

Wprowadzenie. Analiza koszykowa

Systemy Rekomendacyjne 2023/2024

Plan wykładów (wstępny)

1. Wprowadzenie. Analiza koszykowa
2. Wieloręcy bandyci
3. Podział użytkowników na grupy zainteresowań (segmentacja)
4. Rekomendacje dla portali informacyjnych
5. Rekomendacje dla grup użytkowników
6. TBD (rekomendacje grafowe)
7. TBD (jak implementować systemy rekomendacyjne)

Plan laboratoriów

1. Analiza koszykowa
2. Wieloręcy bandyci
3. Podział użytkowników na grupy zainteresowań (segmentacja)
4. Rekomendacje dla portali informacyjnych
5. Rekomendacje dla grup użytkowników
6. TBD (rekomendacje grafowe)
7. Konsultacje, wystawienie ocen

Oceny

- Za każde laboratorium można otrzymać 1 punkt
 - Laboratorium oceniane jest na tych samych lub następnych zajęciach
- Ocena z laboratoriów (i równocześnie ocena końcowa) wystawiana jest na podstawie sumy punktów, zgodnie ze skalą AGH

Czym są
systemy rekomendacyjne?

NETFLIX

Nagrodzone Emmy wciągające seriale i programy

Wybrane dla: Piotr

Moja lista

Nowości

←

Q jak zostać

Filmy, seriale i programy

Polecane

Jak zostać gwiazdą

Jak zostać królem

Dragons' Den - jak zostać milion...

NETFLIX

Nagrodzone Emmy wciągające seriale i programy

Wybrane dla: Piotr

Moja lista

Nowości

NETFLIX

Nagrodzone Emmy wciągające seriale i programy

Wybrane dla: Piotr

Moja lista

Nowości

←

Q jak zostać

Filmy, seriale i programy

Strona główna

Anime

Dramaty

Kino niezależne

Thrillery

Filmy i seriale historyczne

Horror

Filmy rodzinne i dla dzieci

Seriale

NETFLIX

Nagrodzone Emmy wciągające seriale i programy

Wybrane dla: Piotr

Moja lista

Nowości

←

Q jak zostać

Filmy, seriale i programy

Strona główna

Anime

Dramaty

Kino niezależne

Thrillery

Filmy i seriale historyczne

Horror

Filmy rodzinne i dla dzieci

Seriale

Not Sure What to Watch?

Choose Play Something and we'll pick things for you to watch based on your tastes.

Play Something

TV Dramas

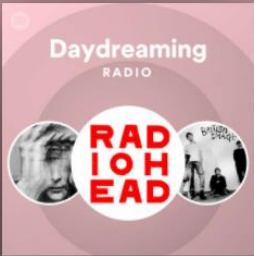


PLAYLISTA

Odkryj w tym tygodniu

Tygodniowa dawka świeżych brzmień. Nowości wybrane specjalnie dla Ciebie. Aktualizacja co poniedziałek.

Spotify • 1 polubienie • 30 utworów, 2 godz. 25 min











PLAYLISTA

Daydreaming Radio

Wraz z Radiohead, UNKLE, Damon Albarn, Jeff Buckley i innymi


Spotify • 264 polubienia • 50 utworów, 3 godz. 47 min



#	TYTUŁ	ALBUM
1	 Daydreaming Radiohead	A Moon Shaped Pool
2	 Ingenue Atoms For Peace	AMOK
3	 Suspirium Thom Yorke	Suspiria (Music for the Luca Guadagnino Film)
4	 By This River - 2004 Digital Remaster Brian Eno	Before And After Science
5	 I'm So Tired Fugazi	Instrument Soundtrack


Przygotowano dla Piotr

Im więcej słuchasz, tym lepsze kawałki Ci polecamy




Daily Mix 1

Herbert von Karajan,
Hallé, Jacqueline du Pré ...




