

# Systemy rekomendacyjne

Sebastian Korniewicz

Dawid Karczewski

Marcin Sobański

# Czego się dowiedzieliśmy

**Podział ze względu na podejście:**

Content based

Collaborative Filtering

Demographic Filtering

Knowledge-Based Filtering

Hybrid

# Czego się dowiedzieliśmy

## Problemy występujące w różnych metodach:

Cold-start

Data sparsity

Accuracy

Scalability

Diversity

oth..

# Zbiór danych

MovieLens

10M ocen

10k filmów

72k użytkowników

# Jak ocenić rekomendacje?

- Model zwraca oceny - Explicit
  - Mean Absolute Error (MAE)
  - Root Mean Square Error (RMSE)
- Gdy model uczy się na zdarzeniach - Implicit
  - Mean reciprocal rank - musimy znać oceny

$$\frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

recommender system precision:  $P = \frac{\text{\# of our recommendations that are relevant}}{\text{\# of items we recommended}}$

recommender system recall:  $r = \frac{\text{\# of our recommendations that are relevant}}{\text{\# of all the possible relevant items}}$

# Content-base:

Toy Story (1995) Adventure Animation Children Comedy Fantasy

## Plusy

- Nie potrzeba danych innych użytkowników
- Rekomendacja niszowych produktów
- Wyniki są interpretowane
- Brak problemu cold startu dla produktów

Antz (1998)  
Toy Story 2 (1999)  
Adventures of Rocky and Bullwinkle, The (2000)  
Emperor's New Groove, The (2000)  
Monsters, Inc. (2001)

## Minusy

- Wyznaczenie odpowiedniej reprezentacja produktów (cechy)
- Problem porównania różnych obiektów
- Możliwość rekomendacji tego samego zbioru w kółko

# KNN Collaborative-filtering

## Plusy

- łatwość uzyskania danych
- Interpretowane wyniki

## Recommendations for Toy Story:

- 1: Willy Wonka & the Chocolate Factory (1971)
- 2: Mission: Impossible (1996),
- 3: Star Wars: Episode VI - Return of the Jedi (1983)
- 4: Independence Day (a.k.a. ID4) (1996)
- 5: Star Wars: Episode IV - A New Hope (a.k.a. Star Wars) (1977)

