Exploring Writing Pattern with Pop Culture Ingredients for Social User Modeling

Chiyu Cai[1,2] Linjing Li[1] Daniel Zeng[1] Hongyuan Ma[3]
[1] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing, China
[3] CNCERT/CC, Beijing, China
{caichiyu2014, linjing.li, dajun.zeng}@ia.ac.cn, mahongyuan@foxmail.com

***Abstract*—Social networks have significantly altered the behavior patterns of netizens all around the world. Therefore, accurate and expressive model of social users is increasingly demanded as it pose great value in a variety of scenarios, such as e-commerce, cyber security, and entertainment to name a few. In this paper, we propose the *Pop Culture Attention Writing Model* (PAWM) to explore the writing patterns of social users by explicitly capturing the influence of Internet pop culture ingredients with an attention mechanism. The writing pattern representations are learned by a memory network through storing and updating historical latent patterns. We then develop the *Deep Social User Model* via jointly modeling basic properties of social users, temporal contents, and the learned writing patterns based on PAWM. This paper is the first trial, to the best of our knowledge, which captures Internet pop culture information and applies deep neural network to model user writing pattern. A series of experiments conducted on social bot detection and social user identification demonstrate and validate the effectiveness of the proposed models.**

Keywords—social user modeling, writing pattern, Internet pop culture, attention mechanism, memory network

I. Introduction

Social networks have become the main platforms of messaging for not only individual netizens but commercial companies, enterprises, institutions, and government agencies as well. Social users, the core ingredients of all social networks, are anonymous, independent, distributed, and unsupervised. Thus, their behaviors could cause unexpected impact on the whole society. For instance, social bots are used by some unscrupulous organizations or groups to manipulate discussions on hot social issues, produce fake reviews or rating [1] to satisfy their own interests. Meanwhile, stolen or compromised accounts may be used to make fake connections or dialogues that may threat the privacy or property security of the users [2, 3]. Therefore, it is of great significance to strengthen the management of social user.

User modeling is an essential and fundamental task in social user management. First, from the point of view of security, it can help detect malicious behavior [4]. Second, from a commercial perspective, it can be adopted to identify the interests of social users then conduct targeted advertising [5] or personalized recommendation [6, 7]. Third, from a research point of view, it is interesting to understand the properties of online users and it poses potential to improve related models, approaches, and algorithms [8, 9]. All in all, how to effectively and efficiently model social users has been regarded increasingly as a highly demanded and valuable research challenge.

As such, many methods and algorithms have already been presented previously. Early methods for social user modeling rely on linkage or content information heavily. For text, researchers often employ lexical patterns, part-of-speech, special symbols, and so on [10]. In the aspect of network structure, network of friends and related information are often used [11]. Nowadays, machine learning based methods are widely used [12], and approaches focused on topic [6], interest, sentiment, and online behavior [13, 14] are proposed. Inspired by the success of neural networks, user embedding and related concepts have also been proposed and applied to various tasks [15, 16]. To achieve better performance, recent researches have especially paid attention to incorporate relevant features which can be extract from social platforms [17, 18].

Among which, writing pattern is one of the most personalized and anti-counterfeit feature. Nevertheless, Web 2.0 has altered the pace of popular culture and the Internet pop culture is beginning to prevail, especially on social platform. With the growing influence of Internet pop culture in people's daily life, the writing pattern of netizens, are gradually changing, especially in social platforms where a large number of Internet pop culture ingredients are integrated into the messages, blogs, and tweets. Web-derived word [19] is one of the main Internet pop culture ingredients which have been widely used in daily life, such as *Turnt* and *orz*. The former is the variation of "turned up", the latter is a hieroglyphic used to mean frustrate or worship.

In this paper, we concentrate on exploring writing pattern via learning deep representation then construct a joint model to handle social user. The social user model is applied in two tasks, social bot detection and social user identification. We propose a novel *Pop Culture Attention Writing Model* (PAWM) for mining writing pattern. PAWM learns writing pattern representation by a deliberated memory network instead of conventional methods. Moreover, PAWM captures pop culture ingredients and utilizes these information in learning writing pattern representation via attention mechanism. Then, we build the *Deep Social User Model* (DSUM) by jointly modeling user basic information, temporal content, and writing patterns. The main contributions of our work can be summarized as follows:

- We first consider the role of pop culture ingredients in user writing pattern and propose a pop culture attention mechanism to focus on the influence of Internet pop culture ingredients on writing mode. The proposed attention-based model are able to attend different parts of a sentence when different pop culture ingredients are concerned. Experimental results show that our attention mechanism is effective.
- We devise a memory network and propose the *Pop Culture Attention Writing Model* for exploring writing pattern. PAWM extracts latent writing patterns from text and constantly updates their representations while analyzing new input. To our best knowledge, this is the first step toward utilizing deep learning in exploring writing patterns of social users.
- We present a novel *Deep Social User Model* which jointly models user basic information, temporal content, and writing patterns.
- We conduct a series of experiments on two important tasks, social bot detection and social user identification, with real world datasets to demonstrate and validate the effectiveness of our models.

