Expert Systems With Applications 198 (2022) 116744

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Coupled social media content representation for predicting individual socioeconomic status

Tao Zhao [a,*], Lu Tang [a], Jinfeng Huang [a], Xiaoming Fu [b]

[a] National University of Defense Technology, Changsha, China
[b] University of Goettingen, Goettingen, Germany

ARTICLE INFO

Keywords:
Socioeconomic status
Coupled social media content representation
Structure-aware social media text representation
Coupled attribute representation

ABSTRACT

Predicting individual socioeconomic status (SES) from social media content benefits various applications in economic and social fields. Most previous works adopt machine learning methods with predefined features to infer SES. Nevertheless, they ignore some important information of social media content, such as order, structure and relation information, which leads to limited performance. In this paper, we propose a COupled social media content REpresentation model (CORE) for individual SES prediction, which efficiently exploits latent complex couplings of social media content. CORE devises a structure-aware social media text representation method to incorporate the order and the hierarchy of social media text, and leverages a coupled attribute representation method to take into account intra-coupled and inter-coupled interaction relationships among user level attributes. Our experiments on a real data set of a Chinese microblogging platform demonstrate that our approach significantly outperforms benchmark methods, which validates its efficiency and robustness. The proposed model could be applied to improve the SES prediction and other user profiling tasks.

1. Introduction

Nowadays, predicting individual socioeconomic status (SES) from social media content has become an important research area. SES characterizes a person's economic and social position in relation to others, which is typically divided into three levels (high, middle, and low).[1] As a person's SES allows different levels of accesses to financial, social and human capital resources, inferring individual SES not only assists governments and research institutions in public policymaking and socioeconomic research on a large scale population, but also helps promote online marketing and personalized services. In addition, it benefits a wide range of other fields, such as health (Propper, Damiani, Leckie, & Dixon, 2007), education (Sirin, 2005) and transportation (Carlsson-Kanyama & Linden, 1999). Traditionally, national statistical offices measure socioeconomic information of a population by a large number of personal or household surveys, which is highly expensive and time-consuming (Soto, Frias-Martinez, Virseda, & Frias-Martinez, 2011).

With the worldwide ubiquity of online social media, such as Twitter, Facebook and Sina Weibo, online social media content has been used in recent research for population informatics in demographics (Burger, Henderson, Kim, & Zarrella, 2011; Golbeck, Robles, & Turner, 2011; Rao, Yarowsky, Shreevats, & Gupta, 2010), economics (Bollen, Mao, & Zeng, 2011; Preoţiuc-Pietro, Volkova et al., 2015; Wang, Gao, Liu, Yang, & Zhou, 2019), social science (Aletras & Chamberlain, 2018; Huang & Wong, 2016; Lampos, Aletras, Geyti, Zou, & Cox, 2016; Tchokni, Séaghdha, & Quercia, 2014) and other research domains (Culotta, 2010; Lampos, Miller, Crossan, & Stefansen, 2015; Lampos, Preoţiuc-Pietro, & Cohn, 2013). In consideration of the great significance of SES, this work focuses on predicting social media users' SES from their own social media contents. For the generalization like previous related work (Lampos et al., 2016; Preoţiuc-Pietro, Lampos et al., 2015), we regard social media text (i.e., microblog text) and platform-based user level attributes (e.g., the number of followers, the number of followees, etc.) as a user's social media content. Meanwhile, this work takes microblogging platforms as a use case study.

Previous related works have looked into predicting individual socioeconomic information based on social media content, such as inferring occupation (Preoţiuc-Pietro, Lampos et al., 2015), SES (Lampos et al., 2016) and income (Abitbol & Morales, 2021; Preoţiuc-Pietro, Volkova et al., 2015). These studies devote to manually extract several kinds of textual features from social media text, and then feed the extracted features and platform-based user level attributes into a classic machine learning algorithm for the specific prediction task. However,

* Corresponding author.
E-mail addresses: taozhao@nudt.edu.cn (T. Zhao), lutang@nudt.edu.cn (L. Tang), hjf@nudt.edu.cn (J. Huang), fu@cs.uni-goettingen.de (X. Fu).

