User-centric Explanation Strategies for Interactive Recommenders

Extended Abstract

Berk Buzcu
Özyeğin University
Istanbul, Türkiye
HES-SO Valais-Wallis
Sierre, Switzerland
berk.buzcu@hevs.ch

Emre Kuru
Özyeğin University
Istanbul, Türkiye
emre.kuru@ozu.edu.tr

Reyhan Aydoğan
Özyeğin University
Istanbul, Türkiye
Delft University of Technology
Delft, Netherlands
reyhan.aydogan@ozyegin.edu.tr

ABSTRACT

With the pervasive usage of recommendation systems across various domains, there is a growing need for transparent and convincing interactions to build a rapport with the system users. Incorporating explainability into recommendation systems has become a promising strategy to bolster user trust and sociability. This study centers on recommendation systems that leverage varying explainability techniques to cultivate trust by delivering comprehensible customized explanations for the given recommendations. Accordingly, we propose two explanation methods aligning with a cluster-based recommendation strategy.

KEYWORDS

Explainable Recommendation; Explanation Strategy

ACM Reference Format:
Berk Buzcu, Emre Kuru, and Reyhan Aydoğan. 2024. User-centric Explanation Strategies for Interactive Recommenders: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

We find ourselves increasingly reliant on algorithm-driven recommender systems for various decision-making processes, ranging from content recommendations on movie streaming platforms to product suggestions on e-commerce platforms. Most systems in this field focus on the selection and the presentation of the recommendation; they disregard the curiosity of the system user for "why" the current recommendation is made. Symeonidis *et. al.* point out that providing an explanation along with movie recommendations will increase the likelihood of a user estimating its movie ranking while also increasing the number of correct estimations to predict a user's favorite movie by boosting the user's confidence in providing information to the system [9]. Furthermore, many other studies from the literature have shown that the ability to express itself for the recommender system would make it more digestible from a user perspective, potentially leading to better goal-oriented results [4, 7, 10] while increasing user satisfaction [2, 5].

The primary objective is to bolster the reliability and credibility of a recommendation system by furnishing them with the capability to provide informative explanations for their recommendations. Accordingly, we propose a basic recommendation strategy with corresponding model-agnostic explanation generation strategies namely *clustered-based item explanations* and *cluster-based contrastive explanations*. We adopted model-agnostic explanations in this work because of the rising accuracy power of the black-box predictors [1, 3, 8, 11, 12].

2 ADOPTED RECOMMENDATION STRATEGY

Our approach relies on extracting informative features of items to be recommended, which affect user's satisfaction (e.g., ingredients of the recipe in a food recommendation, or the genre and actors of a movie) and evaluates the score of each items based on the learnt user preferences. It is assumed that the preferences of the users are estimated by the system (i.e., weights/importance of each feature and evaluation values). The Algorithm 1 indicates the steps to determine a recommendation. Initially, an overall score of each potential item is calculated based on the estimated user satisfaction on each feature. Here, we calculate the *final score* as the weighted sum of each individual feature evaluation (Lines 1–3). Note that the weights are sum normalized between [0, 1], so the resulting range is between zero and one where the score of "1" denotes the most desired item for the user. Our approach always selects the item with the highest *final score*, that has not been previously offered to the user (Line 4). After determining the recommended item, we eliminate it from future potential recommendations (Line 5).

3 CLUSTER-BASED EXPLANATIONS

Clustering is a crucial technique in recommendation systems as it enables grouping similar items based on shared attributes [6]. This organization enhances recommendation quality by suggesting items within the same cluster when a user engages with a particular item, improving relevance and user satisfaction. While generating model-agnostic explanations, it might be intuitive to consider important descriptive features of the current recommendation. Consider that the system recommends a item. Why does the system recommend that item but not another one? There should be some distinguishable features that the user prefers. How can we detect those features? We can cluster the items with respect

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Algorithm 1 Baseline Recommendation Strategy

Require:
F: Feature set of items;
W_u: User's feature weights;
R: Set of candidate items;
r: Recommended item;
Ensure: r
for each r of R **do**
  $R_{finalScore} = \sum_f^F Score_f * W_u(f)$
end for
$r \leftarrow \text{argmax}_{R_{finalScore}}$ (R)
$R \leftarrow R - r$
return r

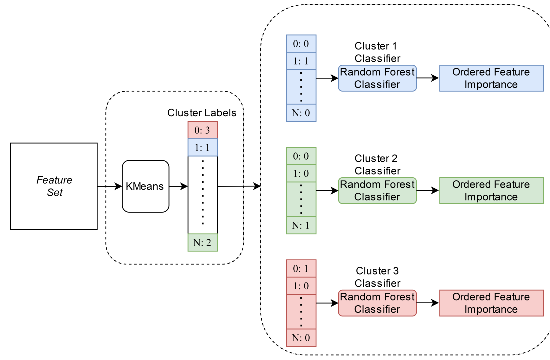

Figure 1: Process of Clustered-Based Explanation

to their features. As usual, it is expected to have similar behavior or pattern in the same cluster, and someone can inquire which features distinguish those items in the same cluster with the current recommendation. Ultimately, we can utilize those features in our explanation generation approach. We employ random forest classifiers to distinguish the cluster features in our approach.

As illustrated in Figure 1, we apply a clustering algorithm (e.g., K-means) to determine distinguishable item groups concerning users' preferences. First, we determine how an item can be represented by means of a vector of features capturing those preferences. For instance, a movie could be defined as a collection of movie titles and actors. Next step is to determine what features play a key role in the cluster distributions. In other words, what features differentiate items from other groups (i.e., clusters). To achieve this, we employ a classifier (Random Forest) for each cluster to detect the important features, which could be utilized to select the features to be used in the explanation. For each classifier, the items in terms of a vector of informative features are labeled as "1" if they belong to the underlying cluster; otherwise, labelled as "0" as a binary classification. Random forest algorithm can assess the contribution of each feature for the classification task (i.e., feature importance score). The most important feature is used for the explanation. Our cluster-based explanation method first determines which cluster the given recommendation belongs to and its most important feature. The chosen feature is used to generate the explanation. If the system identifies the *genre* of a movie as the most important feature, it provides an explanation following a grammar structure [2] similar to: "This movie was recommended because you liked horror movies."

