SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

December 2021, Vol. 64 222104:1–222104:13
https://doi.org/10.1007/s11432-020-3116-x

Sentiment classification via user and product interactive modeling

Xiabing ZHOU, Zhongqing WANG*, Min ZHOU, Qifa WANG, Shoushan LI, Min ZHANG & Guodong ZHOU

School of Computer Science and Technology, Soochow University, Suzhou 215006*, China*

Received 26 May 2020/Revised 11 August 2020/Accepted 26 October 2020/Published online 25 November 2021

Abstract Sentiment classification aims to identify the polarity of a given review. Most existing methods consider each review as an individual while ignoring the importance of the user and product information of the given review. A direct way to integrate user and product information is to employ an attention mechanism to learn the local interaction between them. However, local interactions cannot capture the global optimization among user and product information. Therefore, we propose a novel interactive model to integrate both local and global interactions between users and products. In particular, we employ an attention mechanism to learn local interactions between users and products, and construct user and product interactive graphs to model the global interaction of users and products. Empirical evaluation shows that our model outperforms previous state-of-the-art methods significantly by learning the local and global interactions among users' preferences, product characteristics, and reviews.

Keywords product review analysis, sentiment classification, interactive graph, graph convolutional network

Citation Zhou X B, Wang Z Q, Zhou M, et al. Sentiment classification via user and product interactive modeling. Sci China Inf Sci, 2021, 64(12): 222104, https://doi.org/10.1007/s11432-020-3116-x

1 Introduction

Sentiment classification is a task of identifying the sentiment polarity of a given text. In recent years, sentiment classification has drawn much attention in the field of natural language processing (NLP) and has many useful applications, such as opinion mining, product evaluation [1–3]. Therefore, sentiment classification is one of the most active and critical researches for the industry and customers.

Previous enormous studies in sentiment analysis mainly analyze texts individually by using machine learning algorithms. The performances of these models are heavily dependent on features, which either focus on hand-craft features [4, 5] or use neural networks to discriminate features from data [6, 7].

However, these researches ignore users who express the sentiment and products which are evaluated. Both of them have great influences on predicting the sentiment of texts. Therefore, several studies focus on analyzing the influences of users and products for sentiment classification. For example, Chen et al. [8] built a hierarchical long short-term memory (LSTM) model with a user product joint attention mechanism. Wu et al. [9] applied two individual hierarchical neural networks to generate two representations with user attention or product attention. Kim et al. [10] imposed category-specific weights instead of a single weight for user, product, and review document. All of them show the importance of user and product information. However, most of them either consider auxiliary information according to the user or product ID, which is inadequate for using the information, or simply consider the local interaction by using an attention mechanism.

Local interaction means considering the review individually, while global interaction means building connections among user's reviews, product reviews, and the main review document as comprehensively as possible, and obtains global optimization. As shown in Table 1, from the user's point of view, local

* Corresponding author (email: wangzq.antony@gmail.com)

Table 1 An example of comments

	Review	Sentiment
Main review	These shoes are plain in color and concise in style.	Positive
User-review 1	This plain color shoe is very versatile, no matter what I wear looks good with it.	Positive
User-review 2	I prefer this concise style. It feels very clean and refreshing.	Positive
User-review 3	The style of this dress is too exaggerated. It's so bright that I can't wear it at work.	Negative
Product-review 1	I really like these shoes, the colors are simple and generous, and the feet are very comfortable.	Positive
Product-review 2	The plain color is very versatile. These shoes look good no matter how you wear them.	Positive

limited user information cannot analyze the user's preference exactly, because this user has both positive and negative reviews. Thus, we should consider the user's preference from a global view and get the information that this user prefers plain colors and concise styles. In addition, combining the other product reviews further explains that these shoes are comfortable and liked by most people. Thus, the sentiment of the main review is positive.

Therefore, it is worthwhile to model user's reviews and product reviews for sentiment classification using the local and global views collectively. In this paper, we propose a user product interactive model (UPIM) to learn both local and global interaction of users and products. Firstly, we learn the representations of the main review document, user's review documents, and product review documents. Secondly, we utilize the multi-head attention mechanism to capture the local interaction between the user/product information and the main review, which can aggregate the representations of those user/product-aware words to the main review. Thirdly, we construct a user interactive graph and a product interactive graph to learn the global influence of the user and product information. Specifically, we model the contents of user's reviews and product reviews, which are a more comprehensive reflection of user's preferences and product characteristics than ID modeling. Besides, we incorporate the interactive graphs into the convolutional neural network, which can better capture implicit interaction influences and achieve global optimization. To prove the efficiency of our model, we conduct experiments on the Yelp dataset and Amazon product data. We compare our model with several other models. The experimental results show that our proposed model outperforms previous state-of-the-art methods significantly.

2 Related work

Sentiment classification is a fundamental problem in sentiment analysis, which is a research hotspot in natural language processing.

2.1 Sentiment classification

Traditional studies are based on statistical machine learning models, which mainly rely on text with labeled sentiment polarity to build classifiers. For example, Liu et al. [11] used a multiclass support vector machine (SVM) model to solve sentiment classification, and extended it to adapt to sentiment words on a specific topic. Zhang et al. [12] first used word2vec to cluster the similar features, and then implemented sentiment classification by SVM^{perf}, which is extended by original SVM for optimizing multivariate performance measures. Song et al. [13] calculated the weights to reflect the different number of positive and negative words and utilized multinomial Naïve Bayes algorithm for feature selection and sentiment classification.