Daily Mix 2

Dinu Lipatti, Hilary Hahn,
Radu Lupu i więcej



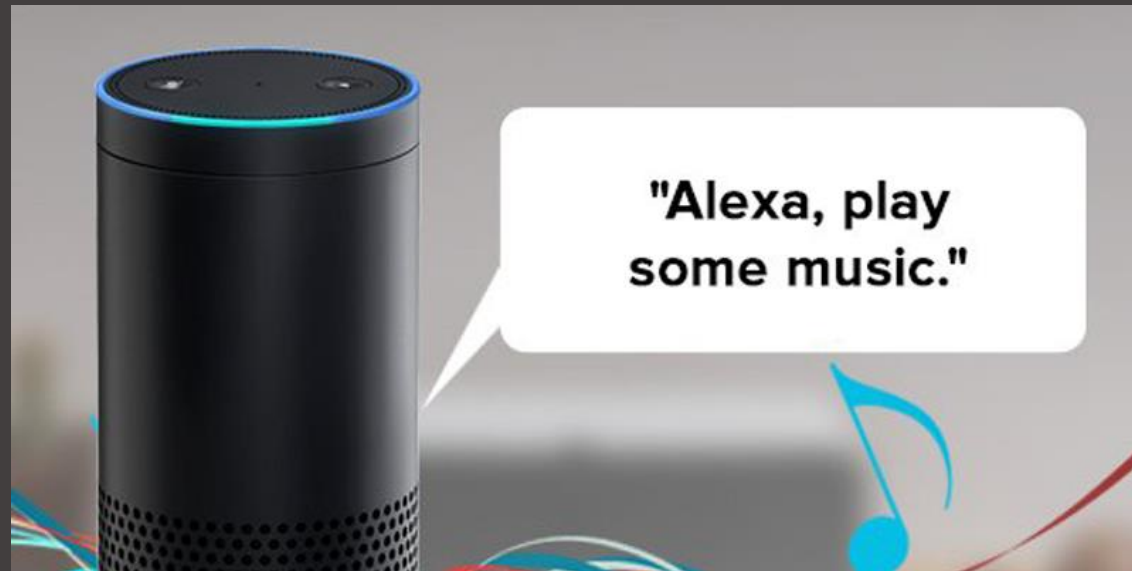
Daily Mix 3

Ennio Morricone, Bear
McCreary, Kevin MacLe...



Daily Mix 4

Víctor Pellegrini, John
Williams, Orchestra Del...



Josephine Williams

Watch

Events

Friends

Memories

See More

Shortcuts

Undiscovered Eats

Weekend Trips

Jasper's Market

Red Table Talk Group

Best Hidden Hiking Trails

See More

+

Add to Story

Tom Russo

Betty Chen

Dennis Han

Cynthia Lopez

What's on your mind, Josephine?

Photo/Video

Tag Friends

Feeling/Activity

Fiona Ozeri

5 hrs · 🌐

This has some great healthy recipes

Sponsored

Lebo's Pizza

Experience the trendy pizza spot in Palo Alto being called the next big thing.

Birthdays

Jessica, Erica and 2 others have birthdays today

Contacts

Dennis Han

Eric Jones

Cynthia Lopez

Betty Chen

Tina Lim

Molly Carter

Osoby, które możesz znać

Alice

4 wspólnych znajomych

Dodaj do znajomych

Usuń

Bob

2 wspólnych znajomych

Dodaj do znajomych

Usuń

Carol

17 wspólnych znajomych

Dodaj do znajomych

Usuń

Dave

8 wspólnych znajomych

Dodaj do znajomych

Usuń

Pokaż wszystkich

Proponowane dla Ciebie

Zobacz więcej

Grupy, które mogą Cię zainteresować

#błm

Praca dla programistów i programistek powyżej 20000zł, 30000zł i 40000zł

24 tys. członków grupy

30 znajomych jest członkami grupy

Dolóż do grupy



WYBRANE DLA CIEBIE



WIADOMOŚCI



Były agent z Korei Północnej przerywa milczenie "To kwestia dni. Tusk złoży propozycję"

"Internetowa apokalipsa". Burza uderzy w Ziemię

Jutro spadnie krupa śnieżna i śnieg w Polsce

7-latek terroryzuje rówieśników w szkole

Gowin pisze list do wyborców

Spicie dziennikarza i posła PiS ws. wyroku TK

Wir polarny jest zakłócany, co zwiastuje zimę

Za kilka dni możemy poznać nowych ministrów

- Miał władzę i był lojalnym, komunistycznym sługą, ale i tak musiał uciec z Korei Północnej w 2014 r.
- Od tego czasu mieszka w Seulu i pracuje dla wywiadu Korei Południowej
- Teraz opowiada BBC o tym, co musiał zrobić, by spełnić oczekiwania północnokoreańskich przywódców
- Dla nich zlecał zabójstwa, a także stworzył profesjonalne laboratorium metaamfetaminy
- Więcej takich tematów znajdziesz na stronie głównej Onetu



