APPLIED ARTIFICIAL INTELLIGENCE
2023, VOL. 37, NO. 1, e2166226 (153 pages)
https://doi.org/10.1080/08839514.2023.2166226

Taylor & Francis
Taylor & Francis Group

ARTICLE

A Review Helpfulness Modeling Mechanism for Online E-commerce: Multi-Channel CNN End-to-End Approach

Xinzhe Li[a], Qinglong Li[a], and Jaekyeong Kim[a,b]

[a]Department of Big Data Analytics, Kyunghee University, Seoul, Korea; [b]School of Management, Kyunghee University, Seoul, Korea

ABSTRACT

Review helpfulness prediction aims to provide helpful reviews for customers to make purchase decisions. Although many studies have proposed prediction mechanisms, few have introduced consistency between the review text and star rating information in the review helpfulness prediction task. Moreover, previous studies that have reflected such a consistency still have limitations, including the star rating facing information loss, and the interaction between review text and star rating not extracted effectively. This study proposes the CNN-TRI model to overcome these limitations. Specifically, this study applies a multi-channel CNN model to extract semantic features in the review text and convert star ratings into a high-dimensional feature vector to avoid information loss. Next, element-wise operation and multilayer perception are applied to extract linear and nonlinear interactions to learn interaction effectively. Results measured by real world online reviews collected from Amazon.com show that CNN-TRI significantly outperforms the state-of-the-art. This study helps e-commerce websites with marketing efforts to attract more customers by providing more helpful reviews and thus, increasing sales. Moreover, this study can enhance customers' attitudes and purchase decision-making by reducing information overload and customers' search costs.

ARTICLE HISTORY
Received 5 October 2022
Revised 28 December 2022
Accepted 4 January 2023

Introduction

With the rapid information and communication technology (ICT) development, the e-commerce market has experienced remarkable growth. Since e-commerce endows customers with low cost and convenience, it plays an essential role in our daily life. Thus, online reviews have become a significant information source in the customer purchase decision process (Ahmad and Laroche 2017; Gottschalk and Mafael 2017; Sun et al. 2019). As defined, an online review is a type of customer feedback that can reflect their evaluation of the product (Hossain and Rahman 2022). Liu et al. (2007) found that 8 out of 10 customers refer to online reviews written by other consumers to explore the information in purchasing decisions. Referring to online reviews can reduce

CONTACT Jaekyeong Kim jaek@khu.ac.kr Department of Big Data Analytics, Kyunghee University, Seoul 02447, Korea

consumers' perceived risk for more effective purchase decision-making (Hossain et al. 2022). However, customers face an information overload problem when the increasing volume of online reviews. Therefore, customers cannot effectively explore helpful information when purchasing decisions (Jones, Ravid, and Rafaeli 2004; Yin, Bond, and Zhang 2014).

To address the information overload problems, many e-commerce websites, such as Amazon, Yelp, and TripAdvisor have introduced a review voting system that allows customers to evaluate their explored reviews. Such a voting system can be seen as a review helpfulness feedback mechanism for online reviews. Meanwhile, e-commerce websites recommend customized reviews based on the number of helpful votes. For example, Amazon.com and Yelp.com provide a "Most Helpful First" option in sorting reviews. Upon reading a review, the customer is then asked a question like: "Was this review useful?" and customers can vote "yes" to evaluate the review's helpfulness. Since review helpfulness can reflect the subjective evaluation and perceived utility of the review as judged by customers, a review voting system is widely used to provide helpful reviews for customers' purchase decision-making (Cao, Duan, and Gan 2011; Li et al. 2013). However, recently written reviews need sufficient time to accumulate helpful votes, making it difficult for them to refer to purchasing decisions (Ghose and Ipeirotis 2010; Ngo-Ye and Sinha 2014; Pan, Hou, and Liu 2020). Additionally, voted reviews are a lower percentage of the total review, and it is challenging to be recommended to customers even if they contain a piece of rich information (Chen et al. 2022). Thus, e-commerce websites must introduce a system that can automatically predict review helpfulness to address several such problems.

Review helpfulness prediction aims to recommend helpful reviews to consumers, which allows them to make purchase decisions through the information contained in the review. Most studies predicted review helpfulness based on review text and numerical rating (Chiriatti et al. 2019; Haque, Tozal, and Islam 2018; Huang et al. 2015). Review text information contains the qualitative evaluation of the product, providing rich and specific information for customers (Ge et al. 2019; Korfiatis, García-Bariocanal, and Sánchez-Alonso 2012). Meanwhile, star rating information represents the customer's quantitative evaluation of the product's characteristics. Since the star rating information is an objective meta-data acquired from each customer, it can be seen as statistical information about the product (Pashchenko et al. 2022). Additionally, rating valence and extremity affect customers when evaluating review helpfulness (Agnihotri and Bhattacharya 2016). Some scholars argue that consistency between review text and star rating information is essential when customers evaluate online review helpfulness (Quaschning, Pandelaere, and Vermeir 2015; Schlosser 2011; Shen et al. 2019). The text and ratings of online reviews are the customer's qualitative and quantitative evaluations of the product, respectively. Customers expect that review text and star ratings

indicate consistent information when making purchase decisions by referring to online reviews (Huang et al. 2015). Inconsistency of information could cause customers to become disabled when making purchase decisions effectively, decreasing the perceived review's credibility and helpfulness (Du et al. 2020). Therefore, this study argues the consistency of review text and star rating is a significant factor in review helpfulness prediction.

Although previous studies that predict review helpfulness through the consistency of review text and star rating information have improved the prediction performance, limitations still exist. Some studies convert review text into high-dimensional feature vectors and utilize star ratings as scalars (Ghose and Ipeirotis 2010; Hu and Chen 2016). Since scalar representations limit the capacity of star rating information, it is challenging to effectively extract the interaction between review text and star rating information. Deep learning models equipped with a deep neural network structure have been widely used in several domains in recent years. In particular, a convolutional neural network (CNN) that can extract deep latent features has illustrated excellent performance in natural language processing (NLP) tasks (Chen et al. 2019; Zhao et al. 2018). Qu, Li, and Rose (2018) applied CNN model and treated the star ratings as the last word of the review texts to enlarge the representation capacity of the star rating information. Although such an approach converts the star rating into a feature vector combined with the review texts, the star rating only interacts locally with the last few words of the review texts. Additionally, since the capacity to represent star rating information can be limited in the max-pooling layer, it is challenging to learn interactions effectively. To address the such problems, Du et al. (2020) proposed a novel approach that independently embedded review text and star ratings and utilized the element-wise operation to capture the interaction. Although such an approach can capture the interaction without information loss, it still exists limitations in the capture interaction process. Such element-wise operation only captures a linear relationship between review text and star ratings to predict review helpfulness. Nevertheless, such a simple linear relationship approach challenges capturing the complex interaction (He et al. 2017). Thus, this study summarizes the limitations of existing studies as follows: (1) Since the star rating representation capacity limitation, it is challenging to effectively capture the interaction between the review text and star ratings. (2) Since interaction between review text and star ratings is based on linear relationships, it is challenging to capture complex nonlinear relationships.

To address the limitations of star rating representation capacity and complex nonlinear interaction, this study proposes a multi-channel CNN-based review helpfulness modeling that utilizes text and rating interaction (CNN-TRI). This study applies a multi-channel CNN mechanism to extract latent representation features contained in the review text. Compared to single-channel CNN, multi-channel CNN has the advantage of applying several

size kernels, which can extract latent representation features of several lengths simultaneously (Chung and Shin 2020; Han, Kou, and Snaidauf 2019). This study independently maps the review text and star ratings into high-dimensional feature vectors to effectively utilize the presentation capacity of star rating information and minimize information loss problems. Meanwhile, this study converts review text and star rating feature vectors into the same dimension to ensure equivalent representation capacity. To extract the interactions effectively, this study utilizes a multi-layer perceptron (MLP) and element-wise operation to learn nonlinear and linear interactions. Based on this, our proposed mechanism can model review helpfulness effectively based on linear and nonlinear interactions. To evaluate the proposed mechanism performance in this study, this study utilized real-world online reviews collected from Amazon.com. The experimental results indicate that the proposed CNN-TRI mechanism outperforms better the state-of-the-art method. This study aims to introduce consistency of review text and star rating information to improve review helpfulness prediction performance. The proposed mechanism contributes both theoretically and methodologically in demonstrating how to enhance the prediction model by learning review text-star rating interaction. Our findings can help e-commerce websites to create effective marketing strategies and improve recommendations to provide customers with credible review information when making purchase decisions.

The rest of this paper is organized as follows. Section 2 describes related work in detail. Section 3 defines the problem statement of the CNN-TRI mechanism, and section 4 describes the experimental setting. Section 5 discusses the experimental results in detail, and Section 6 concludes the study, and lays out the theoretical and practical implications, and limitations.

