Cognitive Computation (2019) 11:459–468
https://doi.org/10.1007/s12559-019-9624-y

Improving User Attribute Classification with Text and Social Network Attention

Yumeng Li[1] · Liang Yang[1] · Bo Xu[1] · Jian Wang[1] · Hongfei Lin[1]

Received: 14 December 2017 / Accepted: 3 January 2019 / Published online: 19 January 2019

Abstract

User attribute classification is an important research topic in social media user profiling, which has great commercial value in modern advertisement systems. Existing research on user profiling has mostly focused on manually handcrafted features for different attribute classification tasks. However, these research has partly overlooked the social relation of users. We propose an end-to-end neural network model called the social convolution attention neural network. Our model leverages the convolution attention mechanism to automatically extract user features with respect to different attributes from social texts. The proposed model can capture the social relation of users by combining semantic context and social network information, and improve the performance of attribute classification. We evaluate our model in the gender, age, and geography classification tasks based on the dataset from SMP CUP 2016 competition, respectively. The experimental results demonstrate that the proposed model is effective in automatic user attribute classification with a particular focus on fine-grained user information. We propose an effective model based on the convolution attention mechanism and social relation information for user attribute classification. The model can significantly improve the accuracy in various user attribute classification tasks.

Keywords User attribute classification · Social media · Convolution neural network · Attention mechanism

Introduction

In recent years, online social media, such as Twitter, Facebook, and Sina Weibo, have developed rapidly, and the amount of user data on social media increases accordingly. These user data, as an important resource, can accurately characterize social media users for user profiling. User profiling has thus become one of the hottest research topics in social media analysis, and have been applied to various domains, such as precision marketing, precision medical care, and financial risk forecasting. User attribute classification is a crucial problem in user profiling, which has significant commercial value in modern society. Since social media provide abundant user information for analyzing and inferring user attributes [1, 2], social media analysis based on user attribute classification has therefore become an effective way to discover potential business customers and generate intelligent marketing reports for different brands.

Previous research on user attribute classification has mostly focused on feature engineering [3–6]. Feature engineering generally requires much manual labor to design and extract task-specific features, which may partly limit the scalability of learned classification models. Different features are extracted for different attributes to achieve the ideal performance. Moreover, most studies using feature engineering were based on traditional machine learning classifiers [1, 7–9]. These studies may ignore semantic information in text and social information in user relationships.

In this paper, we propose a novel user attribute classification method to capture the social and semantic information of users and reduce manual labor from feature

✉ Hongfei Lin
hflin@dlut.edu.cn

Yumeng Li
lym199286@mail.dlut.edu.cn

Liang Yang
liang@dlut.edu.cn

Bo Xu
xubo@dlut.edu.cn

Jian Wang
wangjian@dlut.edu.cn

[1] Dalian University of Technology, Dalian, China

engineering. The proposed method makes the most of user-generated text and user social networks by combining text and social relation embedding in an end-to-end model. The learned model can automatically generate fine-grained features and extract keywords in social text for different attribute classification tasks. The attention mechanism in a neural network has been applied in different tasks [10, 11], such as document classification [12, 13] and machine translation [14]. We propose a convolution attention mechanism to identify the importance of each word using its context word embedding in the proposed model. We conduct extensive experiments to evaluate the effectiveness of this method on the dataset from the SMP CUP 2016 competition. Experimental results show that the proposed method can achieve the state-of-the-art performance for different user attribute classification tasks. This model largely improves the classification accuracy of user attributes influenced by social relations by incorporating social network information.

We summarize the contributions of this paper as follows.

1. We propose an end-to-end neural network model based on the convolution attention mechanism to model social media users. The network extracts significant words and valuable microblogs for attribute classification without any handcrafted features.
2. We integrate social relation information into the proposed attention neural network. The integrated network enriches user representations and improves the accuracy of social relation-oriented attribute classification.
3. We conduct extensive experiments to evaluate the effectiveness of the proposed method for different user attribute classification tasks. The experimental results show that the proposed model outperforms state-of-the-art models in user profiling.

The remainder of this paper is organized as follows. Section "Related Works" introduces the related work on user attribute classification; "Social Convolution Attention Neural Network" details the proposed model with text and social network attention for user attribute classification; "Experiments" provides experimental results for three different attribute classification tasks and the attention visualization; and "Conclusion" concludes the paper and provides information regarding future work.

Related Works

Recently, the study of user attribute classification has attracted much attention in the research area of user profiling. Most early research focused on long text generated from social media platforms, such as blogs. For example, Schler et al. [15] built handcrafted feature sets to classify social media users in terms of age and gender by analyzing user writing styles differences and generating user contents for different ages and genders. Mukherjee et al. [16] combined feature selection methods and part-of-speech information to enhance the accuracy of blogger gender prediction.