Więźnia w Korei Południowej były kiedyś wypelnione szpiegami aresztowanymi przez dziesięciolecia za różnego rodzaju prace szpiegowskie na rzecz Korei Północnej. Jednak od 2017 r. jest ich mniej, bo kraj Kim Dzong Una stawia na nowe technologie. Uciekinierzy z Pjongiang ostrzegają, że Korea Północna ma armię sześciu tys. wykwalifikowanych hakerów. I jest to prawdopodobne, bo to północnokoreańska grupa Lazarus miała stać za cyberatakami, który sparaliżował część NHS (National Health Service - red.) i innych organizacji na całym świecie w 2017 r.

- Polecamy: [Koszmar złapanych uciekinierów. Kim Dzong Un przykręcił śrubę](#)



Wypadek miał miejsce w niedzielę na skrzyżowaniu ulic Armii Polskiej i 30 stycznia. Najpierw doszło tam do zderzenia dwóch samochodów osobowych – chevroleta i BMW. Pierwszy z nich odbił się i potrącił 4-letniego chłopca. Na skutek odniesionych obrażeń dziecko zmarło.

ZOBACZ RÓWNIEŻ



Kierowca śmiertelnie potrącił 4-latkę i uciekł. Policja ma "swoje typy"



Tragedia w Gorzowie Wielkopolskim. Rozpędzone auto wjechało w 4-latkę

– Bezpośrednio po wypadku kierowca chevroleta pieszo oddalił się z miejsca zdarzenia. Obecnie poszukują go policjanci ze wszystkich pionów gorzowskiej komendy oraz z oddziałów prewencji. Nie wykluczamy, że zbiegł, gdyż mógł prowadzić auto pod wpływem alkoholu, bądź narkotyków – tłumaczył mediom w poniedziałek rano Marcin Maludy, rzecznik lubuskiej policji.

Google

recommender |

- recommender **systems**
- recommender **systems machine learning**
- recommender **meaning**
- recommender **synonym**
- recommender **system is an example of**
- recommender **system python**
- recommender **system algorithms**
- Recommender Systems Handbook
- recommender **common app**
- recommender **engine**

Google Search

I'm Feeling Lucky

Report inappropriate predictions

Google

spotify

All Images Videos Maps News More

Tools

About 350,000,000 results (0.57 seconds)

Ad · <https://www.spotify.com/>

Spotify® Premium - Wypróbuj miesiąc za darmo

Twoi ulubieni wykonawcy i muzyka bez przerw. Potem 19,99 zł / mies. Obowiązują Warunki. Odkryj świetnych muzyków. Playlisty na każdą chwilę.

Śłuchaj gratis w Spotify®

Odtwarzaj utwory, które kochasz za darmo. Zarejestruj się już dziś

Spotify® dla studentów

Spotify® Premium za 50% ceny 1 miesiąc za darmo. 9,99 zł/mies.

Spotify® dla rodziny

Podaruj bliskim muzykę bez końca. 1 miesiąc za darmo. 29,99 zł/mies.

Spotify® Premium Duo

2 ludzi, 2 konta Premium. 1 miesiąc za darmo. 24,99 zł/mies.

<https://www.spotify.com>

Spotify

Spotify is all the music you'll ever need.

Download Spotify

Spotify is a digital music service that gives you access to millions ...

Sign up

Sign up with your email address · By clicking on sign-up, you ...

More results from [spotify.com](https://www.spotify.com)

Premium

Get one month free. Ad-free music listening. Get over 50 million ...

Free Mobile App

Spotify is a digital music service that gives you access to millions ...

People also ask

Is Spotify really free?

How can I use Spotify for free?

Is Spotify music illegal?

