

Matematyczne modele wykorzystywane w systemach rekomendacji.

Anita Kudaj

5 lutego 2019

Spis treści

1	Wstęp	3
2	Preliminaria	4
2.1	Definicje związane z rachunkiem prawdopodobieństwa	4
2.2	Definicje związane z algebrą liniową	4
2.3	Eksploracja danych i metody statystyczne użyte w regułach rekomendujących	4
2.4	Definicje związane z metodami	4
3	Modele tworzenia rekomendacji	5
3.1	Filtrowanie kolaboratywne (Collaborative filtering)	5
3.1.1	Problem	5
3.1.2	Metoda	5
3.1.3	Filtrowanie kolaboratywne oparte na użytkowniku	6
3.1.4	Filtrowanie kolaboratywne oparte na elementach	7
3.1.5	Wady i zalety filtrowania kolaboratywnego	8
3.1.6	Porównanie filtrowania kolaboratywnego opartego na użytkownikach i filtrowania kolaboratywnego opartego na elementach: . . .	9
3.2	Systemy rekomendujące oparte na treści (Content-based recommender systems):	10
3.2.1	Problem	10
3.2.2	Metoda i Algorytm	10
3.2.3	Przykład	13
3.2.4	Wady i zalety systemu rekomendującego opartego na treści . . .	15
3.3	Systemy rekomendujące kontekstowe (Context – aware recommender systems):	16
3.3.1	Problem	16
3.3.2	Metoda i Algorytm	16
3.3.3	Przykład	17

3.3.4	Wady i zalety systemów rekomendujących kontekstowych	18
3.4	Hybrydowe systemy rekomedyjacje ():	19
3.5	Systemy rekomendujące oparte na modelach():	19
4	Eksperymenty / część praktyczne	20
5	Podsumowanie	21

Rozdział 1

Wstęp

Rozdział 2

Preliminaria

2.1 Definicje związane z rachunkiem prawdopodobieństwa

Przykładowe?

2.2 Definicje związane z algebrą liniową

Przykładowe?

2.3 Eksploracja danych i metody statystyczne użyte w regułach rekomendujących

1.Odległość euklidesowa 2.Podobieństwo kosinusów 3.Współczynnik korelacji Pearsona
4.Faktoryzacja macierzy.

5.Term frequency 6.Term frequency inverse document frequency

Co ponadto? -i Czy wymagane jest opisanie, np.:

-Grupowania: k-średnich

-Klasyfikacje: Regresja liniowa, Metoda KNN, Drzewa decyzyjne

2.4 Definicje związane z metodami

1.Przedmiot 2.Użytkownik 3.Preferencje 4.Cecha/własność

Rozdział 3

Modele tworzenia rekomendacji

3.1 Filtrowanie kolaboratywne (Collaborative filtering)

3.1.1 Problem

Rozważmy sytuację w której szukamy opinii na temat ostatnio wydanej książki znanego pisarza. Otrzymujemy dobre rekomendacje co prawdopodobnie zachęci nas do jej przeczytania. Poznajemy stwierdzenia, że ta sama książka jest katastrofą co powoduje, że nie chcemy tracić pieniędzy na jej zakup oraz czasu na oddanie się lekturze. W trzecim przypadku otrzymujemy dwie sprzeczne opinie na temat tej samej pozycji co powoduje, że wybór staje się trudniejszy. We wszystkich przypadkach analizujemy wartość uzyskanych ocen i ostatecznie podejmujemy na ich podstawie najbardziej odpowiednią dla nas decyzję.

3.1.2 Metoda

Metody rozwiązywania podobnych problemów często spotykane są w matematyce. Jednym z przykładów jest **filtrowanie kolaboratywne**. W rozważanej metodzie wyróżniamy dwa podstawowe typy:

- filtrowanie kolaboratywne oparte na użytkowniku (ang. user-based)
- filtrowanie kolaboratywne oparte na elementach (ang. item-based)

Cechą wspólną dla obu powyższych metod jest fakt, że oceny jednych użytkowników są podstawą do tworzenia rekomendacji dla innych.