Related Works

Review Content Feature-Based Helpfulness Modeling

Review helpfulness can be seen as a signal of a customer's endorsement of a specific review (Metzger, Flanagin, and Medders 2010), as well as reflect consumers' positive attitudes toward products (Kim, Maslowska, and Malthouse 2018). Various previous studies have illustrated that helpful reviews strongly impact consumer decision-making (Chen and Xie 2008; Topaloglu and Dass 2021). Since voted reviews are a lower percentage of the total review and e-commerce websites recommend reviews based on the number of helpful votes, most reviews are less likely to be recommended to customers. Therefore, e-commerce websites must introduce a review helpfulness prediction system for customers. Over the last decade, the topic of predicting review helpfulness has attracted increasing attention from scholars (Bilal et al. 2021). Previous studies mainly predict review helpfulness by utilizing features contained in the

review contents. Ghose and Ipeirotis (2010) indicated the subjectivity, informativeness, readability, and linguistic correctness features in the review contents affect perceived helpfulness. Additionally, they applied random forests (RF) model to predict review helpfulness. The results showed a significant improvement in review helpfulness prediction performance by utilizing reviewer characteristics, review subjectivity, and review readability. Lu et al. (2010) proposed a methodology that utilizes reviewer identity and social network contextual information in addition to textual content features to predict review helpfulness. The results showed the additional contextual features improved prediction performance. Martin and Pu (2014) proposed a methodology that extracts emotional features to predict review helpfulness. They analyzed the prediction performance through support vector machine (SVM), RF, and naive Bayes (NB) models, which are widely utilized in review helpfulness prediction. The results indicated that such emotional features improved prediction performance by 9% compared to the structure-based methodology that utilizes features such as review length or readability. Yang et al. (2015) applied linguistic inquiry and word count (LIWC) and general inquirer (INQUIRER) semantic features to overcome the problems of interpreting review helpfulness. The results showed the proposed methodology not only improved prediction performance but also can provide the semantic interpretation of review helpfulness. Mauro, Ardissono, and Petrone (2021) proposed a novel review helpfulness prediction methodology by combining star ratings, review lengths, and polarity deviations of reviews, with results indicating star ratings and review length deviations have a significant impact on review helpfulness. Such studies mainly predict review helpfulness by utilizing various features contained in the review contents. However, feature selection criteria differ depending on the products or domains. Therefore, scholars may cost a lot of time acquiring knowledge for a specific domain. Additionally, the more extracted features, the higher correlation between information, which causes multicollinearity to reduce prediction performance.

Deep learning models such as CNN generally refer to algorithms with deep layers of neural networks. Deep learning models can automatically extract latent features compared to traditional machine learning models (Liu, Qiu, and Huang 2016; Nguyen and Le Nguyen 2019). CNN model has been effectively applied in NLP tasks such as text classification, sentiment analysis, and relation classification (Dos Santos and Gatti 2014; Kim 2014; Nguyen and Grishman 2015). Kim (2014) applied CNN with multiple filters for text classification and attracted attention for its excellent performance. Since such a multi-channel CNN model equips multiple filters with different kernel sizes, it can extract various lengths of semantic features contained in the review text information at the same time (Chung and Shin 2020; Han, Kou, and Snaidauf 2019). Zhang, Roller, and Wallace (2016) proposed a CNN-based model to apply various embedding sizes. Guo et al. (2019) indicated that one

term has different significance in a different class of document. To overcome the limitations of one term just having one weight in previous studies, they applied a multi-channel CNN model that gives each term multiple weights. In recent years, multi-channel CNN has been applied in review helpfulness prediction tasks. Saumya, Singh, and Dwivedi (2020) proposed a multi-channel CNN-based model to improve review helpfulness prediction performance. To extract various lengths of semantic features, they applied filters of sizes 3, 4, and 5 to learn tri-gram, four-gram, and five-gram features of review text information. Olmedilla, Martínez-Torres, and Toral (2022) first applied 1D-CNN to the review helpfulness prediction task by utilizing multiple sizes of filters to perform review helpfulness prediction and identify clusters of review helpfulness according to the most extracted important contextual features. Such studies indicated that multi-channel CNN model improves prediction performance compared to single-channel CNN model. Specifically, the advantages of multi-channel CNN can be summarized as follows: (1) Multi-channel CNN can apply multiple sizes of filters to convolution layer to extract semantic features. (2) Multi-channel CNN has strong robustness in multiple channels. (3) Multi-channel CNN can improve learning ability and overcome limitations on feature extraction. Therefore, this study applies a multi-channel CNN model as an encoder to extract semantic features contained in the review text information effectively.

Review Text and Star Rating Interaction-Based Helpfulness Modeling

Previous studies have combined star rating information with the review texts in review helpfulness prediction. Hu and Chen (2016) indicated the significance of the interaction effect between star rating and hotel star class information. They applied linear regression, model tree (M5), and SVM to indicate that interaction affects review helpfulness. The results showed such an interaction effect improves the prediction performance indeed. Chiriatti et al. (2019) extracted various categories of features in the review that can affect review helpfulness and applied SVM to predict review helpfulness. The results showed that combining star ratings-related features with others improved the model prediction performance. Mauro, Ardissono, and Petrone (2021) applied RF and SVM to analyze the impact of star ratings, review length, and polarity deviation on review helpfulness. The results indicated that even though the polarity deviation is less important than star ratings, review length, and unigram, they clearly improved review helpfulness prediction performance. Additionally, several studies set star ratings as a moderating variable that affects the relationship between review texts and helpfulness. Malik and Hussain (2017) proposed a deep neural network (DNN) model that utilized star ratings as a visibility feature to improve the prediction performance. In previous studies, scholars also utilized the extreme star rating information.

Many studies have indicated that extreme star ratings are more helpful than moderate star ratings (Cao, Duan, and Gan 2011; Ghose and Ipeirotis 2007). Korfiatis, García-Bariocanal, and Sánchez-Alonso (2012) applied extreme star ratings to Tobit regression, and the results showed that reviews with higher star ratings had a more significant impact on review helpfulness than other star ratings. Although such studies utilized review text and star rating information to predict review helpfulness, the consistency of review texts and star ratings also affects review helpfulness (Schlosser 2011). Since review texts and star ratings are information written by the same consumer based on their own experience, considering the consistency of information is necessary (Hazarika, Chen, and Razi 2021).

Some studies applied CNN model to capture the interaction between review text and star rating information. Qu, Li, and Rose (2018) proposed combination method (CM) model to improve CNN's review helpfulness prediction performance. They treated star ratings as the last word of review texts. Therefore, the CM model combined review text and star rating information into a feature vector to capture the interaction through convolution and max-pooling layers. However, such a model has limitations: (1) Star ratings only interact with the last few words. (2) Star ratings face information loss problem in the max-pooling layer. Therefore, such limitations cause limited representation capacity of learned interaction. Fan et al. (2018) proposed multi-task neural learning (MTNL) model for review helpfulness prediction and star rating prediction tasks. The star rating prediction aims to avoid overfitting to improve review helpfulness prediction performance. However, such a study predicts review helpfulness by assuming customers are unaware of star rating information when voting for reviews, which is unconventional. Du et al. (2020) proposed text – rating interaction (TRI) model to learn the interaction. Such a model utilized the element-wise operation to capture the linear relationship-based interaction between review text and star rating information. However, the linear interaction captured by the element-wise operation cannot represent the complex interaction structure between review text and star rating information. Thus, such a study also causes limited representation capacity of interaction. Liu, Yuan, and Ma (2022) proposed a recommender system that performs multi-task learning of review helpfulness and star rating prediction. The learning process can be seen as learning implicit interaction between review text and star rating information to improve prediction performance. However, the limitation of such a model is similar to that of the MTNL model, which assumes customers vote for reviews unaware of star rating information.

The CNN-TRI model proposed in this study independently embeds review text and star rating information and converts it into a high-dimensional feature vector. Therefore, the CNN-TRI model can address the star rating information loss problems and enlarge the representation capacity. To address

the limited representation capacity of interaction, the CNN-TRI model utilizes the element-wise operation and MLP to extract the linear and nonlinear interactions between review text and star rating information.

CNN-TRI Framework

This study proposes the CNN-TRI model that predicts review helpfulness based on the interactions between review text and star rating information. As Figure 1 shows, the CNN-TRI model consists of 3 parts: Review Text Encoder (RTE), Star Rating Encoder (SRE), and Text-Rating Interaction (TRI). The RTE and SRE are applied to extract features contained in the review text and star rating information, and the TRI is applied to learn the interactions between review text and star rating information. The RTE applies a multi-channel CNN model to extract semantic features contained in review text information and convert semantic features into a feature vector. The SRE converts star rating features into a high-dimensional vector instead of utilizing raw ratings. Such a methodology can enlarge the representation capacity of star rating information compared to utilizing raw ratings. The TRI learns the interaction through the feature vectors output from RTE and SRE for predicting review helpfulness. First, the TRI extracts the linear and nonlinear interactions by utilizing the element-wise operation and MLP. Second, the TRI combines linear and nonlinear interaction feature vectors to obtain an overall interaction feature vector. Finally, the TRI predicts review helpfulness based on the overall interaction feature vector.