Explanations could also be generated in a "contrastive" manner, where we identify an item that exemplifies the positive aspects of a recommendation and then explain the current recommendation by contrasting it negatively with that item (See Algorithm 2). First, we choose the most similar item to the recommended item from another cluster (Line 1). Then, we compare the values of each feature of the contrastive item with those of the recommended item. When the score of the chosen recommendation is higher for a given feature, we consider it to be a positively contrastive feature, whereas vice-versa is applied for the negatively contrastive features (Lines 3–9). To generate an explanation, we promote the recommended item with the positively contrastive features, whereas negative features indicate why the system does not suggest the contrastive example. For instance, if *Movie 1* was the chosen recommendation, and we selected *Movie 2* as a contrastive example, with *genre* being the most vital differing feature, the system produces: "We could recommend you *Movie 1* due to its high rating; however, we belive you prefer horror movies".

Algorithm 2 Cluster-Based Contrastive Explanation Approach

Require:
ϵ^+, ϵ^-: Positively and negatively contrastive features;
F: Feature set;
R'_c: Set of scored items in the other clusters as r;
r, r': Recommended and contrastive items, respectively;
1: $r' \leftarrow \text{argmin}_{c \in R'_c}\ distance(c, r)$
2: **for each** feature f in F **do**
3:   **if** score(r[f]) > score(r'[f]) **then**
4:     $\epsilon^+ \leftarrow \epsilon^+ \cup f$
5:   **else**
6:     $\epsilon^- \leftarrow \epsilon^+ \cup f$
7:   **end if**
8: **end for**
9: return ϵ^+, ϵ^-, r, r'

4 CONCLUSION

In conclusion, this study seeks to partake in the ongoing discussion on integrating explanation generation strategies in recommendation. As we search to enhance the transparency and effectiveness of recommendation systems, it is clear that user-centric simplicity and clarity are essential components of successful explanations. Despite these findings, it is important to acknowledge that the effectiveness of explanation strategies may vary depending on the specific user, rather than the collective user opinion on recommendation items guiding explanations. The preliminary experimental results show that most participants appreciate the proposed explanation types and prefer simple, to-the-point explanations to contrastive ones.

ACKNOWLEDGMENTS

This work has been supported by the CHIST-ERA grant CHIST-ERA-19-XAI-005, by the Scientific and Research Council of Turkey (TÜBİTAK, G.A. 120N680), and the Swiss National Science Foundation (G.A. 20CH21_195530).

REFERENCES

[1] Marco Ancona, Enea Ceolini, A. Cengiz Öztireli, and Markus H. Gross. 2017. A unified view of gradient-based attribution methods for Deep Neural Networks. In *NIPS Workshop on Interpreting, Explaining and Visualizing Deep Learning - Now What? (NIPS 2017)*.

[2] Berk Buzcu, Melissa Tessa, Igor Tchappi, Amro Najjar, Joris Hulstijn, Davide Calvaresi, and Reyhan Aydoğan. 2024. Towards interactive explanation-based nutrition virtual coaching systems. *Autonomous Agents and Multi-Agent Systems* 38, 1 (20 Jan 2024), 5. https://doi.org/10.1007/s10458-023-09634-5

[3] Deniz Cemiloglu, Maris Catania, and Raian Ali. 2021. Explainable Persuasion in Interactive Design. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*. 377–382.

[4] Fatih Gedikli, Mouzhi Ge, and Dietmar Jannach. 2011. Understanding Recommendations by Reading the Clouds. In *E-Commerce and Web Technologies*. Springer Berlin Heidelberg, Berlin, Heidelberg, 196–208.

[5] Joris Hulstijn, Igor Tchappi, Amro Najjar, and Reyhan Aydoğan. 2023. Metrics for Evaluating Explainable Recommender Systems. In *AAMAS, EXTRAAMAS 2023, London, England, May 29, 2023*. Springer.

[6] Chih-Lun Liao and Shie-Jue Lee. 2016. A clustering based approach to improving the efficiency of collaborative filtering recommendation. *Electronic Commerce Research and Applications* 18 (2016), 1–9. https://doi.org/10.1016/j.elerap.2016.05.001

[7] Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *International Conference on Intelligent User Interfaces, Proceedings IUI*, Vol. 2006. 93–100. https://doi.org/10.1145/1111449.1111475

[8] Mirka Saarela and Susanne Jauhiainen. 2021. Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences* 3 (02 2021).

[9] Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos. 2009. MoviExplain: A Recommender System with Explanations. In *Proceedings of the Third ACM Conference on Recommender Systems* (New York, New York, USA) *(RecSys '09)*. Association for Computing Machinery, New York, NY, USA, 317–320.

[10] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender systems handbook*. Springer, 353–382.

[11] Yifan Xu, Joe Collenette, Louise Dennis, and Clare Dixon. 2022. Dialogue-Based Explanations of Reasoning in Rule-based Systems. In *3rd Workshop on Explainable Logic-Based Knowledge Representation*.

[12] Xiubin Zhu, Dan Wang, Witold Pedrycz, and Zhiwu Li. 2023. Fuzzy Rule-Based Local Surrogate Models for Black-Box Model Explanation. *IEEE Transactions on Fuzzy Systems* 31, 6 (2023), 2056–2064.