With the development of neural network models, a series of models have been successfully applied to sentiment analysis. These models are leveraged as feature extractors to learn the representation of text, including convolutional neural network (CNN), recurrent neural network (RNN), and LSTM.

CNN uses a convolutional operation to capture features [14]. Wei et al. [15] proposed a two-layer CNN for cross-domain product review sentiment classification. Nguyen, Kavuri, and Lee [16] integrated the CNN in the fuzzy logic domain, which can benefit from the use of fuzzy logic and obtain more refined outputs. Liu et al. [17] utilized a gated recurrent unit (GRU) to obtain the compositional semantics of the document and CNN to capture more dependencies between sentence features. Then, they employed an attention mechanism in these two parts to distinguish the importance of words, sentences, and features in the document.

RNN forms the representation gradually from the previous historical context [18]. LSTM is a variant of RNN, which can learn long short-term dependency and address the problem of gradient disappearance [19]. Wang et al. [20] used LSTM for Twitter sentiment prediction, which can handle interactions between words. Qian et al. [21] utilized LSTM for sentiment classification, and also attempted to model the linguistic role of sentiment lexicons, negation words, and intensity words. Lei et al. [22] proposed a novel sentiment lexicon enhanced attention-based LSTM model for sentence-level sentiment analysis. Feng et al. [23] incorporated preceding tweets for context-aware sentiment classification and developed a context attention based LSTM network.

Transformer is one of the state-of-the-art model architectures in NLP [24]. It benefits from the powerful multi-head self-attention mechanism, which learns the token dependencies. Long et al. [25] combined LSTM with multi-head attention to improve the performance of sentiment analysis. BERT stands for bidirectional encoder representations from transformers, which can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks and obtains new state-of-the-art results, including sentiment analysis [26–28].

Despite the effective performance these neural network models have, the improvement is still limited, because they consider the text individually.

2.2 Sentiment classification with external information

Apart from the text, user's preferences and most customers' impressions about the product make a significant effect on the ratings. In recent years, some researchers consider additional information (e.g., user/product information) for sentiment classification. For example, Tang et al. [29] modeled each user and each product by using a vector space model, and incorporated user- and product- level information into a neural network approach for sentiment classification. Gui et al. [30] made use of a heterogeneous network to model the shared polarity in product reviews and learn representations of users, products they commented on, and words they used simultaneously. Ma et al. [31] cascaded the user and product information to influence the generation of attention on the word and sentence layers when judging the sentiment of a document. Dou [32] proposed a deep memory network for document-level sentiment classification, which could capture the user and product information at the same time. Wu et al. [9] used user attention and product attention for training and final sentiment classification. Kim et al. [10] studied customized text classification and proposed to use basis vectors to effectively incorporate categorical metadata on various parts of a neural-based model, which can customize classifiers based on possibly multiple different known categorical metadata information (e.g., user/product information for sentiment classification).

Different from previous studies, they either constructed an auxiliary information model according to the user and product ID, or considered user and product joint attention, which lead to underutilization of information. In addition, most studies employ the attention mechanism, but do not consider the global interaction. Therefore, we propose a user product interactive model to capture the local and global interaction among reviews, users, and products for sentiment classification.

3 User and product interactive modeling

In this paper, we aim to detect the sentiment of product reviews. Previous studies focus on analyzing the text individually. However, there are interactions among users, products, and reviews. In order to incorporate user preferences and product characteristics reasonably and effectively, we propose a UPIM. In order to understand the symbols in the paper more clearly, we describe the main symbols in Table 2.

Our task can be formalized as follows: let U, P, and D represent the sets of users, products, and reviews respectively. A user u writes a review document d about product p, and we should predict the sentiment y for the review d, where $u \in U$, $p \in P$ and $d \in D$. In order to better use the user and product information, the input of our model is user's reviews $U(d)$, product reviews $P(d)$ and the main review document d, where $U(d) = \{d_u | d_u$ is the review that written by the user $u, d_u \neq d\}$ and $P(d) = \{d_p | d_p$ is the review about the product $p, d_p \neq d\}$.

Figure 1 illustrates the overview of our proposed model. From the figure, the framework can be separated into three main components: the representation of review, users and products, the local interaction with multi-head attention mechanisms, and the global interaction with graph convolutional neural network. In the representation part, we employ LSTM and self-attention to learn the representation of user's

Table 2 Symbol description

Symbol	Description
U	The set of all the users
P	The set of all the products
D	The set of all the reviews
d	The main review that is used for sentiment analysis
d_u	The review is written by the user u
d_p	The review for the product p
$U(d)$	The set of d_u
$P(d)$	The set of d_p