3.1.3 Filtrowanie kolaboratywne oparte na użytkowniku

Idea rozważana pod tym hasłem mówi, że jeżeli użytkownicy A i B wykazują podobieństwo oraz użytkownik A zaopiniuje pewien przedmiot, którego użytkownik B jeszcze nie ocenił, to prawdopodobnie opinia użytkownika B będzie podobna do opinii użytkownika A.

3.1.3.1 Algorytm

Stworzenie rekomendacji opartej na użytkowniku wykonamy w kolejnych krokach:

1. Znalezienie podobieństwa między czytelnikami opierającego się na informacji o przeczytanych książkach. Najczęstszymi stosowanymi podejściami do obliczania szukanego podobieństwa są Metryka Euklidesowa i Współczynnik Korelacji Pearsona.
2. Wyestymowanie ocen, które czytelnicy (w szczególności osoba dla której tworzymy rekomendację) mogliby wystawić dla nieprzeczytanych książek z rozważanego zbioru.

W celu dokładniejszego zrozumienia tego typu filtrowania został przedstawiony poniższy przykład.

3.1.3.2 Przykład

Zakładamy, że tabela przedstawia oceny czytelników dla kilku wybranych książek oraz, że każda z zapytanych osób mogła wystawić ocenę z zakresu 1 -10. Istotne jest, że nie wszyscy zapytani wystawili ocenę dla każdej z książek:

Książka/Czytelnik	Anna	Maciej	Bartek	Ewa	Sandra	Kacper
Władza Absolutna	6	3		6	4	
Patrioci		6	6	5	6	
Proxima	7	7	8	7	8	9
Strażacy	8	10	10	7	6	8
Gra o Śmierć	9	6	6	6	6	
Cyrk	5	7	7	5	4	2

W tym przykładzie zastosujemy pierwsze ze wspomnianych rozwiązań. Używając wzoru:

$$d_e(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}$$

obliczymy szukane odległości:

(tabela)

Bazując na podobieństwach między poszczególnymi użytkownikami, przez obliczenie średniej ważonej, zostaje przewidziana ocena jaką Bartek zaproponuje dla książki „Władza Absolutna”. W poniższym równaniu wartość podobieństwa między Bartkiem i innymi użytkownikami została pomnożona przez ocenę jaką dany użytkownik wystawił dla książki „Władza Absolutna”. Następnie, w celu normalizacji, wynik został podzielony przez sumę wartości wszystkich podobieństw.

(obliczenie)

Ostatecznie, gdy znane są oceny dla wszystkich książek dokonana zostaje rekomendacja dla naszego użytkownika.

3.1.4 Filtrowanie kolaboratywne oparte na elementach

W przypadku filtrowania kolaboratywnego opartego na elementach wartości podobieństwa między użytkownikami zostaje zastąpiona przez wartości podobieństwa między elementami.

W tym przypadku założenie mówi, że jeżeli użytkownik wybrał przedmiot A w przeszłości oraz przedmiot B jest podobna do A, to użytkownik będzie skłonny wybrać również przedmiot B.

3.1.4.1 Algorytm

Podobnie jak w przypadku opartym na użytkowniku, w tym również należy wykonać dwa kroki.

1. Pierwszym etapem jest znalezienie podobieństw występujących między elementami. Najczęstszą miarą podobieństwa w tym przypadku jest podobieństwo kosinusów. Miara ta wyraża podobieństwo między n-wymiarowymi wektorami poprzez kąt między nimi w przestrzeni wektorowej. Wraz ze wzrostem wartość kąta rośnie podobieństwo.
2. Następnie, na podstawie wydanych przez użytkownika ocen, należy wyestymować noty dla elementów przez niego nieocenionych.