With the development of short text-based social media, many studies have focused on social data, such as microblogs [17]. User attribute classification has become one of the hottest research topics [4, 18]. Burger et al. [19] extracted n-grams and other handcrafted features from microblog texts, user personal descriptions and user names. They modified the balance winnow algorithm with the extracted features to improve the performance of user gender classification. Miller et al. [20] combined the perceptron with the Bayes model using n-gram features for user gender classification. Ludu et al. [8] modeled the interests of users using the information of followed celebrities and then summarized the characteristics of different users in different fields. Their study implied that men paid more attention to technical celebrities, and women were more interested in family celebrities. Based on this information, they improved the performance of gender classification. Mueller et al. [3] used user names in Twitter to handcraft a variety of features based on word structures for user gender classification. Sesa-Nogueras et al. [9] inferred user genders using the dynamic information from user online writings on digital tablets. These studies have indicated that information from social media can benefit user attribute classification with handcrafted features.

Existing studies have also focused on the extraction of different user attributes. For example, Bo et al. [21] incorporated information gain and the maximum entropy into feature selection for user geographical classification. Their method built a related geographical vocabulary to reduce feature dimensions for accelerating the classification. Ahmed et al. [22] applied the Chinese restaurant process to the recognition of geometrical name entities based on probability models and used the hierarchical relation information of each geographical location for geographical attribute extraction. Rahimi et al. [7] integrated the social relation information into extracting the information of the character "@" in microblogs, and they used the label propagation algorithm and the logistic regression model to improve the prediction of user geographical attributes. Park et al. [2] measured personalities and psychological richness using language models in social media. Their study summarized the ways people expressed themselves and recorded their change over time in terms of words, phrases and topics, which provided a clear portrait of unfolding mental life. Sloan et al. [6] developed a set of pattern-matching rules to infer the distributions of user ages and occupations on Twitter for better attribute extraction performance.

Volkova et al. [1] employed generated texts in social networks to analyze emotions and sentiments for the latent online personality of users based on log-linear models and feature engineering. Li et al. [23] incorporated prior sentiment information at both word level and document level to investigate the influence of each word on the sentiment label for learning effective word representations. Alradaideh et al. [24] proposed to use the rough set-based methods for sentiment analysis to classify tweets written in the Arabic language. Asgarian et al. [25] investigated the impact of NLP tools, various sentiment features, and sentiment lexicon generation approaches for sentiment polarity classification of internet reviews written in Persian language. Mukhtar et al. [26] used different evaluation measures for validation of best classifiers in Urdu sentiment analysis. Peng et al. [27] provided a comprehensive review on sentiment analysis in Chinese language.

Another important research topic in user attribute classification is how to represent users as vectors for measuring user similarities based on representation learning techniques. Some recent studies have focused on this topic. For example, Peng et al. [28] proposed a new method to simultaneously identify the feature dimension of the learned subspace and learn the underlying subspace in the presence of Gaussian noise to boost classification accuracy, robustness and efficiency. They also bridged the gap between Frobenius-norm-based representation (FNR) and nuclear-norm-based representation (NNR) for providing new insights to understand FNR and NNR under a unified framework [29]. Word representation methods have also been used in user attribute classification tasks [30–32]. Furthermore, Le et al. [33] represented sentences and documents as low-dimensional dense vectors to capture abundant semantic information. Some studies have also attempted to develop effective representations of nodes in a large-scale graph, such as social networks [34–36]. These methods have been successfully useful in many tasks, such as visualization, node classification, and link prediction [37–40]. Network embedding has been proven to be effective in node classification particularly on social network-based graphs [34, 35, 39, 40]. These studies treated users in social media as nodes to embed user relations in a continuous vector space. However, to our knowledge, few studies have used network embedding for user attribute classification. We believe that network embedding would benefit user attribute classification by considering the relationships of social media users.

Even though previous studies in user attribute classification have improved the performance of different tasks, their performance may be further enhanced by incorporating more semantic information from user relations. Moreover, feature engineering and feature selection in these studies are relatively expensive in manual labor. Automatically generated features may largely reduce the manual labor and improve the efficiency of attribute extraction and classification.

To this end, we introduce pre-trained word representations to capture the semantic information from public social media texts and propose an end-to-end neural network model for different user attribute classification tasks to reduce the workload of feature engineering. The proposed model introduces the social convolution attention mechanism to assign distinct weights for different microblogs and words; the model automatically distinguishes the content contributions in different tasks. This mechanism enriches detailed information, avoiding the removal of low relevant words in other feature engineering methods [41]. We adopt the large-scale information network embedding [42] as the pre-trained representations to incorporate the social relation information into the text-based neural network model, which may largely enhance the existing one-hot representations in previous models [5].

Our previous work has used microblog data to detect and extract hot topics [43] and learned to rank expansion terms [44]. The difference between this work and our previous works lies in the tasks and methods. This work mainly focuses on the user attribute classification task, whereas our previous work focused on the hot topic detection task. We adopt neural network models in this work instead of using handcrafted features as in our previous works.

