

Association for Information Systems

AIS Electronic Library (AISeL)

Proceedings of the 2022 Pre-ICIS SIGDSA
Symposium

Special Interest Group on Decision Support and
Analytics (SIGDSA)

12-12-2022

Exploiting Topic Modeling and Neural Word Embeddings for Interpretable Retail Item Recommendations

Sambit Tripathi

University of Massachusetts Lowell, sambit_tripathi@student.uml.edu

Xiaobai Li

University of Massachusetts Lowell, xiaobai_li@uml.edu

Amit V. Deokar

University of Massachusetts Lowell, amit_deokar@uml.edu

Follow this and additional works at: <https://aisel.aisnet.org/sigdsa2022>

Recommended Citation

Tripathi, Sambit; Li, Xiaobai; and Deokar, Amit V., "Exploiting Topic Modeling and Neural Word Embeddings for Interpretable Retail Item Recommendations" (2022). *Proceedings of the 2022 Pre-ICIS SIGDSA Symposium*. 1.

<https://aisel.aisnet.org/sigdsa2022/1>

This material is brought to you by the Special Interest Group on Decision Support and Analytics (SIGDSA) at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the 2022 Pre-ICIS SIGDSA Symposium by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Exploiting Topic Modeling and Neural Word Embeddings for Interpretable Retail Item Recommendations

Research-in-Progress

Sambit Tripathi

University of Massachusetts Lowell
Sambit_tripathi@student.uml.edu

Xiao-bai Li

University of Massachusetts Lowell
Xiaobai_li@uml.edu

Amit Deokar

University of Massachusetts Lowell
Amit_deokar@uml.edu

Abstract

Digital platforms have used recommender systems to recommend relevant products to their users based on their historical interactions. Recently, neural network-based recommender systems that generate embedding vectors have gained popularity in both research and practice and show improved performance over traditional methods. However, it is often difficult to explain why and how the recommended items are provided to specific users by these black box systems. In this study, we propose a novel user-centric approach to recommending retail items by exploiting the latent intent of the users from transaction histories. The latent theme is learned using a Latent Dirichlet Allocation topic modeling method. The proposed method can explain the intent of the focal user and other similar users. A preliminary evaluation study shows our method outperform the baseline methods in both the accuracy and the interpretability of the recommended items.

Keywords

Explainable machine learning, Latent Dirichlet Allocation, word embeddings, recommender systems

Extended Abstract

Recommendation systems are a crucial component of modern-day digital platforms. The objective of these systems is to recommend relevant products to the customers. Recommendation systems help to convert browsing into buying, promote cross-selling, and improve user loyalty and retention. Around 35 percent of consumers' purchases on Amazon and 75 percent of streamed content on Netflix come from product recommendations based on machine learning algorithms (MacKenzie et al. 2013). One of the important applications of recommendation systems is in e-commerce platforms for retail products such as grocery items. An important characteristic of a retail item recommender system is personalization. For example, a user can show patterns in purchase types and purchase frequencies and exhibit specific preferences for product characteristics like brand or product mix.

A retail store/website visit often happens with an intent, such as preparing a recipe, stocking up necessities, or getting pet supplies. In other words, item purchases are highly user-centric and consist of a hidden theme that could be well captured by the interaction between the items that were purchased together in a basket. The objective of this study is to exploit this latent property of grocery baskets to generate recommendations to a user. Topic modeling, a machine learning technique for natural language processing (NLP), can be used to identify the latent theme and intent behind these purchase transactions. This latent characteristic could help in explaining the recommendations to the users as well.

E-commerce platforms have exploited user's purchase history to recommend relevant products to their users (Linden et al. 2003). Traditional recommendation systems use matrix factorization techniques on the item-item interaction and user-item interaction to recommend items to users. However, these methods fail to handle non-linear user interaction information, real-time changes in recommendation, and the cold-start problem (Lu et al. 2015). Currently, neural probabilistic models have gained popularity in recommender systems. Such models generate user and item embeddings and measure the closeness of items or users in embedding space. They have been successful in item recommendations in the e-commerce applications, outperforming traditional matrix factorization techniques for recommendations (Barkan and Koenigstein 2016; Grbovic et al. 2015; Trofimov 2018).