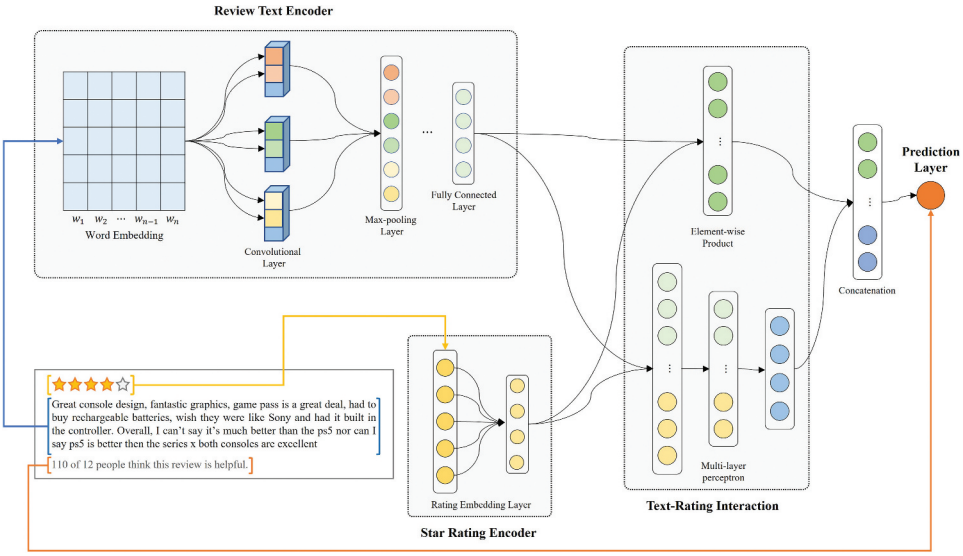

Figure 1. A proposed CNN-TRI Framework.

In this study, each review d contains 4 attributions $[s, r, c, l]$, where s denotes review text information, r denotes star rating information, c denotes helpfulness score calculated by the ratio of helpful votes to the total number of votes and l denotes review helpfulness label for review helpfulness classification (Krishnamoorthy 2015; Malik and Hussain 2017). This study approached review helpfulness prediction as a binary classification problem to classify helpful and unhelpful reviews. The CNN-TRI model predicts review helpfulness score c based on the information s and r. The l is labeled as "1"(Helpful) or "0"(Unhelpful) by comparing c and a helpfulness threshold θ as:

$$l = \begin{cases} 1, & \text{if } c \geq \theta \\ 0, & \text{if } c < \theta \end{cases} \tag{1}$$

Review Text Encoder

The RTE applies a multi-channel CNN model that has illustrated excellent performance in NLP tasks (van Dinter, Catal, and Tekinerdogan 2021). The multi-channel CNN model can extract semantic features contained in the review text information through the convolution layer. Since multi-channel CNN model applies multiple filters of different sizes on the same layer, it can extract various lengths of semantic features compared to the single-channel CNN model.

Let a review text $s = \{x_1, x_2, \ldots, x_N\}$ be a sequence of N tokenized words. This study first converts each word in the review texts into a vector through the word embedding layer. This study applies word embedding $f : x_i \in R^D$ for each word x_i, where D denotes the dimension of the embedded word vector, then each word is represented as a dense vector. Therefore, a review text s can be represented by an embedding matrix $X \in R^{N \times D}$. Next to the word embedding layer, a convolution layer applies multiple filters of different sizes to extract semantic features contained in the embedding matrix X. Let W_j be a filter with a sliding window to perform the convolution operation. Here, $W_j \in R^{K \times D}$ denotes the kernel size is $K \times D$. Following Li et al. (2021), a specific semantic feature c_j is generated as:

$$c_j = relu(X * W_j + b_J) \tag{2}$$

Here, * denotes the convolution operation and b_j denotes the convolution bias. Since the relu activation function has no gradient vanishing problem and the computational complexity of relu is much less than sigmoid and tanh, this study utilized the relu as the activation function (Liu, Yuan, and Ma 2022). The relu activation function is defined as:

$$relu(x) = max(0, x) \tag{3}$$

The max-pooling layer utilizes the maximum pool operation to take the maximum value of the feature map for obtaining the main semantic feature. Following Du et al. (2021), the maximum pool operation is defined as:

$$o_j = max([c_1, c_2, \ldots, c_{N-K+1}]) \tag{4}$$

The CNN-TRI model utilizes m multiple filters to obtain m maximum feature values. Following Liu, Yuan, and Ma (2022), semantic features contained in the review text information are defined as a fixed-size vector O by:

$$O = [o_1, o_2, \ldots, o_m] \tag{5}$$

The CNN-TRI model utilizes the element-wise operation to extract the linear interaction between review text and star rating information. Thus, this study needs to convert feature vectors of the review text and star rating information into the same dimension. Such a methodology can prevent bias against the specific interaction and ensure both interactions play the equivalent role in review helpfulness prediction. Following He et al. (2017), this study adds the hidden layers to the output feature vector to generate dimensionally reduced feature vector O' as:

$$\begin{aligned} \phi_O(O) &= relu(W_O O + b_O), \\ &\ldots\ldots \\ O' = \phi_{O_L}(O_L) &= relu(W_{O_L} O_L + b_{O_L}) \end{aligned} \tag{6}$$

Here, $\emptyset_{o_x}$, W_{Ox}, and b_{Ox} denote the mapping function, weight matrix, and bias of the x-th layer's perceptron.

Star Rating Encoder

The SRE converts star rating information into a high-dimensional feature vector to enlarge the representation capacity of the star rating information. Similar to the embedding process of RTE, this study applies star rating embedding $f : r_i \in R^K$ for each star rating r_i. The range of star rating information is 1 to K that representing consumer evaluation of a specific product. To enlarge the representation capacity of star rating information, this study maps the star rating information to an M-dimensional feature vector. Therefore, a star rating r can be represented by an embedding matrix $E \in R^{M \times K}$. The representation capacity of the star rating feature vector is M times larger than raw ratings. Additionally, since such a methodology can suppress the noise of raw ratings distributed into individual dimensions, it has strong robustness. Finally, this study converts the star rating feature vector into the same dimension as the review text feature vector by:

$$E' = relu(W_E\ E + b_E) \tag{7}$$

Here, E', W_{Ox}, and b_{Ox} denote the dimensionally reduced star rating feature vector, weight matrix, and bias. Although review text and star rating information are represented with different feature vectors, both are written by the same consumer based on their own experience. Therefore, the methodology of converting review text and star rating information into the same dimension is effective.

Text-Rating Interaction

The TRI learns the linear and nonlinear interactions between feature vectors output from the RTE and SRE to predict review helpfulness. The TRI utilizes the element-wise operation to extract linear interaction. Since the element-wise operation has a small number of parameters and a fast training speed, it is widely utilized to extract interaction between information Quan et al. (2022). The linear interaction feature vector h is defined as:

$$h = E' \otimes O' \tag{8}$$

Here, $\otimes$ denotes the element-wise operation. Meanwhile, the TRI concatenates the review text and star rating feature vectors and applies the MLP to extract nonlinear interaction. He et al. (2017) proposed a methodology that indicates the MLP can extract complex interactions between latent features contained in the information. Thus, the TRI first concatenates the feature vectors as:

$$h' = O' \oplus \mathrm{R}' \tag{9}$$

Here, $\oplus$ and h'denote the concatenation operation and concatenated feature vector. However, such a simple concatenated feature vector does not consider any interactions between review text and star rating latent features. Therefore, the concatenated vector h'cannot represent the nonlinear interaction between review text and star rating information. To extract the nonlinear interaction, the TRI applies MLP to the concatenated vector as:

$$\begin{aligned} \phi_{h'}(h') &= relu(W_{h'}h' + b_{h'}), \\ &\cdots\cdots \\ T_{h'} &= \phi_{h'_L}(h'_{h'_L}) = relu(W_{h'_L}h'_L + b_{h'_L}) \end{aligned} \tag{10}$$

Here, $T_{h'}$ denotes the extracted nonlinear interaction vector. Similar to the embedding process of review text and star rating information, the TRI set $T_{h'}$ with the same dimension as h. Such a methodology can ensure linear and nonlinear interaction vectors play the equivalent representation capacity for

review helpfulness prediction. Finally, the overall interaction feature vector h is defined as:

$$H = h \oplus T_{h'} \tag{11}$$

This study extracts the overall interaction by concatenating the linear and nonlinear interaction feature vectors. Following Li et al. (2021), the overall interaction is utilized for review helpfulness prediction as:

$$\hat{y} = sigmoid(W_H \bullet H + b_H) \tag{12}$$

Here, the sigmoid activation function is utilized to classify the review helpfulness label based on the predicted review helpfulness score. Additionally, this study utilizes a cross-entropy minimization to minimize the error between the predicted and actual review helpfulness label. Such a methodology is widely utilized for binary classification. Following Filieri and Mariani (2021), the training process of CNN-TRI is defined as:

$$L = -\sum_{i=1}^{N} (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \tag{13}$$

Here, $\widehat{y_i}$ and $\hat{y}$ denote the predicted and actual review helpfulness labels of N training samples. In the learning process, this study utilized the Adaptive Moment Estimation (Adam) methodology to update the parameters of CNN-TRI. The Adam methodology is faster than the Stochastic Gradient Descent (SGD) and can effectively obtain the optimal learning rate value (Kingma and Ba 2014).

