Chinese Journal of Electronics
Vol.32, No.4, July 2023

Rating Prediction Model Based on Causal Inference Debiasing Method in Recommendation

NAN Jiangang[1], WANG Yajun[1], and WANG Chengcheng[1]

(1. *School of Electronic Information Engineering, Liaoning University of Technology, Jinzhou 121001, China*)

Abstract — The rating prediction task plays an important role in the recommendation model. Most existing methods predict ratings by extracting user and items characteristics from historical review data. However, the recommended strategies in historical review data are often based on partial observational data, which having the problems of unbalanced distribution, lack of robustness, and inability to obtain unbiased prediction results. Therefore, a novel rating prediction model based on causal inference debiasing (CID) method is proposed. The model can mitigate the negative effects of context bias and improve the robustness by studying the causal relationship between review information and user ratings. The proposed CID rating prediction model is plug-and-play and is not limited to one baseline prediction method. The proposed method is tested on four open datasets. The results show that the proposed method is feasible. Compared with the most advanced models, the prediction accuracy of the CID rating prediction model has been further improved. The experimental results show the debiasing effectiveness of the CID rating prediction model.

Key words — Recommendation, Rating prediction, Context interaction, Counterfactual analysis, Causal inference.

I. Introduction

Recommendation systems provide users with personalized information services. It plays an indispensable role in e-commerce, video media, and other interactive platforms. The task of recommendation system can be divided into rating prediction, top-K recommendation and click prediction tasks. Many e-commerce platforms allow users to review and rate items they have purchased. With the continuous iteration of Internet data, a large amount of review and rating data is accumulated on these platforms. The rating prediction task mainly uses these data to make recommendations, so it has become a popular research direction.

Many researchers exploit the semantic information contained in reviews instead of using natural sparse user-item ratings to develop rating prediction models, which results in significant performance benefits. The earlier models use matrix factorization and topic modeling to model potential topics in the review text [1]. The modeling methods are gradually phased out with the development of deep learning. Deep learning models have significantly advanced technology in vector representation of review text and automatic extraction of semantic information from context, such as DeepCoNN [2], D-Attn [3], TransNet [4], ANR [5], and CARP [6].

Most of the methods mentioned above focus on modeling contextual interaction and combining these interaction cues with historical review cues for rating prediction. Although these efforts have brought significant performance improvement, there are some inherent data bias problems. These problems have a negative impact on many representational learning models. From the data point of view, the historical data in the training data set often cannot contain all the decision situations. The data proportion of positive reviews is far greater than that of negative reviews. Therefore, distribution deviations are produced by using historical data. From the perspective of methods, representational learning models are susceptible to imbalanced data distribution, which makes model decisions more inclined to events existing in historical data.

To solve the above problems, a novel rating prediction model based on causal inference debiasing method (CID) is proposed in this paper. Inspired by causal inference methods [7], [8], this paper uses a counterfactual intervention to investigate the causal relationship between review information and prediction ratings. Un-

Manuscript Received Apr. 7, 2022; Accepted Sept. 26, 2022. This work was supported by the National Natural Science Foundation of China (61503169, 61802161) and the Natural Science Foundation of Liaoning Province (2020-MS-291).

DOI:10.23919/cje.2022.00.076

like the traditional causal inference method, it is further extended to the training process of prediction model optimization. Specifically, a real causality diagram is firstly constructed by using prior knowledge, whose nodes include review information, contextual interactions, and ratings. As shown in Fig.1, context interaction may produce negative confounding due to the bias between positive and negative review distributions in the training data set. Then, counterfactual intervention is conducted to review information to sever the dependency between context and review. Since text review information is mapped to a continuous real-valued vector, inspired by [9], counterfactual reviews are used to replace the review information features, such as the average vector or zero vector of reviews. The counterfactual prediction represents the adverse effects of context interaction bias. Finally, the difference between the original prediction and the counterfactual prediction is used to represent the causal prediction. At this point, the negative effects of confounding factors are alleviated. Therefore, the proposed CID rating prediction model is plug-and-play and is not limited to one baseline prediction method, such as capsule network-based CARP [6] and CNN-based DeepCoNN [2]. The experemental results show that the proposed method is feasible. Compared with the state-of-the-art model, the prediction accuracy of proposed CID rating prediction model has been further improved.

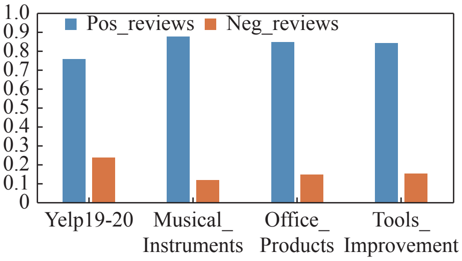


Fig. 1. The bias between positive and negative reviews in four public datasets.