## Wady

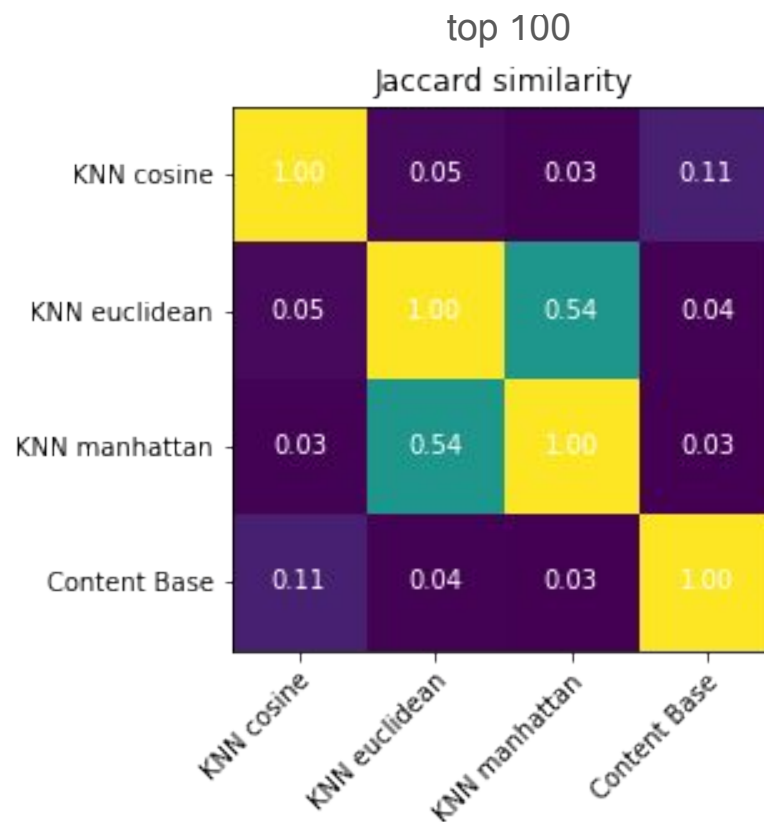
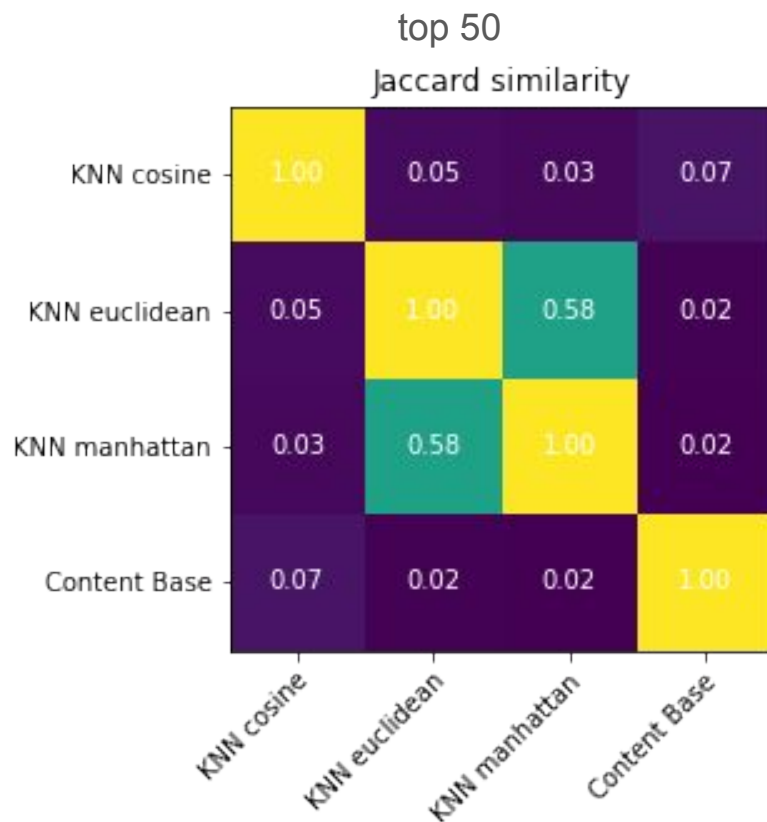
- bias popularności
- cold-start
- powtarzalność rekomendacji
- model instance-based ??

## Recommendations for Toy Story 2:

- 1: Monsters, Inc. (2001)
- 2: Shrek (2001)
- 3: Sixth Sense, The (1999)
- 4: Toy Story (1995)
- 5: Bug's Life, A (1998)

# Porównanie rekomendacji

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$





# Collaborative-filtering

# surpr!se

Dla naszego przypadku:

Dane uczące:

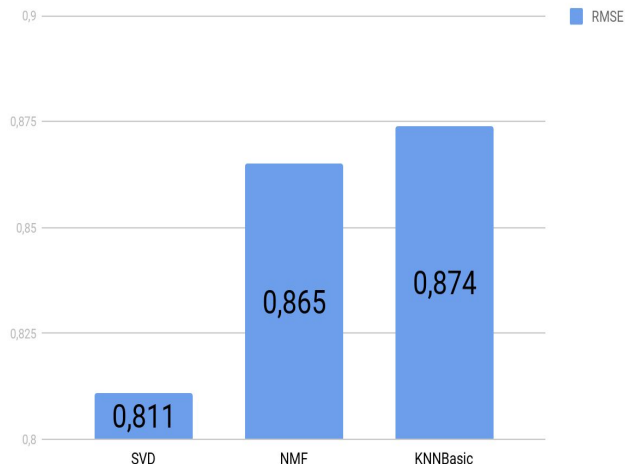
- Oceny filmów przez użytkowników

Wejście:

- użytkownik
- film

Wyjście:

- ocena



Plusy

- Proste do automatycznego pozyskiwania dane o zachowaniach (oceny, wejścia, obejrzenie)
- Łatwa ewaluacja

Minusy:

- Problem z poszerzaniem o nowych użytkowników, filmów
- Wymagana jest odpowiednio duży zbiór ocen
- Duże nakłady: obliczeniowe dla faktoryzacji, pamięciowe dla pamięciowych