Experiments

Dataset

This study utilizes the public Amazon 5-Core dataset to evaluate the performance of the proposed mechanism. The original dataset collected between May 1996 and July 2014 (He and McAuley 2016). Amazon is the largest e-commerce platform, which consists of various domains and covers amounts of online reviews written by customers. Since the customers of Amazon have actively traded various products, it has accumulated amount of data to overcome the information sparsity problem. Additionally, since collected online reviews include customer evaluation of the specific product, it is widely utilized in review helpfulness prediction tasks (Ghose and Ipeirotis 2010; Kim et al. 2006; Malik and Hussain 2018). This study utilizes the Amazon Books dataset, which contains 8,898,041 reviews from 603,668 customers on 367,928 products. Book datasets are a popular domain adopted in several studies, enabling a fair comparison with previous studies (Du et al. 2020;

Table 1. An example review of the amazon book dataset.

Attribute	Value
Reviewer ID	A2PK3NTC9RMEF4
Product ID	1894063171
Total number of votes	20
Number of helpful votes	18
Star rating	2
Review time	03, 11, 2009
Review title	Disappointing
Review text	There are eleven stories in this "grimoire;" stories where Holmes encounters crimes and/or events beyond the usual scope of the rational detective. Some are good, some are not …

Qu, Li, and Rose 2018). The Amazon Books dataset contains many online reviews, and it can be seen as a typical dataset of Amazon datasets. Table 1 shows an example review of the Book Datasets. The experimental dataset includes various information such as customer ID, product ID, helpfulness vote information, numerical star rating, review time, review title, and review text. This study mainly utilizes the review text, star rating, and review helpfulness information.

Figure 2 demonstrates the distribution of review helpfulness votes. Overall, the number of helpfulness votes in most reviews is between 1 and 10. Previous studies argued that there is a word of few mouths (WOFM) phenomenon in review helpfulness vote, which means few customers vote on most reviews (Roy, Datta, and Mukherjee 2019; Zhang, Tran, and Mao 2012). This study filters reviews that receive more than 10 votes of helpfulness to prevent

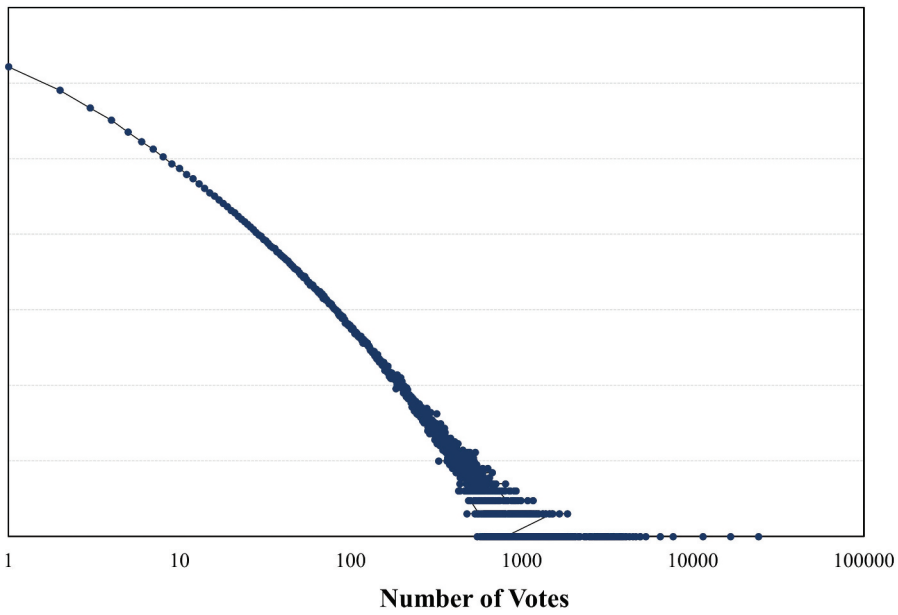

Figure 2. Review helpfulness vote distribution.

WOFM bias according to the strategies of previous studies (Tay, Zhang, and Karimi 2020). Since this study utilizes the CNN model to extract semantic features contained in the review text information, the review length as a parameter affects the model performance (Kiran, Kumar, and Bhasker 2020). Figure 3 demonstrates the distribution of review length and indicates the length of most reviews is less than 100. As shown in Table 2, the maximum length of overall reviews is 17 times longer than 90% of the rest, indicating a review length bias. To accelerate the model training process, this study utilizes 90% of the rest reviews as the maximum length to conduct the model training process effectively.

To improve model helpfulness prediction performance, this study adopts the NLTK (Natural Language Toolkit) package to perform review text preprocessing. The process of review text preprocessing is summarized as follows: (1) This study removed space and non-English reviews. (2) The dataset size is reduced by unifying upper-case and lower-case letters. (3) This study removed stopwords such as "is," "the," and "an," and special characters. (4) This study

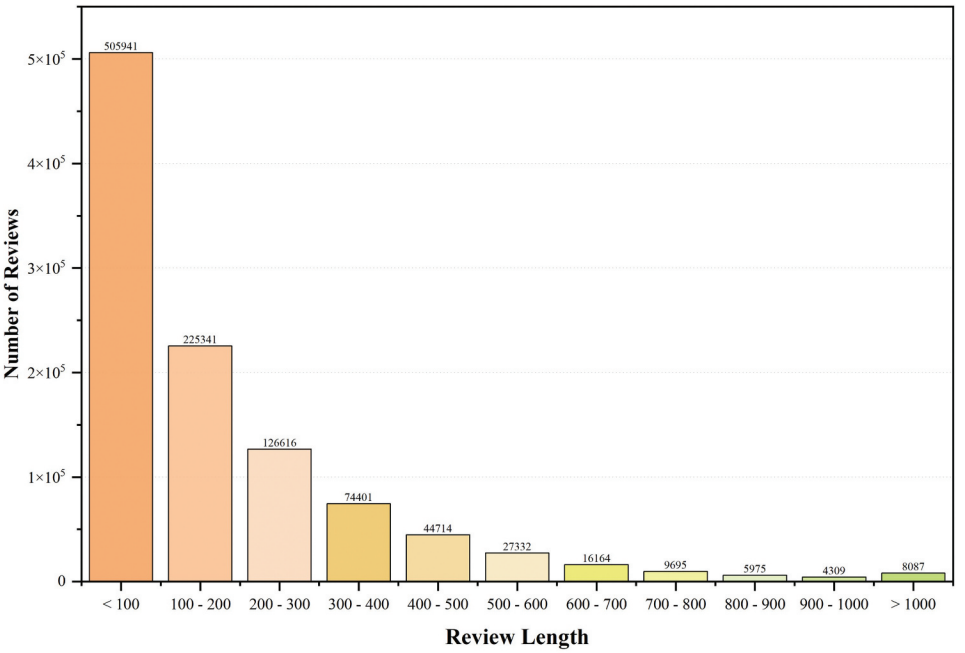

Figure 3. Review length distribution.

Table 2. Review length difference.

Review	Length
100% (Maximum)	6,154
90%	358
Multiples	17.19

adopted Stemming and Lemmatizing methodology to unify the words of the review texts.

Following the previous study strategy, preprocessed reviews are labeled as follows. This study labeled it as helpful if the ratio of helpful votes to total votes is more than 60% and labeled it as unhelpful otherwise (Mitra and Jenamani 2021; Yang, Yao, and Qazi 2020). This study also randomly sampled the same number of 180,000 reviews for each label to prevent prediction bias.

Evaluation Criteria

This study approached review helpfulness prediction as a binary classification problem (Li et al. 2021; Park et al. 2012). Therefore, this study utilized Accuracy, Precision, Recall, and F1-Score as evaluation criteria to evaluate review helpfulness prediction performance. Accuracy is the most commonly utilized evaluation criteria for prediction performance that represent the ratio of accurate classification results to overall classification results. Precision represents the ratio of the actual helpful reviews to the classified helpful reviews. Recall represents the ratio of classified helpful reviews to actual helpful reviews. F1-Score represents the balance weight average of the Precision and Recall values. The Accuracy, Precision, Recall, and F1-Score are defined in Equation 14-Equation 17 as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

Baselines Methodologies

To compare the proposed mechanism performance, this study has adopted deep learning methods (e.g., CM and CNN) and traditional machine learning methods (e.g., SVM and NB), which are utilized widely in review helpfulness prediction tasks. This study also constructed nonlinear and linear interaction models between review text and star ratings, respectively, and compared them with the proposed mechanism. The baseline methods applied in this study are as follows:

- **CNN** (LeCun et al. 1998): CNN has illustrated excellent performance in image processing and NLP tasks, and recent studies have employed CNN to extract semantic features for review helpfulness prediction. For a fair comparison, this study utilized the multi-channel CNN model consistent with the CNN-TRI to compare the performance.
- **CM** (Qu, Li, and Rose 2018): The CM model improves prediction performance, which treats the star ratings as the last word of the review texts. Such a study is the first to combine review text and star rating information for review helpfulness prediction, which learns the interaction through convolution and max-pooling layers.
- **TRI** (Du et al. 2020): The TRI model is proposed to avoid information loss problems to improve review helpfulness prediction performance. Such a study independently embeds review text and star rating information and utilizes the element-wise operation to learn the interaction.
- **SVM** (Yang, Yang, and Wang 2009): The SVM model aims to maximize the input margin to classify the data's label. Therefore, previous studies employed SVM to classify review helpfulness through the features contained in the reviews. Although such a model is efficient in geometry, it has limitations in analyzing feature distribution, outliers, and influential points.
- **NB** (Mohammad, Alwada'n, and Al-Momani 2016): The NB model classifies with probability by assuming the specific feature in a class is unrelated to the other features. Since NB has a simple structure and excellent classification performance that has utilized to classify review helpfulness. However, it still has limitations in the zero-frequency problem, which assigns zero probability to a categorical variable.

