IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 16, 2020, accepted December 24, 2020, date of publication December 29, 2020,
date of current version January 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047947

Detecting Fake Reviews Using Multidimensional Representations With Fine-Grained Aspects Plan

MEILING LIU[1], (Member, IEEE), YUE SHANG[1], QI YUE[1], AND JIYUN ZHOU[2]
[1]School of Information and Computer Engineering, Northeast Forestry University, Harbin 150006, China
[2]Lieber Institute, Johns Hopkins University, Baltimore, MD 21218, USA

Corresponding author: Qi Yue (yueqi@nefu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702091, and in part by the Fundamental Research Funds for the Central Universities under Grant 2572018BH06.

ABSTRACT Due to the rapid growth of network data, the authenticity and reliability of network information have become increasingly important and have presented challenges. Most of the methods for fake review detection start with textual features and behavioral features. However, they are time-consuming and easily detected by fraudulent users. Although most of the existing neural network-based methods address the problems presented by the complex semantics of reviews, they do not account for the implicit patterns among users, reviews, and products; additionally, they do not consider the usefulness of information regarding fine-grained aspects in identifying fake reviews. In this paper, we propose an attention-based multilevel interactive neural network model with aspect constraints that mines the multilevel implicit expression mode of reviews and integrates four dimensions, namely, users, review texts, products and fine-grained aspects, into review representations. We model the relationships between users and products and use these relationships as a regularization term to redefine the model's objective function. The experimental results from three public datasets show that the model that we propose is superior to the state-of-the-art methods; thus showing the effectiveness and portability of our model.

INDEX TERMS Fake reviews detection, multidimensional representations, relationship modeling, fine-grained aspects.

I. INTRODUCTION

Currently, the Internet is not only a tool by which people acquire knowledge but also a platform on which people can express their views and disseminate information. In the realm of e-commerce, review information has a significant impact on both users' purchasing decisions [1] and enterprises' development on online platforms. According to the latest data from the social commerce platform Bazaarvoice, more than 50% of users discontinue their purchasing behavior and lose trust in brands after discovering fake reviews of a product. Fake reviews may not only damage the entire online review system, but ultimately cause a loss of credibility [2]. Therefore, it is important to automatically identify fake reviews on online platforms and provide users with more truthful information.

Fake review detection was first named specifically as the opinion spam detection by Jindal and Liu [3]. Due to the important research implications of this work, a number of fake review detection methods have been proposed during recent years. The early work on this subject focused on manual design features in combination with machine learning methods. For example, semantic features of the text include the length of the review text [4], [5], its lexical features [6], and its affective polarity [7]. Users' behavioral features include the number of good or bad reviews that they publish [4] and the frequency of these reviews [8]. Driven by profits, spammers are enhancing and disguising their schemes in accordance with the corresponding detection methods. During recent years, along with the development of deep learning, a number of fake review detection methods based on deep learning have been developed [9]–[11], [19], [20]. Compared to feature-based methods, these methods have a greater ability to automatically capture semantic information implicit within text without a manual design and have a stronger domain adaptable and effective.

The existing methods have achieved good results, but most of them are only from a single perspective, such as that of review texts or users; additionally, they ignore some user implicit expression patterns and the influences among users,

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi.

TABLE 1. Examples of data analysis of Yelp reviews. Examples of fine-grained aspects in reviews are marked in red and the general items are marked in blue.

Prod	Reviews	Rating	Label	User
prod3	The Spicy Thai Chicken Sausage with Sesame Seaweed...	5	TRUE	user1
prod1	..the French Toast was covered in sliced pears...	4	TRUE	
	Classic.Breakfast.Cool space.Cool staff. Totally cool...	4	FAKE	user2
prod2	..something very basic (eggs & chicken sausage)...	2	TRUE	user1
	...The food stinks and the place was a mess...	1	FAKE	user2
prod4	...Food was totally average and the place smelled...	2	FAKE	

products and texts [11]. In addition, we find when users express their true feelings, whether their reviews are positive or negative, their descriptions will include some details (such as the taste of a dish in a resturant) that enhance their emotional expression. Their expression are far more descriptive. However, a spammer cannot describe a product in detail because he or she is not describing a personal experience or an actual use. Fine-grained aspect is a set of terms used to describe a topic in a related domain, which can be the features of a product or attributes of a service [12], that is the ''details'' mationed above. Thus, we assume that fine-grained aspects can be used as a plan to detect fake reviews.

To illustrate this issue, we analyze real data derived from the yelpCHI dataset. As shown in Table 1, if a review's rating is greater than 3, this means that the review is positive. A label of ''FAKE'' denotes a fake review. The real user, ''user1'', comments on fine-grained aspects (i.e., French toast and chicken sausage) regardless of whether the review is positive or negative. The spammer, ''user2'', regardless of whether he or she is leaving a positive or negative review, provides a general evaluation (i.e.,food, place).