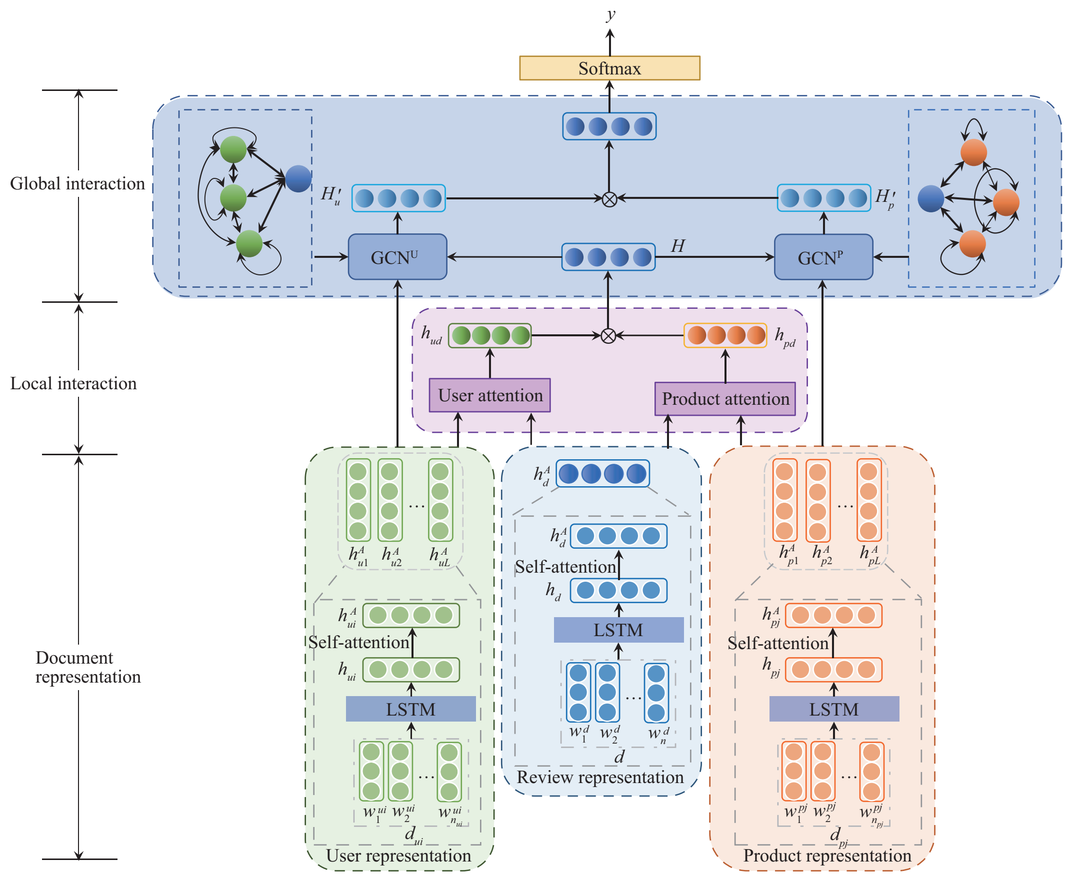

Figure 1 (Color online) Overview of our proposed model.

reviews $U(d)$, product reviews $P(d)$, and the main review d. The local interaction mainly learns the words in the main review that show the strong user's preferences and indicate the product's characteristics. The global interaction learns the implicit interaction influence among user's reviews, product reviews, and the main review. In the following, we will illustrate the model in detail.

3.1 Review representation

In general, we denote a review document d with n words $\{w_1, w_2, \ldots, w_n\}$. Firstly, we transform each token w_i into a real-valued vector x_i using the word embedding vector of w_i, obtained by looking up a pre-trained word embedding table V via the skip-gram algorithm to train embeddings. Then, we use a standard LSTM model to learn the shared document representation, and generate a hidden vector sequence $\{h_1, h_2, \ldots, h_n\}$. At each step t, the hidden vector h_t of the LSTM model is computed based on the current vector x_t and the previous vector h_{t-1} with $h_t = \mathrm{LSTM}(x_t, h_{t-1})$. Thus, we obtain the review document d initial representation $h_d = \{h_{d1}, h_{d2}, \ldots, h_{dn}\}$.

In order to obtain the syntactic and semantic features of the review document for sentiment analysis, we use the multi-head self-attention mechanism [24]. For each review d, weight distributions over tokens are calculated, allowing the model to flexibly encode reviews in different representation subspaces by attending to different words. For each head $z \in 1, \dots, m$, we have

$$\text{head}_d^z = \sum_{i=1}^{n} v_i^z \beta_i^z, \tag{1}$$

$$\alpha_i^z = W_\alpha^z h_{di}, \tag{2}$$

$$\beta_i^z = W_\beta^z h_{di}, \tag{3}$$

$$v_i^z = \exp(\alpha_i^z) / \sum_j \exp(\alpha_j^z), \tag{4}$$

where $W_\alpha^z \in \mathbb{R}^{1\times k}$ and $W_\beta^z \in \mathbb{R}^{k_{\text{head}}\times k}$ are weights. $k_{\text{head}} = k/m$ is the dimension of each head. k is the dimension of input hidden vector. We jointly attend to information from different representation subspaces and get the review representation vector:

$$h_d^A = \text{Concat}(\text{head}_d^1, \text{head}_d^2, \dots, \text{head}_d^m). \tag{5}$$

3.2 User and product representations

In our method, we model the contents of user's reviews and product reviews. Thus, we first transform each token of the user's review and the product review into a real-valued vector by looking up the embedding vector from a pre-train word embedding table. And then, we employ the LSTM model to capture contextual information, and obtain the initial representation $h_u = \{h_{u1}, h_{u2}, \dots, h_{un}\}$ and $h_p = \{h_{p1}, h_{p2}, \dots, h_{pn}\}$ of the user's review and the product review respectively.

From the user's point of view, not all words reflect user's preference equally for sentiment classification. Thus, we use multi-head self-attention to learn the user-specific representation. Similar to the review representation, the final user's review representation vector is obtained formally as

$$h_u^A = \text{Concat}(\text{head}_u^1, \text{head}_u^2, \dots, \text{head}_u^m). \tag{6}$$

It is also true for the product review, and we can obtain the product review representation h_p^A in the same way.

3.3 Local interaction with user and product attention

It is obvious that not all words contribute equally to the sentence meaning for different users and products. Thus, we employ a multi-head attention mechanism to extract the important words for sentiment classification and to aggregate the representations of those informative user/product-aware words.