II. RELATED WROKS

A. *Social User modeling*

Besides a few statistical methods, the majority of works for social user modeling depend on machine learning. Among the various methods, content based model is the most popular one, and it is also the basis of many effective methods. Pennacchiotti and Popescu [20] proposed a machine learning framework relies on four general features, including user profile, tweeting behavior, linguistic content, and network. Yin et al. [21] focused on the influence of user interest and temporal context and proposed a temporal context-aware mixture model to model user rating behaviors. Costa et al. [13] modeled online activity patterns of user and found the distribution of postings inter-arrival times. Xu et al. [14] analyzed the influence of breaking news and social friends on user behavior and presented a mixture topic model based on collapsed Gibbs sampling which modeling user posting behavior. There are also some works which consider writing pattern. Gao et al. [22] analyzed the usage of hashtags and URLs in messages from Twitter and Sina Weibo. Zhang et al. [23] applied the FP-growth algorithm to analyze the user writing pattern and constructed the hierarchical analysis model to calculate the weight of different patterns. They choose four writing pattern features, lexical, syntactic, structural, and content-specific. Same features were employed in work [24], and ensemble methods were proposed to estimate the social users' reputations. Besides, Dewang et al. [25] presented some new features like self-reference words, emotiveness, function word ratio, attractive text ratio, and employed some machine learning based models. Recently, deep learning have shown promising results. Amir et al. [8] encoded latent aspects of social users and learned user embedding via projecting similar users into nearby regions in embedding space. Yuan et al. [16] developed a multi-source long short term memory to model user behaviors by using a variety of information, including historical reversion information, edit page titles, and categories.

B. *Social Bot Detection*

Bot detection is an active research topic in recent years. It devotes to discovering bots among a number of social accounts. Distinguishing human and bots can help users get effective information, focus on valuable social accounts, avoid network traps, and ensure their own security.

Early method for bot detection exploits social honeypots [26]. Recently, machine learning based methods have been successfully adopted. Many works focus on extra information beyond traditional content and network features, which mainly includes user profiles, sentiment, topic, and user behavior. BotCatch [27] proposes a multi-feedback approach and discovers malicious bot by using signature and behavior analysis. It correlates signature and behavior results and feds back all results to dynamically adjust detecting modules. Zafarani and Liu [28] explored individual behavioral patterns when selecting username and introduced a set of consistent behavioral patterns to identify social users. Dickerson et al. [29] developed a feature collection including network, linguistic, and application oriented variables. They found that a number of sentiment related factors were key to the identification of bots. Many researches took multiple features into account and combined them through some machine learning methods. Chu et al. [30] proposed a classification system consisted of four components: entropy component, machine learning component, account properties component, and decision maker. They used conditional entropy to detect regular timing and employed Bayesian classification to detect text patterns. Wang [31] developed a machine learning based method using three graph-based features and three content-based features. Lee et al. [32] employed some ensemble learning methods, such as boosting and bagging of Random Forest classifier, to integrate user demographics, friendship networks, content and history information. Stweeler [33] took user data and tweet content as input to identify bots through its internal components, which using entropy, account properties, NLP classification, and ranking algorithms.

C. *Social User Identification*

User Identification is a classical task which can be traced back to author identification. The goal of social user identification is to recognize particular users among a number of accounts. Recently, owing to the prevalence of social media, the importance of social user identification have been emphasized.