To generate multidimensional dense sentence representations containing information regarding users, products, texts, and fine-grained aspects, we design a Multilevel Interactive Attention neural Network with Aspect plan (MIANA). There are two levels in our MIANA model: a Word-level Fusion Module (WFM) and a Sentence-level Interactive Attention Neural network module (SIAN). In the WFM, we obtain user (product) sentence representation that contains unique user (product)-related patterns and aspects-related sentence representation from the words of reviews. In the SIAN, first, the output of the WFM is orthogonally decomposed to obtain user (product)-related sentence representations. Then, the original review sentence representation and the aspects-related sentence representation are across activated by gate mechanism. To encode the relationship between users and products, we treat users and products as entities and reviews as the relationships between these entities. In accordance with the TransD [13], the relationship between these three items is modeled as a regularization term to redefine the model's objective function and is incorporated into the model.

In summary, our contribution is threefold:

1) We propose a new scheme to detect fake reviews using fine-grained aspects. In order to verify our scheme, we propose an attention-based multilevel interactive neural network model with fine-grained aspect constraints for fake review detection; this model can produce multidimensional dense sentence representations that incorporate user expression patterns, product fine-grained attributes, and contextual semantic information at the word and sentence levels.

2) We model the relationship between users, review texts and products, use it as a regularization term to optimize the model's objective function, and incorporate the implicit relationship into the model.

3) The experimental results with three public datasets are significantly better than those of the state-of-the-art methods, which demonstrates the usefulness and portability of the MIANA model for the identification of fake reviews.

The rest of this paper is structured as follows. Section 2 discusses related work. Section 3 illustrates the internal structure and calculation process of our proposed model. Section 4 give a summary on the datesets and describes experiment detils, then extensive experiments are presented to justify the effectiveness of our proposals. Section 5 presents the results and discussion and finally section 6 concludes this work and future direction.

II. RELATED WORK

Since the review spam detection task proposed, the early research, based on feature engineering and machine learning, has mainly focused on the analysis of user behavioral features, structural features and text semantic features.

After analyzing reviews and users on Amazon.com, Jindal and Liu [14] classified spam reviews into three categories: untruthful opinions, reviews of brands only, and nonreviews such as advertisements. Additionally, they proposed a total of 36 text-centric, user-centric, and product-centric features that could be combined with logistic regression methods to identify spam reviews. Li *et al.* [6] combined semisupervised machine learning methods to identify fake reviews based on multiple text- and user-related features and analyzed the impact of each feature. Li *et al.* [15] identified the differences in language usage between truthful and fake reviews. Wang *et al.* [16] performed a tensor decomposition of 11 relationships that exist between users and products based on reviews and classified them according to the SVM model.

Melleng *et al.* [17] combined sentiments and emotions to form review representations, and used these representations to identify fake reviews. In the study of [18], a rule-based feature-weighting scheme is proposed that combines various features of reviews, reviewers and products.

Although feature engineering can obtain great results in spam review detection, it cannot characterize global semantic information, which limits its detection ability [9]. It is easily detected by spammers. Inspired by the excellent performance of deep learning methods in the field of NLP, Ren and Zhang [9] applied the CNN to fake review detection for the first time to obtain dense high-dimensional representations, which verified the superiority of their method. Wang *et al.* [10] incorporated text features and behavioral features into sentence representations using the CNN model to solve the problem of cold start in review spam detection. Yuan *et al.* [11] used a hierarchical fusion attention mechanism to model the relationship between users, products, and reviews and generated fused text representations of users and products. In the study of [19], word vectors are combined with three emotional expression features to form sentence representations, this method relied on a multilayer perceptron neural network with two hidden layers for classification.

Although the previous methods have achieved good detection results, most of them are based on the textual information or behavioral information contained in reviews. The work of Yuan *et al.* [11] has solved the potential patterns among users, products, and reviews, but some useful contextual information is lost during the calculation. Furthermore, the fine-grained aspects that contain product attributes can be seen as a plan for identifying fake reviews, as these attributes are rarely evaluated by spammers but are described by real users. The existing studies ignore the impact of aspect-level information.

In this paper, to enable the integration of user-level expression patterns, textual context semantic information, and fine-grained product attributes into review representations from a global perspective, under the constraints of fine-grained aspect information, we propose the MIANA model. The relationship between users and products is modeled based on TransD, and this relationship is employed as a regularization term to optimize the fake review detection model, integrated into this model, and used to enhance its performance.

III. PROPOSED MODEL

In this section, we provide a detailed description of our proposed MIANA model. The structure of MIANA is shown in Fig. 1.

Let us first define some notations. Every review $S = \{S_1, S_2, \ldots, S_\ell\}$ contain ℓ sentences. Each of these sentences $S_\ell = [word_1, word_2, \ldots, word_m]$ contains m words. Reviews in all datasets are mined for fine-grained aspects $A = (A_1, A_2, \ldots, A_n)$. Given the user ID U_x, the product ID P_x, the review S_x, and the fine-grained aspects $(A_1, A_2, \ldots, A_x)$ involved in S_x, our goal is to determine if S_x is a fake review.

A. FINE-GRAINED ASPECT EXTRACTION

We extract the fine-grained aspects regarding the things that users care about from the reviews. As we defined in the section 1, fine-grained aspects are the product attributes contained in user reviews. For example, *"Awesome food. I came here with friends on a Friday night and we were seated outside, where they have a cute eating area with lights and umbrella tables. We ordered the Papa Rellena to start, which is a* ***potato*** *stuffed with ground* ***beef****. It was delicious- the outside of the* ***potato*** *was crispy, I'm assuming fried, and inside it was like a mashed* ***potato*** *with the* ***beef****…"*, in this review, "potato", "beef", etc. refer to fine-grained aspects about the food in this restaurant.