In general, the main contributions of the proposed CID rating prediction model are summarized as follows:

1) In this study, the traditional rating prediction model is analyzed by using the structure causal model, and it is found that context interaction is a confounder.

2) A CID rating prediction model is proposed, and the counterfactual analysis method is used to make causal intervention on the review information, which alleviates the excessive dependence on context bias in traditional recommendation models.

3) The proposed CID rating prediction model is plug-and-play and is not limited to one baseline prediction method. It further improves the performance of many baseline models.

II. The Proposed Method

1. Review-based recommendation system

In recent years, many scholars have studied and proposed a series of methods to improve the performance of the recommendation model by using semantic information contained in the reviews [1]–[4], [10], [11].

In the early days, researchers mainly use topic modeling techniques to extract semantic information contained in reviews. These studies focus on integrating the latent semantic topics into the factor learning models [1], [3]. TLFM [12] proposes two separate factor learning models by utilizing the sentiment-consistency and text-consistency of users and items. Then they combined these two views to make ratings predictions. CDL [13] proposes to couple SADE over reviews and PMF [14]. However, due to the lack of contextual information, these methods of using the bag-of-word representation patterns may lose a lot of information.

In recent years, researchers have been interested in using deep learning techniques to model contextual information in reviews. Convolutional neural networks (CNN) [15] and recurrent neural network (RNN) [16] can extract the semantic context information in reviews and represent it as a continuous real-valued vector. ConvMF shares the same architecture with CDL and uses CNN to extract item characteristics from the item description [11]. DeepCoNN uses parallel CNN networks to uncover semantic features from user and item documents [2]. TransNet adds an additional transform layer to DeepCoNN to infer the representation of the target user-item review that is unavailable during the rating prediction [4]. D-Attn leverages global and local attention mechanisms to identify important words in the review documents for rating prediction. More recently, ANR proposes a co-attention mechanism to automatically estimate aspect-level ratings and aspect-level importance in an end-to-end fashion [5]. Capsule network is a hierarchical architecture designed to deal with complex relationships among latent features. The dynamic routing (Routing by Agreement) mechanism attached to the capsule can selectively aggregate low-level features into high-level features [17]. CARP-RA [6] used the structure based on the capsule network mentioned above to extract users' viewpoints and items' aspects at the same time and combined with users' emotions to complete the rating prediction. CARP [6] added a new routing by bi-agreement (RBiA) mechanism to CARP-RA to achieve multiple objectives jointly. RBiA enables capsule output to take into account both inter-capsule and intra-capsule agreements. However, the above approach does not consider the inherent bias between the distribution of positive and negative reviews in the dataset. To address this problem, a CID rating prediction model approach is proposed which encourages the model to focus more on the reviews themselves than on biased contextual interactions.

2. Recommendation model based on causal inference

Although the method of supervised learning has achieved good results, it requires a large amount of training data to cover a variety of recommendations. In practice, online data retention records often follow only one or more recommended policies. It doesn't cover all the recommendations. Therefore, if trained on such samples, the recommended strategies obtained from supervised learning will have a certain bias. Algorithms are more easily influenced a priori by historical policies. In the field of recommendation systems, traditional machine learning can only find correlations between data based on correlations. But after learning the correlation, it cannot give an accurate recommendation result. The purpose of causal inference [7], [18] is to enhance the model's competence to pursue causal effects: it can get rid of the spurious bias, disentangle the desired model effects [19], and modularize reusable features that generalize well [20]. Although Rubin's framework of potential outcome [21] is essentially equivalent [22] to Pearl's structural causal model [23], we use Pearl's structural causal model. Because Pearl's causality can be clearly modeled in our rating prediction model—each node in the structural causal model can be located in the rating prediction model. However, when we cannot model causalities explicitly, we can try Rubin's theory, such as using the propensity ratings [24]. In recent years, in the field of trajectory prediction [25] and computer vision [26], many researchers have tried to combine causal inference method with deep neural network to improve model performance. In the recommendation area, there is some work to model recommendation models through causal inference. The clear causalities can improve model transparency. DICE [27] assign users and items with separate embeddings for interest and conformity, and make each embedding capture only one cause by training with cause-specific data which is obtained according to the colliding effect of causal inference. PDA [28] removes the confounding popularity bias in model training and adjusts the recommendation score with desired popularity bias via causal intervention.

III. CID Rating Prediction Model

In this section, the proposed CID rating prediction model is introduced. Fig.2 shows the overall architecture of the CID rating prediction model.

1. Problem definition

The input of the rating prediction task in the recommendation system includes review information and contextual interaction. Note that although many models do not explicitly specify context interactions as inputs, their complex design is all about learning cues from contextual interactions.