3.1.4.2 Przykład

Chcąc zastosować ten rodzaj filtrowania kolaboratywnego do rozważanego problemu i przewidzieć ocenę użytkownika dla pewnej wybranej książki należy na wstępie zdefiniować wszystkie książki podobne do wybranej. Można to zrobić używając wspomnianego wcześniej, podobieństwa kosinusów. W tym przypadku elementami wektorów są oceny wystawione przez użytkowników dla poszczególnych książek, np.: (7,7,8,7,8,9) dla książki "Proxima". Poniższa tabela przedstawia natomiast podobieństwa występujące między wszystkimi książkami obliczone za pomocą wzoru:

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|}$$

	Władza Absolutna	Patrioci	Proxima	Strażacy	Gra o Śmierć	Cyrk
Władza Absolutna	1	0.6339	0.7372	0.7195	0.8935	0.7599
Patrioci	0.6339	1	0.7951	0.8150	0.7977	0.8898
Proxima	0.7372	0.7951	1	0.9780	0.8586	0.9200
Strażacy	0.7195	0.8150	0.9780	1	0.8860	0.9681
Gra o Śmierć	0.8935	0.7977	0.8586	0.8860	1	0.9413
Cyrk	0.7599	0.8898	0.9200	0.9681	0.9413	1

Znając podobieństwa między pozycjami oraz noty wystawione przez Kacpra możemy więc wyestymować jak Kacper oceni inne pozycje z naszego spisu.

Rozważmy książkę "Patrioci". Ustalone podobieństwa między pozycją "Patrioci" a każdą z innych książek ocenianych przez Kacpra wymnożymy przez oceny, które nadał pozycjom. Następnie sumę iloczynów dzielimy przez sumę wszystkich podobieństw.

$$\frac{(0.7951 \cdot 9 + 0.8150 \cdot 8 + 0.8898 \cdot 2)}{(0.7951 + 0.8150 + 0.8898)} = 6.16$$

Na podstawie przeprowadzonych obliczeń zakładamy, że ocena jaką wystawiłby po przeczytaniu Kacper książce "Patrioci" to 6.16. Powtarzając powyższe obliczenia dla każdej z pozycji nieocenionych przez Kacpra otrzymamy wszystkie brakujące opinie. Następnie bazując na zdobytych danych z łatwością odnajdziemy pozycję najbardziej odpowiednią do zarekomendowania naszemu użytkownikowi.

3.1.5 Wady i zalety filtrowania kolaboratywnego

Mając przed sobą dwa typy filtrowania kolaboratywnego możemy zadać pytanie o efektywność, czy precyzyjność tego rozwiązania.

Poniżej kilka wniosków i informacji, opartych na rozważaniach Christian Desrosiers i George Karypis [44] [file:///C:/Users/akuda/Downloads/NbrRSsurvey2011.pdf] [strona 7] [A comprehensive survey of neighborhood-based recommendation

methods][Christian Desrosiers, George Karypis] które pozwalają dostrzec zalety i wady tego rozwiązania.

Zalety:

- Opisywane podejście tworzenia rekomendacji jest intuicyjne i łatwe implementacji zarówno w przypadku metody opartej na użytkownikach jak i metody opartej na elementach.
- Metody filtrowania kolaboratywnego pozwalają ponadto na zwięzłe i intuicyjne wyjaśnienie obliczeń prognostycznych, które wykonujemy.
- W rozważanych metodach filtrowania nie są wykorzystywane informacje o wartości produktów, czy informacje o profilu użytkownika. Kiedy więc wzrośnie liczba ocen dla konkretnego produktu zmiana ulegnie jedynie wartość podobieństwa między elementami.

Z drugiej strony:

- Filtrowanie kolaboratywne jest kosztowne obliczeniowo, ponieważ wykorzystywane są tu informacje o użytkownikach, produktach oraz ocenach produktów przez użytkowników.
- Podejście to zawodzi, kiedy istnieje potrzeba stworzenia rekomendacji dla nowego użytkownika o którego ocenach nie ma informacji.
- Zarówno metoda oparta na użytkownikach, jak i metoda oparta na elementach jest mało wiarygodna kiedy zasób danych na którym bazujemy jest mały.