We adop the method proposed by Zhang *et al.* [20] to build a Fine-grained Aspect lexicon $\bar{A}$ from the whole review datasets, which contains a total of 1224 words. **Note that**, getting the fine-grained aspects is not our focus. There are many methods to obtain aspects (e.g. [21]–[23]). Instead, assuming fine-grained aspects are available, we use aspects as a plan to propose a new method of fake review detection. Our problem is to verify whether the fine-grained aspect can be used as an effective solution to detect fake reviews.

B. WORD-LEVEL FUSION MODULE

As shown in Figure 1, based on the method of Yuan *et al.* [11], we designed a word-level fusion module (WFM) to fuse the hidden features of users (products) with their corresponding original review texts and to identify the hidden patterns of texts related to users (products) from a global perspective, and to obtain aspects-related information in the sentence. In order to reduce the impact of word segmentation on the subsequent calculations, a word-related domain k is used. For each user, we use the attention mechanism to calculate on the $(word_{i-k}, word_{i+k})$ to obtain a representation of the user-related features $V_{uc} \in \mathbb{X}^{1\times d}$ at the word level as follows:

$$\begin{aligned} V_{uc} &= \sum_{i=1}^{2k+1} \alpha_i X, \\ \alpha_i &= \frac{\exp(u_i)}{\sum_{n=1}^{2k+1} \exp(u_n)}, \\ u &= \tanh\left(XW_x + U_c W_u\right), \end{aligned} \tag{1}$$

where X is the embedding of $(word_{i-k}, word_{i+k})$, α_i is the score function that determines the importance of user-level words in an entire sentence, W_x, W_u are transformation matrices, and $U_c \in \mathbb{R}^{(2k+1)\times d}$ is $2k+1$ copies of the user embedding U_c. The representations V_{pc} of the product-related features are generated in the same way.

In order to obtain the fine-grained aspects information in each review, we determine if each word in S_ℓ contains fine-grained aspects based on certain simple rules: we perform lemmatization on $word_i, i \in (0, m)$, and determine if $word_i \in \bar{A}$. The attention mechanism is used to calculate the suspiciousness of each fine-grained aspect V_{aspect}

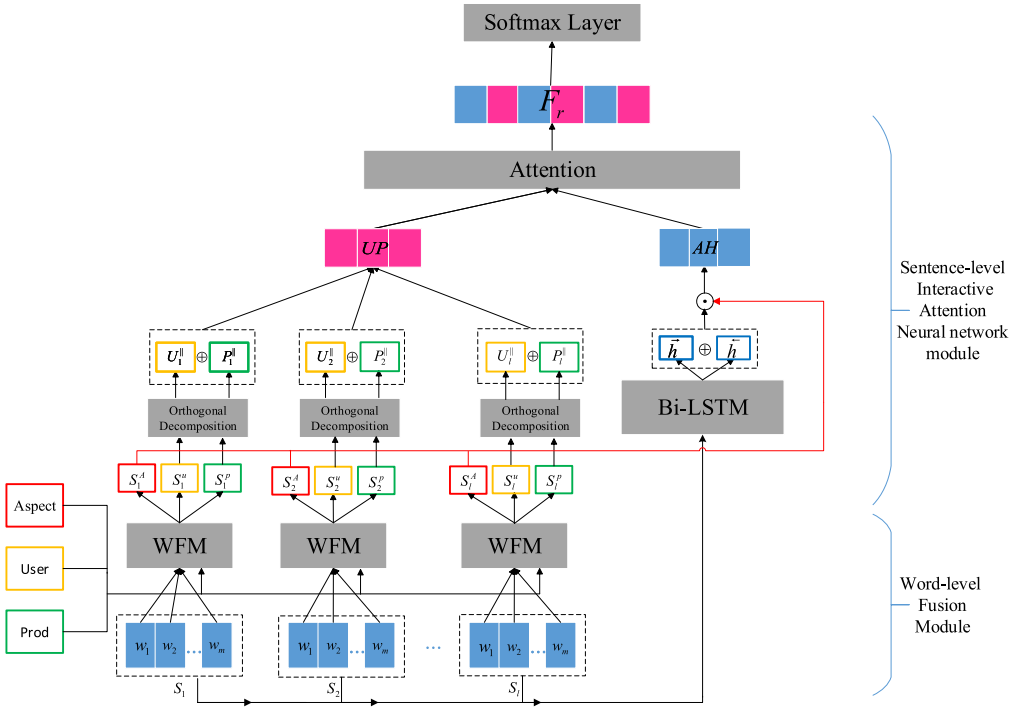


FIGURE 1. MIANA model. *Aspect* denotes fine-grained aspects, *User* is the embedding of users who posted the corresponding reviews, and *Prod* is the embedding of the reviewed products.

in the reviews as:

$$\begin{aligned} V_{\text{aspect}} &= \beta_t X_t, \\ \beta_t &= \text{softmax}\left(ap_t\right), \\ ap_t &= \tanh\left(X_t W_x + A_t W_u\right), \end{aligned} \tag{2}$$

where X_t is the embedding of $word_i$, A_t is the embedding of aspects in the S_ℓ, and β_t is the score function.