Social Convolution Attention Neural Network

In this section, we describe the proposed model using text and social network attention for user attribute classification. We first present our text-based attention model for user attribute classification. Then, we introduce the convolution attention mechanism in detail. Finally, we extend text-based attention with social network embedding. We illustrate the structure of our model based on the text and social attention neural network (SA-NN) in Fig. 1.

The network consists of two-level attention structures as follows: the word-level structure (Word Attention in Fig. 1) and the text-level structure (Text Attention in Fig. 1). The word-level structure uses word2vec [30] to pre-train word embedding, which is taken as inputs for the attention layer.[1] The attention layer learns the semantic weights for different words and encodes word vectors into microblog representations. The text-level structure takes the

[1]We have also used an extra convolutional neural network (CNN) and long-short-term memory (LSTM) layers to encode word representations before the attention layer, but no improvement was achieved with greater time cost. We keep the CNN and LSTM layers as comments in our codes for future optimization.

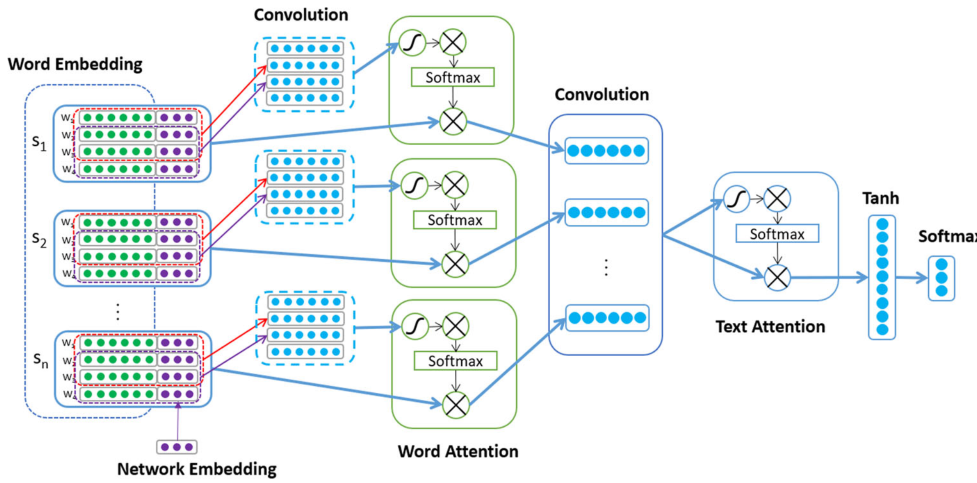

Fig. 1 Model structure based on the text and social attention network (SA-NN)

representations of microblogs as inputs. We regard each microblog as a sentence, because microblogs are mostly short and each has a single topic. In this way, we portray each user in social media as a collection of sentences based on user-generated microblogs. Then, we use the attention layer to learn the semantic sentence representations with respect to each user in a user representation.

User Attribute Classification Based on Text Attention

We use neural network-based embedding vectors in two levels to capture the semantic information in words and sentences. Neural networks with word embedding have been used in many classification tasks to generate sentence representations based on different structured neural networks. In these networks, term frequency and inverse document frequency (TF-IDF)-based features are always applied to measure the importance of different words in sentences. However, the performance may be hindered by the inconsistency between the goal of different tasks and their corresponding target loss functions. In this paper, we incorporate more neural cells to generate different word weights using context-based semantics and social network structure-based information. The proposed model adopts the attention mechanism to automatically assign distinctive attentions for the significant words in different tasks. We name our model as the text attention neural network model (TA-NN).

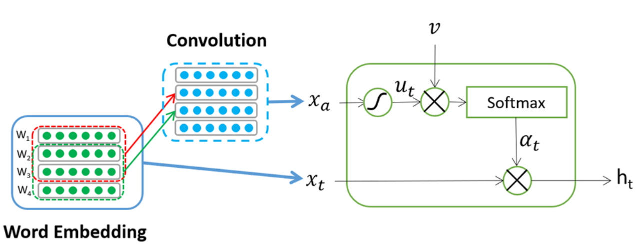

Fig. 2 Convolution attention layer used in our two-level semantic representations

We illustrate the proposed network with attention layers in Fig. 2, which contains three input units and one output unit. In the figure, x_t is a matrix of word vectors in one sentence. x_a and v are used to calculate the weights of different words. h_t is the output unit of the attention layer, which seeks to capture the semantic information of the sentence.