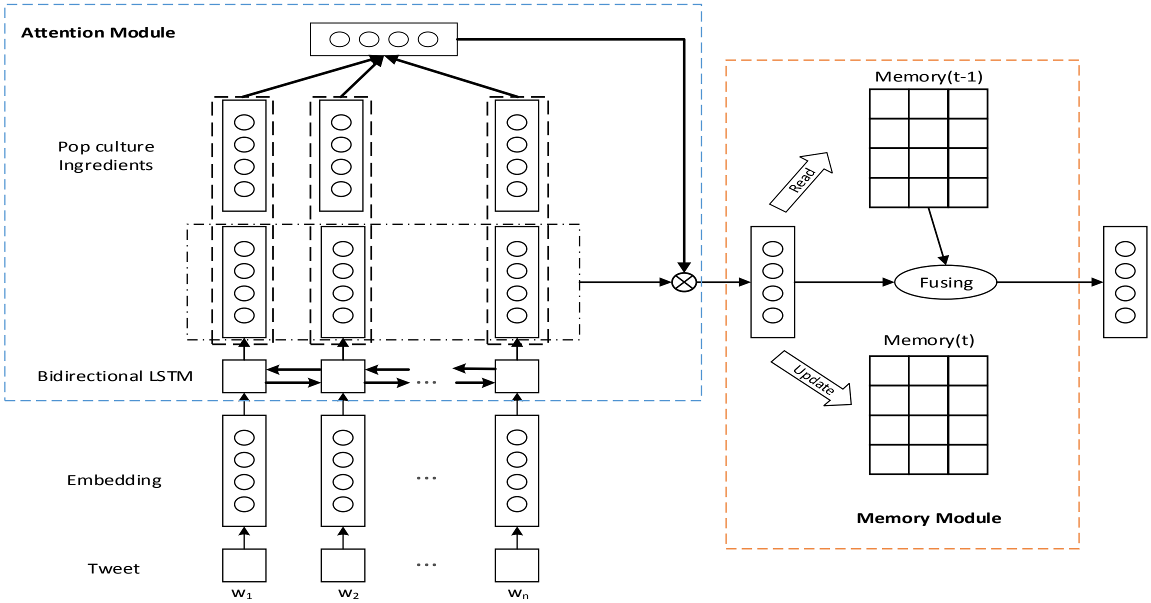

Fig. 1. The architecture of Pop Culture Attention Writing Model

The classical methods for user identification relied on linguistic features and network structure. Brocardo et al. [18] selected stylometry features, including lexical features, syntactic features, semantic features, and developed model combines supervised learning and n-gram analysis. EnTwine [17] integrates profile features, content features, and network features, and presented a unsupervised approach used for candidate selection which assigns match probabilities to each user. Lapenok et al. [34] developed an algorithm of social user identification relied on the analysis of text attributes and social relationship. Lesaege et al. [35] proposed a topic model extending the Latent Dirichlet Allocation using a hidden variable to represent the active user and assuming consumption times to be generated by latent temporal topics. With the development of location-based social network, many studies have been proposed in this network. DLUF [36] defines a new distance metric that measures the diversity between local users from physical dimension using location based social network and identifies a set of local users with maximum diversity. Meanwhile, many research concentrate on identifying users in specific role in network [37]. Recently, deep learning methods have been employed in this task. Shao et al. [38] employed a deep belief network to find the underlying relationships among the behavioral characteristics of network users. Bagnall [39] employed a multi-headed recurrent neural network to model the language of several authors concurrently. Miura et al. [40] proposed an attention based model which integrates word information and character information with multiple neural network layers.

III. Writing Pattern Representation Learning

In this section, we propose the novel *Pop Culture Attention Writing Model* for exploring writing pattern of social users. PAWM captures pop culture information via attention mechanism and encodes the writing pattern representation depend on memory network, as shown in Fig.1. PAWM is comprised of three modules, input module, pop culture attention module, and the memory module. The input module receives tweets from each social user and converts tweets to tweet matrices using word embedding method. The pop culture attention module fetches the influence of pop culture ingredients in writing mode and learns latent writing pattern representation of tweets using contextual information. The memory module stores writing pattern representations that learned from historical information and keeps updating according to new latent writing pattern mined.

A. Pop Culture Attention Module

1) Internet Pop Culture Ingredients. In this paper, we capture Internet pop culture ingredients from three main aspects, web-derived word, emoticon, and hot-spot.

a) Web-derived Word: The so-called web-derived words cannot be found in ordinary dictionaries, they are completely derived from the Web. In general, these words are used to convey ideas in a meaningful or interesting manners, such as, abbreviations, hieroglyphs, homophones, and mixed language. Taking the word “Thicc” as example, it is used to describe a person who create sexy curves. We collected web-derived words from social media according to [41].

b) Emoticon: Emoticons can convey strong sentiment and help social users express mood when they post tweets. Emoticons are popular in social platform, in recently years, more and more emoticons are created and widely used in messages. Different users may have their own favorite emoticons and apply emoticons in different ways and frequencies. Therefore, emoticon can be selected as a pop culture feature.

c) Hot-spot: Hot-spot is a common manifestation of pop culture. Different users may focus on different topics according to their preferences. The concrete form we employed in this paper to depict hot-spot is *hashtag* which widely used in many social platforms.

2) Tweet Encoder. For better modeling the context information of a whole sentence, we adopt the bidirectonal long short term memory (Bi-LSTM) model to extract the high level representation. The content used to describe user u can be treated as a sequence of tweets $C_u = [S_{u1}, S_{u2}, \cdots, S_{un}]$, where n is the number of tweets posted by user u. Considering an input tweet $S_{ui} = [w_1, w_2, \cdots, w_\omega]$ with ω words. The input module maps a word w_t into a vector x_t, and converts S_{ui} into a tweet matrix $\boldsymbol{S} \in \mathbb{R}^{e \times \omega}$, where e is the dimension of word embedding. Then we feed the word embedding $[x_1, x_2, \cdots, x_\omega]$ into the Bi-LSTM layer, and it produces a sequence of hidden states $[h_1, h_2, \cdots, h_\omega]$, where h_t summaries the contextual semantic information of word x_t. The Bi-LSTM transition functions are

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(\mathrm{x_t}), t \in [1, \omega] , \quad (1)$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(\mathrm{x_t}), t \in [\omega, 1] , \quad (2)$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}], t \in [1, \omega] . \quad (3)$$