Therefore, we obtain user-related feature representations V_{uc}, product-related feature representations V_{pc}, and aspect-related feature representations V_{aspect}.

C. SENTENCE-LEVEL INTERACTIVE ATTENTION NEURAL NETWORK MODULE

1) GENERATE RELATED KNOWLEDGE MATRIX

As shown in Figure 1, in the first step of SIAN, we further process the output of the WFM. We concatenate the embedding of V_{uc} and $word_i$ to obtain their user-relevant word representations, and then we concatenate all the user-relevant word representations together to obtain the user-level sentence matrix: $V_u = [word_1 \oplus V_{u1}, word_2 \oplus V_{u2}, \ldots, word_m \oplus V_{um}]$, where $V_u \in \mathbb{R}^{m \times 2d}$. Then, we select the features that contain the most information in the current sentence and transform them into user-relevant sentence representations S_c^u as follows:

$$\begin{aligned} S_u &= \tanh\left(V_u W_v + b\right), \\ S_c^u &= \max_{dim=1}\left(S_u\right), \end{aligned} \tag{3}$$

where $W_v \in \mathbb{R}^{2d \times d}$ is the transformation matrix. User-relevant sentence representations $\left[S_1^u, S_2^u, \ldots, S_\ell^u\right]$ contain user-related information as well as some additional information that is not relevant to users. To obtain pure user sentence representations $U_c^{\|}$ (including only user-related information), we use orthogonal decomposition to decompose S_c^u according to user embedding parallel direction decomposition:

$$U_c^{\|} = \frac{S_c^u U_e^T}{U_e U_e^T} U_e. \tag{4}$$

Therefore, we obtain the user representation User $= \left[U_1^{\|}, U_2^{\|}, \ldots, U_\ell^{\|}\right]$, where User $\in \mathbb{R}^{\ell \times d}$. The product representation Prod $= \left[P_1^{\|}, P_2^{\|}, \ldots, P_\ell^{\|}\right]$, which contains information about products only, is obtained in the same way. No orthogonal decomposition of aspect-level sentence representations $A = \left[S_1^A, S_2^A, \ldots, S_\ell^A\right]$ is done, $A \in \mathbb{R}^{\ell \times d}$. User, Prod and A can be viewed as global perspective features for fake review detection. The user sentence representation User and the product sentence representation Prod are spliced together to create a related knowledge matrix UP $= \left[U_1^{\|} \oplus P_1^{\|}, U_2^{\|} \oplus P_2^{\|}, \ldots, U_\ell^{\|} \oplus P_\ell^{\|}\right]$.

2) GENERATE ORIGINAL TEXT INFORMATION

User, Prod, and A contain a wealth of information focused on different points, but the review text itself contains rich semantic information, which maybe ignored in them (Proved in Section 5). We directly use the embedding of the original review sentence set $S = \{S_1, S_2, \ldots, S_\ell\}$ as the input of bi-directional LSTM (Bi-LSTM) during the second step. The output of Bi-LSTM, namely, its forward hidden layer $\overrightarrow{h}$ and backward hidden layer $\overleftarrow{h}$ are concatenated to obtain the original review sentence representation $h \in \mathbb{R}^{1 \times 2n}$, n is the

hiddden size of Bi-LSTM:

$$E_s = \text{Embedding}(S),$$
$$h = \text{Bi-LSTM}\,(E_s) = \overrightarrow{h} \oplus \overleftarrow{h}. \tag{5}$$

3) MUTUAL ACTIVATION BETWEEN TEXT INFORMATION AND ASPECT INFORMATION

During the third step, in order to further extract and integrate information, the gated mechanism is applied on $\tilde{h}$ and $\tilde{A} \in \mathbb{R}^{\ell \times d}$ to cross activate each other.

$$\begin{aligned}\tilde{h} &= hW_1 + b_1,\\ \tilde{A} &= AW_2 + b_2,\\ h' &= \tilde{h} \odot sigmoid(\tilde{A}),\\ A' &= \tilde{A} \odot sigmoid(\tilde{h}),\end{aligned} \tag{6}$$

Then we concatenate h' and A' to get the compound information representation $AH \in \mathbb{R}^{2\ell \times d}$.

4) FUSION INFORMATION

The last step is to use the attention mechanism to calculate the interaction between UP and AH. In the context of fake review detection, this is done by determining the number of suspicious features in the user-product information given the compound information; additionally, the number of suspicious features in the text and aspect information is determined given the user-product features. The softmax layer is used to normalize the score between user-product information and compound information to obtain the attention weight ${\partial_{\text{up}}}^i$ for user-product representation UP_i as follows:

$$\begin{aligned} q &= \frac{1}{2\ell}\sum_{i=1}^{2\ell} \text{AH}^i,\\ \gamma_i &= \tanh\left(\text{UP}_i^T W_{\partial 1} q\right),\\ {\partial_{\text{up}}}^i &= \frac{\exp(\gamma_i)}{\sum_{i=1}^{2\ell}\exp(\gamma_i)},\end{aligned} \tag{7}$$

where q is the average value of AH, γ_i is the value of association between UP_i and AH, and $W_{\partial 1} \in \mathbb{R}^{d\times d}$ is the transformation matrice. The attention weight ${\partial_{ah}}^i$ for AH_i is generated in the same way. Next, we multiply UP_i with ${\partial_{up}}^i$ and AH_i with ${\partial_{ah}}^i$ to get the weighted representations UP_f and AH_f:

$$\begin{aligned}\text{UP}_f &= \sum_{i=1}^{2\ell} {\partial_{up}}^i \text{UP}_i,\\ \text{AH}_f &= \sum_{i=1}^{2\ell} {\partial_{ah}}^i \text{AH}_i.\end{aligned} \tag{8}$$