In the attention layer, x_a is an attention matrix, which consists of attention representations of each word. x_{ai} is the word attention representation for the i^{th} word, which is correlated with the weight of the word in the sentence. In the proposed model, the convolutional results of each word are used as the attention representation in which nearby words comprehensively capture the semantic contribution of each word in the context. We visualize the effects of the proposed attentions in the experiments in "Visualization of Attention". We formulate this method as follows.

$$x_{ai} = relu(w[x_{i-1}, x_i, x_{i+1}] + b) \tag{1}$$

$$u_{ti} = tanh(wx_{ai} + b) \tag{2}$$

$$\alpha_{ti} = \frac{exp(u_{ti}^T v)}{\sum_i exp(u_{ti}^T v)} \tag{3}$$

$$h_t = \sum_i \alpha_{ti} x_{ti} \tag{4}$$

where u_{ti} denotes the result of the nonlinear transformation of the attention vector x_{ai} for the ith word. v is a randomly initialized vector affecting the word weight distribution. The same word in different sentences shares the same vector. Specifically, v is used to learn the weights of each dimension of word representations. h_t is the output of the attention layer used as the sentence representation or the user representation.

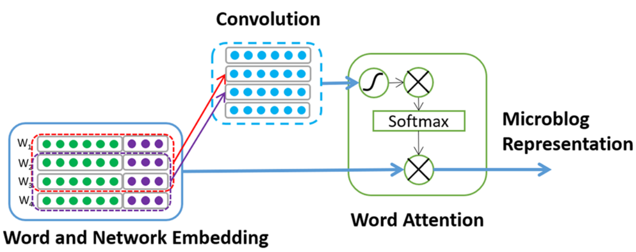

Fig. 3 Convolution attention layer with text and social network embedding

User Attribute Classification Based on Text and Social Network Attention

We further consider the *followed relations* of users in social media to represent users in terms of multiple views based on the proposed TA-NN model in this section. We propose the text and social network attention neural network (SA-NN). In the SA-NN model, we regard users as nodes and relations of users as edges in social networks. We pre-train node presentations of users in a social network with 550 millions *followed relations* using the LINE network embedding method [36]. Users with *followed relations* in social media have similar vector representations in the learned vector space. Since users in different social groups have different habits in word expressions, the probability distribution of words tends to be different among microblogs of different users. In classification tasks, the difference influences the weights of words in attention layers. Therefore, we provide the network embedding results as the prior information for our model. We illustrate the structure of the proposed network based on text and social network attention in Fig. 3.

We calculate the nonlinear transformation of the attention vector u_{ti} and the output of the attention layer h_t as follows.

$$u_{ti} = tanh(w_t x_{ai} + w_r x_r + b) \quad (5)$$

$$h_t = \sum_i \alpha_{ti}[x_{ti}, x_r] \quad (6)$$

where x_r represents the node vectors of users in social networks. α_{ti} is still calculated based on Eq. 3. We implement the vectors both in the word-level and text-level attention layers. According to these two-level attention layers, the low-dimensional dense representations of users encode the semantics and social relation information simultaneously. Finally, we use the user vector representations to classify different user attributes.

We use a nonlinear transformation layer to embed the user vector representations u into a new vector representation u' in a user attribute space. Then, we connect the layer with a Softmax layer to obtain the probability distribution of user attribute values and output the classification results. We formalize this method as follows.

$$u' = tanh(wu + b) \quad (7)$$

$$p_c = \frac{exp(u'_c)}{\sum_{k=1}^{C} exp(u'_k)} \quad (8)$$

$$\zeta = -\sum_{u \in U} \sum_{c=1}^{C} p_c^g(u) \cdot log(p_c(u)) \quad (9)$$

where C is the number of user attribute values in different attribute classification tasks. p_c is the probability that the model predicts the user attribute value as c. U represents the user set. $p_c^g(u)$ is 1 if the ground truth value of u's attribute is c; otherwise, $p_c^g(u)$ is 0. We adopt the cross-entropy as the loss function. The optimization goal is to minimize the classification loss defined in Eq. 9.

Table 2 Statistics of datasets

	#Users	#Microblogs	#Social users
Train	3200	241568	1561
Test	1240	95936	616
Total	4440	337504	2177

Experiments

In this section, we evaluate the proposed model in three different attribute classification tasks. Then, we introduce the experimental settings and report the experimental results in the following subsections.

Table 1 Snapshot of the used dataset

User ID	Age	Gender	Geography	Weibo content
1	1980-1989	Female	Northeast	不知道这边有没有猫屎喝呢？我在沿江路 Don't know if there is Luwak Coffee in the neighborhood? I am at YanJiang Street. 一上班就收到玲姐的圣诞礼物，很开心[微笑] I receive a gift from Ling when I go to work. I am very happy[Smile]

Table 3 Attribute values

Attribute	Label
Age	−1979/1980-1989/1990+
Gender	Male / Female
Geography	Northeast/ North China/ Central/ East China/ Northwest/ Southwest/ South China/ Overseas

Experimental Settings

We use the dataset from the SMP CUP 2016 competition.[2] The competition includes three subtasks in attribute classifications based on the same dataset. There are more than 350 teams participating in the competition. The dataset is constructed based on Sina Weibo, the largest social network website in China, which contains abundant information on social media users. The dataset contains three types of information, including social relations, microblogs, and user attributes. Social relation information contains about 550 million followed relations of 25.67 million users. The followed relations are either one-directional or bi-directional. Microblogs provide the text information from 4.6 million users in the social network. User attributes include the age, gender and geographical attributes of about 5000 users from the above 4.6 million users with microblogs. These 5000 users are chosen by the organizer of the competition to compare the performance of different models. Readers can refer to the official website of the competition for more details. We provide a sample snapshot of the used dataset in Table 1. We use the 5000 users with attribute information in our experiments. We divide the data into the training set and the test set as shown in Table 2.