3.1.6 Porównanie filtrowania kolaboratywnego opartego na użytkownikach i filtrowania kolaboratywnego opartego na elementach:

Warty rozważenie jest również fakt wyboru między rekomendacją opartą na użytkowniku, a rekomendacją opartą na elementach. Według Christian Desrosiers, George Karypis [44][<file:///C:/Users/akuda/Downloads/NbrRSsurvey2011.pdf>] [strona 7] [A comprehensive survey of neighborhood-based recommendation methods][Christian Desrosiers, George Karypis] jest kilka obszarów, które należy rozważyć przed ostatecznym wyborem toku postępowania:

- **Precyzyjność:** Metodę wybieramy w zależności od stosunku między użytkownikami a przedmiotami w rozważanych danych. Mianowicie, jeżeli rozważany zbiór

zawiera dużą liczbę użytkowników i jednocześnie małą liczbę elementów preferowanym rozwiązaniem jest metoda oparta na elementach.

- **Sprawność:** Złożoność rozważanych algorytmów zależy od stosunku między liczbą użytkowników, a liczbą elementów. Przyjmując O , U , E jako liczbę odpowiednio ocen, użytkowników i elementów zdefiniujemy $p = O/U$ i $q = O/E$. Wtedy też złożoność metody opartej na użytkownikach wyrażona zostaje przez p^2/E , a złożoność metody opartej na elementach przez q^2/U .
- **Stabilność:** Rozważając ten aspekt przed wyborem metody należy rozważyć co wzrasta szybciej – liczba użytkowników, czy liczba elementów. Jeżeli liczba elementów wydaje się bardziej statyczna wtedy też lepszym wyborem jest metoda oparta na elementach i odwrotnie.
- **Uzasadnienie:** Pod tym względem lepszym wyborem będzie system rekomendacji oparty na elementach. W przypadku bowiem potrzeby wyjaśnienia naszej rekomendacji przedstawienie listy rozważanych elementów jest łatwiejsze niż przedstawienie listy użytkowników.
- **Serendipity:** Patrząc pod kontem możliwości wyszukiwania zaskakujących rekomendacji lepszym wyborem byłby system oparty na użytkowniku. Pozwala on bowiem dojść do znacznie ciekawszych wniosków niż system oparty na elementach.

3.2 Systemy rekomendujące oparte na treści (Content-based recommender systems):

3.2.1 Problem

Wspomniany wcześniej system rekomendacji filtrowania kolaboratywnego opiera się na informacji o ocenach przyznanych poszczególnym elementom w rozważanym zbiorze. Analizując jednak przypadek osoby, która przyznała ocenę 5 dla wybranej książki zauważmy, że użytkownik ten miał na uwadze wiele czynników, na przykład: zawartą historię, gatunek, styl pisanie autora.

3.2.2 Metoda i Algorytm

W tym przypadku rozwiązaniem matematycznym są systemy rekomendujące oparte na treści, ukierunkowane na spersonalizowany poziom użytkownika, rozważają jego

indywidualne preferencje oraz treść produktu. Analizie zostają poddane zestawy dokumentów i opisy przedmiotów, które wcześniej były oceniane przez użytkowników. Następnie bazując na cechach obiektów ocenianych przez kandydata zostaje zbudowany jego profil. Wspomniane metody opierają się na obliczaniu podobieństw oraz wykorzystują zadania uczenia maszynowego, takie jak klasyfikacja.

W typie rekomendacji opartym na treści stworzenie rekomendacji i wygenerowanie listy elementów, które mogą być odpowiednie użytkownikowi możemy przedstawić w trzech krokach:

1. Wygenerowanie profilu produktu

W większość systemów rekomendacji opartych na treści można zauważyć użycie prostych modeli wyszukiwujących. Jednym z najbardziej popularnych jest Model Przestrzeni Wektorowej (*ang. Vector Space Model*) z algorytmem TF-IDF (*ang. TF – term frequency, IDF – inverse document frequency*). Jest to przestrzenna forma reprezentacji dokumentów. W modelu tym dokument jest reprezentowany przez wektor w przestrzeni n -wymiarowej, gdzie każdy z n wymiarów stanowi rozważaną cechę produktu.

Formalnie rzecz ujmując każdy z dokumentów jest przedstawiony za pomocą wektora wag, gdzie waga w odpowiedni sposób wyraża zależność między dokumentem, a badanym terminem.