3) Pop Culture Attention Mechanism

Conventional methods calculate the average value of all hidden states, thus all words will contribute equally in writing patterns which is obviously not the case. In this paper, we introduce a pop culture attention mechanism to learn the latent writing pattern representation of tweets. The attention mechanism learns to allocate high weights for significant elements in a tweet, such as the pop culture ingredients. It emphasizes the influence of pop culture in writing pattern and suppresses the interference of common words. Therefore, we incorporate a pop culture representation v_{pc} in the attention mechanism. We first map each hidden state h_t and pop culture representation v_{pc} through a fully connected network into a hidden space and obtain the representation c_t. Then we measure the significance of the state as the similarity between c_t and a context vector c_s and calculate the attention weight α_t. Finally, we get the weighted representation p though a weighted summation.

$$c_t = \tanh\left(W_h h_t + W_{pc} v_{pc} + b_{pc}\right) , \quad (4)$$

$$\alpha_t = \frac{exp(c_s^{\mathrm{T}} c_t)}{\sum_t exp(c_s^{\mathrm{T}} c_t)} , \quad (5)$$

$$p = \sum_t \alpha_t h_t , \quad (6)$$

where c_s is a context vector which indicates the informative words for writing patterns, it will be randomly initialized and learned during the training phase. W_h, W_{pc}, b_{pc} are weight matrices and bias that will be learned as well.

B. Memory Module

To store the knowledge captured in previous turns and constantly update the writing pattern representations, we design the memory module which combines RNN with an external memory bank. The memory bank is a matrix $\mathbf{M} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}}$ which consists of n memory block and the size of each block is m.

1) Attention Mechanism. In this paper, we employ a scoring function as the attention mechanism in the memory module. In each turn, we compute the match between the current representation p and a previous step memory vector $M_{t-1}(i)$, the weight $u_t(i)$ is calculated as

$$u_t(i) = \frac{exp\left(S(M_{t-1}(i), p)\right)}{\sum_i exp\left(S(M_{t-1}(i), p)\right)} . \quad (7)$$

The scoring function S takes the mixture feature vector $\mathrm{z}(me, p)$ as input. We define the feature vector that captures a variety of similarities between memory vector me and input latent pattern vector p, i.e.

$$S(M_{t-1}(i), p) =$$

$$W_m^3 \tanh(W_m^2 \tanh(W_m^1 \mathrm{z}(M_{t-1}(i), p) + b_m^1) + b_m^2) + b_m^3 \quad (8)$$

$$\mathrm{z}(me, p) = [me, p, |me - p|, me \circ p, me^{\mathrm{T}} W_z p] , \quad (9)$$

where $\circ$ is the element-wise product.

2) Memory Updating. When we update the memory unit, new information is written into all locations to different extents at once. The attention distribution we calculated describes the amount we write at every location. The updated value of a position in memory unit can be computed through a convex combination of the newly written value and the old value which depends on the attention weight $u_t(i)$, which can be formulated as

$$M_t(i) = u_t(i) a_t + (1 - u_t(i) e_t) M_{t-1}(i) , \quad (10)$$

$$a_t = \tanh(W_a p + b_a) , \quad (11)$$

$$e_t = \mathrm{sigmoid}(W_e p + b_e) , \quad (12)$$

where W_a, W_e, b_a, b_e are parameters to be optimized. a_t is an add vector represents the current latent writing pattern information learned from new input and $u_t(i) a_t$ denotes the new knowledge which will be added into the memory unit according to new input. e_t can be viewed as an erase vector, thus $(1 - u_t(i) e_t) M_{t-1}(i)$ indicates the amount of old knowledge saved from the previous step.

3) Memory Reading. We get historical writing pattern representation from the memory module via a read operation. How much knowledge is read out depend on the attention weight $u_t(i)$ and the result of the read operation is a weighted summation

$$r_t(i) = \sum_i u_t(i) M_{t-1}(i) . \quad (13)$$

4) Representation Encoding. The memory writing pattern r_t and current latent writing pattern p are tied together for encoding history and current information, the output representation tp is the fusion of r_t and p. We calculate the sum of vector r_t and current embedding p and feed the vector into a fully connected neural network. The writing pattern representation tp of the input tweet is calculated as

$$tp = W_o(r_t(i) + p) , \quad (14)$$

where W_o is a weight matrix.

IV. Deep Social User Model

This section introduces a joint model for social user modeling which fuses the features of user basic information, content information, and writing patterns. The model consists of four components, writing pattern encoder, user basic encoder, temporal content encoder, and fusing module. The user basic encoder encodes basic information. The temporal content encoder produces temporal content representation. The writing pattern encoder learns user writing pattern representation. Then the fusing module jointly generates the social user representation via incorporating all the above representations. The proposed model is illustrated in Fig. 2.