We examine the effectiveness of our models in three attribute classifications. We provide details about the attribute labels used in these tasks in Table 3. The parameter values are empirically set in our model. Specifically, we pre-train 300 dimension word embeddings on a 5 GB Weibo dataset from SMP CUP 2016 with random initialization. We choose the dimension of word embedding following the existing models [12, 45, 46]. We tune the hyper parameters of our model using 10% of the training data as the development set. We switch the dimension of network embedding from 64 to 512 to obtain the optimal value. We observe that the accuracy tends to be higher with the dimension increase, but the cost on memory and time increases accordingly. We also switch the number of attention units from 64 to 256 but observe no significant difference in terms of accuracy. We select the best configuration based on the performance on the development set and report the selected parameters in Table 4.

[2]https://biendata.com/competition/smpcup2016/

Table 4 Model parameters

Parameters	Values
Word embedding	300
Network embedding	512
Convolution units	128
Word attention units	128
Weibo attention units	128
Tanh units	100
Dropout	0.5

Baselines

We compare our models with several strong baseline methods for user attribute classification. The baseline methods include the most effective models used by the top-5 teams in the SMP Cup 2016 and state-of-the-art models for document classification.

Majority concerns the attribute category with the largest proportion in the training set as the predicted attributes for each user in the test set.

TF-IDF involves term frequency and inverse document frequency, which are built in the ensemble models used by competition participants. We train a logistic regression

Table 5 Accuracy of the geographical attribute classification task

Models	Accuracy (%)
Models with text information	
Majority	28.24
TF-IDF	56.78
AvgWordVec	44.35
HCNN	59.25
PVDM	58.08
HN-ATT	63.55
CNN-GRU	50.81
TA-NN	*69.27*
Models with text and social relation information	
KNN	64.92
XGBoost	66.45
SA-NN	*71.94*
Ensemble models	
HLT-HITSZ	72.70
DUTIR-TONE	69.76
3rd-Team	68.06
SA-NN-Ensemble	*73.55*

The italic numbers indicate the models that achieve the best performance using each feature set classifier to learn the TF-IDF features for the microblog texts of each user. We adopt a 1-gram TF-IDF with the minimum document frequency of 3.

AvgWordVec averages the word embeddings of each microblog to obtain user representations. The user representations are then fed into a support vector machine (SVM)-based classifier as features. We apply the gensim tool[3] to pre-train the word embedding in our experiments.

HCNN uses a hierarchical convolution neural network (CNN) layer and a pooling layer to encode pre-trained word representations as user representations. The filter length is empirically set to be 5, and each CNN layer has 128 filters. We employ relu as the activation function and adadelta as the optimizer.

PVDM merges user-oriented microblogs and adopts the PVDM model [33] to train document vectors. Then, we take user representations as features for SVM-based classification.

HN-ATT is a hierarchical attention network for document classification, which has been proven to be effective in various text classification tasks [12].

CNN-GRU first learns sentence representations using convolution neural network and then encodes semantics of sentences in high-level representations with a gated recurrent neural network [42].

KNN trains a K-nearest neighbor (KNN) algorithm with handcrafted features. We set the parameter k as 20 and use the same feature set as Feature-based XGboost.

XGBoost trains an XGBoost classifier [47] with handcrafted features. The model has been successfully used by the teams in the competition. We implement the model and upload our code on Github.[4]

HLT-HITSZ is the model by the champion team of the SMP CUP 2016 competition. They handcrafted a location-based lexicon and extracted 70 features based on the lexicon integrated with graph embedding-based features. The SVM-based stacking model is adopted to obtain the final outputs.

DUTIR-TONE is the model by the second place team of the SMP CUP 2016 competition. They use text features, temporal features and handcrafted location lexicon-based features. An ensemble of several full-connection neural networks using KNN is adopted for model stacking to obtain the final outputs.

TA-NN is the proposed model in "User Attribute Classification Based on Text Attention", which only uses text information with the attention mechanism. We use TA-NN as a single model for user attribute classification.

SA-NN is the proposed model in "User Attribute Classification Based on Text and Social Network Attention", which uses text and social network information with the attention mechanism. We apply SA-NN as a single model for user attribute classification. The parameters of our models are shown in Table 4.

SA-NN-Ensemble is our ensemble model. Both of the top two winner teams in SMP CUP 2016 competition used the timestamp of microblogs to capture the activity schedules of users for more useful information. The final performance is then enhanced by an ensemble of various models. We follow the ensemble method using our text and social attention model integrated with KNN and temporal features.