[1] https://en.wikipedia.org/wiki/Socioeconomic_status

https://doi.org/10.1016/j.eswa.2022.116744
Received 31 August 2020; Received in revised form 27 January 2022; Accepted 22 February 2022
Available online 8 March 2022
0957-4174/

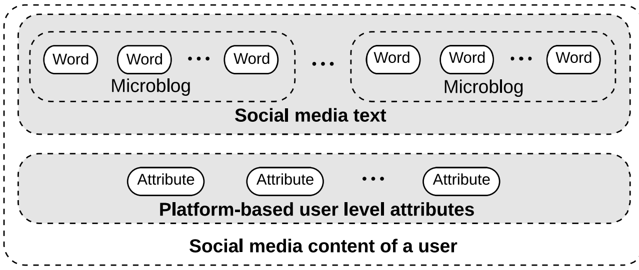


Fig. 1. The architecture of social media content.

the prediction performance heavily depends on these extracted features, which needs effective feature engineering. Although extracting features as comprehensively as possible for representing the social media content, existing methods still ignore the following implicit important information for the social media content representation, which limits the prediction performance.

- **Order of social media text.** Previous related approaches represent social media text with predefined features, such as n-grams or word embedding based features (Preoţiuc-Pietro, Lampos et al., 2015). However, these textual features cannot capture the order of social media text, which is an important information for representing long text sequence. Especially for microblogging, the order relationships among words and microblogs are ignored.
- **Structure of social media text.** Most existing methods first aggregate social media text and then directly extract user level textual features. However, they ignore the hierarchical structure of social media text, leading to information loss. As shown in Fig. 1, words form microblogs and microblogs form social media text of a microblogging user.
- **Relations among user level attributes.** In real world, attributes are more or less interacted and coupled via explicit or implicit relationships (Wang, She, & Cao, 2013). Therefore, user level attributes may share the same statistical information on social media. Nevertheless, to our knowledge, previous related methods on SES prediction regard the user level attributes independent and feed them into classifiers without considering their coupling relations.

In this paper, we propose a novel COupled social media content REpresentation model (CORE) for individual SES prediction based on social media content, which is able to sufficiently explore the complex coupling information of social media content. Specifically, to better represent the social media content, CORE takes into account three types of significant information ignored by previous works. First, in order to account for the ordered couplings of text, CORE applies a recurrent neural network to represent text sequences in social media content due to its representational power and effectiveness in capturing dependencies of a sequence. Second, since each social media text has a hierarchical structure, CORE further presents a structure-aware social media text representation method based on recurrent neural networks, i.e., constructing a social media text representation by first building representations of microblogs with corresponding words and then aggregating them into a social media text representation. Third, to take account of latent coupling relations among user level attributes, CORE devises a coupled user level attribute representation method, which incorporates intra-coupled interaction (i.e., the correlations between attributes and their own powers) and inter-coupled interaction (i.e., the correlations between attributes and the powers of others). Finally, the proposed model learns a coupled social media content representation by aggregating social media text representation and coupled user level attribute representation to predict individual SES.

Our work takes Sina Weibo, a popular microblogging platform in China, as a use case study, and creates a real Sina Weibo dataset. To demonstrate the efficiency and robustness of the proposed model on individual SES prediction, the proposed model is applied to this dataset. Extensive experiments demonstrate that CORE substantially outperforms previous work on SES-related prediction by 11%–27% in terms of accuracy, and provides consistent improvements (about 2%–4%) compared with previous social media content representation methods. It is worth to note that the proposed model can be applied not only to the SES prediction problem but also to other similar tasks, such as user profiling and sentiment analysis.

To summarize, the main contributions of this paper are as follows:

- We propose a novel coupled social media content representation model for individual SES prediction, which efficiently incorporates complex couplings of social media content. To our best knowledge, this is the first work in this area.
- We present a structure-aware social media text representation method to take into account the order and hierarchical structure information of social media text.
- We devise a coupled attribute representation method to explore the intra-coupled and inter-coupled interaction relationships among user level attributes, which can effectively capture the intrinsic couplings of attributes for the SES prediction.
- We build a real dataset of Sina Weibo users with rational SES labels and apply CORE to this dataset. Extensive experiments demonstrate that the proposed model significantly outperforms the benchmark models, which validates its efficiency and robustness.

The rest of this paper is organized as follows. In Section 2, we briefly introduce the related work including socioeconomic-related information prediction based on social media content and representation learning of social media content. Section 3 presents the proposed model in details. In Section 4, we introduce the data collection and preprocessing. The efficiency and robustness of our proposed model is demonstrated with experimental evaluation in Section 5. Finally, we conclude this work in Section 6.