Training and Hyperparameters

In the experiment, this study set the ratio of the training and test set as 8:2. Meanwhile, this study utilized the 20% of the training set as the validation set to avoid overfitting in the training process. Next, this study would introduce the detailed setting of the training process. This study utilized the Adam methodology as the optimizer for the deep learning-based models. Since this study approached the review helpfulness prediction as a binary classification problem, this study utilized cross-entropy as the loss function. Additionally, this study utilized the early stopping method to address the problem of selecting an appropriate number of epochs. This study set the value of the early stopping method as 10 epochs, which means the training process would stop when the validation loss is no longer decreasing in 10 epochs (Du et al. 2021). Finally, this study conducted the experiments five times and reported the mean and the standard deviation of model performance. Additionally, this study determined the parameters of the CNN-TRI proposed in this study by

Table 3. Results on Affect Word Embedding Dimension on Performance.

Word Embedding Dimension	Accuracy ± SD	Precision ± SD	Recall ± SD	F1-Score ± SD
100	0.692 ± 0.002	0.679 ± 0.028	0.731 ± 0.013	0.704 ± 0.012
200	0.683 ± 0.005	0.678 ± 0.013	0.701 ± 0.034	0.688 ± 0.011
300	0.719 ± 0.800	0.709 ± 0.010	0.743 ± 0.027	0.725 ± 0.012
400	0.699 ± 0.009	0.686 ± 0.016	0.737 ± 0.027	0.710 ± 0.009
500	0.697 ± 0.016	0.676 ± 0.018	0.759 ± 0.017	0.715 ± 0.011

Note: SD (Standard Deviation)

conducting various experiments as shown in Tables 3–5. Since the CNN-TRI applies the multi-channel CNN model to extract semantic features contained in the review text information, the parameters of word embedding dimension, dropout rate, and vocabulary size affect the CNN-TRI model performance. Therefore, this study selects the word embedding dimension, dropout rate, and vocabulary size as the representative parameters. This study obtained the optimal values of the word embedding dimension from 100 to 500, the dropout rate from 0.1 to 0.9, and the vocabulary size from [90131, 90000, 80000, 70000, 60000, 50000, 40,000, 30000, 20000, 10000]. According to the experimental results, when the word embedding dimension is 300, the dropout rate is 0.4, and the vocabulary size is 40000, the CNN-TRI model shows the best performance. For the other parameters of CNN-TRI, this study set the parameters according to previous studies (Khan and Niu 2021; Kim 2014; Saumya, Singh, and Dwivedi 2020). Such studies applied the CNN model to NLP task and showed superior prediction performance. Therefore, according to previous studies, this study utilized three different kernel sizes of 3, 4, and 5 to extract multiple lengths of semantic features. Additionally, this study set the feature map size as 100 and the batch size as 256. In this study, the experimental program was written in Python 3.6, which utilized the TensorFlow of version 2.4.0. Meanwhile, the experiments were conducted in an environment with CPU Intel Core i9-11900F, 128 G of memory, and double GeForce RTX 3080 Ti.

Results

Impact of Word Embedding Dimension

To show the best prediction performance of the CNN-TRI model proposed in this study, this study conducted various experiments to determine the optimal values of the parameters. In this study, the word embedding dimension, dropout rate, and vocabulary size were selected as representative parameters of the CNN-TRI model. This study first conducts fine-tuning to the word embedding dimension. Since this study applied the multi-channel CNN model to extract semantic features contained in the review text information, the prediction performance of the CNN-TRI changes with the word embedding dimension. The word embedding dimension denotes the size of the vector in

the word embedding layer. This study can represents each word as a fixed-length vector and then group semantically similar words (Bengio, Ducharme, and Vincent 2000). Therefore, this study evaluated the CNN-TRI model performance according to the word embedding dimension from 100 to 500, as shown in Table 3.

This study compared model prediction performance mainly based on the Accuracy and F1-Score. With the increase of the word embedding dimension, the performance decreased at the beginning. However, when the value of the word embedding dimension is 300, the model performance increases to the best performance with an Accuracy of 0.719 and an F1-Score of 0.719, then decrease. Such a result indicates the parameter of the word embedding dimension affects the performance of CNN-TRI. When the word embedding dimension is 100 or 200, CNN-TRI cannot group semantically similar words effectively. Meanwhile, when the word embedding dimension is 400 or 500, it causes unnecessary noise of review text to the learning process. However, when the word embedding size is 300, the CNN-TRI model can group semantically similar words effectively and show the best prediction performance.

Impact of Dropout Rate

Next, this study conducted fine-tuning to the dropout rate. The dropout is a regularization methodology to address overfitting problem in the training process (Wu and Gu 2015). Meanwhile, the dropout rate means randomly dropping out units according to a specific probability in the fully connected layer. As shown in Table 4, this study evaluated the CNN-TRI model performance according to the dropout rate from 0.1 to 0.9.

The CNN-TRI model performance changes with the increase in the dropout rate. When the dropout rate is 0.4, the CNN-TRI model shows the best performance with an Accuracy of 0.720 and an F1-Score of 0.725. Such a result indicates when this study set the dropout rate as 0.4, the CNN-TRI model can avoid the overfitting problem effectively.

Table 4. Results on Affect Dropout Rate on Performance.

Dropout Rate	Accuracy ± SD	Precision ± SD	Recall ± SD	F1-Score ± SD
0.1	0.696 ± 0.009	0.703 ± 0.017	0.682 ± 0.049	0.691 ± 0.019
0.2	0.714 ± 0.009	0.706 ± 0.019	0.737 ± 0.037	0.720 ± 0.011
0.3	0.698 ± 0.013	0.683 ± 0.014	0.741 ± 0.017	0.710 ± 0.012
0.4	0.720 ± 0.004	0.712 ± 0.017	0.741 ± 0.034	0.725 ± 0.008
0.5	0.715 ± 0.005	0.708 ± 0.024	0.739 ± 0.053	0.721 ± 0.013
0.6	0.716 ± 0.016	0.736 ± 0.025	0.681 ± 0.077	0.704 ± 0.035
0.7	0.711 ± 0.027	0.704 ± 0.026	0.727 ± 0.039	0.715 ± 0.029
0.8	0.701 ± 0.010	0.699 ± 0.012	0.705 ± 0.736	0.702 ± 0.009
0.9	0.702 ± 0.004	0.709 ± 0.007	0.686 ± 0.016	0.697 ± 0.006

Note: SD (Standard Deviation)

Table 5. Results on Affect Vocabulary Size on Performance.

Vocabulary Size	Accuracy ± SD	Precision ± SD	Recall ± SD	F1-Score ± SD
90131 (Maximum)	0.710 ± 0.021	0.714 ± 0.020	0.701 ± 0.026	0.707 ± 0.022
90000	0.716 ± 0.018	0.712 ± 0.021	0.726 ± 0.013	0.719 ± 0.015
80000	0.712 ± 0.022	0.702 ± 0.022	0.736 ± 0.028	0.719 ± 0.022
70000	0.706 ± 0.011	0.707 ± 0.017	0.707 ± 0.026	0.707 ± 0.012
60000	0.715 ± 0.008	0.719 ± 0.013	0.708 ± 0.025	0.713 ± 0.011
50000	0.720 ± 0.007	0.717 ± 0.010	0.729 ± 0.037	0.722 ± 0.014
40000	0.726 ± 0.008	0.711 ± 0.017	0.758 ± 0.028	0.733 ± 0.007
30000	0.718 ± 0.015	0.714 ± 0.008	0.725 ± 0.038	0.719 ± 0.021
20000	0.710 ± 0.013	0.708 ± 0.016	0.715 ± 0.030	0.711 ± 0.016
10000	0.721 ± 0.015	0.728 ± 0.014	0.706 ± 0.039	0.716 ± 0.024

Note: SD (Standard Deviation)

Impact of Vocabulary Size

Finally, this study conducted fine-tuning to the vocabulary size. The vocabulary size denotes the number of words this study utilizes in the word embedding layer. With the increase in the number of words, the review length gets longer, which affects the speed of the training process and prediction performance. Therefore, the methodology that utilizes the overall number of words as vocabulary size is ineffective. To obtain optimal vocabulary size, this study evaluated the CNN-TRI model performance according to the vocabulary size from 90,131 to 10,000 as shown in Table 5. This study started with the maximum number of words and repeated for every 10,000 words according to the occurrence frequency.