Finally, we join UP_f and AH_f together to get the final review representation F_r as:

$$F_{\text{r}} = \text{UP}_f W_{f1} + \text{AH}_f W_{f2}, \tag{9}$$

where $W_{f1} \in \mathbb{R}^{d\times d}$ and $W_{f2} \in \mathbb{R}^{d\times d}$ are transformation matrices. F_r contains rich features from the four perspectives.

D. CLASSIFICATION AND REGULARIZATION

We regard the F_r as the review feature for the fake review detection task, convert it into a nonlinear layer in the vector, and finally use the softmax layer to determine and calculate the falseness of the corresponding review as:

$$\begin{aligned}\mathbf{y} &= \tanh\left(F_r W_r + b_r\right),\\ p\left(c_j \mid \theta\right) &= \frac{\exp\left(\mathbf{y_j}\right)}{\sum_{j=1}^{n_0}\mathbf{y_j}},\end{aligned} \tag{10}$$

where c_j is the predicted category, n_0 is the classified category, and θ is all the conversion parameters mentioned above.

As mentioned in Section 2, different from Yuan *et al.* [11], we coded the relationship between the user, the product, and the review text based on TransD and as regularization terms to optimize the objective function of our model. If a user and a product are viewed as head and tail entity respectively, the corresponding review text can be viewed as the relationship between these two entities, creating a triad in which there are not only one-to-many and many-to-one relationships, but the head entity and the tail entity should be mapped to different vector spaces, because the head entity user and the tail entity product belong to different categories

Therefore, we perform a knowledge representation of triples based on TransD, map the different entity properties to different matrices, and construct two projection matrices, namely, M_{ru} and M_{rp}, to map users and products from the entity space to the relationship space as follows:

$$\begin{aligned}M_{ru} &= \underset{dim=1}{mean}\left(\text{User}'\right)^T F_r' + I^{d\times d},\\ M_{rp} &= \underset{dim=1}{mean}\left(\text{Prod}'\right)^T F_r' + I^{d\times d},\end{aligned} \tag{11}$$

where User′, Prod′ and F_r' represent the projection vector of User, Prod and F_r, respectively. $I^{d\times d}$ is the identity matrix that initializes the projection matrix. The distance between a user and a product in the relationship space is calculated as:

$$\ell_j(u,p) = \left\|M_{ru}\text{User} + F_r - M_{rp}\text{Prod}\right\|_2^2. \tag{12}$$

On this basis, the tail entity product is negatively sampled to obtain the negative sampling distance $\ell_j(u,p)'$. The difference $L_j(u,p)$ between $\ell_j(u,p)$ and $\ell_j(u,p)'$ is viewed as the loss of the triple relationship and Margin Ranking Loss is used to optimize it. $L(u,p)$ is considered as the regularization term of our model.

Finally, regarding the model training, the goal is to minimize cross-entropy loss after optimization as follow, λ is a hyperparameter:

$$\mathcal{L}(\theta) = -\sum_i \log\left(c_i \mid \theta\right) + \lambda\sum_{j=1}^{M} L_j(u,p). \tag{13}$$

IV. EXPERIMENTS

A. DATASETS AND THE EVALUATION METRICS

To evaluate the effectiveness of our model, we conducted experiments with three public datasets: YelpChi [4] contains real business reviews of restaurants and hotels from Chicago on the Yelp website. YelpNYC and YelpZIP [5] contain

TABLE 2. Statistical information of the datasets.

	YelpChi		YelpNYC		YelpZIP	
	Non-Fake	Fake	Non-Fake	Fake	Non-Fake	Fake
Reviews	58476	8919	322167	36885	528132	80466
%All	86.70%	13.20%	89.70%	10.30%	87.00%	13.00%
average length	170	120	141	95	140	102
Rating$\geq$ 4	74.29%	70.34%	77.18%	75.24%	73.94%	70.05%
Rating<4	25.71%	29.66%	22.82%	24.76%	26.06%	29.95%
Users	30325	7738	131721	28504	198045	62232

only restautants reviews from different regions of the Yelp. Table 2 shows the statistical information related to the datasets. Skewness can be observed between the fake and the actual reviews. After our analysis of the review data in the dataset and the frequency of text words, we find that because real reviews have a high probability of describing details, the average length of the actual reviews are longer than the fake reviews. There are no significant differences in sentiment between the fake reviews and the actual reviews when they are examined with sentence-level sentiment analysis.

Evaluation metrics: for the unbalanced datasets, we employ average precision (AP) and area under the curve (AUC) as evaluation metrics.

B. BASELINES

To illustrate the effectiveness of the proposed method, we selected several advanced methods to be used for comparison, including feature engineering methods and deep learning algorithms.

TensorD [16]: It automatically generates 11 relationships based on two basic rules from the perspectives of users and products. This method uses tensor decomposition to map users and products to a vector space and uses SVM to classify the embedding of the reviews.

