

ИЗПОЛЗВАНЕ НА СТРУКТУРИРАНА ИНФОРМАЦИЯ ОТ ПОТРЕБИТЕЛСКИ ТАГОВЕ ЗА ПРЕПОРЪЧВАНЕ НА КНИГИ

Резюме

От Мелания Бербатова, магистърска програма “Изкуствен интелект”
Факултет по математика и информатика,
Софийски университет “Св. Климент Охридски”

Февруари 2020

Препоръчващите системи са софтуерни системи, базирани най-често на алгоритми за машинно самообучение, които се използват за препоръчване на продукти и онлайн съдържание на онлайн потребители, като препоръките се базират на научаването на потребителския профил и преференциите на потребителите. Препоръчващите системи навлизат все по-често в нашия живот, като до голяма степен са движещата сила на големите сайтове за продажби и разпространение на онлайн съдържание, като *Amazon*, *Youtube* и *Netflix*. В настоящата дипломна работа разглеждаме препоръчващи системи за книги, като базираме работата си на свободно наличните данни *goodbooks-10k* от сайта *GoodReads.com*, от които разглеждаме най-вече потребителските тагове. Потребителските тагове отразяват най-разнообразна неструктурирана информация, като мнение и отношение на потребителя към книгата, и характеристики на книгата като поредица, автор, жанр и други. Целта на работата ни е да извлечем структурирана информация от таговете под формата на граф, която чрез алгоритми за графови ембединги да приведем във векторно представяне и да използваме за създаването на препоръчваща система, базирана на съдържание. За тази цел провеждаме експерименти с няколко подхода за научаване на структура от таговете - научаване на връзки между таговете чрез ембединги на думи, клъстеризиране и съпоставяне с експертна онтология. Експериментите показват подобрение на препоръките при използването на графова структура спрямо използването на данните в суров формат. Резултатите ни дават надежда, че разработването на и разширяването на методи, подобни на изследваните, може да доведе до по-добри препоръки и да бъде използвано в препоръчващи системи, използвани за реални цели.

USING STRUCTURED INFORMATION FROM TAGS FOR BOOK RECOMMENDATIONS

Abstract

by Melania Berbatova, Master program "Artificial Intelligence"
Sofia University "St. Kliment Ohridski",

February 2020

Recommendation systems are software systems, most often based on machine learning algorithms, that are used to recommend products and online content to online users, with recommendations based on learning user profile and user preferences.

Recommendation systems are becoming more common in our lives, and to a large extent are the driving force behind major online retail and broadcasting websites such as *Amazon*, *Youtube* and *Netflix*.

In this master thesis, we explore recommendation systems for books. We base our work on the goodbooks-10k dataset from the website Goodreads.com, from which we mainly observe the customer tags data. Customer tags consist of a variety of unstructured information, such as the opinion and attitude of the user towards the book, and characteristics of the book such as series, author, genre, and more. The goal of our work is to extract structured information from tags in the form of a graph, which through graph-based embedding algorithms can be turned into vector representation and used to create a content-based recommendation system. To achieve this, we conduct experiments with several approaches for learning tag structure - learning tag relationships by using word embeddings, clustering, and mapping to an expert ontology. The experiments show an improvement in the recommendations when using a graph structure, compared to using the data in a raw format. The results obtained give us hope that developing and extending methods similar to the ones examined in this master thesis can lead to better recommendations and can be used in recommender systems for real-life tasks.