

Interdyscyplinarne Studia Doktoranckie - Technologie
Informacyjne: Badania i ich Interdyscyplinarne
Zastosowania

ROZPRAWA DOKTORSKA

mgr Paweł Ładyżyński

Modelowanie niepewnych preferencji w systemach rekomendacyjnych

Promotor
prof. dr hab. Przemysław Grzegorzewski

Streszczenie

Głównym celem systemów rekomendacyjnych jest proponowanie produktów w jak najlepszym stopniu odpowiadających zainteresowaniom lub aktualnym potrzebom użytkowników. Aby cel ten mógł być osiągnięty, jednym z kluczowych zagadnień związanych z tworzeniem rekomendacji jest określenie podobieństwa pomiędzy preferencjami poszczególnych użytkowników. Zadanie to staje się utrudnione w wyniku szeroko pojętej niepewności dostępnych danych, co jest sytuacją bardzo często spotykaną w praktycznych zastosowaniach systemów rekomendacyjnych.

W pracy zbadano, czy reprezentacja niepewnych systemów preferencji w formie intuicjonistycznych zbiorów rozmytych Atanassova pozwala modelować niepewne systemy preferencji w wielu typach systemów rekomendacyjnych i usprawniać różne aspekty ich działania. Poddano analizie dostępne w literaturze miary podobieństwa zdefiniowane dla intuicjonistycznych zbiorów rozmytych Atanassova pod kątem ich zastosowania w systemach rekomendacyjnych oraz zaproponowano nową miarę podobieństwa posiadającą pożądane własności. Zaproponowano nowatorską graficzną metodę oceny rekomendacji pozwalającą na analizę typów rekomendacji generowanych przez różne systemy rekomendacyjne. Przedstawiono modyfikację dostępnych w literaturze algorytmów rangowania etykiet wykorzystujących modele probabilistyczne, która zaowocowała znacznym skróceniem czasu obliczeń. Zaproponowano system rekomendacyjny łączący metody automatycznego przetwarzania tekstu z analizowanym modelem reprezentacji systemów preferencji, który rekomenduje produkty bazując na zagregowanej wie-

dzy płynącej z opinii o produktach w formacie tekstowym, uwzględniając jednocześnie indywidualne preferencje danego użytkownika.

W rozprawie badania przeprowadzono zarówno pod kątem teoretycznym jak i metodami eksperymentalnymi, wykorzystującymi rzeczywiste dane pochodzące z popularnych serwisów internetowych (np. Amazon, Netflix).

Abstract

The main goal of a recommender system is to generate meaningful recommendations for items or products that might be interesting for a user. Measuring similarity between preferences is a crucial problem for recommender systems. This task becomes significantly harder when preferences are incomplete or somehow vague which is very common in practical cases when it comes to generating recommendations.

In this thesis it is shown how to model vague preferences using Atanassov's intuitionistic fuzzy sets (IF-sets) and this representation is used to improve various aspects of the most common types of recommender systems. Several comparison measures between IF-sets are considered to find those possessing properties desirable in recommender systems. A new measure that might be useful in finding other customers somehow similar to a new user of a recommender system and promoting those customers, who have a piece of extensive knowledge on many products not yet familiar to this new user, is proposed. A novel graphical method for comparing different types of recommendations is suggested.

The scope of the thesis considers also learning to rank problems. The new approach of using IF-sets representation of preference systems applied to some algorithms, based on probabilistic models, is proposed. It is shown that the time of calculations on benchmark datasets is reduced significantly. The recommender system combining automatic text processing algorithms with analyzed preference systems representation models is designed within the scope of this thesis. It allows generating recommendations based on automatic summaries of products

customer reviews, which are available in a text format and including individual user preferences in the recommendation process.

The research in this thesis considers both the theoretical background of proposed methods, as well as experiments performed on real datasets coming from popular internet services (i.e. Amazon, Netflix).

Spis treści

Wprowadzenie	9
1 Typy systemów rekomendacyjnych	15
2 Modelowanie niepewnych preferencji	19
2.1 Niepewne systemy preferencji	19
2.1.1 Model reprezentacji niepewnych systemów preferencji	21
2.1.2 Porównywanie niepewnych systemów preferencji	25
2.1.3 Miary podobieństwa i korelacji	36
2.2 Tworzenie rekomendacji	39
2.3 Rekomendacje i entropia	45
2.3.1 Modyfikacja miary podobieństwa	49
2.3.2 Strategia budowania rekomendacji	50
2.3.3 Graficzna metoda wspomaganie decyzji użytkownika	52
2.4 Wyniki eksperymentalne	55
3 Probabilistyczne rangowanie etykiet i rekomendacje oparte na metadanych	63
3.1 Rangowanie etykiet z wykorzystaniem modelu Mallowsa	64
3.1.1 Podstawowe oznaczenia	64
3.1.2 Model Mallowsa	65
3.2 Niepewne preferencje w rangowaniu etykiet	67
3.3 Niekompletna wiedza w modelu Mallowsa i rekomendacje	70
3.3.1 Główna idea	70

3.3.2	Algorytmy	71
3.3.3	Wyniki eksperymentalne	73
3.4	Rangowanie a prawdopodobieństwo wygranej	75
3.4.1	Wielowymiarowe prawdopodobieństwo wygranej	75
3.4.2	Prawdopodobieństwo wygranej w niepewnych systemach preferencji	78
4	System wspomaganie decyzji wykorzystujący automatyczne przetwarzanie opinii w formacie informacji tekstowych	89
4.1	Wprowadzenie	89
4.2	Zbiór danych i wybrane algorytmy	90
4.3	Wydobywanie cech produktów z opinii	92
4.3.1	Utajona alokacja Dirichleta	92
4.3.2	Reguły asocjacyjne	93
4.3.3	Wyniki	95
4.4	Automatyczne podsumowania opinii	96
4.4.1	Klasyfikacja opinii	96
4.4.2	Tworzenie automatycznych podsumowań opinii	97
4.5	Tworzenie rekomendacji	100
	Podsumowanie	107
	Bibliografia	115
	Źródła online	123
	Indeks	125