As shown in Fig.2, the rating prediction framework that models both user reviews and item reviews can be divided into three main parts. They are review information coding module, context interaction module and rating prediction module respectively. Given N users, M items, we can be the u-th user reviews documents defined as $D_u = (w_1, w_2, \dots, w_k)$. Define the reviews document for m-th item as $D_m = (w_1, w_2, \dots, w_l)$, where the k and l are the number of words in the document. These documents extract user features and item features through the review information coding module. The context interaction module models the interaction between user features and item features. Get a collection $X = \{X_1, X_2 \dots, X_{NM}\}$ of user-item interaction reviews information that contains context information. The interactive review information for i-th user-item pair is $X_i = \{(v_u^t,\ a_m^t) \mid t = 1, 2, \dots, W\}$, where the v_u^t, a_m^t respectively denotes user u and item m's t-th review features. W indicates the number of features extracted by the review information coding module. The ground-truth rating of i-th user-item pair can be defined as $Y_i = \{r_i \mid (v_u^t, a_m^t)\}$.

In the traditional rating prediction framework, given review information X_i and rating Y_i, the rating prediction process can be modeled as

$$\hat{Y}_i = F_\theta (X_i, C) \tag{1}$$

where C represents a context interaction that has always existed but has not been discovered. In causal science it is called the confounding factor. Y_i represents the prediction rating. Given a set of data $\{(X_i, Y_i)\} \in \Phi$, the predictor can be optimized using the L_2 loss function:

$$L_{\text{rat}}(\theta, \varphi \mid \Phi) = L_{L2}\left(Y_i, \hat{Y}_i\right) \tag{2}$$

where θ, φ represent the parameters of the rating prediction model.

In the next three sections, first, we describe how we use structural causality modeling (SCM) to model the causality of rating prediction models. Then we introduce the application of counterfactual analysis to our CID rating prediction model. Finally, the implementation details of the CID rating prediction model are introduced.

2. Structural causal model

In this section, we describe how we formulate the causalities among review X, context interaction C, and

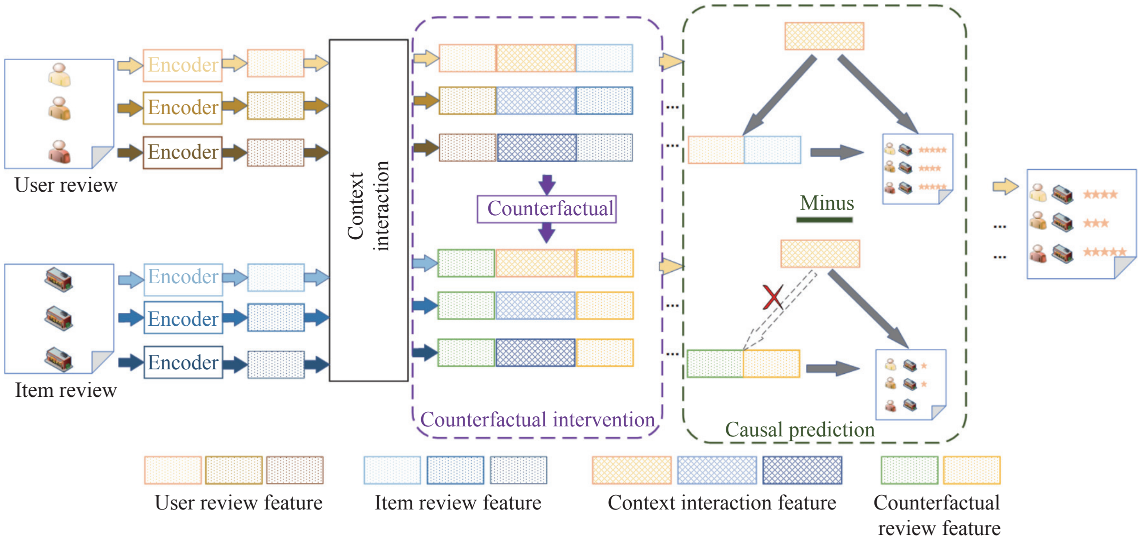
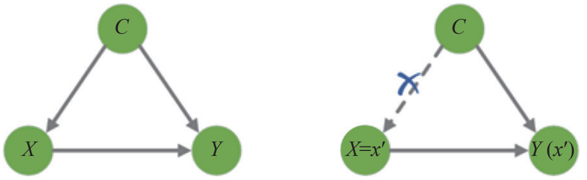


Fig. 2. The overall architecture of CID rating prediction model.

