi opisanymi powyżej algorytmami oceny. W celach praktycznych dokonano normalizacji wartości uzyskanych błędów do przedziału $<0,1>$ poprzez podzielenie wszystkich rezultatów przez maksymalny z nich. Poniższe wykresy prezentują otrzymane wyniki.

Należy zauważyć, że podobne wartości błędów w metodach oceny Top list wynikają z faktu, że zbiór ocen testowych jest dużo większy niż lista zarekomendowanych filmów. Dlatego pomimo, że względne różnice są małe, to wyniki mogą być statystycznie istotne.



Rysunek 5 Ogólne wyniki

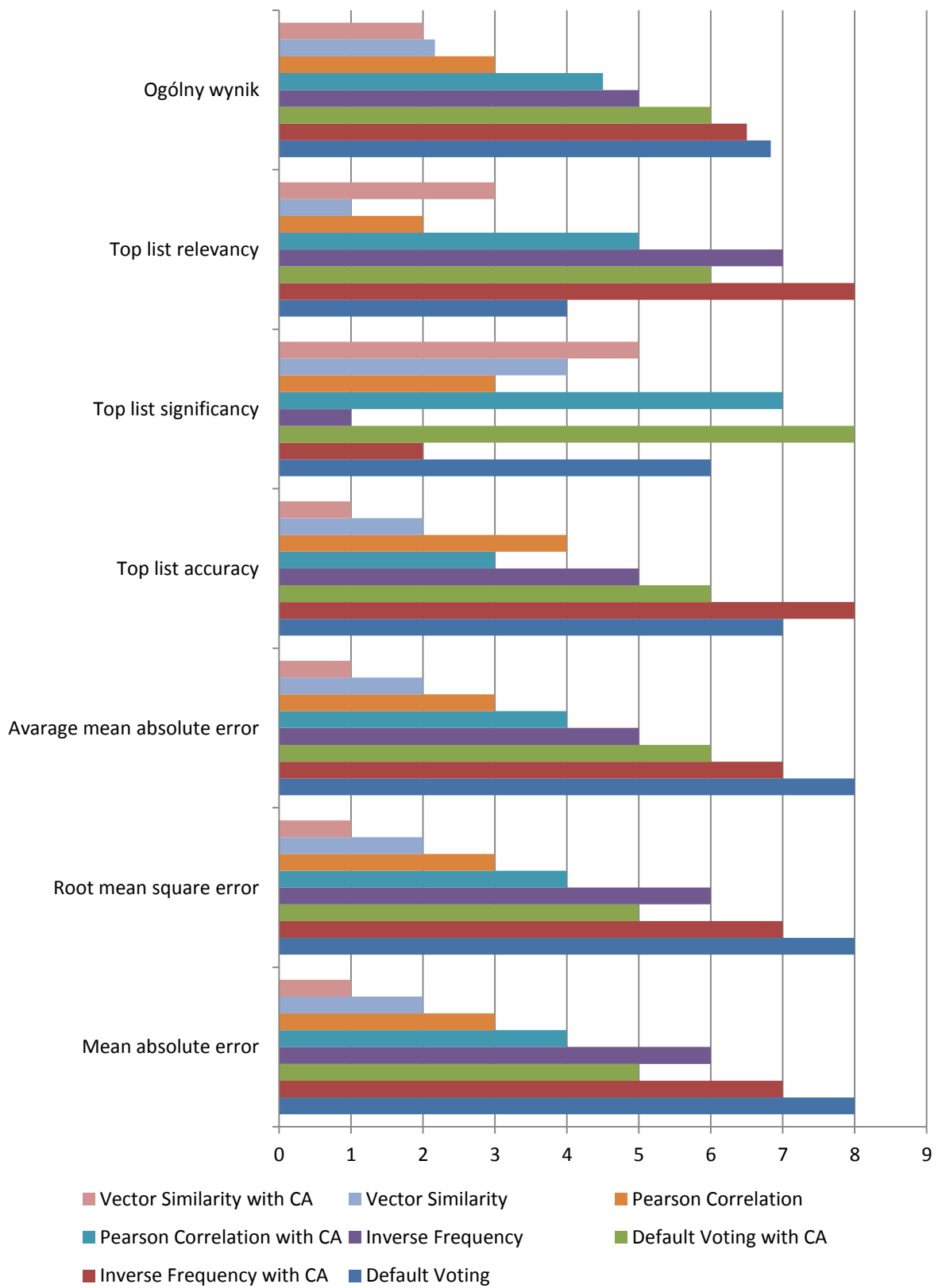
Poniższy wykres prezentuje sumy błędów uzyskanych we wszystkich eksperymentach przez każdą z metod rekomendacji.



Rysunek 6 Ogólna suma błędów

Kolejny wykres przedstawia metody porównane według kolejności ich wyników w eksperymentach. Jako ogólny wynik przyjęto średnie miejsce uzyskane przez algorytm w eksperymentach.

Miejsca algorytmów



Rysunek 7 Metody rekomendacji z podziałem na miejsca uzyskane w eksperymentach

6. Wnioski

Podsumowując, wyniki uzyskane w trakcie eksperymentów pozwalają sformułować następujące wnioski:

- Najlepszą spośród zaimplementowanych metod rekomendacji okazała się metoda Vector Similarity. Niewiele gorszą, jeśli chodzi o uzyskane wartości błędów jest metoda Pearson Correlation.
- Technika Case Amplification może poprawić uzyskiwane rezultaty, ale nie jest to regułą. Dodatkowo poprawa nie jest bardzo istotna, choć zauważalna w pewnych przypadkach.
- Zaprezentowane metody rekomendacji dobrze nadają się do zrównoleglenia, dzięki czemu można oczekiwać, że na komputerze klasy serwerowej lub rozwiązaniu rozproszonym użytkownik może otrzymać listę zarekomendowanych filmów w czasie zbliżonym do rzeczywistego.