SAE [17]: It represents reviews based on a combination of emotion and sentiment, using three sentiment dictionaries and an emotion analysis API that combines sentiments and emotional features to create review representations while using a random forest algorithm for fake review detection.

SPR2EP [24]: It is a semisupervised fake review detection framework. The Node2vec algorithm is used to represent users and products in a vector, and Doc2vec is used to represent review texts. These two algorithms are combined to identify fake reviews.

ABNN [25]: It is an attention-based neural network that uses MLP to identify user behavior features and CNN to identify textual language features, combining these two based on attention to identify review spam.

HFAN [11]: It is a hierarchical fusion attention that combines users, reviews and products to obtain a review representation that classifies reviews.

DFFNN [19]: It is a deep feedforward neural network that combines bag-of-words, n-gram features of comment text, word embedding, and multiple emotion indicators to create representations.

TABLE 3. Review-only classification results.

Dataset	yelpChi		yelpNYC		yelpZIP	
Model	AP	AUC	AP	AUC	AP	AUC
Bi-LSTM	30.21	56.20	24.70	54.65	**30.24**	55.63
CNN	27.63	57.18	19.79	52.98	28.99	55.07
RCNN	**31.27**	**62.40**	**30.55**	**62.91**	28.63	**60.61**
LSTMATT	29.26	57.67	22.35	54.04	29.63	56.27

C. DEVELOPMENT EXPERIMENTS

To select the underlying framework of the SIAN, we use multiple neural networks to classify the text of the raw reviews. Four neural networks, namely, the LSTM+attention (LSTMATT) framework, Bi-LSTM, CNN, and RCNN (Bi-LSTM+max-pooling layer) were used. The classification results are shown in Table 3. Note that during this experiment, only the pretrained embedding of the preprocessed original reviews is used as the model input. As seen from the results (the best results are marked in red), Bi-LSTM and RCNN are better at classifying when only reviews are employed. Therefore, we use these two neural networks as the basic framework of the SIAN for model building. Through experiments, we found that MIANA based on Bi-LSTM and MIANA based on RCNN perform equally well, due to space limitations, we only analyzes the MIANA based on Bi-LSTM model in this paper.

D. IMPLEMENTATION DETAIL

We use Pytorch to implement our model and most of preprocessing procedure is the same as previous works.

After analyzing the data, the maximum sentence length is set to 200. We employ word2vec[1] to pretrain the word vector with the embedding dimension set to 300. In the WFM, the word-related domain k is set to 3; in the SIAN, the hidden layer number of Bi-LSTM is set to 2 and the hidden layer dimension n is set to 150. The regularization coefficient λ of MIANA is set to 0.5, dropout rate is set to 0.5. The Adam [26] algorithm is used to train the model, and the learning rate is initialized to e^{-3}. The other hyperparameters in the model are determined through experiments on the training set and the validation set.

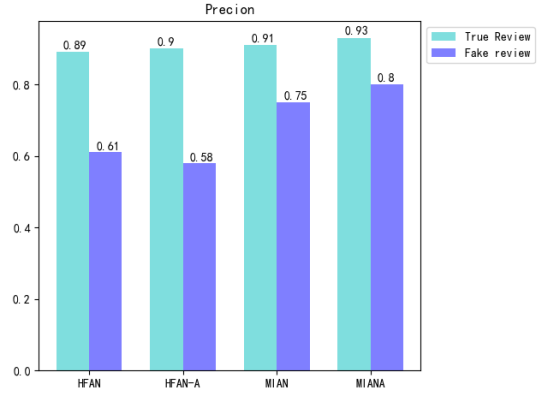
E. RESULTS AND ANALYSIS

The experimental results are shown in Table 4 below, and the precion and recall comparisions are shown in Figures 2a and 2b, from which we can make observations as follows.

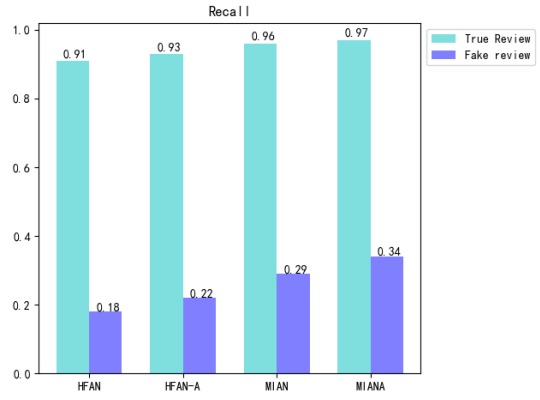
[1] https://code.google.com/p/word2vec/

TABLE 4. Experimental results.

Dataset	yelpChi		yelpNYC		yelpZIP	
Model	AP	AUC	AP	AUC	AP	AUC
TensorD	35.65	77.86	36.47	79.05	48.04	80.97
SAE	32.89	76.27	33.69	77.46	38.87	79.33
SPR2EP	33.51	80.71	32.02	81.31	42.28	83.28
ABNN	34.48	78.62	36.23	78.86	48.36	80.74
DFFNN	35.03	78.76	35.87	79.43	49.29	81.57
HFAN	49.26	83.24	54.48	84.96	63.13	87.63
HFAN-A	51.75	84.80	56.58	85.01	65.20	88.11
MIAN	52.71	85.34	63.83	90.83	70.11	92.64
MIANA	**53.22**	**86.65**	**64.27**	**91.89**	**71.82**	**93.26**



(a) Precion Comparision



(b) Recall Comparision

FIGURE 2. Precion and Recall Comparision of Different Models on the YelpChi Dataset.