rating Y, with a structural causal model (SCM) [8]. And reconstruct the above rating prediction framework. As illustrated in Fig.3(a), we construct an SCM of node C (context interaction), X (review) and Y (rating). The links between two nodes represent the hidden causalities and how these nodes interact with each other. For example, the link $X \rightarrow Y$ indicates that review X is the reason for the rate Y. Note that the node C and the corresponding links are the ever-overlooked causalities, they are not imposed on the original rating prediction model. Context interaction C is not explicitly visible information. It exists in the dataset. When we train the model, context interaction becomes a confounder that affects the fairness of the model. The reason is that the encoders of most deep models include CNN, RNN or attention modules. These modules implicitly use the context in the encoder by updating their own parameters. This contextual interaction information is an important reason why the model works. But it also introduces some statistical biases that make the model make biased decisions. Now we detail the high-level rationale behind the SCM.

$C \rightarrow X$: Context interaction C determines what to picture in a review X. By "context interaction," we adopt the general meaning in language: the relationships among words in a language scene. Therefore, C tells us that "cheap" and "expensive" are used to describe price,



(a) Factual causal graph (b) Counterfactual intervention

Fig. 3. Structural causal model, $Y_{\text{Causal}} = Y - Y(x')$.

and "delicious" is used to describe "food". When context interaction C is absent, the model cannot truly understand the meaning of words in the reviews.

$X \rightarrow Y \leftarrow C$: $C \rightarrow Y$ denotes an obvious causality: the contextual constitution of a review affects the ratings. It is worth noting that even if we do not explicitly take C as an input for the rating prediction model, $C \rightarrow Y$ still holds. The evidence lies in the fact that review semantic contexts will emerge in higher-level layers of model when training rating prediction models, which essentially serve as a semantic feature encoder for natural language processing that highly relies on contexts, such as CNN and RNN. To think conversely, if $C \rightarrow Y$ in Fig.3(a) does not exist, the only path left from C to Y: $C \rightarrow X \rightarrow Y$, is cut off conditional on X, then no contexts are allowed to contribute to the ratings by training $P(Y \mid X)$, and thus we would never uncover the context information that causes a particular score, e.g., the semantic feature. So, it's impossible to predict ratings by modeling user or items reviews.

So far, we have pinpointed the role of context C played in the causal graph of rating prediction in Fig.3(a). Thanks to the graph, we can clearly see how C confounds X and Y via the backdoor path $X \leftarrow C \rightarrow Y$: even if some features in X have nothing to do with Y, the backdoor path can still help to correlate X and Y, resulting the model bias. Next, we propose a counterfactual analysis method to remove the confounding effect.

3. Counterfactual analysis

This section describes how we use counterfactual analysis for causal intervention to eliminate the adverse effects of contextual interaction bias on rating prediction models.

For traditional prediction models based on probability, the causal relationship between different observational cues and predicted results is not transparent. Prediction models can easily be “tricked” by shortcuts between biased contextual interactions and the final predicted rating. Causal inference [29] is a method of trying to analyze the causalities between different cues. Inspired by it, in order to mitigate the bias in context interaction and encourage the model to focus on the real cause of rating, we apply counterfactual intervention to the rating prediction model. Counterfactual intervention is the replacement of an original cue with one that doesn’t actually exist. In our CID rating prediction model, the counterfactual vector is used to replace the original comment embedding vector. In the causal inference methods [8], the intervention is defined as do(·). When we intervene in a variable, all links into the variable in CAM are cut off. The value of this variable is given independently and all variables outside this variable remain unchanged. The best way to perform operation do(·) is a “physical” intervention, i.e., a randomized controlled trial. For example, if we can collect positive ratings in all contexts, then we have $P(\text{pos} \mid \text{do}(X)) = P(\text{pos} \mid X)$. But such “physical” intervention is impossible. In the method that we proposed, we do not use expensive “physical” interventions. Instead, an actual “virtual” intervention is performed only from the observed data set (the training data itself). As shown in Fig.3(b), we replaced the review information embedding vector in the system with an imaginary review embedding vector, cutting the link from $C \rightarrow X$ in Fig.3(b). Under this intervention, we define the results as counterfactual predictions:

$$\hat{Y}_{X_i=x'} = F_\theta\left(\text{do}\left(X_i = x'\right), C\right) \tag{3}$$

where $X_i = x^{'}$ represents the counterfactual review vector. We can use different vectors for counterfactual interventions, such as a zero vector, a random vector or an average vector of all review information.

