

ABSTRACT

Recommendations are an essential part of the personalized user experience for any digital brand. Leading companies, notably Amazon, YouTube and Netflix, have proven their value and radically changed customer expectations of a digital experience. That is why a lot of work has been and continues to be invested in the development of recommender systems to optimally recommend the products offered to the customers.

The subject of this thesis is the implementation of a recommender system for products of the REWE Online Shop using a state-of-the-art algorithm. In order to do so, in a first step the two-dimensional input data containing information about all customers and products is reduced in size. For this purpose, the input matrix gets splitted into two narrower matrices. Each of these matrices contains information about all customers or about all products of the system respectively. The second dimension is formed by the so-called latent factors. Taken together, both matrices represent an approximation of the input data. After multiple optimization of both matrices, they are able to make predictions about the preferences of individual customers or to find further complementary items for a specified product. This optimization process is performed with Alternating Least Squares (ALS). After proving the mathematical correctness, a parallel implementation is presented and the results are evaluated. In particular, the results obtained are compared with a second approach, the association rules. The latter is currently generating the recommendations in the REWE Online Shop.

The results of the evaluation clearly show that the quality of personalized recommendations goes hand in hand with the number of latent factors used and the use of normalized data. To reach this conclusion, different approaches and parameter sets were tested on the available data. In spite of dealing with sparse input data, ALS algorithm generates much more appealing recommendations than the association rules. This result was corroborated by applying behavioral analytics on 65 test subjects. In addition, the ALS algorithm generates the recommendations in adequate time by parallelization. Furthermore it offers the possibility to calculate personalized recommendations. This approach delivers remarkably more accurate results than the naive approach, which always recommends the ten most sold products.

In summary, the experiments and analyses prove the in this paper intensively discussed ALS algorithm to be a very efficient approach for implementing suitable recommender systems within the context of grocery online shops.

KURZFASSUNG

Empfehlungen sind ein wichtiger Bestandteil der personalisierten Benutzererfahrung für jede digitale Marke. Führende Unternehmen, insbesondere Amazon, YouTube und Netflix, haben ihren Wert unter Beweis gestellt und die Erwartungen der Kunden an ein digitales Erlebnis radikal verändert. Aus diesem Grund wurde und wird weiterhin viel Arbeit in die Entwicklung von Empfehlungssystemen investiert, um den Kunden die angebotenen Produkte bestmöglich zu empfehlen.

Gegenstand dieser Arbeit ist die Realisierung eines Empfehlungssystem für Produkte des REWE Online Shops durch einen State-of-the-Art Algorithmus. Dazu werden die zweidimensionale Eingabedaten, die Informationen über alle Kunden und Produkte enthalten, zunächst in ihrer Dimension reduziert. Hierfür werden sie in zwei schmalere Matrizen aufgespalten. Je eine dieser Matrizen enthält Informationen über alle Kunden beziehungsweise alle Produkte des Systems. Die jeweils zweite Dimension bilden die sogenannten latenten Faktoren. Zusammengenommen stellen beide Matrizen eine Approximation der Eingabedaten dar. Nach mehrfacher Optimierung beider Matrizen sind diese imstande, Vorhersagen über die Präferenzen einzelner Kunden zu treffen oder für einen Artikel komplementäre Produkte zu finden. Dieser Optimierungsprozess wird mit ALS vorgenommen. Nach dem Beweis der mathematischen Korrektheit wird eine parallelisierte Implementierung vorgestellt und die Ergebnisse werden evaluiert. Dabei werden die erzielten Resultate insbesondere mit einem zweiten Ansatz, den Assoziationsregeln, in Vergleich gestellt. Letzterer generiert derzeit die Empfehlungen im REWE Online Shop.

Die Resultate der Evaluation zeigen eindeutig, dass die Güte personalisierter Empfehlungen mit der verwendeten Anzahl an latenten Faktoren sowie der Verwendung von normalisierten Daten einhergeht. Um zu diesem Schluss zu gelangen, wurden verschiedene Ansätze und Parametereinstellungen auf den vorliegenden Daten getestet. Trotz spärlicher Eingabedaten erzeugt der ALS-Algorithmus deutlich ansprechendere Empfehlungen als die Assoziationsregeln. Dieses Ergebnis wurde durch die eine Verhaltensanalyse von 65 Probanden bestätigt. Zudem generiert der ALS-Algorithmus die Empfehlungen durch die Parallelisierung auch in adäquater Zeit. Darüber hinaus bietet er die Möglichkeit neben unpersonalisierten Empfehlungen auch personalisierte zu berechnen. Dabei liefert er auffallend genauere Ergebnisse als der naive Ansatz, bei dem stets die zehn Topseller empfohlen werden.

Zusammengefasst ergeben die Experimente und Analysen, dass sich der in dieser Arbeit intensiv thematisierte ALS-Algorithmus als sehr effizient für Empfehlungssysteme in Online Shops des Einzelhandels erweist.