<https://twitter.com/Spotify>

Spotify (@Spotify) · Twitter

Drop your comfort album 📀
#WorldMentalHealthDay

Twitter · 19 hours ago



It's okay to take a break 📀
We hope this World Mental Health Day, you take a beat and tune into you.

Twitter · 2 days ago



@aespa_official are always leveling up and their first Mini Album #Savage is proof
spotify.link/savage

Twitter · 3 days ago

Top stories

Sydney Morning Herald

IVF on us: Spotify leads the way with family benefits for Australian staff

1 day ago



The Indian Express

Spotify Premium gets Diwali discount: Here are the details

8 hours ago



View all

Systemy rekomendacyjne:

- wybierają podzbiór dostępnych elementów
- optymalizują zadaną miarę
 - zysk
 - zadowolenie użytkowników
- automatyzują proces doboru treści na stronach
- zastępują człowieka
 - nie męczą się
 - analizują większą ilość danych
 - są bezstronne (w pewnym sensie)

Co możemy optymalizować?

- Miary syntetyczne:

- accuracy,
- precision,
- recall,
- F1,
- ...

- Miary biznesowe:

- CTR (click-through ratio),
- czas spędzony na czytaniu/słuchaniu/oglądaniu,
- głębokość scrolla,
- przychody z reklam,
- liczba subskrypcji,
- powracalność użytkowników,
- długość i liczba sesji,
- ...

Jakimi danymi dysponujemy?

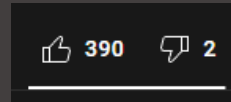
Dane od użytkowników (feedback)

- Feedback bezpośredni:

- Thumb-up



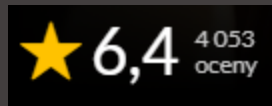
- Polubienia



- Reakcje



- Oceny



- Udostępnienia

- Komentarze

- Feedback pośredni:

- Impresje

- Kliknięcia

- Brak kliknięć

- Czas spędzony

- Głębokość scrolla

- Powracalność (np. do kolejnych odcinków)

Dane dotyczące treści

- Dane o autorach, czasie utworzenia, miejscu pochodzenia
- Analiza tekstu (podobieństwo tekstów, named entities, sentiment)
- Tagi (gatunek filmu, muzyki, tematyka artykułu, rodzaj piwa)
- Parametry (taneczność lub instrumentalność muzyki, mroczność lub komediowość filmu)

Algorytmy naiwne

Losowy

- Losowa próbka ze zbioru elementów rekomendowanych
- Należy rozważyć, czy prawdopodobieństwo wylosowania każdego elementu powinno być takie samo
- Warto wyeliminować niektóre elementy (np. już przeczytane/obejrzane)

Naj-

- Najczęściej czytany, najlepiej oceniany, najbardziej kontrowersyjny, ...
- Warto rozważyć zawężenie grupy docelowej (np. "Najpopularniejsze wśród fanów Black Mirror")
- Prosty i efektywny
- Zaniedbuje nowe materiały

Algorytm-pułapka

- Najbardziej podobne do tych, które użytkownik chętnie czytał/oglądał/słuchał
- W jaki sposób możemy porównywać elementy między sobą?
- W krótkiej perspektywie może dać świetne rezultaty
- Użytkownik w końcu się znudzi

Analiza koszykowa

Definicja problemu

- Naszym biznesem jest platforma e-commerce
 - np. sklep internetowy albo platforma typu Allegro
- Zbieramy historyczne dane o koszykach użytkowników - co i kiedy było kupowane
- Na podstawie bieżącego, jeszcze niesfinalizowanego koszyka użytkownika chcemy wygenerować rekomendację - co jeszcze ten użytkownik może kupić?

Wskaźniki

Support

- Jaki odsetek wszystkich transakcji zawierał (wszystkie) elementy ze zbioru A

$$\text{supp}(A) = \frac{|\{t \in T; A \subseteq t\}|}{|T|}$$

Confidence

- Jaka jest szansa, że użytkownik kupi produkt B, jeśli kupił zbiór produktów A

$$\text{conf}(A \Rightarrow B) = \frac{\text{supp}(A \cup B)}{\text{supp}(A)}$$

Lift

- Jaka jest istotność obserwacji wskazanej w *confidence*
 - Innymi słowy - czy kupno produktów ze zbioru A i produktu B jest niezależne

$$\text{lift}(A \Rightarrow B) = \frac{\text{supp}(A \cup B)}{\text{supp}(A) \times \text{supp}(B)}$$