Raw real predictions $\hat{Y}_i$ depend on the interaction of reviews and context. The counterfactual prediction $\hat{Y}_{X_i=x'}$ depends only on the interaction of context. Because the original review was replaced by a counterfactual review. To study the true impact of critical information itself, we define causal prediction by calculating the difference between a true prediction and a counterfactual prediction:

$$\hat{Y}_{\text{Causal}} = \hat{Y}_i - \hat{Y}_{X_i=x'} \tag{4}$$

Compared with the original likelihood prediction, causal prediction is more reliable because it avoids the bias effect of the context confounder. In the training process, causality prediction is carried out through L_2 loss function optimization network:

$$L_{\text{Causal}}(\theta, \varphi \mid \varPhi) = L_{L2}\left(Y_i, \widehat{Y}_{\text{Causal}}\right) \tag{5}$$

4. Implementation details of CID rating prediction model

This section describes in detail how we combine causal reasoning with the baseline rating prediction model to implement our CID rating prediction model. As shown in Fig.2, the rating prediction system framework generally consists of three main modules. They are review information coding module, context interaction analysis module and rating prediction module respectively.

Aiming at the traditional likelihood rating prediction model, counterfactual intervention was conducted on some clues. We define the prediction based on the original review and the counterfactual review as the true prediction result and the counterfactual prediction result respectively. And by calculating the difference between the two prediction results, a CID rating prediction model is established. Our CID method is simple yet effective for the dependability of the rating prediction system. To evaluate the generality of our CID rating prediction model, we applied this approach to six different baseline models including DeepCoNN [2] and ConvMF [11] based on CNN, D-Attn [3] and ANR [5] based on the attention mechanisms, CARP [6] and CARP-RA [6] based on capsule network. Let’s briefly introduce three of these implementations below.

CID-DeepCoNN: DeepCoNN models user and project review documents using two parallel convolutional structures. And introduce a shared layer at the top to couple the two parts together. The shared layer interacts with the underlying factors learned from users and projects in a manner similar to a factorizer. For review information, we use counterfactual intervention to replace the embedding vectors (mean vector, zero vector or random vector of all review embedding) of user and project reviews respectively. Then the original and counterfactual connectivity features of the review information section are predicted respectively. Finally, we took the difference between the factual and counterfactual prediction results as our CID-DeepCoNN prediction.

CID-CARP: The coder of CARP maps the text review information to a continuous real value vector by using word embedding technology. Then, convolutional operation and self-attention mechanism are used to extract user views and project aspects from user documents and project documents respectively. Logical units are responsible for connecting user perspectives to project aspects to extract interaction characteristics.

Emotion analysis and rating prediction were then performed as input to the capsules. At this baseline, we replace the original review embedding vector with a counterfactual intervention (the average, zero, or random vector of all review embedding). Then the original review embedding vector and counterfactual review embedding vector are predicted respectively. Finally, the prediction result of CID-CARP adopts the difference of the above two results.

CID-ANR: ANR performs aspect-based representation learning for users and items by designing an attention mechanism to focus on the relevant parts of these reviews while learning the representation of aspects on the task. Furthermore, they estimate aspect-level user and item importance in a joint manner using the idea of co-attention. For ANR, we do the same for the review embedding as for CID-DeepCoNN and CID-CARP. We predicted the ratings of factual and counterfactual reviews separately. The difference between the two results serves as the final prediction for CID-ANR.

IV. Experiments

In this section, we conducted a comprehensive experiment on four data sets from two different sources to evaluate the performance of the CID rating prediction model against the baseline model and other state-of-the-art models.

1. Experimental settings

Datasets: Four different datasets are used in the experiment including Yelp19-20, Musical_Instruments, Office_Products, and Tools_Improvement. And among them, Yelp19-20 is the data from 2019 to 2020 extracted from Yelp challenge website. The other three datasets are from the Amazon-5cores. To be clear, we processed the Yelp19-20 dataset to have at least five reviews per user and item.

All datasets were preprocessed in the same way as reference [1]. Then, the rating data that does not contain review information are deleted. The details of the four data sets after final processing are shown in Table 1. It can be seen that the proportion of positive reviews in each dataset is much larger than the proportion of negative reviews ("Pos/Neg ratio" in Table 1). Then each dataset is randomly split in the ratio 8:2 to construct the training set and testing set. In addition, we ensure that at least one interaction data for each user-item pair was included in the training set, and the review data corresponding to the testing set was not included in the training set for the purpose of simulating real world scenarios.

Table 1. Statistics of the four datasets

Datasets	Users	Items	Ratings	Words per review	Words per user	Words per item	Pos/Neg ratio	Density
Yelp19-20	51986	9240	420384	39.26	135.33	149.71	3.17	0.007%
Musical_Instruments	1429	900	10261	32.45	141.32	200.12	7.28	0.798%
Office_Products	4905	2420	53228	48.15	197.93	229.52	5.73	0.448%
Tools_Improvement	16638	10217	134345	38.75	162.53	212.48	5.42	0.079%

Baseline: The proposed CID rating prediction model is compared with the baseline models which include CARP [6] based on capsule network, DeepCoNN [2] based on convolutional network and the review-based deep learning solution, D-Attn [3], ANR [5], and ConvMF [11].