2. Related work

In this section, we first discuss related work in socioeconomic information prediction based on social media content and then discuss existing work in social media content representation learning.

2.1. Socioeconomic information prediction based on social media content

Socioeconomic information prediction based on social media content has been studied in the past few years (Gao, Zhang, & Zhou, 2019). Preoţiuc-Pietro, Lampos et al. (2015) focuses on the occupational class for a Twitter user, in which they first extract user level features and textual features such as word clusters and embeddings, and then employ linear and non-linear methods (including logistic regression with Elastic Net regularization, Support Vector Machine and Gaussian Process) for classification. (Abitbol & Morales, 2021; Preoţiuc-Pietro, Volkova et al., 2015) studies whether the user behaviors on Twitter can be used to build a predictive model of income. (Preoţiuc-Pietro, Volkova et al., 2015) designs different feature categories and use a Gaussian Process-based method for user income prediction. Abitbol and Morales (2021) develops a model which uses Twitter users' mobility patterns and hashtags to predict user income. The work most similar to our work is (Lampos et al., 2016). They also use a similar methodology, where they utilize several kinds of textual features extracted from social media text and platform-based user level attributes in Twitter to represent each user and then use a composite Gaussian Process model to infer the SES of Twitter users. However, these methods only leverage predefined features with feature engineering, which cannot capture the heterogeneous couplings among social media text and platform-based user level attributes.

There are also several related works based on additional data sources, such as social networks (Aletras & Chamberlain, 2018) and geo-location information (Borges, Almeida, Pappa, et al., 2014; Ding, Huang, Zhao, & Fu, 2019), which are different from our task here since our work only focuses on general social media content, i.e., social media text and user level attributes. Some other works also studies potential factors for SES attribute prediction. For example, (Hasanuzzaman, Kamila, Kaur, Saha, & Ekbal, 2017) first employs a weakly supervised learning method to automatically identify the temporal orientation of tweets on Twitter and quantify a user's income based on the overall temporal orientation. Ding, Gao, Dong, Tong, and Fu (2021) develops a factorization machine based multi-task learning method using an attention mechanism to predict a person's socioeconomic attributes (income level, family income level, occupation type, and education level) from the area-level average income, points of interest, and housing prices near the person's home location.

2.2. Social media content representation learning

Nowadays there are a large body of existing works on text-based social media content representation learning for various specific applications, such as sentiment classification, rumor detection, and user profiling. Early approaches devote to design effective textual feature as representation and adopt machine learning algorithms for classification. Representative textual features include word n-grams (Wang & Manning, 2012), text topic (Ganu, Elhadad, & Marian, 2009), bag-of-opinions (Qu, Ifrim, & Weikum, 2010), sentiment lexicon features (Kiritchenko, Zhu, & Mohammad, 2014). However, such methods are labor intensive and unable to extract the adequate information from textual data for text representation.

Motivated by the great success of deep learning in many fields, such as computer vision (Krizhevsky, Sutskever, & Hinton, 2012) and natural language processing (Bengio, Ducharme, Vincent, & Jauvin, 2003), recent studies leverage neural networks to learn social media content representation without any feature engineering and mostly achieve significantly higher performance compared with traditional machine learning methods. Nevertheless, existing relevant social media content representation approaches only provide limited improvement for individual SES prediction as they only consider a part of couplings among heterogeneous social media data. Meanwhile, a large number of methods mainly focus on pure social media text representation. For instance, Tang, Qin, and Liu (2015) proposes a gated recurrent neural network to learn vector-based document representation in a unified and bottom-up fashion for sentiment classification. Yang et al. (2016) presents a hierarchical attention network for document classification inspired by the hierarchical structure of documents, which only captures the hierarchical couplings of textual data. With neural network based methods, these works only capture a part of couplings of textual data, but ignore the effect of platform-based user level attributes on social media. Considering the potential contributions of these user level attributes, recent efforts introduce several attributes into text data. For example, Jin, Cao, Guo, Zhang, and Luo (2017) proposes using recurrent neural network (RNN) to fuse textual features and platform-based attributes through directly concatenating them. Like most of existing fusion methods, they regard the attributes to be independent of each other, which do not consider complex coupling relationships among them. Besides, several recent methods propose to fuse additional data sources like image and video into social media content (Farnadi, Tang, De Cock, & Moens, 2018; Jin et al., 2017; Zadeh, Chen, Poria, Cambria, & Morency, 2017). The key difference between our work and this strand of previous work lies in that we focus on learning the representation of general social media content, i.e., social media text and platform-based user level attributes.