We report the experimental results of all models in Table 5. The table shows that the TA-NN model based on text information achieves significant improvement compared with other text-based models. For the models based on text and social relation information, KNN and XGBoost improve the performance of text-based models, and the proposed SA-NN model further enhances the performance with an improvement in terms of classification accuracy by 5.49%. The reason is that the proposed model incorporates text and social relation information simultaneously for comprehensive user representations. The two proposed models, as single models, significantly improve the attribute classification performance without lexicon or handcrafted features, and the models even outperform certain ensemble models. It benefits from the automatically generated features using word embedding and network embedding with the convolution attention neural network. Our ensemble model SA-NN-Ensemble achieves the best performance using a SA-NN model with temporal features, which demonstrates the effectiveness of the proposed model.

Effects of Attention Layers

In this section, we replace the two proposed attention layers with average pooling or max-pooling layers to demonstrate the effectiveness of the attention layers, respectively. The experimental results are presented in Table 6.

Table 6 shows that the word-level attention layer increases the accuracy by 10.73% compared with the average pooling layer. The text-level attention layer

Table 6 Comparison of two-level attention layers

Word level	Text level	Accuracy (%)
Mean pooling	Attention	58.54
Attention	Max pooling	64.48
Attention	Mean pooling	67.34
Attention	Attention	69.27

[3]https://radimrehurek.com/gensim/

[4]https://github.com/liyumeng/SmpCup2016

Table 7 Comparison of the text attention and social network attention in terms of accuracy

Attribute	Text attention (%)	Social network attention (%)
Gender	87.42	87.42
Age	65.24	65.56
Geography	69.27	71.94

increases the accuracy by 2.03% compared with the average pooling layer. The two-level attention layers achieves the best performance. It indicates that the attention layers are useful in user attribute classification; meanwhile, the word-level attention layer is more useful than the text-level attention layer. The reason is that users with more microblogs are modeled accurately with abundant text by different layers. In contrast, users with less microblogs cannot be well modeled with average pooling, and the attention layer is able to weight different microblog vectors for better performance.

Effects of Text Attention and Social Network Attention

We compare the effects of text attention and social network attention model in three different user attribute classification tasks in Table 7. The results indicate that the social network attention improves the accuracy by 2.67% in the geography classification task and improves the accuracy by 0.32% in the age classification task. However, it has no effect in the gender classification task. That means that users in the same area or in similar social network structures are more likely to establish social relationships; users in the same age group are relatively easy to establish social relationships. However, gender attributes have little influence in establishing social relationships. This finding is consistent with our intuition. For those user attributes influenced by social relations, the proposed model can combine social network vectors to further improve the classification accuracy.

Visualization of Attention

Our social attention neural network model can capture the important words under different attribute labels. We take the geography classification task as an example in Table 8. The table lists the five most important words of the attention layer and their corresponding attention weight w_i^c in each category of geographical labels. The weights are calculated as follows.

$$w_i^c = \frac{\sum_{j\in c} \alpha_{dj}\alpha_{wi}^{dj}}{Z^c} \tag{10}$$

$$Z^c = \sum_{i\in M}\sum_{j\in c} \alpha_{dj}\alpha_{wi}^{dj} \tag{11}$$

where w_i^c is the weight of the word w_i in the attribute tag c. α_{dj} is the weight of the current microblog d_j in all the microblogs of the user. $\alpha_{wi}^{d_j}$ is the weight of the word w_i in the microblog d_j. z^c is the normalization factor of the word in the attribute tag c, and M is the number of words in the vocabulary.

Table 8 shows that most attentional words are in cities or provinces belonging to the target region. The proposed model can automatically discover the geographical relationships between cities or provinces and their corresponding regions, even though the training set with only regional labels does not contain any specific cities or provinces. In addition, our model can automatically extract local dialects.

Table 8 The top 5 words weighted by attention layers in each label of the geography attribute

Northeast		North		Central		East	
Shenyang	5.6	Beijing	6.7	Wuhan	4.7	Shanghai	55
Ha'erbin	3	Tianjin	1.8	Zhengzhou	3.8	Beijing	0.8[a]
Dalian	2.4	Shijiazhuang	1	Hubei	1.7	Jiangsu	0.7
Changchun	2.1	Shenyang	0.5[a]	Changsha	0.7	Nanjing	0.6
Liaoning	1.2	Shanghai	0.4[a]	Beijing	0.7[a]	Hangzhou	0.5
Northwest		Southwest		South		Overseas	
Xi'an	11	Chengdu	6	Guangzhou	3	Korea	3.6
Lanzhou	3.2	Chongqing	3	Shenzhen	2.3	Xi'an	1.57[a]
Shanxi	2	Sichuan	2.8	Wu	0.8	Shenyang	1.49[a]
Yinchuan	1	Kunming	1	Ghuangdong	0.8	Geshen	0.95[a]
Xinjiang	0.7	Guizhou	0.7	Dongguan	0.7	Guangzhou	0.92[a]

The number after each word is the weight in percentage. The words followed by [a]are those cities not belonging to the geographical label

Table 9 Examples in attention network model

User ID	Microblog contents
1	不知道这边有没有猫屎喝呢？我在**沿江路** Don’t know if there is Luwak Coffee in the neighborhood? I am at YanJiang Street.
2	铁路**珲春段**进展有点慢 Railway Construction in Hunchun Section is a bit slow.
3	我正在收听**四川**财富广播@DJ雨小轩老师主持的财富社会 I’m listening to the Fortune Plaza Program which is presented by DJ Yu of Sichuan Fortune Plaza Radio Station.