Evaluation metric: We use the same metrics as the baseline model. The mean square error (MSE) index evaluates model performance by calculating the mean square error of the predicted rating and the real rating. It is widely used in the rating prediction task.

2. Performance evaluation

Comparison with baseline: To verify the effectiveness of our method, we plug the CID to all the review-based deep learning solutions. Such as D-Attn, ANR, ConvMF, CARP-RA, DeepCoNN and CARP. The performance comparison is summarized as Table 2. Because of the different equipment environments, the performance shown in the original baseline model paper is better than the performance we reproduce. To make a fair comparison, the data and hyperparameter settings in our methods are consistent with the baseline model settings.

As shown in Table 2, our improved model showed a performance improvement over the baseline model in all datasets. Specifically, in terms of mean square error index obtained on average on four data sets, the CID-ConvMF improved by 9.06% compared with baseline ConvMF, CID-DeepCoNN improved by 7.31% compared with baseline DeepCoNN, CID-D-Attn improved by 8.77% compared with baseline D-Attn, CID-ANR improved by 8.51% compared with baseline ANR, CID-CARP-RA improved by 8.70% compared with baseline CARP-RA, while CID-CARP improved by more than 8.63% compared with baseline CARP. We believe that this is due to the fact that the counterfactual intervention eliminates the negative effects of context bias, improves the robustness of the model, and forces the baseline model to pay more attention to the real reasons for user ratings.

Table 2. Baseline and our CID methods

Method	Performance (MSE)							
	Musical_Instruments		Office_Products		Tools_Improvement		Yelp19-20	
	Original	CID	Original	CID	Original	CID	Original	CID
ConvMF	0.962	0.949	0.937	0.880	1.514	1.422	2.836	2.432
DeepCoNN	0.887	0.881	0.913	0.874	1.461	1.394	2.488	2.179
D-Attn	0.903	0.898	0.911	0.855	1.372	1.287	2.437	2.090
ANR	0.886	0.879	0.911	0.857	1.373	1.287	2.144	1.839
CARP-RA	0.896	0.883	0.909	0.848	1.249	1.174	2.140	1.837
CARP	0.884	0.872	0.901	0.846	1.223	1.147	2.055	1.762

Comparison with several models: the proposed optimal method CID-CARP based on causality is also compared with other state-of-the-art methods. Table 3 shows the results of all the models on four datasets. First, PMF performed worst on four datasets because it did not take into review information. This is consistent with the conclusions of many studies [2], [3], [6].

Second, of all the review based baseline models, ConvMF performed worst on the four datasets. This is reasonable because it only uses convolution to extract review features in combination with probability matrix decomposition. In contrast, DeepCoNN achieved better performance due to its use of text reviews for joint modeling of user views and item attributes. Because D-Attn uses the dual word-level attention mechanisms, it achieves better prediction results than DeepCoNN in most datasets. ANR achieves relatively good performance on different datasets due to aspect level modeling of semantic information in reviews. Using capsule networks with single routing by agreement mechanisms to model user perspectives and project aspects, CARP-RA achieved results comparable to the best baselines. In contrast, CARP adopts capsule network with routing by bi-agreement mechanism. It performed best on all four datasets by simultaneously calculating coupling coefficients between and within capsules, combining user viewpoint, item aspect and emotion simulation rating behavior.

Table 3. Comparisons of CID-CARP to other models on four datasets, the best results are highlighted in bold

Method	Performance (MSE)			
	Musical_Instruments	Office_Products	Tools_Improvement	Yelp19-20
PMF	1.329	1.241	2.001	3.261
ConvMF	0.962	0.937	1.514	2.836
DeepCoNN	0.887	0.913	1.461	2.488
D-Attn	0.903	0.911	1.372	2.437
ANR	0.886	0.911	1.373	2.144
CARP-RA	0.896	0.909	1.249	2.140
CARP	0.884	0.901	1.223	2.055
CID-CARP	**0.872**	**0.846**	**1.147**	**1.762**

Finally, our optimal method CID-CARP based on causal intervention consistently achieved the best performance on four data sets. Compared with two recently proposed advanced models (ANR, CARP), the relative improvement was 12.9% and 8.6%, respectively. The validity of the CID rating prediction model is verified.

For the causal approach in, Table 2 and Table 3, it is the counterfactually intervene in the review information by replacing the original review embedding with a zero vector. Here are some other attempts.

