Counterfactual Explanations of Time Varying Rankings (Student Abstract)

Ryusei Ohtani[1], **Yuko Sakurai**[1], **Satoshi Oyama**[2]

[1] Nagoya Institute of Technology,
[2] Nagoya City University
{r.otani.638@stn., sakurai@}nitech.ac.jp, oyama@ds.nagoya-cu.ac.jp

Abstract

Counterfactual explanations in Explainable AI (XAI) identify which features to change to alter an outcome, but existing methods adjust only the features of a single agent. We present a new approach to re-evaluating rankings that is based on predictions of future features of the other agents in a ranking system. It uses an algorithm that provides a more realistic counterfactual explanation of changing the ranking of a particular agent. Computer experiments demonstrated that the proposed algorithm can capture the time variation of the entire ranking system in the inference results.

Introduction

Explainable AI (XAI) has garnered attention as a crucial means of understanding how AI models operate and why they produce specific outputs (Barredo Arrieta et al. 2020). One XAI method, counterfactual explanations, identifies which features and to what extent they should be modified to change a specific outcome (Verma, Dickerson, and Hines 2020). Previous research on counterfactual explanations mainly focused on classification problems. One of the few exceptions tried to explain the minimum feature change required to change the ranking of a specific agent representing an entity in the ranking to the desired position (Saliminparsa 2023). However, this approach has a limitation : the ranking changes of the other agents are not considered, so the overall shift in ranking cannot be accurately reflected. For example, if one retailer improves its marketing plan in accordance with to the counterfactual explanation to achieve a target sales ranking, it may not achieve the target ranking if other retailers take similar measures. This limitation may be one factor that reduces the reliability and utility of the explanation. Furthermore, it is impractical for conventional counterfactual explanatory methods to compute sequential changes in rankings for all agents during feature searches. This is because performing a feature search for all agents in a large real-world data set may lead to a combinatorial explosion.

We present a new approach to re-evaluating the entire ranking process that is based on predictions of all agents' future features in the ranking system. This approach is expected to make counterfactual explanations for changing the ranking of a particular agent more realistic and practical. Specifically, the auto regressive integrated moving average (ARIMA) model is used to predict future values of features on the basis of past data, and counterfactual explanations for changing an agent's ranking are applied on the basis of the re-evaluated ranking. Our proposed algorithm, called Counterfactual Explanations of Time Varying Rankings (CFETVR), can prevent combinatorial explosion and improves the effectiveness of the explanations. A more realistic and practical counterfactual explanation is provided by identifying the optimal way to modify features to achieve a change in ranking for a particular agent and by considering dynamic changes in the overall ranking system.

Proposed Algorithm: CFETVR

The CFETVR algorithm consists of three functions: initial ranking generation, prediction of feature variation using the ARIMA model, and feature search, as shown in Figure 1. The algorithm is aimed at determining the minimum change required for a data instance to achieve a different rank. It is based on the greedy algorithm, which searches for optimal features. The input data set contains a unique ID for each agent, multiple features, specific result labels, and time-series data. The output is a set of features with minimum modifications and a final ranking result.

The procedure for generating counterfactual explanations is as follows: (1) generate the initial ranking from the input data, (2) use the ARIMA model and, the time-series data to estimate the future features of all agents, (3) adjust the optimal features of the agent whose rank is to be changed in accordance with the greedy algorithm until the desired rank is achieved. The ranking is updated after each feature adjustment. Counterfactual explanations require that the changes in features required to achieve the desired outcome are minimized. When the desired rank is obtained on the basis of the greedy algorithm, the altered features and their ranking results become counterfactual explanations.

Algorithm 1 shows the pseudocode of our algorithm. Mlmodel is a machine learning model for generating rankings. In our experiments described in Section , we used XGBoost as MlModel. The input data represents a dataset.

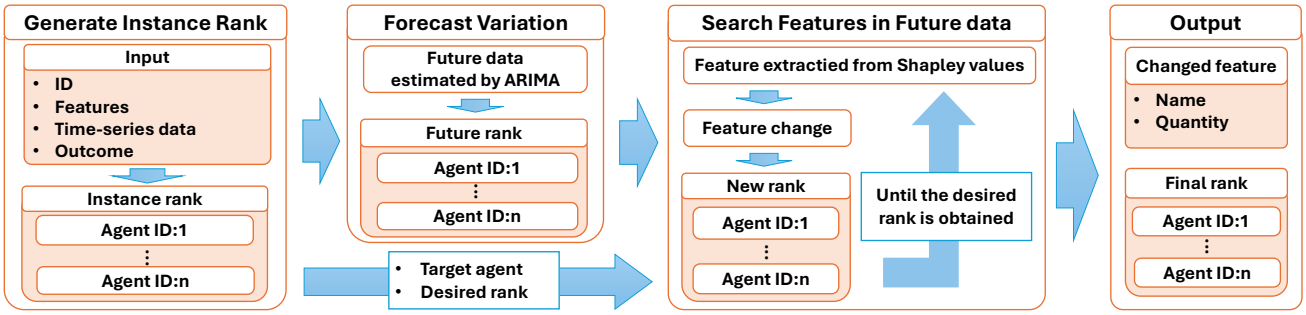

Figure 1: Flow of CFETVR Algorithm

Algorithm 1: Counterfactual Explanations of Time Varying Rankings (CFETVR)