Rekomendacje

Idea

- Użytkownik ma w koszyku zbiór artykułów A
- Dla każdego podzbioru artykułów A' chcemy znaleźć produkty B o wysokim *confidence* i *lift*

Idea

- Użytkownik ma w koszyku zbiór artykułów A
- Dla każdego podzbioru artykułów A' chcemy znaleźć produkty B o wysokim *confidence* i *lift*
- *Problem jest NP-trudny*

Ograniczenie rozmiaru zbioru A

- Możemy podjąć autorytatywną decyzję, że nie będziemy analizować zbiorów produktów większych niż k
- Obliczamy i -wymiarowe macierze *support* kolejno dla zbiorów $1, 2, 3, \dots, i, \dots, k$ -elementowych
- Na podstawie wartości *support* możemy obliczyć *confidence* i *lift* dla każdego co najwyżej k -elementowego podzbioru towarów w koszyku użytkownika
- Wybieramy najbardziej wartościowe rekomendacje produktów

Ograniczenie wartości *support*

- Zakładamy, że zbiory produktów, których *support* $< \varepsilon$ są nieistotne z punktu widzenia rekomendacji
- Zauważmy, że jeśli *support*($\{a, b, c\}$) $< \varepsilon$, to także *support*($\{a, b, c, d\}$) $< \varepsilon$
- Zaczynając od jednoelementowych zbiorów produktów, obliczamy *support* dla wszystkich możliwości i jeśli jest $> \varepsilon$ - zapisujemy w strukturze danych

Struktury danych

- Słownik
 - Kluczem jest lista produktów
 - Może być lista ID
 - Może być hash z listy
 - Warto posortować produkty leksykograficznie
 - Wartością jest *support*
 - Jeśli w słowniku nie ma klucza dla danego zbioru produktów, to zakładamy *support == 0.0*

Struktury danych

- Graf acykliczny skierowany
 - Każdy wierzchołek grafu zawiera listę produktów i wartość *support*
 - Krawędzie skierowane prowadzą do węzłów zawierających o jeden produkt więcej
 - Np. z węzła opisującego zbiór (mąka, mleko, jajka) do węzła (mąka, mleko, jajka, cukier)
 - Warto dodać specjalny wierzchołek "zerowy"

Jak ocenić jakość
rekomendacji?

Dlaczego to nie jest takie proste?

- Standardowa procedura - podział na zbiór treningowy i testowy, porównanie predykcji wygenerowanych przez model ze zbiorem testowym
- Skąd mamy wiedzieć, jak bardzo nasz system się myli?
- Bezwzględne wartości niewiele nam mówią, musimy je porównać z jakimiś punktami odniesienia (ang. *baseline*)
- Dobry punkt odniesienia stanowią naiwne algorytmy (zwłaszcza losowy jest powszechnie używany, ale trzeba to robić z głową)

Rekomendacje pozytywne i negatywne

- Podczas oceny wytrenowanego modelu, obliczamy liczbę rekomendacji prawdziwie dodatnich (TP), fałszywie dodatnich (FP), prawdziwie ujemnych (TN) i fałszywie ujemnych (FN)

Miary syntetyczne

- $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$
- $\text{Precision} = (\text{TP}) / (\text{TP} + \text{FP})$
- $\text{Recall} = (\text{TP}) / (\text{TP} + \text{FN})$
- $\text{F1} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$

Rekomendacje nie są jednoelementowe

- Chcielibyśmy ocenić jakość rekomendacji na podstawie pierwszych k elementów spośród N wskazanych przez algorytm

Miary dla wieloelementowych rekomendacji

- $Relevant@k$ – liczba wyników TP spośród pierwszych k elementów zwróconych przez algorytm
- $Precision@k = Relevant@k / k$
- $Recall@k = Relevant@k / Relevant@N$

Podsumowanie

- Czym są systemy rekomendacyjne?
- Jakie dane możemy zbierać?
- Jakie miary możemy optymalizować?
- Analiza koszykowa – prosty przykład rekomendacji opartych o statystykę
- Jak porównać jakość algorytmów?