3. Analysis of CID method

Performance comparison of different counterfactual interventions: We tried counterfactual interventions for reviews using the zero vector, the mean vector and the random vector respectively, where the random vector was sampled randomly from the uniform distribution $[-0.1, 0.1]$. These three different counterfactual interventions ("Zero", "Mean", "Random") can be easily implemented and do not require a lot of computing resources. In the actual training and testing process, the counterfactual intervention using zero vector and mean vector has always been consistent. However, for the counterfactual intervention using random vector, in order to avoid introducing bias in the model testing stage, the vectors randomly sampled from uniform distribution $[-0.1, 0.1]$ are used in the training process of the model. However, in the model test process, the expectation vector (i.e., zero vector) of uniformly distributed $[-0.1, 0.1]$ are used. As shown in Table 4, application of three different counterfactual intervention vectors to the baseline model improved model performance. It is proved that the counterfactual analysis method is effective. In addition, 1) The performance of the model with three different counterfactual vectors is close, which indicates the robustness of the counterfactual intervention method. 2) Among the three counterfactual intervention vectors, the zero vector obtained the greatest performance improvement. We think it's probably because the zero vector is a more intense intervention. Thus, the causal relationship between reviews and ratings can be highlighted.

Table 4. Performance comparison of CID-CARP with different counterfactual interventions

Method	Intervention vector	Performance (MSE)
CARP	–	2.055
CID-CARP	Zero	1.762
	Mean	1.776
	Random	1.769

Evaluation of model size: We calculate the changes in the number of parameters of CARP and DeepCoNN before and after the addition of counterfactual analysis. As shown in Table 5, our CID rating prediction model does not add additional training parameters to the baseline model because the original prediction parameters are shared with the counterfactual prediction parameters in the model.

V. Conclusions

In this paper, how to integrate causal inference into the rating prediction model is researched. Context interaction is a confounder by analyzing the generation process of score prediction through structural causal model. Rating prediction models can easily be "tricked" by shortcuts between biased contextual interactions and the final predicted rating. We have presented the CID rating prediction model to investigate the causal relationship between reviews and ratings in the rating prediction task. A counterfactual intervention is used to replace the original review embedding vector with a counterfactual embedding vector. Then subtract the counterfactual prediction rating from the original prediction rating as the final causal prediction rating. Causal prediction encourages the model to discover the real reason for the user's rating behavior during training. The negative influence of positive and negative review distribution bias in the model is reduced. The proposed CID rating prediction model is plug-and-play and is not limited to one baseline prediction method. In the experiments, we tried to implement a variety of counterfactual analysis methods and apply CID to two different rating prediction models. The experiments are conducted on four real data sets and achieved consistent performance improvements. This work explores the limitations of purely data-driven rating prediction models. We believe that this paradigm is universal for solving the problem of rating prediction task bias caused by uneven distribution of data sets. In addition, the future work will plan to explore the influence of other confounder and construct a more general causal model in the future.

Table 5. The parameter size of our model compared to the baseline model

Method	CARP	CID-CARP	DeepCoNN	CID-DeepCoNN
Parameters count	18772284	18772284	20131451	20131451

References

[1] Y. Bao, H. Fang, and J. Zhang, "TopicMF: Simultaneously exploiting ratings and reviews for recommendation," in *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, Québec City, Canada, pp.2–8, 2014.

[2] L. Zheng, V. Noroozi, and P. S. Yu, "Joint deep modeling of users and items using reviews for recommendation," in *Proceedings 10th ACM International Conference on Web Search and Data Mining*, Cambridge, United Kingdom, pp.425–433, 2017.

[3] S. Seo, J. Huang, H. Yang, *et al.*, "Interpretable convolutional neural networks with dual local and global attention for review rating prediction," in *Proceedings of the 11th ACM Conference on Recommender Systems*, Como, Italy, pp.297–305, 2017.

[4] R. Catherine and W. Cohen, "TransNets: Learning to transform for recommendation," in *Proceedings of the 11th ACM Conference on Recommender Systems*, Como, Italy, pp.288–296, 2017.

[5] J. Y. Chin, K. Q. Zhao, S. Joty, *et al.*, "ANR: Aspect-based neural recommender," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino, Italy, pp.147–156, 2018.

[6] C. L. Li, C. Quan, L. Peng, *et al.*, "A capsule network for recommendation and explaining what you like and dislike," in *Proceedings of the 42th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris, France, pp.275–284, 2019.

[7] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer*, John Wiley & Sons, Chichester, UK, pp.73–80, 2016.

[8] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*, Penguin, London, UK, pp.322–331, 2019.

[9] K. H. Tang, Y. L. Niu, J. Q. Huang, *et al.*, "Unbiased scene graph generation from biased training," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR 2020*, Seattle, WA, USA, pp.3713–3722, 2020.