3.3.1 *User based attention*

In order to exchange information across users and review documents from the local interactive view, we apply the multi-head attention mechanism to $U(d)$ and d, which can collect information from the user's preferences to analyze the sentiment of the main review.

Similar to self-attention, it allows user's reviews to attend to the main review document by calculating an attention distribution. Because we consider the L user's review representation h_u^A, we first average them and get the representation $(h_u^A)'$. Then, for each head, we can obtain the user attention by calculating the equations:

$$Q^z = W_Q^z h_d^A, \tag{7}$$

$$K^z = W_K^z (h_u^A)', \tag{8}$$

$$V^z = W_V^z h_d^A, \tag{9}$$

$$h_{ud}^z = \text{softmax}\left(\frac{Q^z (K^z)^{\mathrm{T}}}{\sqrt{k_{\text{head}}}}\right) V^z, \tag{10}$$

where W_Q^z, W_K^z, $W_V^z \in \mathbb{R}^{k_{\text{head}}\times k}$ are weights [24]. Multi-head attention allows the model to jointly attend to information from different representation subspaces, thus, $h_{ud} = \text{Concat}(h_{ud}^1, h_{ud}^2, \dots, h_{ud}^m) W^o$ represents the review document hidden vector generated by an attention operation over the user's reviews, where $W^o \in \mathbb{R}^{k\times k}$.

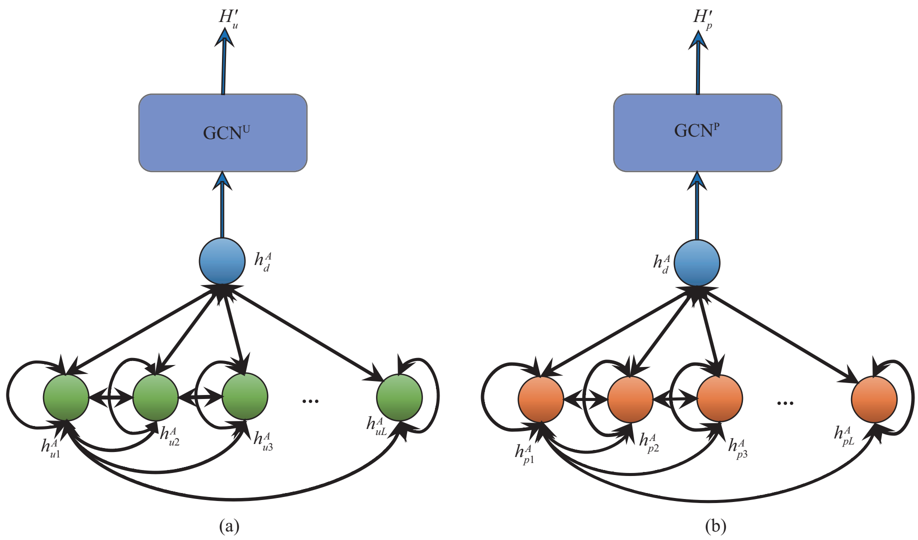

Figure 2 (Color online) Overview of user and product graph convolutional neural network. In the figure, the circles represent the reviews, and the directed edges represent the associations of two nodes.

3.3.2 *Product based attention*

It is also true that each word or sentence for different products provides different information to the main review document. Based on common sense, the product multi-head attention, which incorporates product information into the main review document, is similar to the user multi-head attention. In the view of product, we replace K^z in (8) with $\hat{K}^z = \hat{W}^z_K h^A_p$, and get h_{pd}.

Finally, we concatenate h_{ud} and h_{pd}, and get $H = h_{ud} \oplus h_{pd}$.

3.4 Global interaction with user and product graphs

In order to better learn the connections among the user's reviews $U(d)$, the product reviews $P(d)$ and the main review document d, we construct the user interactive graph and product interactive graph, which take more account of the influence of user's preferences and product characteristics. Figure 2 shows two interactive graphs and the graph convolutional operation. In the figure, green circles represent the other reviews $U(d)$ written by the user d, where L is the size of $U(d)$, and orange circles represent the other reviews $P(d)$ written about the product p, where L is the size of $P(d)$. The blue circle represents the main review d. The directed edges represent the association of two nodes, and the thickness of an edge represents the degree of association. The details are introduced below.

3.4.1 *User interactive graph*

The user interactive graph is used to model the global connections between the user's reviews and the main review. Thus, for the user interactive graph, the vertex sets are the user's reviews $U(d) = \{d_{u1}, d_{u2}, \ldots, d_{uL}\}$ and the review document d, where L is the number of connections.

To represent the graph, the adjacency matrix A needs to be introduced. The elements of the adjacency matrix indicate whether pairs of vertices are adjacent or not in the graph. We employ the adjacency matrix A_u to construct the user interactive graph to capture the global influence among the user's reviews. As shown in Figure 2(a), we choose L user's reviews to connect the main review vertice, and the user's reviews are related to each other. If the user's review vertice i connects with the main review vertice (we set it the label 0), the $(i, 0)$-entry and the $(0, i)$-entry of adjacency matrix A_u will be assigned a non-zero positive value.

3.4.2 *Product interactive graph*

The product interactive graph is used to model the global connections among the product reviews and the main review. It is true for the product interactive graph, the vertex sets are the product reviews $P(d) = \{d_{p1}, d_{p2}, \ldots, d_{pL}\}$ and the review document d.