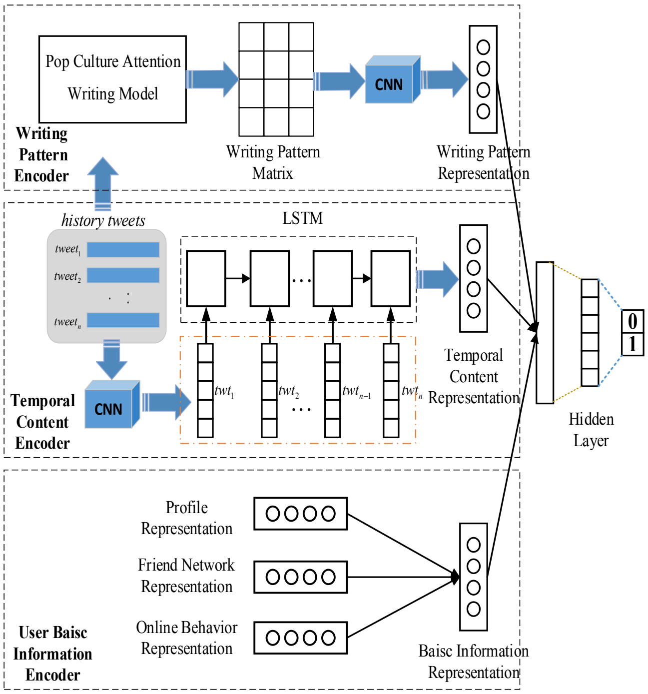

Fig. 2. The architecture of Deep Social User Model

A. Writing Pattern Encoder

An input tweet of user u can be converted, by PAWM, into a writing pattern representation tp. For user u, we map all his historical tweets into vectors and obtain a writing pattern matrix. A CNN is employed to extract high level representation wp_u which indicates the user writing pattern representation.

B. Basic Information Encoder

We consider the following features when learning user basic information representation to catch social influence.

Profile is often presented by a set of textual fields that describe users' basic information such as username, description, verified account, location, gender, occupation, etc. In its general form, description is a short self-introduce or a few words of tags. The profile is thought to be an outline of a personal life which is relatively stable. For profile based information, we use the word embedding model and CNN network to convert the text as a vector. The profile vector is denoted as $v_{profile} = [v_{username}, v_{location}, v_{description}]$.

Friend network refers to the social interactions of a user with other ones which is an intuitive demonstration of the user's social influence. We adopt features including the number of followers, mutual followers, the rate of mutual followers in friends, etc. The network vector is denoted as $v_{network}$.

Online behavior is an important index which reflects the difference of users' activity characteristic. We calculate the average number of posting (or retweeting) per day, compare the posting behavior between working days and weekend days by calculating the ratio between the average number during the weekends (Saturday-Sunday) and the one during the week (Monday-Friday). In addition, we take the circadian rhythm into consideration which is an individual difference in personality. We calculate the average number during four-hour intervals (early morning, morning, midday, afternoon, night, midnight). Finally, we build an online behavior vector $v_{behavior}$.

We integrate profile vector, network vector, and behavior vector, map the fused vector to user basic information representation bi_u using a fully connected network, i.e.

$$bi_u = \mathrm{sigmoid}(W_b v_{pn} + b_b), \tag{15}$$

$$v_{pn} = [v_{profile}, v_{network}, v_{behavior}]. \tag{16}$$

C. Temporal Content Encoder

It is obvious that correlation existed between tweets posted by the same user. Therefore, the historical tweets can be regarded as temporal information instead of being simply treated as plain text in nearly all conventional methods. We construct a temporal content encoder that learns content representation which not only captures semantic information but also learns the temporal patterns. The temporal content encoder is comprised of CNN and LSTM where CNN maps each tweet into embedding space and forms a representation, while LSTM transfers the historical tweet representation sequence of a user u into a temporal representation tc_u.

By adopting the same method introduced in the above pop culture attention module, we obtain a tweet matrix $\mathbf{S}_{uj}$. Then it is fed into the following convolutional layer. In this layer, $\mathbf{S}_{uj}$ will be convoluted with a series of filters $\mathbf{F} \in \mathbb{R}^{e \times m}$, where m is the width of the filter. In this paper, we denote the result of the convolution as a vector $c_l \in \mathbb{R}^{\omega+m-1}$, then the convolution operator can be formulated as

$$c_{lk} = (\mathbf{S} * \mathbf{F}_l)_k = \sum_{m\omega} \left(\mathbf{S}_{[:,\boldsymbol{k-m+1:k}]} \odot \mathbf{F}_l\right)_{m\omega}, \tag{17}$$

where $\odot$ denotes the *Hadamard Product*, i.e., the element-wise multiplication, $\mathbf{S}_{[:,\boldsymbol{k-m+1:k}]}$ is a block of size $e \times m$. In order to speed up the train phase, ReLU is adopted as the non-linear function. A max pooling layer is then connected to the convolutional layer, its output $twt_{uj} \in \mathbb{R}^{1 \times s}$ is the final pooled representation of tweet $\mathbf{S}_{uj}$. For a user u, the historical tweet representation sequence is denoted as

$$ts_u = [twt_{u1}, twt_{u2}, \cdots, twt_{un}]. \tag{18}$$

We feed the sequence ts_u as input of LSTM and learn the temporal content representation tc_u.