Our proposed models, namely, 1) MIAN without aspect plan and 2) MIANA, both achieve better classification results than other models. On the three datasets, compared to the HFAN, the APs increased by approximately 4%, 9%, and 8%, respectively, and the AUCs increased by approximately 3%, 7%, and 6%, respectively.

1) MULTIDIMENSIONAL REPRESENTATIONS

From results of Table 4, in contrast to the feature engineering approach (TensorD, SAE, SPR2EP), the effectiveness of the deep learning model is illustrated. Compared to the base model (Table 3), which is only applicable when classifying textual information, the improvement is even more obvious. This shows that integrating users' information regarding expression mode, context semantics and fine-grained attributes, as well as the related product information and the relationship between these factors, into a model to represent reviews is very helpful for the detection of fake reviews.

2) ASPECT PLAN

HFAN-A is the HFAN [10] with aspect plan. We use aspect information as a gate mechanism to further constrain the process of user-related representation and product-related representation. MIAN is our final model MIANA without aspect plan, this means that in Formula 6-9 we use the original review sentence representation h and the user-product information UP to perform attention calculations and generate the final review representation. In the Figure 2a and 2b, compared with HFAN and MIAN, the results of fake review of HFAN-A and MIANA improve the recall rate while ensuring precion. This shows that the fine-grained aspects information can distinguish between true reviews and fake reviews. The experimental results of HFAN-A and MIANA represent an improvement upon those of HFAN and MIAN (In Table 4), showing that fine-grained aspect information is falsely discriminatory in the context of reviews and confirming our assumption in this paper that fine-grained aspects can used as a plan to identify fake reviews.

3) CONTEXTUAL SEMANTIC INFORMATION

By comparing the results of HFAN and MIAN, AP and AUC values are improved on the three datasets, proving our conjecture in Section 3.3, that is, some useful contextual semantic information will be ignored during the generation of user-related and product-related sentences. From result of Figure 2a and 2b, the precion and the recall of MIAN's fake reviews have been improved compared to HFAN,HFAN-A, which further shows that the contextual information in review text contains useful information for classification. It is necessary to combine user-product information with contextual text information when identifying fake reviews.

For the task of detecting fake reviews, the cost of identifying fake reviews as true is much greater than the cost of identifying true reviews as fake. Therefore, more attention should be paid to the recall rate. Referring to Figure 2a and 2b, MIANA gets 19%, 16% Recall over the other modoles for the true reviews and the fake reviews on the YelpChi dataset. The above results demonstrate the effectiveness and transferability of the method proposed in this paper for the task of fake review detection.

V. CONCLUSION

In this study, we focused on the task of identifying spam reviews. After analyzing the reviews in the datasets, we propose a hypothesis that fine-grained aspect information can be used as a new scheme for fake review detection and reconstructed the representation of reviews from four perspectives: users, products, reviews text, and fine-grained aspects. We proposed a multilevel interactive attention neural network

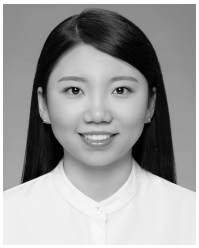
model with aspect plan; to optimize the model's objective function, we transformed the implicit relationship between users, reviews and products into a regularization term. To verify the effectiveness of the MIANA, we conducted extensive experiments on three public datasets. Our experiments showed that the classification effect has been significantly improved, that the MIANA outperforms the state-of-the-art methods for fake review detection tasks, and proved the effectiveness and feasibility of our proposed scheme.

In this paper, the fine-grained aspect terms are for restaurants and hotels. When it comes to cross-domain issues, you only need to further obtain fine-grained aspects in the relevant domain. This is the current limitation of our proposed method, and it is also the content of our future research. Our futher work includes: (a) validate the performance of our proposed method on cross-domain datasets, (b) build a joint model that can automatically extract fine-grained aspects and identify fake reviews.