In order to construct the product interactive graph, we employ the adjacency matrix A_p to capture the global influence among the product reviews. As shown in Figure 2(b), we choose L product reviews to connect the main review vertice, and the product reviews are related to each other. If the product review vertice i connects with the main review vertice (we set it the label 0), the $(i, 0)$-entry and the $(0, i)$-entry of adjacency matrix A_p will be assigned a non-zero positive value.

In order to consider the different influences of user and product reasonably, we set the different weights in the adjacency matrix, which are represented in the figure by the edges of different thicknesses. The final weights are determined by the experiment.

3.4.3 *Graph convolutional networks over user and product interactive graphs*

We adopt the graph convolutional operation to model the connections among user's reviews and product reviews with the main review document by converting the interactive graph into its corresponding adjacency matrix A [33]. Given an $n \times n$ adjacency matrix, which represents a graph with n nodes, the output vector of node i at the l-layer can be written as

$$h_i^{(l)} = \sigma\left(\sum_{j=1}^{n} A_{ij} W^{(l)} h_j^{(l-1)} + b^{(l)}\right), \tag{11}$$

where $W^{(l)}$ is a linear transformation, $b^{(l)}$ is a bias term, and σ is a nonlinear function (e.g., ReLU). Intuitively, during each graph convolution, each node gathers and summarizes information from its neighboring nodes in the graph.

To make use of the representations for sentiment classification with user's preference and product characteristic information, we obtain the graphs based on the representation as follows:

$$H'_u = f(\mathrm{GCN}(H_{ud})), \tag{12}$$

$$H'_p = f(\mathrm{GCN}(H_{pd})), \tag{13}$$

where $H_{ud} = \{H, h^A_{u1}, \ldots, h^A_{uL}\}$ denotes the collective hidden representations of the main review and the user's reviews, $H_{pd} = \{H, h^A_{p1}, \ldots, h^A_{pL}\}$ denotes the collective hidden representations of the main review and the product reviews. f is a max pooling function that maps from L output vectors to the post vector. GCN is the graph convolutional operation as (11). At last, we concatenate H'_u and H'_p and get $R = H'_u \oplus H'_p$.

3.5 Objective function and training

In the end, we use a linear and softmax function to classify the hidden vector R, and get the predicted sentiment

$$\hat{y} = \mathrm{softmax}(WR + b), \tag{14}$$

where W and b are model parameters.

Our training objective is to minimize the cross-entropy loss over a set of training examples $(d_i, y_i)|_{i=1}^{N}$,

$$J(\theta_y) = -\sum_{i=1}^{N} [y_i \log \hat{y}_i^j + (1 - y_i) \log(1 - \hat{y}_i)]. \tag{15}$$

4 Experimentation

4.1 Experimental settings

In this study, we derive the dataset from the Yelp dataset in 2013 and Amazon product data (i.e., Movies and CDs). In the experiment, we first carry on a series of preprocessing to the data, and then, we choose 16000/2000 reviews as training/test set respectively. We keep the ratio of positive and negative samples 1:1. The number of users' reviews and product reviews are both 8.

The hyper parameters in the experiment are set as follows. The learning rate in the model is 0.0001 and the optimizer is Adam; The batch size, embedding size, and hidden size are 32, 128 and 128 respectively;

Table 3 Comparison of baselines on the Yelp dataset

Method	Positive			Negative			Accuracy
	Precision	Recall	F1-score	Precision	Recall	F1-score	
LSTM	0.854	0.859	0.864	0.862	0.857	0.852	0.828
Transformer	0.858	0.855	0.855	0.857	0.855	0.856	0.856
BERT	0.904	0.877	0.852	0.860	0.884	0.910	0.881
NSC+UPA	0.921	0.830	0.756	0.793	0.858	0.935	0.846
HUAPA	0.836	0.876	0.919	0.910	0.863	0.820	0.870
BCLSTM	0.965	0.894	0.833	0.853	0.908	0.970	0.902
LIM	0.925	0.912	0.899	0.901	0.914	0.927	0.913
UPIM	**0.930**	**0.921**	**0.911**	**0.913**	**0.922**	**0.932**	**0.922**

Table 4 Comparison of baselines on the Movies dataset

Method	Positive			Negative			Accuracy
	Precision	Recall	F1-score	Precision	Recall	F1-score	
LSTM	0.824	0.814	0.808	0.813	0.817	0.824	0.816
Transformer	0.837	0.833	0.831	0.832	0.835	0.838	0.834
BERT	0.874	0.858	0.843	0.848	0.863	0.879	0.861
NSC+UPA	0.786	0.846	0.916	0.899	0.818	0.750	0.833
HUAPA	0.861	0.839	0.818	0.827	0.847	0.868	0.843
BCLSTM	0.852	0.843	0.848	0.845	0.854	0.844	0.849
LIM	0.958	0.929	0.902	0.907	0.933	0.960	0.931
UPIM	**0.974**	**0.947**	**0.921**	**0.925**	**0.950**	**0.976**	**0.948**

The number of heads in multi-head attention mechanism is 8 and the maximum length is 250. The weights of the adjacency matrix are set as follows: the weight from the user's reviews or the product reviews to the main review is set to 10, and the reverse is set to 6; the weight is set to 3 between user's reviews and product reviews, and the weight of the edge pointing to itself is 5. We mainly use standard accuracy to measure the overall sentiment classification performance [8, 10, 32], and also use precision, recall and F1-score to further prove the effectiveness of the model. All the results are the average of five repeated experiments.

4.2 Experimental results

In this subsection, we compare the proposed models with the following baselines.