D. Joint Model

The fusing module jointly models the information from three component using a fully connected layer, including representations of user basic information, temporal content, and writing pattern, we build the social user representation as

$$U_u = \mathrm{ReLU}(W_u[bi_u, tc_u, wp_u] + b_u), \tag{19}$$

where W_u and b_u are the weight and bias. Then a classifier is employed to solve the corresponding tasks.

E. Training

At the training phase, we select cross-entropy as the loss function and mini-batch Adagrad as the optimizer. The main advantage is that the learning rate can be self-adapt. The dimension of word and memory embedding is set as 100. The memory size is set as 600 to store carried knowledge. During the training process, the dropout rate is set to be 0.5 to avoid overfitting.

V. Experiments

We conduct experiments on social bot detection and social user identification. The models used in this section are trained independently, while the architecture remains the same.

A. Social Bot Detection

1) Dataset

To assess the proposed model in social bot detection task, we choose a public dataset [42]. This dataset is consisted of a large number of Twitter accounts and their labels. For each account, we collected the 1000 most recent tweets via Twitter API. The information we collected including user profile, tweets, online behaviors, etc. We discarded accounts with tweets posted less than 200. The final dataset contains 2742 bot accounts and 2916 human accounts. Table I provides the details on the data.

TABLE I. Summary of Dataset For Bot Detection

	#Accounts	**#Tweets**
Bot	2742	2,487,000
Human	2916	2,635,000

2) Baselines and Evaluation Metrics

The evaluation metrics we selected are precision, recall, and F1 score, which are widely used in bot detection. Evaluation was performed through 10-fold cross-validation. For each split, we fit the model with 80% data, tuned the model with 10% data and tested on the remaining data. To validate the effectiveness of the proposed model, we compared it with the following baselines, *EWS*[24], *DW*[25], *RSC* [13], *Boosting* [32], *BoostOR* [42], and *Stweeler* [33].

3) Experimental Results

Table II shows the experimental results. It can be seen that our model outperforms all baselines, it achieves the best performance with precision 0.9163 and F1 value 0.8951 for detecting bots. The RSC model gets the worst performance as it only mines the online activities patterns. Stweeler and BoostOR perform better than RSC by combine several features including basic information, network structure, text, etc. The empirical results indict that social user modeling is a comprehensive task, thus features from different aspects should be taken into consideration, single or a few features perform poorly in this task. In addition, comparing with the variant model B+C in Table III, which use the similar feature information with Stweeler and BoostOR, our model performs better than all baslines. The results validate the effectiveness of our deep social user modeling framework.

TABLE II. Performance of proposed model and baselines on Bot Detection

Methods	Precision	Recall	F1
EWS	83.41	82.73	83.07
DW	84.65	82.68	83.65
RSC	79.52	77.79	78.65
Boosting	85.23	84.32	84.77
Stweeler	83.38	88.23	85.74
BoostOR	83.16	89.25	86.10
Proposed model	**91.63**	**87.49**	**89.51**

To further analyze the impact of each component in our proposed model, we compared our model with several variants. Table III depicts the corresponding experimental results. Model WP can achieve the performance with F1 value 0.858 with only one feature and perform better than EWS and DW which also model user writing pattern. Meanwhile it performs better than some methods which fuse several features. The results validated the effectiveness of the proposed PAWM. Surprisingly, we found that the basic information can improve the performance. Compared to the models without basic information, performance of B+WP, B+C+WP are increased by approximately 0.01 for F1 value. This is possibly due to the relatively stability and reliability of explicit user features in the profile. The results of model C+WP and B+C+WP demonstrated the positive effect of using writing pattern. The performance comparison demonstrated that our model is an effective method. This further implies that jointly modeling basic, content and writing pattern information could provide a positive effect on bot detection.

TABLE III. Performance of several variants on Bot Detection

Methods	Precision	Recall	F1
WP	84.01	87.67	85.80
B+C	89.15	83.47	86.22
B+WP	88.29	86.21	87.24
C+WP	90.57	86.48	88.48
B+C+WP Proposed model	**91.63**	**87.49**	**89.51**
WP-AT	83.28	85.83	84.54
B+C+WP-AT	89.45	86.36	87.88

(B: basic information, C: temporal content, WP: writing pattern, AT: attention)

To validate the pop culture attention mechanism used in our model, we further removed the pop culture attention mechanism to form two models WP-AT and B+C+WP-AT. Compared to baselines, they still perform better than some baselines. It partly validated that the memory network we designed for learning writing pattern representation performs well. On the other hand, the results that perform worse than model WP and B+C+WP demonstrated the effectiveness of the pop culture attention mechanism we proposed.

B. *Social User Identificaton*

In this task, social users are divided into five roles including star, media professional, traveler, corporate, and e-commerce. Here, the goal is to identify the role of each unknown account.

1) Dataset

To assess the proposed model in social user identification task, we collected data from Sina Weibo, a famous Chinese social media similar to Twitter. The whole dataset contains 882 users and 1,498,700 history posts. The information we collected including user profile, tweets, online behaviors, etc. The summary of the dataset is shown in Table IV.