Niech $P = \{p_1, p_2, \dots, p_n\}, n \in \mathbf{N}$ będzie zestawem dokumentów / analizowanych przedmiotów. Natomiast $C = \{c_1, c_2, \dots, c_n\}, n \in \mathbf{N}$ zestawem cech rozważanych w przedmiotach. Każdy z dokumentów $p_j, j \in \{1, \dots, n\}$ jest reprezentowany jako wektor w przestrzeni wektorowej n -wymiarowej. Zatem $p_j = \{w_{1j}, w_{2j}, \dots, w_{nj}\}$, gdzie w_{kj} jest wagą dla cechy c_k w dokumencie p_j .

Do generowania profilu produktu używany jest wspomniany wcześniej algorytm **TFIDF** pozwalający policzyć względną ważność powiązania cechy z przedmiotem. Zakładamy tu, że:

- rzadkie cechy są równie istotne jak częste (założenie IDF),
- kilkukrotne wystąpienie terminu w rozważanym dokumencie jest równie istotne jak pojedyncze (założenie TF),
- długość dokumentu (filmu, książki) nie ma znaczenia (założenie normalizacji).

Możemy więc powiedzieć, że jeżeli termin występuje często w konkretnym przedmiocie rozważań (TF) i równocześnie rzadko w pozostałych elementach zbioru (IDF) ma większe prawdopodobieństwo stać się jedną z istotnych cech rozważanych w temacie. Ponadto normalizacja wektorów wag pozwala zrównoważyć wartość wyników i umożliwia ich porównywanie w dalszej analizie.

Powyższe założenia odzwierciedla funkcja TFIDF:

$$TFIDF(t_k, d_j) = TF(c_k, p_j) * IDF,$$

gdzie:

- $TF(c_k, p_j)$ (macierz *term frequency*) przedstawia odniesienie każdego z podanych terminów do każdego z badanych elementów:

$$TF(c_k, p_j) = \frac{f_{k,j}}{\max_z f_{z,j}},$$

$\max_z f_{z,j}$ - maksymalna w odniesieniu do wszystkich cech c_z , które pojawiły się w dokumencie p_j częstotliwość wystąpień ($f_{z,j}$)

- IDF (*inverse dokument frequency*) wyraża się formułą:

$$IDF = \log \frac{N}{n_k}$$

N - całkowita liczba dokumentów w zbiorze,

n_k - liczba dokumentów w których cecha c_k wystąpiła przynajmniej raz.

Ponadto w związku z założeniem o normalizacji wagi, które zostały uzyskane w wyniku $TFIDF(t_k, d_j)$ poddane zostaną metodzie transformacji kosinusowej:

$$w_{k,j} = \frac{TFIDF(t_k, d_j)}{\sqrt{\sum_{i=1}^{|T|} TFIDF(t_i, d_j)^2}}.$$

Dodatkowo aby pokazać związki między poszczególnymi dokumentami warto posłużyć się miarą podobieństwa kosinusów:

$$sim(p_i, p_j) = \frac{\sum_{i=k} w_{k,i} \cdot w_{k,j}}{\sqrt{\sum_{k=1} w_{k,i}^2} \cdot \sqrt{\sum_{k=1} w_{k,j}^2}}.$$

2. Wygenerowanie profilu użytkownika

W tym kroku stworzona zostaje macierz preferencji wyrażona za pomocą wektorów wag i dopasowana do treści produktu.

3. Rozpoznanie cech produktu odpowiednich dla użytkownika

Proces rekomendacji bazuje na dopasowaniu cech profilu użytkownika i wartości opisujących treść obiektu. Rezultatem jest stwierdzenie czy rozważany kandydat jest zainteresowany analizowanym przedmiotem. Zainteresowanie użytkownika danym przedmiotem może zostać dostarczone na podstawie porównania podobieństw kosinusów.

3.2.3 Przykład

Aby dokładnie przyjrzeć się metodom opartym na treści rozważmy, podobnie jak poprzednio, przykład oparty na książkach: Tym razem stworzenie rekomendacji wymaga większej liczby faktów.