- LSTM, a basic neural model, which is applied to the twitter sentiment prediction for the first time, obtains the effectiveness result [20].
- NSC+UPA, a hierarchical LSTM with user and product joint attention to generate document representation for sentiment classification [8].
- Transformer, an attention based model [24] which shows strong performance in many NLP tasks.
- HUAPA, which models user attention and product attention separately in hierarchical BiLSTM [9].
- BERT, a pre-trained bidirectional Transformer encoder [26] which achieves the state-of-the-art performance across a variety of NLP tasks. We fine-tune the BERT model to learn the character representation for sentiment classification.
- BCLSTM, a customized classifier based on possibly multiple different known categorical metadata information (e.g., user/product information for sentiment classification) [10], which uses basis vectors to effectively incorporate categorical metadata on various parts of a neural-based model. It reports the state-of-the-art performance in review sentiment classification.
- LIM, a local interactive model with user and product information, which uses the multi-head attention mechanism to capture the local influence between the users' reviews, product reviews, and the main review.

Tables 3–5 give the results of comparison among the UPIM and several strong baselines. The results are separated into two groups: the methods only using review and the methods using both review and user/product information. From the table we have the following conclusions. (1) In general, the second part has better performance than the first part, which indicates the importance of user and product information. (2) The NSC+UPA uses the attention mechanism and obtains better results than the

Table 5 Comparison with baselines on the CDs dataset

Method	Positive			Negative			Accuracy
	Precision	Recall	F1-score	Precision	Recall	F1-score	
LSTM	0.785	0.804	0.828	0.821	0.790	0.768	0.798
Transformer	0.819	0.829	0.839	0.835	0.824	0.814	0.827
BERT	0.881	0.868	0.855	0.859	0.872	0.885	0.870
NSC+UPA	0.746	0.831	0.938	0.916	0.781	0.680	0.809
HUAPA	0.773	0.829	0.894	0.874	0.800	0.738	0.816
BCLSTM	0.860	0.915	0.887	0.909	0.851	0.879	0.883
LIM	0.948	0.926	0.905	0.909	0.929	0.951	0.928
UPIM	**0.967**	**0.943**	**0.920**	**0.924**	**0.946**	**0.969**	**0.945**

LSTM model. It shows that the attention mechanism is more effective in incorporating user and product information. However, it uses user and product joint attention to integrate user and product, so the performance of BERT is better than it. (3) The HUAPA and BCLSTM perform better than NSC+UPA because they incorporate user and product information more reasonably and effectively by using user and product information separately. Especially, the BCLSTM optimizes the category-specific weights properly for machine usage. (4) Our proposed model outperforms all of the strong baselines. The LIM model not only considers the effect of user and product information separately but also captures the local interaction information by using multi-head attention. The UPIM model has the best performance, because it captures the global interaction by constructing the user and product interactive graphs on the basis of the LIM model, which incorporates user's preferences and product characteristics more effectively. The same performance results in three datasets show the efficiency of our model.

4.3 Analysis and discussion

4.3.1 *Influence of user and product information*

In this subsection, we analyze the influence of user and product information on three datasets. Since the Transformer model has the effective performance with a multi-head attention mechanism, and graph convolutional network is a part of our model, we consider the Transformer model and GCN as the basic components for analyzing the influence of user and product information. Figures 3–5 show the results.

In Figures 3–5, d means the model only considers the review document individually. $+U$ and $+P$ denote that the model considers the user information and the product information respectively. The depth of color represents the level of performance. We discuss the local interaction of the attention mechanism and the global interaction of GCN separately. The Att+GCN is a model considering both local and global interactions.

From these tables, we can find that models considering either user information or product information perform better than models only considering review documents individually. Since the different influences of user's preferences and product characteristics for review sentiment classification, it is the best to combine both of them into the models. In addition, the multi-head attention mechanism and the graph convolutional operation are both helpful for sentiment classification, because the attention mechanism can extract the user/product-aware words and the graph convolutional operation can interact the information between the user/product and review document effectively. Specifically, the performance of the GCN+U+P model is better than the Transformer+U+P model, which indicates that considering only the attention mechanism to obtain the local limited influence is worse than the global interaction by using a graph convolutional neural network. Our proposed model UPIM learns the local interaction and global interaction collectively, and thus it outperforms all of the other models.

4.3.2 *Influence of graph structure*

In order to choose the best structure of the interactive graph, we compare the performance by changing the number of connections, that is, the number of the user/product vertex sets in the interactive graph. In order to analyze the influence of more connections on performance, we use 2000 training data, which contain more user's reviews and product reviews. The results on three datasets are illustrated in Figures 6–8.

Figure 3 (Color online) Influence of the user and product information on the Yelp dataset.

Figure 4 (Color online) Influence of the user and product information on the Movies dataset.

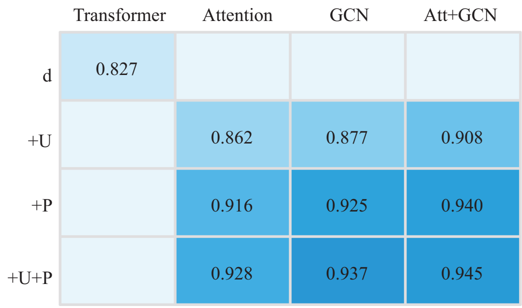

Figure 5 (Color online) Influence of the user and product information on the CDs dataset.