[10] C. Chen, M. Zhang, Y. Q. Liu, *et al.*, "Neural attentional rating regression with review-level explanations," in *Proceedings of 2018 World Wide Web Conference*, Lyon, France, pp.1583–1592, 2018.

[11] D. Kim, C. Park, J. Oh, *et al.*, "Convolutional matrix fac-

torization for document context-aware recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems*, Boston, MA, USA, pp.233–240, 2016.

[12] K. S. Song, W. Gao, S. Feng, *et al.*, "Recommendation vs sentiment analysis: A text-driven latent factor model for rating prediction with cold-start awareness," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, Melbourne, Australia, pp.2744–2750, 2017.

[13] H. Wang, N. Y. Wang, and D. Y. Yeung, "Collaborative deep learning for recommender systems," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, Australia, pp.1235–1244, 2015.

[14] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, Vancouver, Canada, pp.1257–1264, 2007.

[15] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, pp.1746–1751, 2014.

[16] T. Mikolov, M. Karafiát, L. Burget, *et al.*, "Recurrent neural network based language model," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, Makuhari, Japan, pp.1045–1048, 2010.

[17] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, pp.3859–3869, 2017.

[18] D. B. Rubin, "Essential concepts of causal inference: A remarkable history and an intriguing future," *Biostatistics & Epidemiology*, vol.3, no.1, pp.140–155, 2019.

[19] M. Besserve, A. Mehrjou, R. Sun, *et al.*, "Counterfactuals uncover the modular structure of deep generative models," in *Proceedings of the 8th International Conference on Learning Representations*, Addis Ababa, Ethiopia, pp.1–29, 2020.

[20] G. Parascandolo, N. Kilbertus, M. Rojas-Carulla, *et al.*, "Learning independent causal mechanisms," in *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, pp.4036–4044, 2018

[21] D. B. Rubin, "Causal inference using potential outcomes: design, modeling, decisions," *Journal of the American Statistical Association*, vol.100, no.469, pp.322–331, 2005.

[22] R. C. Guo, L. Cheng, J. D. Li, *et al.*, "A survey of learning causality with data: Problems and methods," *ACM Computing Surveys*, vol.53, no.4, article no.75, 2020.

[23] J. Pearl, *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, UK, pp.105–134, 2000.

[24] P. C. Austin, "An introduction to propensity score methods for reducing the effects of confounding in observational studies," *Multivariate Behavioral Research*, vol.46, no.3, pp.399–424, 2011.

[25] G. Y. Chen, J. L. Li, J. W. Lu, *et al.*, "Human trajectory prediction via counterfactual analysis," in *Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp.9804–9813, 2021.

[26] D. Zhang, H. W. Zhang, J. H. Tang, *et al.*, "Causal intervention for weakly-supervised semantic segmentation," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, Canada, article no.56, 2020.

[27] Y. Zheng, C. Gao, X. Li, *et al.*, "Disentangling user interest and conformity for recommendation with causal embedding," in *Proceedings of Web Conference 2021*, Ljubljana, Slovenia, pp.2980–2991, 2021.

[28] Y. Zhang, F. L. Feng, X. N. He, *et al.*, "Causal intervention for leveraging popularity bias in recommendation," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, pp. 11–20. 2021.

[29] J. Pearl, *Causality*, Cambridge University Press, Cambridge, UK, pp.75–96, doi: 10.1017/CBO9780511803161, 2009.

NAN Jiangang was born in Shandong Province, China, in 1995. He received the B.E. degree from School of Artificial Intelligence, Shandong Management University, in 2020. He is now an M.E. candidate in the School of Electronic & Information Engineering, Liaoning University of Technology. His research interests include recommended system and natural language processing.
(Email: nanjiangang@gmail.com)

WANG Yajun (corresponding author) was born in Liaoning Province, China, in 1978. She received the B.S. and M.S. degrees in electronics information engineering from Shenyang Normal University in 2001 and 2004 respectively. She received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2015. From 2004 to 2019, she was a teacher with Liaoning University and Technology, Jinzhou, China. Since 2020, she has been a Professor. She is the author of more than 30 articles. Her research interests include multivariate statistical modeling, process monitoring, and data mining and big data analysis.
(Email: wyjjs2022@163.com)

WANG Chengcheng was born in Shandong Province, China, in 1998. He received the B.E. degree from School of Physics and Physical Engineering, Qufu Normal University, in 2020. He is now an M.E. candidate in the School of Electronic & Information Engineering, Liaoning University of Technology. His research interests include recommended system and natural language processing.
(Email: 3502883354@qq.com)