Książka	Gatunek
Władza Absolutna	Kryminał
Patrioci	Proza współczesna
Proxima	Powieść fantastycznonaukowa
Strażacy	Literatura faktu
Gra o Śmierć	Romans
Cyrk	Komedia

Przez wykorzystanie algorytmu TFIDF stworzymy profil każdej z książek. Pierwszym etapem algorytmu jest stworzenie macierzy „term frequency”. W tym przypadku jej wypełnienie przedstawia odniesienie każdego z podanych terminów do każdej z książek. Załóżmy, że 1 oznacza iż książka reprezentuje cechy danego gatunku, natomiast 0 oznacza brak takich cech.

Książka / Gatunek	Kryminał	Proza współczesna	Powieść fantastycznonaukowa	Literatura faktu	Romans	Komedia	Książka akcji	Thriller
Władza Absolutna	1	0	0	0	0	0	1	0
Patrioci	0	1	0	0	0	0	1	0
Proxima	0	1	1	0	0	0	0	0
Strażacy	0	0	0	1	0	0	1	0
Gra o Śmierć	0	0	0	0	1	0	0	1
Cyrk	1	0	0	0	0	1	0	0

Zbadajmy teraz ”inverse dokument frequency”.

$$IDF = \log \frac{N}{n_k}$$

W rozważanym przypadku N to liczba książek, natomiast n_k to całkowita liczba wystąpień "term frequency", uzyskana dla wszystkich dokumentów.

Kryminał	Proza współczesna	Powieść fantastycznonaukowa	Literatura faktu	Romans	Komedia	Książka akcji	Thriller
1.098612	1.098612	1.791759	1.791759	1.791759	1.791759	0.693147	1.791759

Mając stworzona macierz "term frequency" oraz wektor "inverse dokument frequency" przejdźmy do ostatniego kroku i stwórzmy macierz TFIDF dla rozważanego przypadku:

Książka / Gatunek	Kryminał	Proza współczesna	Powieść fantastycznonaukowa	Literatura faktu	Romans	Komedia	Książka akcji	Thriller
Władza Absolutna	1.098612	0	0	0	0	0	0.693147	0
Patrioci	0	1.098612	0	0	0	0	0.693147	0
Proxima	0	1.098612	1.791759	0	0	0	0	0
Strażacy	0	0	0	1.791759	0	0	0.693147	0
Gra o Śmierć	0	0	0	0	1.791759	0	0	1.791759
Cyrk	1.098612	0	0	0	0	1.791759	0	0

Zakończyliśmy generowanie profilu przedmiotu, przejdźmy więc do wygenerowania profilu użytkownika.

Rozważmy zbiór danych, który przedstawia informacje o czytelnikach i książkach. W poniższym zestawieniu 1 oznacza, że dana osoba przeczytała książkę, natomiast puste miejsce, że nie podjęła się lektury.

Książka/Czytelnik	Anna	Maciej	Bartek	Ewa	Sandra	Kacper
Władza Absolutna	1	1		1	1	
Patrioci		1	1	1	1	
Proxima	1	1	1	1	1	1
Strażacy	1	1	1	1	1	1
Gra o Śmierć	1	1	1	1	1	
Cyrk	1	1	1	1	1	1

Profil użytkownika powinien zawierać jego preferencje dotyczące cech danego przedmiotu, w tym przypadku będą to preferencje dotyczące gatunku książki. Iloczyn skalarny zbudowany na TFIDF i macierzy preferencji użytkowników przedstawi powinowactwo każdego z użytkowników do każdego z rozważanych gatunków książki.

(macierz powinowactwa)

Kolejnym krokiem po wygenerowaniu profilu przedmiotu i profilu użytkownika jest przedstawienie w jakim stopniu każdy z użytkowników będzie zainteresowany każdą z książek. Do tego wykorzystane zostanie, wcześniej już wspomniane, podobieństwo kosinusów.