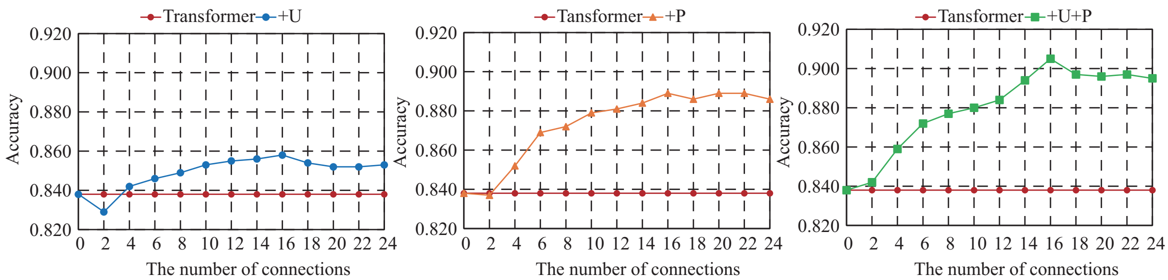

Figure 6 (Color online) Overview of user and product graph convolutional neural network on the Yelp dataset.

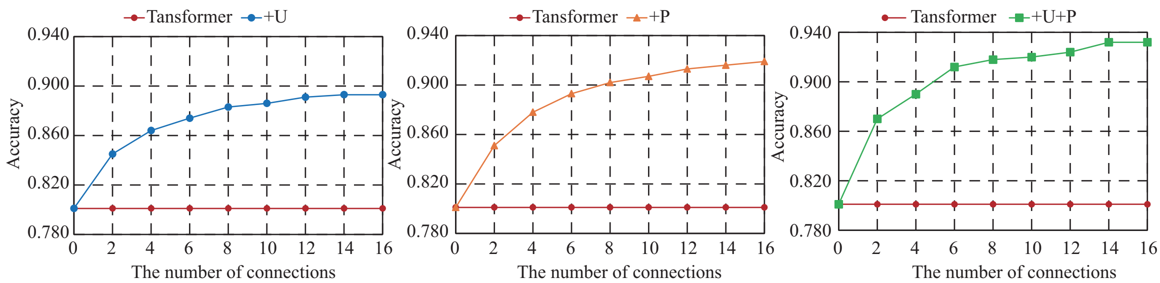

Figure 7 (Color online) Overview of user and product graph convolutional neural network on the Movies dataset.

From the figures, it is obvious that the performances are getting better with change of the number of graph edges, which indicates that in order to capture the user's preferences and product characteristics more accurately, the model needs to fully consider the user's reviews and product reviews. For the Yelp dataset, when the number is 16, the performance of the model reaches the best, and the larger graph structure makes little contribution to the performance. Therefore, 16 is the best number of user's reviews and product reviews to construct the user and product interactive graphs in the model. For the Amazon dataset, including Movies and CDs datasets, we find that the performance is basically stable since the correlation number is 12.

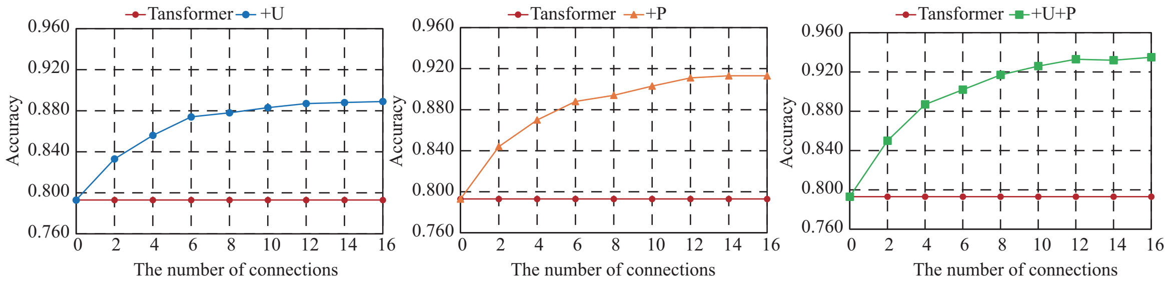

Figure 8 (Color online) Overview of user and product graph convolutional neural network on the CDs dataset.

Table 6 Examples of case study

Reiew	Transformer	BERT	BCLSTM	UPIM
[E1] This isn't a fabulous downtown development by any means, but I did enjoy walking to the movie theater when I lived down the street at Camden Copper Square. I think the "garden" aspect of the development facing Van Buren ST is pretty lovely, but this place is far from what I consider a mall or even a strip mall. This isn't the best that the downtown Phoenix offers.	Positive	Positive	Positive	Negative
[E2] Cool enough venue, not much character but great acoustics and that's what you want when you shell out more than $100 for a concert. How about dropping your fees a bit?	Negative	Negative	Positive	Positive
[E3] I was really confused when I first walked in. There was no menu. I had no idea how they were going to charge me. I didn't know. No one had explained it to me. I found out later that you pay one flat price. I made it through the process fine, got my food and enjoyed it immensely. The food was great, you get a lot for the price.	Negative	Negative	Negative	Positive

5 Case study

We choose three examples to illustrate the effectiveness of the proposed UPIM model in Table 6. In particular, we employ Transformer, BERT, and BCLSTM models to classify sentiment as a baseline model.

From Table 6, we show three examples that are not easy to analyze. Thus, our model can classify the correct sentiment based on capturing the local and global influence of users' review and product review information, while BERT and Transformer fail to classify the sentiment in all of these examples.