The CNN-TRI model performance changes with the decrease in the vocabulary size. When the vocabulary size is 4000, the CNN-TRI model shows the best performance with an Accuracy of 0.726 and an F1-Score of 0.733. Such a result indicates when utilizing 40,000 most frequently occurring words, the CNN-TRI model can show the best prediction performance. Meanwhile, such a methodology can also accelerate the model training process. Since the deep learning-based models such as TRI, CM, and CNN applied similar CNN in review helpfulness prediction, this study utilized fine-tuned parameters to the state-of-the-art.

Comparison with Baseline Methodologies

This study utilized the fine-tuned parameters to compare the model prediction performance with the state-of-the-art, as shown in Figure 4. The results showed the CNN-TRI outperforms the state-of-the-art. Such a result indicates the methodology that predicts review helpfulness based on the linear and nonlinear interactions between review text and star rating information can improve the review helpfulness prediction performance.

More specifically, this study reported the improvement in prediction performance, as shown in Table 6. Compared to the current best benchmark TRI,

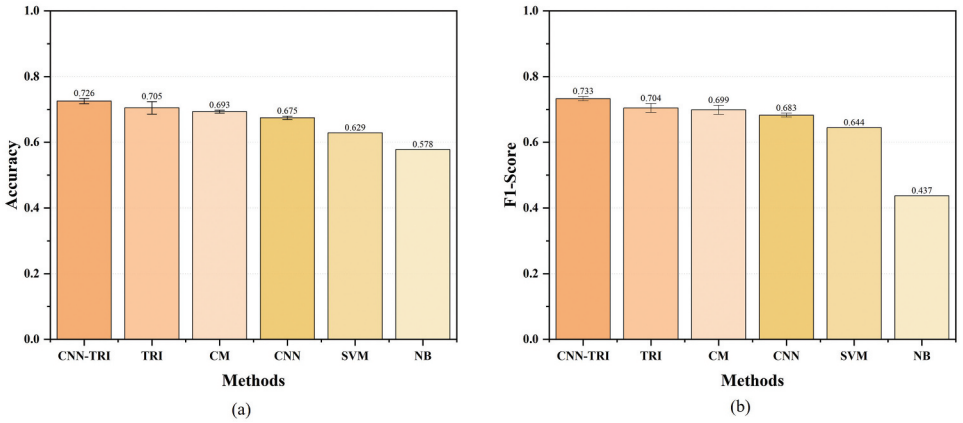

Figure 4. Comparison of CNN-TRI and Benchmarks.

Table 6. Performance Comparison with Benchmarks.

Model	Accuracy	Percentage Change	F1-Score	Percentage Change
TRI	0.705	+2.912%	0.704	+4.05%
CM	0.693	+4.632%	0.699	+4.906%
CNN	0.675	+7.553%	0.683	+7.354%
SVM	0.629	+15.346%	0.644	+13.735%
NB	0.578	+25.608%	0.437	+57.75%

CNN-TRI improved Accuracy by 2.912%, and F1-Score by 4.05%. Compared to the CM that treats star ratings as the last word of review texts, CNN-TRI improved Accuracy by 4.632% and F1-Score by 4.906%. Compared to the basic CNN, CNN-TRI improved Accuracy by 7.553% and F1-Score by 7.354%. Compared to the traditional machine learning models SVM and NB, CNN-TRI improved Accuracy by 15.346% and 25.608%, respectively, and F1-Score by 13.735% and 57.75%, respectively. Such results indicate that the overall interaction feature vector obtained by combining linear and nonlinear interactions can effectively learn the interaction information to improve review helpfulness prediction performance.

The results of the study are as follows: (1) The methodology of independently embedding review text and star rating information (CNN-TRI and TRI) outperforms treating star ratings as the last word of review texts (CM). It indicates that CM significantly creates problems of information loss and limited prediction performance. (2) The methodology of utilizing review text and star rating information outperforms only utilizing the review texts (CNN). It indicates the significance of the methodology that utilizes the review text and star rating information in the review helpfulness prediction task. Specifically, this study can confirm the consistency of review text and star rating information does affect review helpfulness. (3) The methodology of applying deep neural networks outperforms traditional machine learning

(SVM and NB). Since deep neural networks consist of multiple hidden layers, such a methodology can utilize more advanced operations, or multiple activations, in one neuron, as compared to the traditional machine learning model (Janiesch, Zschech, and Heinrich 2021). Meanwhile, it can extract latent semantic features contained in the review text information to improve prediction performance. Therefore, as the results show, deep neural network-based models can provide superior prediction performance. In addition, this study also found that the prediction performance of SVM outperforms NB. Since SVM generalizes positively in high dimensional spaces, it can provide better performance than simple probability-based NB. Overall, the proposed CNN-TRI demonstrates the best performance when compared to the state-of-the-art. It demonstrates the correctness of our ideas and proves the effectiveness of combining linear and nonlinear interactions between review text and star ratings in the review helpfulness prediction task.

Discussion and Conclusion

Conclusions

With the development of e-commerce, online reviews, which play an essential role in customers' purchasing decisions, are becoming more critical. However, as the number of online reviews increases, customers have difficulty exploring the information they need to make purchase decisions. To address information overload problems, online e-commerce websites provide review helpfulness voting services to help consumers make purchasing decisions. However, reviews written a long time ago can receive many votes due to a large number of exposures. In contrast, reviews written recently have a problem of relatively small or missing votes due to a small number of exposures to consumers. To compensate for these problems, it is essential to automatically predict and provide customized online reviews to customers that help them make purchasing decisions. This study has proposed a CNN-TRI mechanism that effectively learns the interaction between online review text and star ratings to explore and provide reviews needed for customer purchase decisions. This study also evaluated the proposed mechanism utilizing Amazon online reviews and confirmed that it outperforms the state-of-art methodologies. The experimental result in Table 6 shows the proposed CNN-TRI mechanism improves by 2.912% to 7.553% on Accuracy and improves by 4.05% to 7.354% on F1-Score compared to the deep neural network-based review helpfulness prediction models. Such a result demonstrates proposed CNN-TRI mechanism can provide superior review helpfulness prediction performance to address information overload problems by recommending helpful reviews for customers.

Theoretical Implications

This study's results provide the following theoretical contributions. First, it fills a gap in current research regarding effectively utilizing the consistency of review text and star rating information in review helpfulness prediction. The results in Figure 4 and Table 6 demonstrate such a consistency significantly affects review helpfulness prediction performance. Specifically, customers prefer reviews that have consistent information when making purchase decisions by referring to online reviews (Huang et al. 2015). This indicates that consistent reviews help reduce the uncertainty surrounding the review's content. Therefore, e-commerce websites should introduce a similar review helpfulness prediction system to explore helpful reviews for customers. Since the proposed mechanism was built with an end-to-end structure, it can automatically explore helpful reviews. Thus, this study can effectively solve information overload problems and recommend customized reviews to customers.

Second, this study addresses the limitation of the star rating information loss problem in previous studies. Although several studies have introduced the consistency of review text and star rating information in review helpfulness prediction, there still exist limitations in star rating representation capacity. Such studies primarily utilize star rating information as a scalar, which limits its representation capacity. Although Qu, Li, and Rose (2018) address star ratings as the last word of review texts in the embedding process to enlarge representation capacity, this encounters information loss problems in the convolution and max-pooling layers. Therefore, this study embedded review text and star rating information independently to avoid information loss and convert them into high-dimensional feature vectors to enlarge representation capacity.

Third, this study addressed the limitation of learning review text-star rating interaction. Previous studies predicted review helpfulness based on simple linear interaction between review text and star ratings. Since such a mechanism disables to capture interaction effectively, it causes limited review helpfulness prediction performance. Therefore, this study additionally applies MLP to capture nonlinear interaction and combine linear interaction to improve prediction performance. The results shows that the proposed mechanism outperforms the state-of-the-art, demonstrating the efficiency of this study.

Practical Implications

The practical implications of this study can be summarized as follows. First, this study helps e-commerce websites to identify the reviews that customers perceive as helpful and focus on the content of those to attract customers. Such helpful reviews often contain richer information than others, which are typically a customer's opinions on specific products. As many customers are concerned about

sustainability and corporate social responsibility, the sellers can use the opinions of customers to improve the quality of the products and increase customer satisfaction, resulting in an effective marketing strategy to create a positive business image.

Second, e-commerce websites can apply the proposed methodology to minimize potential customers' search costs. Through the prediction mechanism, e-commerce websites can recognize helpful reviews when customers write to attract information-seeking customers. Such a friendly automated prediction and recommendation system can enhance customers' attitudes and purchase decision-making based on the prediction mechanism. Moreover, providing rewards, such as coupons, to customers who write helpful reviews can stimulate customers to repeat the behavior and write more helpful reviews, which increases the quality of reviews for the products.