3. The proposed model

In this paper, we propose a novel coupled social media content representation framework to infer users' SES, which incorporates structure-aware social media text representation and coupled user level attribute representation to exploit complex important coupling information of social media content, i.e., the order and hierarchical structure of social media text and implicit interaction relationships among platform-based user level attributes.

3.1. Problem statement

Assume that a social media user $u \in U$ has a set of posted microblogs $B = \{b_1, b_2, \ldots, b_n\}$ and the ith microblog $b_i \in B$ contains a sequence of words $\{w^i_1, w^i_2, \ldots, w^i_{l_i}\}$, where l_i is the length of ith microblog. Additionally, user u has a set of platform-based user level attributes $\{a_1, a_2, \ldots, a_m\}$, where m is the number of user level attributes. For each user, the proposed model aims at projecting the raw social media content into a vector representation, based on which we build a classifier to perform individual SES prediction.

3.2. Coupled social media content representation model

As shown in Fig. 2, we first present the structure-aware social media text representation method and the coupled user level attribute representation method. Then, two representations are aggregated into a vector representation of social media content. Finally, based on the social media content representation, we build a 3-way classifier to assign SES label to each social media user.

Structure-aware Social Media Text Representation. Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997), a variation of recurrent neural network, is widely adopted for textual data modeling due to its excellent performance on sequence modeling. LSTM is able to consider long-term dependencies of a sequence through introducing a memory cell. To model the semantic representation of social media text and consider the order couplings of text sequence, we leverage BiLSTM, Bidirectional LSTM, to represent the social media text both from forward and backward, which can increase the amount of input information available to the network compared with LSTM. It is worth to note that the employed method can be expanded/replaced by other methods. Besides, to take into account the hierarchical structure of social media text, inspired by the principle of compositionality (Frege, 2003), we model a social media user's text through a hierarchical structure composed of three levels, i.e., word-level, microblog-level and user-level.

As shown in Fig. 2, in the word level of the hierarchical structure, each word in a microblog b_i is embedded into a low dimensional semantic space, i.e., each word w^i_j is mapped to its embedding $\boldsymbol{w}^i_j \in \mathbb{R}^d$. The word embedding method and its settings will be described in Section 5.1. At each step, given an input word embedding $\boldsymbol{w}^i_j$, the current cell state $\boldsymbol{c}^i_j$ and hidden state $\boldsymbol{h}^i_j$ can be updated with the previous cell state $\boldsymbol{c}^i_{j-1}$ and hidden state $\boldsymbol{h}^i_{j-1}$ as follows:

$$\begin{bmatrix} \boldsymbol{i}^i_j \\ \boldsymbol{f}^i_j \\ \boldsymbol{o}^i_j \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \end{bmatrix} (\boldsymbol{W}[\boldsymbol{h}^i_{j-1}, \boldsymbol{w}^i_j] + \boldsymbol{b}), \tag{1}$$

$$\hat{\boldsymbol{c}}^i_j = \tanh(\boldsymbol{W}[\boldsymbol{h}^i_{j-1}, \boldsymbol{w}^i_j] + \boldsymbol{b}), \tag{2}$$

$$\boldsymbol{c}^i_j = \boldsymbol{f}^i_j \odot \boldsymbol{c}^i_{j-1} + \boldsymbol{i}^i_j \odot \hat{\boldsymbol{c}}^i_{j-1}, \tag{3}$$

$$\boldsymbol{h}^i_j = \boldsymbol{o}^i_j \odot \tanh(\boldsymbol{c}^i_j), \tag{4}$$

where $\boldsymbol{i}$, $\boldsymbol{f}$, $\boldsymbol{o}$ indicate gate activations, $\odot$ denotes element-wise multiplication, σ is the logistic sigmoid function and $\boldsymbol{W}$, $\boldsymbol{b}$ are the trainable parameters. Therefore, for a sequence of words $\{w^i_1, w^i_2, \ldots, w^i_{l_i}\}$, the forward LSTM reads the word sequence from w^i_1 to $w^i_{l_i}$ and the