The three examples include both positive and negative words. If analyzing the review context individually, it is hard to understand the potential implications. Our model considers the users' preference and product characteristics from the users' history reviews and the other reviews of the product. For the first example, this user did not like eating and shopping in remote places from the user view, and most people did not like this mall from the product view, and thus, our model gives the result of the negative sentiment. The second example is about theatre, and this user thought that the acoustic was great and most users liked the seat layout, food, and location, and thus our model shows the positive sentiment by combining the user and product information. For the third example, although this user gave a confused comment when he first went to the restaurant, he had a positive review on food and other aspects. Combined with the user's historical reviews and other users' reviews on the restaurant, we get a positive sentiment conclusion. Especially, the BCLSTM model gives the correct results in E2, because it uses user and product information. However, in E1 and E3, it fails to classify the sentiment, which means that the model still does not make full use of user and product information.

6 Conclusion

In this paper, we focus on the sentiment classification of review documents, which is a critical research for the industry and customers. Most existing approaches either consider the text individually or learn the local interactive influences of user and product information. Therefore, we propose a user product interactive model. The model has three contributions. Firstly, the model generates representations of user's reviews, product reviews, and the main review document. Secondly, we employ the multi-head attention to learn the local interaction among users, products, and reviews. Thirdly, the user interactive graph and product interactive graph are constructed to learn the global interaction and deep influence of user and product information. Experimental results show the efficiency of the proposed model compared with several strong baselines.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61525205, 61702518).

References

1 Kapočiūtė-Dzikienė J, Damaševičius R, Woźniak M. Sentiment analysis of lithuanian texts using traditional and deep learning approaches. Computers, 2019, 8: 4

2 Roy A, Guria S, Halder S, et al. Summarizing opinions with sentiment analysis from multiple reviews on travel destinations. Int J Synth Emotions, 2018, 9: 111–120

3 Wan X. Co-training for cross-lingual sentiment classification. In: Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics, 2009. 235–243

4 Kiritchenko S, Zhu X, Mohammad S M. Sentiment analysis of short informal texts. J Artif Intell Res, 2014, 50: 723–762

5 Kalaivani P, Shunmuganathan K L. Feature selection based on genetic algorithm and hybrid model for sentiment polarity classification. Int J Data Mining Model Manag, 2016, 8: 315–329

6 Yang M, Tu W, Wang J, et al. Attention based LSTM for target dependent sentiment classification. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017. 5013–5014

7 Chen T, Xu R, He Y, et al. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Syst Appl, 2017, 72: 221–230

8 Chen H, Sun M, Tu C, et al. Neural sentiment classification with user and product attention. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016. 1650–1659

9 Wu Z, Dai X, Yin C, et al. Improving review representations with user attention and product attention for sentiment classification. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018. 5989–5996

10 Kim J, Amplayo R K, Lee K, et al. Categorical metadata representation for customized text classification. Trans Assoc Comput Linguist, 2019, 7: 201–215

11 Liu S H, Li F X, Li F T, et al. Adaptive co-training SVM for sentiment classification on tweets. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, 2013. 2079–2088

12 Zhang D, Xu H, Su Z, et al. Chinese comments sentiment classification based on word2vec and SVM$^{\rm perf}$. Expert Syst Appl, 2015, 42: 1857–1863

13 Song J, Kim K T, Lee B J, et al. A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. KSII TIIS, 2017, 11: 2996–3011

14 Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, 2014. 1746–1751

15 Wei X, Lin H, Yu Y, et al. Low-resource cross-domain product review sentiment classification based on a CNN with an auxiliary large-scale corpus. Algorithms, 2017, 10: 81

16 Nguyen T L, Kavuri S S, Lee M. A fuzzy convolutional neural network for text sentiment analysis. J Intell Fuzzy Syst, 2018, 35: 6025–6034

17 Liu F, Zheng J, Zheng L, et al. Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification. Neurocomputing, 2020, 371: 39–50

18 Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, 2015. 1422–1432

19 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput, 1997, 9: 1735–1780

20 Wang X, Liu Y, Sun C, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, 2015. 1343–1353

21 Qian Q, Huang M, Lei J, et al. Linguistically regularized LSTM for sentiment classification. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017. 1679–1689

22 Lei Z, Yang Y, Yang M. Sentiment lexicon enhanced attention-based LSTM for sentiment classification. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018. 8105–8106

23 Feng S, Wang Y, Liu L, et al. Attention based hierarchical LSTM network for context-aware microblog sentiment classification. World Wide Web, 2019, 22: 59–81

24 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of Advances in Neural Information Processing Systems, 2017. 5998–6008

25 Long F, Zhou K, Ou W. Sentiment analysis of text based on bidirectional LSTM with multi-head attention. IEEE Access, 2019, 7: 141960–141969

26 Devlin J, Chang M, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. 2018. ArXiv:1810.04805

27 Mathew L, Bindu V R. A review of natural language processing techniques for sentiment analysis using pre-trained models. In: Proceedings of the 4th International Conference on Computing Methodologies and Communication (ICCMC), 2020. 340–345

28 Hao Y, Dong L, Wei F, et al. Self-attention attribution: interpreting information interactions inside transformer. 2020. ArXiv:2004.11207

29 Tang D, Qin B, Liu T. Learning semantic representations of users and products for document level sentiment classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, 2015. 1014–1023

30 Gui L, Zhou Y, Xu R, et al. Learning representations from heterogeneous network for sentiment classification of product reviews. Knowl-Based Syst, 2017, 124: 34–45

31 Ma D, Li S, Zhang X, et al. Cascading multiway attentions for document-level sentiment classification. In: Proceedings of the 8th International Joint Conference on Natural Language Processing, 2017. 634–643

32 Dou Z. Capturing user and product information for document level sentiment analysis with deep memory network. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017. 521–526

33 Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. In: Proceedings of International Conference on Learning Representations, 2017