Limitations and Future Research

Although the proposed mechanism outperforms the state-of-the-art, there are several limitations. First, since this study has utilized a single domain dataset to evaluate the model prediction performance, conducting a comparison with multiple domain datasets could reinforce this study's findings. Moreover, many studies built the integrated dataset from multiple domains to evaluate the model performance. Therefore, future research could to compare the review helpfulness prediction performance in multi-domains. Second, many studies have applied the attention mechanism in prediction tasks and demonstrated excellent performance. This study predicted review helpfulness based on linear and nonlinear interactions between review text and star ratings. However, the attention mechanism was built with a more sophisticated structure that can extract latent features and directly compute the attention scores to learn the expressive multimodal features (Han et al. 2022). Therefore, future research could apply a mechanism to verify whether it can improve review helpfulness prediction performance. Finally, this study aimed to predict helpful reviews to recommend to customers automatically. However, it did not formulate the criteria to recommend helpful customer reviews. The methodology of recommending reviews according to the helpfulness score creates a problem that reviews with a high helpfulness score to obtain more votes, and reviews with a low helpfulness score are likely not recommended to customers. Therefore, future research of this study could propose a methodology to address such a problem.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the BK21 FOUR (Fostering Outstanding Universities for Research) funded by the Ministry of Education(MOE, Korea) and National Research Foundation of Korea(NRF)

References

Agnihotri, A., and S. Bhattacharya. 2016. Online review helpfulness: Role of qualitative factors. *Psychology & Marketing* 33 (11):1006–17. doi:10.1002/mar.20934.

Ahmad, S. N., and M. Laroche. 2017. Analyzing electronic word of mouth: A social commerce construct. *International Journal of Information Management* 37 (3):202–13. doi:10.1016/j.ijinfomgt.2016.08.004.

Bengio, Y., R. Ducharme, and P. Vincent. 2000. A neural probabilistic language model. In *Proceedings of the 2000 Neural Information Processing Systems (NIPS) Conference*, Cambridge, MA, USA: MIT Press, 1–7.

Bilal, M., M. Marjani, I. A. T. Hashem, N. Malik, M. I. U. Lali, and A. Gani. 2021. Profiling reviewers' social network strength and predicting the "Helpfulness" of online customer reviews. *Electronic Commerce Research and Applications* 45:101026. doi:10.1016/j.elerap.2020.101026.

Cao, Q., W. Duan, and Q. Gan. 2011. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems* 50 (2):511–21. doi:10.1016/j.dss.2010.11.009.

Chen, C., M. Qiu, Y. Yang, J. Zhou, J. Huang, X. Li, and F. S. Bao. 2019. Multi-domain gated CNN for review helpfulness prediction. In Proceedings of the 2019 World Wide Web Conference, New York, NY, USA: Association for Computing Machinery, 2630–2636.

Chen, G., S. Xiao, C. Zhang, and W. Wang. 2022. An orthogonal-space-learning-based method for selecting semantically helpful reviews. *Electronic Commerce Research and Applications* 53:101154. doi:10.1016/j.elerap.2022.101154.

Chen, Y., and J. Xie. 2008. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science* 54 (3):477–91. doi:10.1287/mnsc.1070.0810.

Chiriatti, G., D. Brunato, F. Dell'Orletta, and G. Venturi. 2019. What Makes a Review helpful? Predicting the Helpfulness of Italian TripAdvisor Reviews. In *Proceedings of the Sixth Italian Conference on Computational Linguistics*, Bari, Italy: CEUR, 1–6.

Chung, H., and K. -S. Shin. 2020. Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. *Neural Computing & Applications* 32 (12):7897–914. doi:10.1007/s00521-019-04236-3.

Dos Santos, C., and M. Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 69–78.

Du, J., J. Rong, H. Wang, and Y. Zhang. 2021. Neighbor-aware review helpfulness prediction. *Decision Support Systems* 148:113581. doi:10.1016/j.dss.2021.113581.

Du, J., L. Zheng, J. He, J. Rong, H. Wang, and Y. Zhang. 2020. An interactive network for end-to-end review helpfulness modeling. *Data Science and Engineering* 5 (3):261–79. doi:10.1007/s41019-020-00133-1.

Fan, M., Y. Feng, M. Sun, P. Li, H. Wang, and J. Wang. 2018. Multi-task neural learning architecture for end-to-end identification of helpful reviews. In *Proceedings of the 2018*

IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, IEEE Press, 343–350.

Filieri, R., and M. Mariani. 2021. The role of cultural values in consumers' evaluation of online review helpfulness: A big data approach. *International Marketing Review* 38 (6):1267–88. doi:10.1108/IMR-07-2020-0172.

Ge, S., T. Qi, C. Wu, F. Wu, X. Xie, and Y. Huang. 2019. Helpfulness-aware review based neural recommendation. *CCF Transactions on Pervasive Computing and Interaction* 1 (4):285–95. doi:10.1007/s42486-019-00023-0.

Ghose, A., and P. G. Ipeirotis. 2007. Designing novel review ranking systems: Predicting the usefulness and impact of reviews. In *Proceedings of the 9th International Conference on Electronic Commerce*, New York, NY, USA: Association for Computing Machinery, 303–310.

Ghose, A., and P. G. Ipeirotis. 2010. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering* 23 (10):1498–512. doi:10.1109/TKDE.2010.188.

Gottschalk, S. A., and A. Mafael. 2017. Cutting through the online review jungle—investigating selective eWOM processing. *Journal of Interactive Marketing* 37:89–104. doi:10.1016/j.intmar.2016.06.001.

Guo, B., C. Zhang, J. Liu, and X. Ma. 2019. Improving text classification with weighted word embeddings via a multi-channel TextCNN model. *Neurocomputing* 363:366–74. doi:10.1016/j.neucom.2019.07.052.

Han, W., H. Chen, Z. Hai, S. Poria, and L. Bing. 2022. SANCL: Multimodal Review Helpfulness Prediction with Selective Attention and Natural Contrastive Learning. In *Proceedings of the 29th International Conference on Computational Linguistics*, Gyeongju, Republic of Korea: International Committee on Computational Linguistics, 5666–5677.

Han, Q., Y. Kou, and D. Snaidauf. 2019. Experimental Evaluation of CNN Parameters for Text Categorization in Legal Document Review. In *Proceedings of 2019 IEEE International Conference on Big Data (Big Data)*, IEEE, 4320–4324.

Haque, M. E., M. E. Tozal, and A. Islam. 2018. Helpfulness prediction of online product reviews. In *Proceedings of the ACM Symposium on Document Engineering 2018*, New York, NY, USA: Association for Computing Machinery, 1–4.

Hazarika, B., K. Chen, and M. Razi. 2021. Are numeric ratings true representations of reviews? A study of inconsistency between reviews and ratings. *International Journal of Business Information Systems* 38 (1):85–106. doi:10.1504/IJBIS.2021.118637.

He, X., L. Liao, H. Zhang, L. Nie, X. Hu, and T. -S. Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web Companion*, Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 173–182.

He, R., and J. McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference Companion on World Wide Web*, Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 507–517.

Hossain, M. S., and M. F. Rahman. 2022. Detection of potential customers' empathy behavior towards customers' reviews. *Journal of Retailing and Consumer Services* 65:102881. doi:10.1016/j.jretconser.2021.102881.

Hossain, M. S., M. F. Rahman, M. K. Uddin, and M. K. Hossain. 2022. Customer sentiment analysis and prediction of halal restaurants using machine learning approaches. *Journal of Islamic Marketing(ahead-Of-print)(ahead-Of-Print)*. doi:10.1108/JIMA-04-2021-0125.

Huang, A. H., K. Chen, D. C. Yen, and T. P. Tran. 2015. A study of factors that contribute to online review helpfulness. *Computers in Human Behavior* 48:17–27. doi:10.1016/j.chb.2015.01.010.

Hu, Y. -H., and K. Chen. 2016. Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management* 36 (6):929–44. doi:10.1016/j.ijinfomgt.2016.06.003.

Janiesch, C., P. Zschech, and K. Heinrich. 2021. Machine learning and deep learning. *Electronic Markets* 31 (3):685–95. doi:10.1007/s12525-021-00475-2.

Jones, Q., G. Ravid, and S. Rafaeli. 2004. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research* 15 (2):194–210. doi:10.1287/isre.1040.0023.

Khan, Z. Y., and Z. Niu. 2021. CNN with depthwise separable convolutions and combined kernels for rating prediction. *Expert Systems with Applications* 170:114528. doi:10.1016/j.eswa.2020.114528.

Kim, Y. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, 1746–1751.

Kim, S. J., E. Maslowska, and E. C. Malthouse. 2018. Understanding the effects of different review features on purchase probability. *International Journal of Advertising* 37 (1):29–53. doi:10.1080/02650487.2017.1340928.

Kim, S. -M., P. Pantel, T. Chklovski, and M. Pennacchiotti. 2006. Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia: Association for Computational Linguistics, 423–430.

Kingma, D. P., and J. Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412 6980*.