Require: `Mlmodel, data`

`Rank` $\leftarrow$ `Mlmodel(data)`
`AgentId` $\leftarrow$ `SearchId(Rank,Option)`
`RChange` $\leftarrow$ `Option`
$\mathcal{F}^*$ $\leftarrow$ `ImportantFeatures (Mlmodel, data)`
$\mathcal{F} \leftarrow \emptyset$, `Fsub` $\leftarrow \emptyset$
`FutureData` $\leftarrow$ `ForecastedValues (data)`
`FutureRank` $\leftarrow$ `Mlmodel(FutureData)`
if `FutureRank` $-$ `Rank` $\geq$ `RChange` **then**
  return `"No need to change features"`
for `f`* in $\mathcal{F}^*$ **do**
  `InteractionList` $\leftarrow$ `GetInteraction (f*,Mlmodel,FutureData)`
  $\mathcal{F} \leftarrow$ `append(`$\mathcal{F}$`, f*, InteractionList)`
  for `f` in $\mathcal{F}$ **do**
    `Fsub` $\leftarrow$ `append(Fsub, f)`
    $\Delta_f \leftarrow$ {`Min(f)`, 0, `Max(f)`}
    for δ in $\prod_{f \in \texttt{Fsub}} \Delta_f$ **do**
      `newInput` $\leftarrow$ `ReplaceValues (AgentId,FutureData,`δ`)`
      `newRank` $\leftarrow$ `Mlmodel(newInput)`
      if `newRank - Rank` $\geq$ `RChange` **then**
        return `newRank, newInput`
return `"No feasible changes"`

Experimental Results

We evaluated the effectiveness of the CFETVR algorithm in three experiments using the Walmart sales prediction data set used in the Kaggle competition (2019). XGBoost was used as the Mlmodel in Algorithm 1 to generate the classifications. SHAP was used to acquire the features that XGBoost considered important and those that interacted well with XGBoost. In all experiments, the hyperparameters of XGBoost were set to optimal values. And, we randomly select the target agent whose rank was to be changed was randomly selected.

Table 1 shows the counterfactual explanation for a 5-rank increase in Instance Rank for the agent with store ID 23. Although The total number of agents was 45, the table shows only the results for the top 10. In this scenario, for the target agent to achieve the desired Final Rank, it was explained that this would require expanding the store size to 87,972 square feet. Notably, all agents' predicted sales values changed between Instance Rank and Final Rank. This change increased the rank of the agent in question from 10th to 5th. The agent initially ranked 3rd in Instance Rank rose to 1st in Final Rank. Although Such changes in ranking were observed in all scenarios, the ranking of each agent did not change drastically. Thus, cases in which the desired rank change was obtained without a feature search were observed only when Rchange was sufficiently small. For example, in the scenario of increasing the rank of the agent with store ID 7 from 43rd by one position, only the prediction of the ARIMA model was found to increase the rank by two places and achieved the target rank before feature search was conducted.

Instance Rank			**Final Rank**		
Rank	Store ID	Sales	Rank	Store ID	Sales
1	13	28,424	1	20	27,603
2	4	27,889	2	13	27,441
3	20	27,431	3	4	27,324
4	2	26,122	4	2	26,451
5	14	24,448	**5**	**23**	**25,048**
6	27	24,020	6	14	24,522
7	39	23,016	7	27	23,629
8	1	22,372	8	39	22,678
9	32	21,795	9	1	21,905
10	**23**	**20,595**	10	32	21,671

Table 1: Experimental result for Walmart sales prediction dataset

Conclusion

We proposed a counterfactual explanation algorithm that takes into account rank variations across all agents. In this study, we assumed that feature changes are independent. To enhance the quality of the explanations, it is important to quantify the dependencies among features and integrate them into the search process. Thus, we plan to generalize the algorithm to incorporate the dependencies among feature changes.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Numbers JP21K19833, JP24K01112, and by JST CREST Grant Number JPMJCR21D1.

References

Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; and Herrera, F. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82–115.

Salimiparsa, M. 2023. Counterfactual Explanations for Rankings. In *Proceedings of the Canadian Conference on Artificial Intelligence*.

Verma, S.; Dickerson, J. P.; and Hines, K. E. 2020. Counterfactual Explanations for Machine Learning: A Review. *ArXiv*, abs/2010.10596.