Although these complex models can yield accurate predictions, they are not easy to interpret and explain (Guidotti et al. 2018). As per General Data Protection Regulation (GDPR) guidelines, it is necessary to explain the logic of the recommendations by machine learning or automated decision-making tools to an individual (Goodman and Flaxman 2016). Also, explaining predictions provides intuition and insights for managers, senior-level management, and organizations to trust and use machine learning tools effectively. In the context of retail item recommendations, interpretability could be achieved by focusing on the latent theme and intent of item purchases. Recommending items that can best represent the latent theme or items that were purchased by other users having similar theme can allow users to understand the reasons behind those recommendations.

In this study, we propose a novel user-centric approach to recommend a set of items for the user to purchase in the next store visit. To use topic modeling, we consider each item as a word and each basket as a document. We represent each basket as a bag of items like a bag of words in NLP. We apply the Latent Dirichlet Allocation (LDA) model to the transaction data and discover the latent topics for each basket. These latent topics can characterize the theme of a specific user's basket. The themes of each basket can be represented by the topic weights. We then use word-embedding models to create vector representations of the baskets and determine the closeness of the baskets using the cosine similarity measure. For a user's most recent baskets, we find the nearest neighbor baskets that shares the same theme as that of the focal user's baskets. The top k items of the nearest neighbor baskets are then recommended to the focal user. By using the latent topics and items from the nearest neighbors, our approach recommends the items that can be explained by the user's hidden theme and the nearest neighbor's purchase behavior.

We compare our approach with the state-of-the-art baseline methods that uses item embeddings and basket embeddings for making recommendation. The results show that our approach has a higher Recall, Precision, F1 Score and NDCG (Normalized Discounted Cumulative Gain) score than the baselines. We also find that the recommendations using our approach are more explainable to the user than the baselines. Such explainable recommendations have implications for both individual users and e-commerce organizations in terms of possible measures and actions to improve shopping experience.

The main contributions of this paper are as follows:

- We propose a novel approach for retail item recommendation that exploits the latent theme and intent of the users from historical purchases using a topic-modeling technique.
- Recommendations from our approach can be easily explained to a user using the latent topics of the user's baskets and the items purchased by the nearest-neighboring user.
- We evaluate the proposed methods on a real-world dataset and show the advantage of the proposed methods over the baselines in both accuracy and interpretability.

References

- Barkan, O., and Koenigstein, N. 2016. "ITEM2VEC: Neural Item Embedding for Collaborative Filtering," in 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP) (Vol. 2016-November), IEEE, September 8, pp. 1–6.
- Goodman, B., and Flaxman, S. 2016. "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" in Proceedings of the ICML Workshop on Human Interpretability in Machine Learning (WHI'16).
- Grbovic, M., Radosavljevic, V., Djuric, N., Bhamidipati, N., Savla, J., Bhagwan, V., and Sharp, D. 2015. "E-Commerce in Your Inbox: Product Recommendations at Scale," in Proceedings of the ACM SIGKDD

- International Conference on Knowledge Discovery and Data Mining (Vol. 2015-Augus), New York, NY, USA: ACM, August 10, pp. 1809–1818.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., and Giannotti, F. 2018. “A Survey of Methods for Explaining Black Box Models,” *ACM Computing Surveys* (51:5), 93:1-93:42.
- Trofimov, I. 2018. “Inferring Complementary Products from Baskets and Browsing Sessions,” in *Proceedings of Workshop on Intelligent Recommender Systems by Knowledge Transfer and Learning (RecSysKTL’18)*, Newyork, NY, USA: ACM, pp. 1–8.