TABLE IV. SUMMARY OF DATASET FOR USER IDENTIFICATION

Role	#Accounts	#Tweets
Star	145	105,365
Media Professional	128	441,072
Travel	186	314,562
Corporate	193	289,045
E-commerce	230	348,712

2) Baselines and Evaluation

We selected micro-F1 value as the evaluation criteria which is widely used in user identification. Evaluation was performed through 10-fold cross-validation. To validate the performance of the proposed model, we compared our model against six previously introduced baselines, *Stylometric* [18], *EWS*[24], *DW*[25], *EnTwine* [17], *DBN* [38], and *WCAT* [40].

3) Experimental Results

Based on the above experiments about bot detection, we further conducted experiments to compare the proposed model with all baselines and the experimental results are shown in Table V.

TABLE V. PERFORMANCE OF PROPOSED MODEL AND BASELINES ON USER IDENTIFICATION

Methods	Mirco-P	Mirco-R	Mirco-F1
EWS	80.53	79.46	79.99
DW	81.03	79.74	80.38
Stylometric	84.34	81.62	82.96
EnTwine	83.61	85.46	84.52
DBN	82.25	81.07	81.66
WCAT	83.14	81.95	82.54
Proposed model	87.31	85.88	**86.59**
WP	84.28	82.34	83.30
B+C+WP-AT	86.10	84.36	85.22

(B: basic information, C: temporal content, WP: writing pattern, AT: attention)

Comparing with all baselines, the proposed model performs the best, it provides at least a 0.02 improvement in micro-F1 score. Moreover, the proposed model which removes pop culture attention mechanism performs worse. It again demonstrated that pop culture attention mechanism can improve the performance of social user modeling. The contrasts among EWS, DW and WP demonstrate the effectiveness of the user writing pattern representation we learned through PAWM. Finally, the experimental results validate that the proposed model is an effective method for social user identification.

VI. CONCLUSION

This paper explored the writing pattern of social network users. The proposed pop culture attention writing model concentrates on the role of Internet pop culture ingredients in writing pattern learning and presents a pop culture attention mechanism. A memory network was designed in PAWM to store and update the representation of the writing patterns to be learned. Then, the deep social user model is developed to model social users which fuses user basic information, temporal contents, and writing patterns together. Experiments on two tasks, social bot detection and social user identification, demonstrated the effectiveness of the proposed models.

ACKNOWLEDGMENT

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0820105 and 2016QY02D0305, the National Natural Science Foundation of China under Grants 71702181, 71621002, as well as the Key Research Program of the Chinese Academy of Sciences under Grant ZDRW-XH-2017-3. Hongyuan Ma is the corresponding author.

REFERENCES

[1] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM,* vol. 59, pp. 96-104, 2016.

[2] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Compa: Detecting compromised accounts on social networks," in *NDSS*, 2013.

[3] X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling Online Social Behaviors for Compromised Account Detection," *IEEE Trans. Information Forensics and Security,* vol. 11, pp. 176-187, 2016.

[4] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy*, et al.*, "Towards Detecting Anomalous User Behavior in Online Social Networks," in *USENIX Security Symposium*, 2014, pp. 223-238.

[5] T. H. Haveliwala, G. M. Jeh, and S. D. Kamvar, "Targeted advertisements based on user profiles and page profile," ed: Google Patents, 2012.

[6] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Analyzing user modeling on twitter for personalized news recommendations," in *International Conference on User Modeling, Adaptation, and Personalization*, 2011, pp. 1-12.

[7] A. M. Elkahky, Y. Song, and X. He, "A multi-view deep learning approach for cross domain user modeling in recommendation systems," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 278-288.

[8] S. Amir, B. C. Wallace, H. Lyu, and P. C. M. J. Silva, "Modelling context with user embeddings for sarcasm detection in social media," *arXiv preprint arXiv:1607.00976,* 2016.

[9] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk*, et al.*, "Modeling the impact of short-and long-term behavior on search personalization," in *Proceedings of the 35th*