(macierz zainteresowań)

3.2.4 Wady i zalety systemu rekomendującego opartego na treści

Badacze Francesco Ricci, Lior Rokach, Bracha Shapira i Paul B. Kantor w swojej publikacji :”Recommender Systems Handbook” wyróżniają następujące zalety filtrowania opartego na treści:

- **Niezależność użytkowników:** Jest metodą ukierunkowaną na indywidualne rozważanie każdego użytkownika, a budowanie jego profilu odbywa się na podstawie ocen, które zostały przez niego wydane. W metodzie filtrowania kolaboratywnego rekomendacja była dokonana po uwzględnieniu ocen innych użytkowników oraz znalezieniu użytkowników najbardziej podobnych pod względem preferencji.
- **Transparentność:** W celu przedstawienia zaproponowanej rekomendacji opartej na treści możemy przedstawić listę kontekstów, które zostały poddane analizie. Stanowią one listę wskaźników na podstawie których możemy ocenić wartość i prawdziwość rekomendacji. W przypadku filtrowania kolaboratywnego jedyną informacją jaka doprowadziła nas do reprezentowanych wniosków jest podobieństwo między nieznanymi użytkownikami, którzy charakteryzują się podobnym gustem.
- **Nowy przedmiot:** Systemy rekomendacji oparte na treści umożliwiają stworzenie rekomendacji przedmiotu, który nie został wcześniej oceniony przez użytkowników. Istnieje bowiem możliwość szybkiego ustalenia cech w rozważanych przez rekomendację aspektach.

Niemniej jednak, rozważając ten rodzaj systemów rekomendujących możemy dostrzec następujące niedociągnięcia:

- **Ograniczona analiza treści:** Techniki rekomendacji oparte na treści posiadają ograniczenia w postaci liczby i typów cech, które są powiązane z rekomendowanymi obiektami. Przyporządkowanie pewnych kontekstów do przedmiotów może okazać się niewystarczające aby zbadać zainteresowania użytkowników.
- **Nadmierne wyspecjalizowanie:** Stosując rekomendacje oparte na treści nie posiadamy możliwości do odnalezienia wyjątkowo nieoczekiwanych wniosków. Systemy te sugerują bowiem przedmioty, których noty są wysokie w stosunku do profilu użytkownika, skąd wynikiem rekomendacji zawsze będą przedmioty podobne do tych przez użytkownika już ocenionych. Na tej podstawie też często zarzuca się rozważanym systemom niski poziom nowości przy dostarczaniu rekomendacji.
- **Nowy użytkownik:** W przypadku użytkowników, gdzie możemy analizować dużą liczbę wystawionych ocen i dobrze zrozumieć ich preferencje stworzenie odpowiedniej rekomendacji nie stanowi problemu. Jednakże, gdy ocen jest tylko kilka lub użytkownik jest nowy system nie będzie w stanie stworzyć niezawodnej rekomendacji.

3.3 Systemy rekomendujące kontekstowe (Context – aware recommender systems):

3.3.1 Problem

Spersonalizowane systemy rekomendujące opierające się na indywidualnym podejściu do użytkownika i tworzą rekomendacje oparte na jego indywidualnych preferencjach. Należy jednak zauważyć, że na wybór tej samej osoby może mieć również wpływ miejsce, czas, nastrój, czy warunki w których aktualnie się znajduje.

3.3.2 Metoda i Algorytm

Systemy rekomendujące kontekstowe są systemami rekomendującymi opartymi na treści w których zostaje uwzględniony dodatkowy wymiar zwany kontekstem.

Definicja 1 (*Kontekst*)

Kontekstem w rozumieniu eksploracji danych nazywamy obecny stan użytkownika. Autorzy książki "Recommender Systems Handbook" definiują kontekst jako wydarzenie charakteryzujące etap życia użytkownika i wpływające na jego preferencje, status. Pod

pojęciem tym kryje się nie tylko miejsce, czas, pogoda, dzień, ale także fakt, że użytkownik spędza czas samotnie lub w gronie innych osób, narodziny dziecka, zmiana pracy, małżeństwo. Wiedza na temat kontekstowych informacji pozwala zbudować wzorce i algorytmy w odniesieniu do konkretnych, istotnych danych.

Przykładem, który dobrze obrazuje podejście kontekstowe w tworzeniu rekomendacji są biura podróży, które w tworzeniu ofert uwzględniają sezon, miejsca, czas, sytuację finansową klienta oraz czas, kiedy oferta zostaje przedstawiona.