Kiran, R., P. Kumar, and B. Bhasker. 2020. Oslcfit (organic simultaneous LSTM and CNN Fit): A novel deep learning based solution for sentiment polarity classification of reviews. *Expert Systems with Applications* 157:113488. doi:10.1016/j.eswa.2020.113488.

Korfiatis, N., E. García-Bariocanal, and S. Sánchez-Alonso. 2012. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications* 11 (3):205–17. doi:10.1016/j.elerap.2011.10.003.

Krishnamoorthy, S. 2015. Linguistic features for review helpfulness prediction. *Expert Systems with Applications* 42 (7):3751–59. doi:10.1016/j.eswa.2014.12.044.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86 (11):2278–324. doi:10.1109/5.726791.

Li, M., L. Huang, C. -H. Tan, and K. -K. Wei. 2013. Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce* 17 (4):101–36. doi:10.2753/JEC1086-4415170404.

Li, Q., X. Li, B. Lee, and J. Kim. 2021. A hybrid CNN-based review helpfulness filtering model for improving e-commerce recommendation Service. *Applied Sciences* 11 (18):8613. doi:10.3390/app11188613.

Liu, J., Y. Cao, C. -Y. Lin, Y. Huang, and M. Zhou. 2007. Low-quality product review detection in opinion summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Prague, Czech Republic: Association for Computational Linguistics, 334–342.

Liu, P., X. Qiu, and X. Huang. 2016. Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv:1605 05101*.

Liu, Z., B. Yuan, and Y. Ma. 2022. A multi-task dual attention deep recommendation model using ratings and review helpfulness. *Applied Intelligence* 52 (5):5595–607. doi:10.1007/s10489-021-02666-y.

Lu, Y., P. Tsaparas, A. Ntoulas, and L. Polanyi. 2010. Exploiting social context for review quality prediction. In *Proceedings of the 19th International Conference on World Wide Web*, New York, NY, USA: Association for Computing Machinery, 691–700.

Malik, M., and A. Hussain. 2017. Helpfulness of product reviews as a function of discrete positive and negative emotions. *Computers in Human Behavior* 73:290–302. doi:10.1016/j.chb.2017.03.053.

Malik, M., and A. Hussain. 2018. An analysis of review content and reviewer variables that contribute to review helpfulness. *Information Processing & Management* 54 (1):88–104. doi:10.1016/j.ipm.2017.09.004.

Martin, L., and P. Pu. 2014. Prediction of helpful reviews using emotions extraction. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, California, USA: AAAI Press, 1551–1557.

Mauro, N., L. Ardissono, and G. Petrone. 2021. User and item-aware estimation of review helpfulness. *Information Processing & Management* 58 (1):102434. doi:10.1016/j.ipm.2020.102434.

Metzger, M. J., A. J. Flanagin, and R. B. Medders. 2010. Social and heuristic approaches to credibility evaluation online. *The Journal of Communication* 60 (3):413–39. doi:10.1111/j.1460-2466.2010.01488.x.

Mitra, S., and M. Jenamani. 2021. Helpfulness of online consumer reviews: A multi-perspective approach. *Information Processing & Management* 58 (3):102538. doi:10.1016/j.ipm.2021.102538.

Mohammad, A. H., T. Alwadaʻn, and O. Al-Momani. 2016. Arabic text categorization using support vector machine, Naïve Bayes and neural network. *GSTF Journal on Computing (JoC)* 5 (1):1–8. doi:10.7603/s40601-016-0016-9.

Ngo-Ye, T. L., and A. P. Sinha. 2014. The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems* 61:47–58. doi:10.1016/j.dss.2014.01.011.

Nguyen, T. H., and R. Grishman. 2015. Relation extraction: Perspective from convolutional neural networks. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, Denver, Colorado: Association for Computational Linguistics, 39–48.

Nguyen, H. T., and M. Le Nguyen. 2019. An ensemble method with sentiment features and clustering support. *Neurocomputing* 370:155–65. doi:10.1016/j.neucom.2019.08.071.

Olmedilla, M., M. R. Martínez-Torres, and S. Toral. 2022. Prediction and modelling online reviews helpfulness using 1D Convolutional Neural Networks. *Expert Systems with Applications* 198:116787. doi:10.1016/j.eswa.2022.116787.

Pan, X., L. Hou, and K. Liu. 2020. Predicting the future increment of review helpfulness: An empirical study based on a two-wave data set. *Electronic Library* 39 (1):59–76. doi:10.1108/EL-06-2020-0130.

Park, D. H., H. K. Kim, I. Y. Choi, and J. K. Kim. 2012. A literature review and classification of recommender systems research. *Expert Systems with Applications* 39 (11):10059–72. doi:10.1016/j.eswa.2012.02.038.

Pashchenko, Y., M. F. Rahman, M. S. Hossain, M. K. Uddin, and T. Islam. 2022. Emotional and the normative aspects of customers' reviews. *Journal of Retailing and Consumer Services* 68:103011. doi:10.1016/j.jretconser.2022.103011.

Quan, D., S. Wang, N. Huyan, J. Chanussot, R. Wang, X. Liang, B. Hou, and L. Jiao. 2022. Element-Wise Feature Relation Learning Network for Cross-Spectral Image Patch Matching. *IEEE Transactions on Neural Networks and Learning Systems* 33 (8):3372–3386. doi:10.1109/TNNLS.2021.3052756.

Quaschning, S., M. Pandelaere, and I. Vermeir. 2015. When consistency matters: The effect of valence consistency on review helpfulness. *Journal of Computer-Mediated Communication* 20 (2):136–52. doi:10.1111/jcc4.12106.

Qu, X., X. Li, and J. R. Rose. 2018. Review helpfulness assessment based on convolutional neural network. *arXiv preprint arXiv:1808 09016*.

Roy, G., B. Datta, and S. Mukherjee. 2019. Role of electronic word-of-mouth content and valence in influencing online purchase behavior. *Journal of Marketing Communications* 25 (6):661–84. doi:10.1080/13527266.2018.1497681.

Saumya, S., J. P. Singh, and Y. K. Dwivedi. 2020. Predicting the helpfulness score of online reviews using convolutional neural network. *Soft Computing* 24 (15):10989–1005. doi:10.1007/s00500-019-03851-5.

Schlosser, A. E. 2011. Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. *Journal of Consumer Psychology* 21 (3):226–39. doi:10.1016/j.jcps.2011.04.002.

Shen, R. -P., H. -R. Zhang, H. Yu, and F. Min. 2019. Sentiment based matrix factorization with reliability for recommendation. *Expert Systems with Applications* 135:249–58. doi:10.1016/j.eswa.2019.06.001.

Sun, Q., J. Niu, Z. Yao, and H. Yan. 2019. Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence* 81:68–78. doi:10.1016/j.engappai.2019.02.004.

Tay, W., X. Zhang, and S. Karimi. 2020. Beyond mean rating: Probabilistic aggregation of star ratings based on helpfulness. *Journal of the Association for Information Science and Technology* 71 (7):784–99. doi:10.1002/asi.24297.

Topaloglu, O., and M. Dass. 2021. The impact of online review content and linguistic style matching on new product sales: The moderating role of review helpfulness. *Decision Sciences* 52 (3):749–75. doi:10.1111/deci.12378.

van Dinter, R., C. Catal, and B. Tekinerdogan. 2021. A Multi-Channel Convolutional Neural Network approach to automate the citation screening process. *Applied Soft Computing* 112:107765. doi:10.1016/j.asoc.2021.107765.

Wu, H., and X. Gu. 2015. Towards dropout training for convolutional neural networks. *Neural Networks* 71:1–10. doi:10.1016/j.neunet.2015.07.007.

Yang, C. -Y., J. -S. Yang, and J. -J. Wang. 2009. Margin calibration in SVM class-imbalanced learning. *Neurocomputing* 73 (1–3):397–411. doi:10.1016/j.neucom.2009.08.006.

Yang, Y., Y. Yan, M. Qiu, and F. Bao. 2015. Semantic analysis and helpfulness prediction of text for online product reviews. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China: Association for Computational Linguistics, 38–44.

Yang, S., J. Yao, and A. Qazi. 2020. Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining. *Information Processing & Management* 57 (2):102179. doi:10.1016/j.ipm.2019.102179.

Yin, D., S. D. Bond, and H. Zhang. 2014. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly* 38 (2):539–60. doi:10.25300/MISQ/2014/38.2.10.

Zhang, Y., S. Roller, and B. Wallace. 2016. MGNC-CNN: A simple approach to exploiting multiple word embeddings for sentence classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics,1522–1527.

Zhang, R., T. Tran, and Y. Mao. 2012. Opinion helpfulness prediction in the presence of "words of few mouths". *World Wide Web* 15 (2):117–38. doi:10.1007/s11280-011-0127-3.

Zhao, S., Z. Xu, L. Liu, M. Guo, and J. Yun. 2018. Towards accurate deceptive opinions detection based on word order-preserving CNN. *Mathematical Problems in Engineering* 2018:1–9. doi:10.1155/2018/2410206.