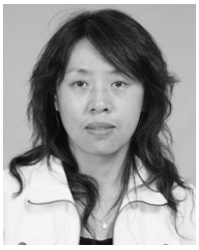
REFERENCES

[1] R. Filieri and F. McLeay, "E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews," *J. Travel Res.*, vol. 53, no. 1, pp. 44–57, Jan. 2014.
[2] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making," *Ind. Marketing Manage.*, vol. 90, pp. 523–537, Oct. 2020.
[3] N. Jindal and B. Liu, "Review spam detection," in *Proc. 16th Int. Conf. World Wide Web*, 2007, pp. 1189–1190
[4] A. Mukherjee, V. Venkataraman, B. Liu, and N. S. Glance, "What yelp fake review filter might be doing," in *Proc. ICWSM*, 2013, pp. 409–418.
[5] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 985–994.
[6] F. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," in *Proc. IJCAI 22nd Int. Joint Conf. Artif. Intell.*, vol. 3, 2011, pp. 2488–2493.
[7] X. Hu, J. Tang, H. Gao, and H. Liu, "Social spammer detection with sentiment information," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 180–189.
[8] S. Kc and A. Mukherjee, "On the temporal dynamics of opinion spamming: Case studies on yelp," in *Proc. 25th Int. Conf. World Wide Web*, Apr. 2016, pp. 369–379.
[9] Y. Ren and Y. Zhang, "Deceptive opinion spam detection using neuralnetwork," in *Proc. 26th Int. Conf. Comput. Linguistics, Tech. Papers COLING*, Dec. 2016, pp. 140–150.
[10] X. Wang, K. Liu, and J. Zhao, "Handling cold-start problem in review spam detection by jointly embedding texts and behaviors," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics (Long Papers)*, vol. 1, 2017, pp. 366–376. [Online]. Available: https://www.aclweb.org/anthology/P17-1034.pdf
[11] C. Yuan, W. Zhou, Q. Ma, S. Lv, J. Han, and S. Hu, "Learning review representations from user and product level information for spam detection," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2019, pp. 1444–1449.
[12] Y. Lu, M. Castellanos, U. Dayal, and C. Zhai, "Automatic construction of a context-aware sentiment lexicon: An optimization approach," in *Proc. 20th Int. Conf. World Wide Web - WWW*, 2011, pp. 347–356.
[13] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process. (Long Papers)*, vol. 1, 2015, pp. 687–696. [Online]. Available: https://www.aclweb.org/anthology/P15-1067.pdf
[14] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proc. Int. Conf. Web Search Web Data Mining WSDM*, 2008, pp. 219–230.
[15] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Long Papers)*, vol. 1, 2014, pp. 1566–1576. [Online]. Available: https://www.aclweb.org/anthology/P14-1147.pdf
[16] X. Wang, K. Liu, S. He, and J. Zhao, "Learning to represent review with tensor decomposition for spam detection," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 866–875.
[17] A. Melleng, A. Jurek-Loughrey, and P. Deepak, "Sentiment and emotion based text representation for fake reviews detection," in *Proc. Int. Conf. Recent Adv. Natural Lang. Process. (RANLP)*, Oct. 2019, pp. 750–757.
[18] M. Z. Asghar, A. Ullah, S. Ahmad, and A. Khan, "Opinion spam detection framework using hybrid classification scheme," *Soft Comput.*, vol. 24, no. 5, pp. 3475–3498, Mar. 2020.
[19] P. Hajek, A. Barushka, and M. Munk, "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining," *Neural Comput. Appl.*, vol. 32, no. 23, pp. 17259–17274, Dec. 2020.
[20] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2014, pp. 83–92.
[21] Y. Jo and A. H. Oh, "Aspect and sentiment unification model for online review analysis," in *Proc. 4th ACM Int. Conf. Web Search Data Mining - WSDM*, 2011, pp. 815–824.
[22] Y. Zhang, H. Zhang, M. Zhang, Y. Liu, and S. Ma, "Do users rate or review?: Boost phrase-level sentiment labeling with review-level sentiment classification," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2014, pp. 1027–1030.
[23] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl.-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016.
[24] C. M. Yilmaz and A. O. Durahim, "SPR2EP: A semi-supervised spam review detection framework," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 306–313.
[25] X. Wang, K. Liu, and J. Zhao, "Detecting deceptive review spam via attention-based neural networks," in *Proc. Nat. CCF Conf. Natural Lang. Process. Chin. Comput.*, 2017, pp. 866–876.
[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15. [Online]. Available: https://arxiv.org/pdf/1412.6980.pdf

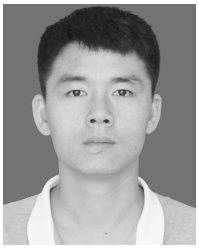
MEILING LIU (Member, IEEE) received the B.S. and M.S. degrees in computer science and technology from Northeast Forestry University, Harbin, China, in 2006, and the Ph.D. degree in computer application technology from Harbin Industrial University, Harbin, in 2012. From 2006 to 2019, she worked with Northeast Forestry University for teaching. From 2014 to 2015, she was a Visiting Scholar with Virginia Tech University, USA, for one year. She has been a member of IEEE CS, since 2017; of ACM, since 2017; and of CCF, since 2012. Her research interests include natural language processing, social media data mining, intelligent city transportation, and artificial intelligence technology application.



YUE SHANG received the B.S. degree from Northeast Forestry University, Harbin, China, in 2018, where she is currently pursuing the M.Sc. degree with the School of Information and Computer Engineering. Her research interests include deep learning, natural language processing, and text classification.



QI YUE received the Ph.D. degree. She is currently a Professor and a Master Tutor with the School of Information and Computer Engineering, Northeast Forestry University, Harbin, China. Her research interests include artificial intelligence, algorithm optimization, and intelligent control. She presided over or participated in ten scientific research projects, published more than 30 academic articles, presided over six provincial and municipal scientific research projects, edited six textbooks, and won the third prize at ministerial-level scientific and technological progress.



JIYUN ZHOU received the M.Eng. degree from the Harbin Institute of Technology, China, and the joint Ph.D. degree from the School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), and the Department of Computing, The Hong Kong Polytechnic University. He is currently a Postdoctoral Researcher with the Lieber Institute, Johns Hopkins University, USA. His main research interests include bioinformatics, natural language processing, and machine learning.