Warto zauważyć również, że uprzednio opisane metody opierały się głównie na rozważaniu problemów dwu-wymiarowych. W tym podejściu, przez dodanie nowego wymiaru, jakim jest kontekst, zaczynamy rozważać problemy trój-wymiarowe:

$$R: \text{Użytkownik} \times \text{element} \times \text{kontekst} \Rightarrow \text{Rekomendacja}$$

W modelu kontekstowym rekomendacje są generowane w dwóch krokach:

1. Metody systemów rekomendujących opartych na treści służące do wygenerowania listy rekomendacji bazującej na preferencjach użytkownika.
2. Odfiltrowanie rekomendacji, które odpowiadają przyjętemu kontekstowi.
Wyróżniamy tutaj dwa warianty. W pierwszym etap filtrowania zostaje dokonany na końcu, natomiast w drugim filtrowanie jest etapem wstępnym do tworzenia rekomendacji.

Filtrowanie jako etap wstępny (ang. Pre-Filtering)

W tym podejściu informacje kontekstowe używane są do odfiltrowania najbardziej istotnych informacji i skonstruowania dwuwymiarowego zbioru danych. Bardzo dużą zaletą tego podejścia jest możliwość implementacji w kolejnym kroku wcześniej opisanych metod rekomendacji.

Filtrowanie jako etap końcowy (ang. Post-Filtering)

Informacje o kontekście są ignorowane w wejściowych danych, a rekomendacja dokonywana jest na całym zbiorze. To w następnym kroku lista rekomendacji stworzona dla użytkownika jest zawężana przez uwzględnienie kontekstu.

3.3.3 Przykład

Wracając do przykładu rozważanego w przykładzie systemów rekomendacyjnych opartych na treści, gdzie oceny dla filmów generowano na podstawie wcześniej stworzonych

profilu przedmiotów i użytkowników dołożymy kontekst. Niech kontekstem w tym przypadku będą procenty odzwierciedlające preferencje użytkownika do poszczególnych gatunków książek, czytanych w różnych porach roku.

(tabela)(zostanie dodana po rozwikłaniu zagadki z metody wyżej)

(tabelka)

Na początku należy utworzyć profil użytkownika uwzględniający każdy z kontekstów i każdy z gatunków książek. Iloczyn skalarny macierzy kontekstu i macierzy profilu użytkownika przedstawia preferencje użytkownika w stosunku do każdego z kontekstów.

(tabela po iloczynie)

Następnym krokiem jest przedstawienie rankingu książek uwzględniającego kontekst dla wybranego użytkownika. Do stworzenia takiego rankingu zostanie użyte, dobrze już znane w zakresie reguł rekomendacyjnych prawdopodobieństwo kosinusów.

Podobieństwo kosinusów(matrix, item profile).

Po uzyskaniu rankingu można przejść do zasugerowania odpowiedniej książki dla wybranego użytkownika. Oczywiście, w tym przypadku, uwzględniając rozważany kontekst.

3.3.4 Wady i zalety systemów rekomendujących kontekstowych

Metody kontekstowe są bardziej zaawansowane niż wcześniej omawiane systemy rekomendujące. Dzięki temu rekomendacje oparte na metodach kontekstowych:

- zawsze pozostają w zgodzie z użytkownikiem i generują informacje uwzględniając jego aktualne potrzeby,
- są najczęściej stosowane przy generowaniu rekomendacji w czasie rzeczywistym.

Podobnie jak w poprzednio rozważanych systemach rekomendujących również i w tym brak jednak serendipity i proponowania nowych, zaskakujących użytkownika rekomendacji.

3.4 Hybrydowe systemy rekometryjace ():

(Czy potrzebny mi jeszcze ten podrozdzial?)

3.5 Systemy rekomendujace oparte na modelach():

(Czy potrzebny mi jeszcze ten podrozdzial?)

Rozdział 4

Eksperymenty / część praktyczne

Rozdział 5

Podsumowanie

Korzystając z [? , Rozdział 3, akapit 4]