international ACM SIGIR conference on Research and development in information retrieval, 2012, pp. 185-194.
[10] H. V. Halteren, "Author verification by linguistic profiling: An exploration of the parameter space," *ACM Transactions on Speech and Language Processing (TSLP),* vol. 4, p. 1, 2007.
[11] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, "Classifying latent user attributes in twitter," in *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, 2010, pp. 37-44.
[12] M. L. Brocardo, I. Traore, S. Saad, and I. Woungang, "Authorship verification for short messages using stylometry," in *Computer, Information and Telecommunication Systems (CITS), 2013 International Conference on*, 2013, pp. 1-6.
[13] A. Ferraz Costa, Y. Yamaguchi, A. Juci Machado Traina, C. Traina Jr, and C. Faloutsos, "Rsc: Mining and modeling temporal activity in social media," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 269-278.
[14] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang, "Modeling user posting behavior on social media," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 545-554.
[15] Y. Yu, X. Wan, and X. Zhou, "User embedding for scholarly microblog recommendation," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 449-453.
[16] S. Yuan, P. Zheng, X. Wu, and Y. Xiang, "Wikipedia Vandal Early Detection: From User Behavior to User Embedding," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2017, pp. 832-846.
[17] N. Chhaya, D. Agarwal, N. Puri, P. Jain, D. Pai, and P. Kumaraguru, "EnTwine: Feature analysis and candidate selection for social user identity aggregation," in *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on*, 2015, pp. 1575-1576.
[18] M. L. Brocardo, I. Traore, S. Saad, and I. Woungang, "Verifying online user identity using stylometric analysis for short messages," *Journal of networks,* vol. 9, p. 3347, 2014.
[19] L. Velikovich, S. Blair-Goldensohn, K. Hannan, and R. McDonald, "The viability of web-derived polarity lexicons," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2010, pp. 777-785.
[20] M. Pennacchiotti and A.-M. Popescu, "A Machine Learning Approach to Twitter User Classification," *Icwsm,* vol. 11, pp. 281-288, 2011.
[21] H. Yin, B. Cui, L. Chen, Z. Hu, and Z. Huang, "A temporal context-aware model for user behavior modeling in social media systems," in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 2014, pp. 1543-1554.
[22] Q. Gao, F. Abel, G.-J. Houben, and Y. Yu, "A comparative study of users' microblogging behavior on Sina Weibo and Twitter," in *International Conference on User Modeling, Adaptation, and Personalization*, 2012, pp. 88-101.
[23] Y. Zhang, Y. Liu, and G. Chen, "A solution of anonymous email identification based on writing structural pattern," in *Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on*, 2015, pp. 1525-1531.
[24] J. H. Suh, "Comparing writing style feature-based classification methods for estimating user reputations in social media," *SpringerPlus,* vol. 5, p. 261, 2016.
[25] R. K. Dewang, P. Singh, and A. K. Singh, "Finding of Review Spam through Corleone, Review Genre, Writing Style and Review Text Detail Features," in *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, 2016, p. 23.
[26] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots+ machine learning," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 2010, pp. 435-442.
[27] Y. Ji, Q. Li, Y. He, and D. Guo, "BotCatch: leveraging signature and behavior for bot detection," *Security and Communication Networks,* vol. 8, pp. 952-969, 2015.
[28] R. Zafarani and H. Liu, "Connecting users across social media sites: a behavioral-modeling approach," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 41-49.
[29] J. P. Dickerson, V. Kagan, and V. Subrahmanian, "Using sentiment to detect bots on twitter: Are humans more opinionated than bots?," in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2014, pp. 620-627.
[30] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Who is tweeting on Twitter: human, bot, or cyborg?," in *Proceedings of the 26th annual computer security applications conference*, 2010, pp. 21-30.
[31] A. H. Wang, "Detecting spam bots in online social networking sites: a machine learning approach," in *IFIP Annual Conference on Data and Applications Security and Privacy*, 2010, pp. 335-342.
[32] K. Lee, B. D. Eoff, and J. Caverlee, "Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter," in *ICWSM*, 2011, pp. 185-192.
[33] Z. Gilani, L. Wang, J. Crowcroft, M. Almeida, and R. Farahbakhsh, "Stweeler: A framework for twitter bot analysis," in *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 37-38.
[34] M. V. Lapenok, A. V. Tsygankova, N. G. Tagiltseva, L. V. Matveyeva, O. M. Patrusheva, and N. V. Gerova, "User Identification in a Variety of Social Networks by the Analysis of User's Social Connections and Profile Attributes," in *International Conference on Smart Education and Smart E-Learning*, 2017, pp. 486-496.
[35] C. Lesaege, F. Schnitzler, A. Lambert, and J.-R. Vigouroux, "Time-aware user identification with topic models," in *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, 2016, pp. 997-1002.
[36] C. Huang, D. Wang, and S. Zhu, "Towards Diversified Local Users Identification Using Location Based Social Networks," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 115-118.
[37] X. Zhou, B. Wu, and Q. Jin, "User role identification based on social behavior and networking analysis for information dissemination," *Future Generation Computer Systems,* 2017.
[38] H. Shao, L. Tang, L. Dong, L. Chen, X. Jiang, and W. Wang, "A Research on the Identification of Internet User Based on Deep Learning," in *International Conference on Machine Learning and Intelligent Communications*, 2018, pp. 73-80.
[39] D. Bagnall, "Author identification using multi-headed recurrent neural networks," *arXiv preprint arXiv:1506.04891,* 2015.
[40] Y. Miura, T. Taniguchi, M. Taniguchi, and T. Ohkuma, "Author Profiling with Word+ Character Neural Attention Network," in *CLEF (Working Notes)*, 2017.
[41] C. Cai, L. Li, and D. Zengi, "Web-derived Emotional Word Detection in social media using Latent Semantic information," in *Intelligence and Security Informatics (ISI), 2017 IEEE International Conference on*, 2017, pp. 95-100.
[42] F. Morstatter, L. Wu, T. H. Nazer, K. M. Carley, and H. Liu, "A new approach to bot detection: striking the balance between precision and recall," in *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2016, pp. 533-540.