For example, the proposed model extracts “Wu” in Cantonese commonly used under the geographical label “South China.” Highly frequent city names, such as “Beijing” and “Shanghai,” obtain more attention under their own regional label than under other labels. The label “overseas” are designated to users who do not have any geographical information. As a result, there exist many noises under this label. The attention-based model can still find the representative word “Korea” for the label “overseas.”

Table 9 provides an example of three users and their most representative microblog identified by the proposed model. The total number of microblogs for each is approximately 80. The attention word in each microblog is marked in bold. It shows that the proposed model selects the most relevant microblog with the geographical information from a large number of microblogs. Although the dataset only contains regional labels, the proposed model learns the provinces and cities within the labeled region, particularly focusing on fine-grained geographical information, such as the names of streets and roads.

Conclusion

In this paper, we propose an end-to-end neural network model with a convolution attention mechanism based on text and social relation information. To capture the semantic information of social media users, the model automatically extracts important words and valuable microblogs without handcrafted features for different attribute classification tasks. User representations are enriched by social relation information and an attention neural network. Extensive experiments are conducted to examine the effectiveness of the proposed model. Experimental results demonstrate that the proposed model reduces manual labor in handcrafted features, outperforms state-of-the-art models, and significantly improves accuracy in three user attribute classification tasks.

Our future work will be carried out in two respects. We will investigate extra social behaviors characterized by the label “message” or the character “@” for better user profiling in social media. We will also explore using the text and social network attention to model hierarchical user attributes that can be merged and cluster unknown user attributes.

Acknowledgments This work is partially supported by grant from the Natural Science Foundation of China (Nos. 61632011, 61572102, 61772103, 61702080, 61602078), the Ministry of Education Humanities and Social Science Project (No. 16YJCZH12), the Fundamental Research Funds for the Central Universities (DUT18ZD102), and the National Key Research Development Program of China (No. 2016YFB1001103). China Postdoctoral Science Foundation (No. 2018M641691).

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflicts of interest.

Informed Consent All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki declaration of 1975, as revised in 2008(5). Additional informed consent was obtained from all patients for which identifying information is included in this article.

Human and Animal Rights This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Volkova S, Bachrach Y, Armstrong M, Sharma V. Inferring latent user properties from texts published in social media. In: AAAI, pp 4296–4297. 2015.
2. Park G, Schwartz AH, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, Ungar LH, Seligman MEP. Automatic personality assessment through social media language. J Pers Soc Psychol. 2015;108(6):934.
3. Mueller J, Stumme G. Gender inference using statistical name characteristics in twitter. arXiv:1606.05467. 2016.
4. Alowibdi JS, Buy UA, Yu P. Language independent gender classification on twitter. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, pp 739–743. ACM. 2013.
5. Chamberlain BP, Humby C, Deisenroth MP. Detecting the age of twitter users. arXiv:1601.04621. 2016.

6. Sloan L, Morgan J, Burnap P, Williams M. Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. Plos one. 2015;10(3):e0115545.
7. Rahimi A, Vu D, Cohn T, Baldwin T. Exploiting text and network context for geolocation of social media users. 2015. arXiv:1506.04803.
8. Ludu PS. Inferring gender of a twitter user using celebrities it follows. 2014. arXiv:1405.6667.
9. Sesa-Nogueras E, Faundez-Zanuy M, Roure-alcobé J. Gender classification by means of online uppercase handwriting A text-dependent allographic approach. Cogn Comput. 2016;8(1):15–29.
10. Wang L, Cao Z, de Melo G, Liu Z. Relation classification via multi-level attention cnns. In: Proceedings of the 54th annual meeting of the association for computational linguistics. Association for computational linguistics. 2016.
11. Rush AM, Chopra S, Weston J. A neural attention model for abstractive sentence summarization. 2015. arXiv:1509.00685.
12. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. In: Proceedings of NAACL-HLT, pp 1480–1489. 2016.
13. Lin Z, Feng M, dos Santos CN, Yu M, Xiang B, Zhou B, Bengio Y. A structured self-attentive sentence embedding. 2017. arXiv:1703.03130.
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Advances in neural information processing systems, pp 6000–6010. 2017.
15. Schler J, Koppel M, Argamon S, Pennebaker JW. Effects of age and gender on blogging. In: AAAI Spring symposium: Computational approaches to analyzing weblogs, vol 6, pp 199–205. 2006.
16. Mukherjee A, Liu B. Improving gender classification of blog authors. In: Proceedings of the 2010 conference on empirical methods in natural language processing, pp 207–217. Association for computational linguistics. 2010.
17. Feng S, Wang Y, Song K, Wang D, Yu G. Detecting multiple coexisting emotions in microblogs with convolutional neural networks. Cogn Comput. 2018;10(1):136–155.
18. Cha M, Gwon Y, Kung HT. Twitter geolocation and regional classification via sparse coding. In: ICWSM, pp 582–585. 2015.
19. Burger JD, Henderson J, Kim G, Zarrella G. Discriminating gender on twitter. In: Proceedings of the conference on empirical methods in natural language processing, pp 1301–1309. Association for computational linguistics. 2011.
20. Miller Z, Dickinson B, Hu W. Gender prediction on twitter using stream algorithms with n-gram character features. Int J Internet Sci. 2012;2(04):143.
21. Bo H, Cook P, Baldwin T. Geolocation prediction in social media data by finding location indicative words. In: Proceedings of COLING, pp 1045–1062. 2012.
22. Ahmed A, Hong L, Smola AJ. Hierarchical geographical modeling of user locations from social media posts. In: Proceedings of the 22nd international conference on world wide web, pp 25–36. ACM. 2013.
23. Li Y, Pan Q, Yang T, Wang S, Tang J, Cambria E. Learning word representations for sentiment analysis. Cogn Comput. 2017;9(6):843–851.
24. Alradaideh QA, Alqudah GY. Application of rough set-based feature selection for arabic sentiment analysis. Cogn Comput. 2017;9(4):436–445.
25. Asgarian E, Kahani M, Sharifi S. The impact of sentiment features on the sentiment polarity classification in persian reviews. Cogn Comput. 2018;10(1):117–135.
26. Mukhtar N, Khan MA, Chiragh N. Effective use of evaluation measures for the validation of best classifier in urdu sentiment analysis. Cogn Comput. 2017;9(4):446–456.
27. Peng H, Cambria E, Hussain A. A review of sentiment analysis research in chinese language. Cogn Comput. 2017;9(4):423–435.
28. Xi P, Lu J, Yi Z, Yan R. Automatic subspace learning via principal coefficients embedding. IEEE Trans Cybern. 2017;47(11):3583–3596.
29. Xi P, Lu C, Yi Z, Tang H. Connections between nuclear-norm and frobenius-norm-based representations. IEEE Trans Neural Netw Learn Syst. 2018;29(1):218–224.
30. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. 2013. arXiv:1301.3781.
31. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543. 2014.
32. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Trans Assoc Comput Linguist. 2017;5:135–146.
33. Le QuocV, Mikolov Tomas. Distributed representations of sentences and documents. In: ICML, vol 14, pp 1188–1196. 2014.
34. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 701–710. ACM. 2014.
35. Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 855–864. ACM. 2016.
36. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. Line: Large-scale information network embedding. In: Proceedings of the 24th international conference on world wide web, pp 1067–1077. ACM. 2015.
37. Dong Y, Chawla NV, Swami A. metapath2vec: Scalable representation learning for heterogeneous networks. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp 135–144. ACM. 2017.
38. Lai Y-A, Hsu C-C, Chen WH, Yeh M-Y, Lin S-D. Prune: Preserving proximity and global ranking for network embedding. In: Advances in neural information processing systems, pp 5263–5272; 2017.
39. Cavallari S, Zheng VW, Cai H, Chang KC-C, Cambria E. Learning community embedding with community detection and node embedding on graphs. In: Proceedings of the 2017 ACM On conference on information and knowledge management, pp 377–386. ACM. 2017.
40. Cao S, Lu W, Xu Q. Grarep: Learning graph representations with global structural information. In: Proceedings of the 24th ACM international on conference on information and knowledge management, pp 891–900. ACM. 2015.
41. Bo H, Cook P, Baldwin T. Text-based twitter user geolocation prediction. J Artif Intell Res. 2014;49:451–500.
42. Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. In: EMNLP, pp 1422–1432. 2015.
43. Yang L, Lin H, Lin Y, Liu S. Detection and extraction of hot topics on chinese microblogs. Cogn Comput. 2016;8(4):577–586.
44. Xu B, Lin H, Lin Y. Assessment of learning to rank methods for query expansion. J Assoc Inf Sci Technol. 2016;67(6):1345–1357.
45. Chen H, Sun M, Tu C, Lin Y, Liu Z. Neural sentiment classification with user and product attention. In: Proceedings of EMNLP. 2016.
46. Cai F, Chen H. A probabilistic model for information retrieval by mining user behaviors. Cogn Comput. 2016;8(3):494–504.
47. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22Nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 785–794. ACM. 2016.