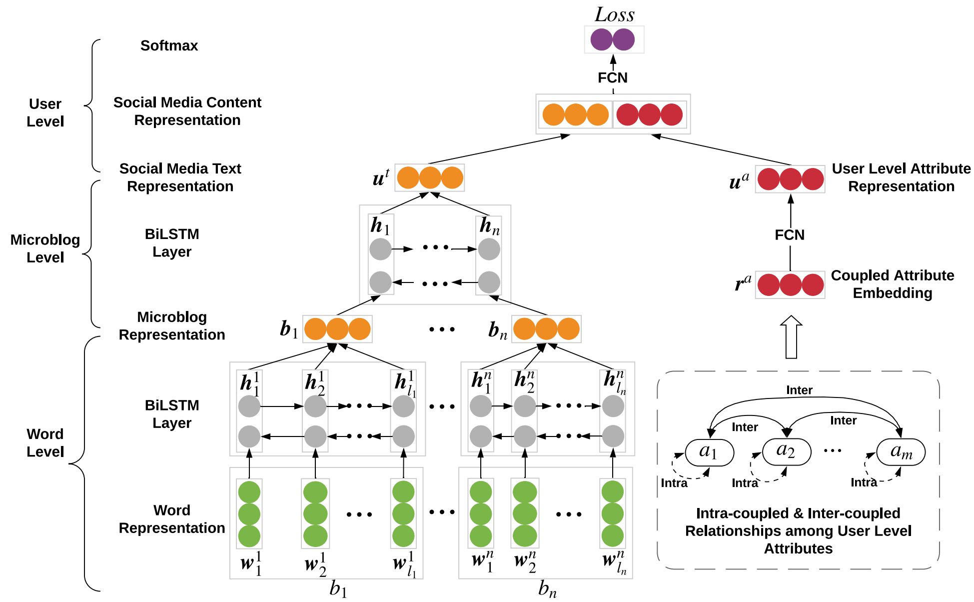


Fig. 2. The architecture of the proposed model.

backward LSTM reads the word sequence from $w^i_{l_i}$ to w^i_1. Then we concatenate the forward hidden state $\overrightarrow{\boldsymbol{h}^i_j}$ and the backward hidden state $\overleftarrow{\boldsymbol{h}^i_j}$, i.e., $\boldsymbol{h}^i_j = [\overrightarrow{\boldsymbol{h}^i_j};\overleftarrow{\boldsymbol{h}^i_j}]$, where $[.;.]$ denotes the concatenation operation. In BiLSTM, the hidden state $\boldsymbol{h}^i_j$ denotes the information of the whole sequence centered around $\boldsymbol{w}^i_j$. As a result, the BiLSTM network receives $[\boldsymbol{w}^i_1, \boldsymbol{w}^i_2, \dots, \boldsymbol{w}^i_{l_i}]$ and generates hidden states $[\boldsymbol{h}^i_1, \boldsymbol{h}^i_2, \dots, \boldsymbol{h}^i_{l_i}]$. Then we feed the hidden states to an average pooling layer to obtain the microblog text representation $\boldsymbol{b}_i$ for microblog b_i.

In the microblog level, given the microblog representation vectors of a user $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$, we also utilize BiLSTM to encode the microblogs as follows:

$$\overrightarrow{\boldsymbol{h}_i} = \overrightarrow{LSTM}(\boldsymbol{b}_i), \tag{5}$$

$$\overleftarrow{\boldsymbol{h}_i} = \overleftarrow{LSTM}(\boldsymbol{b}_i), \tag{6}$$

We then concatenate the forward hidden state $\overrightarrow{\boldsymbol{h}_i}$ and the backward hidden state $\overleftarrow{\boldsymbol{h}_i}$, i.e., $\boldsymbol{h}_i = [\overrightarrow{\boldsymbol{h}_i};\overleftarrow{\boldsymbol{h}_i}]$. $\boldsymbol{h}_i$ summarizes the neighbor microblogs around the ith microblog but still focus on the ith microblog. Then we feed the hidden states to an average pooling layer to obtain the final social media text representation $\boldsymbol{u}^t$ for user u.

Coupled User Level Attribute Representation. Social media users generally have various platform-based user level attributes, e.g., the number of followees, the number of microblogs, etc, which could make contributions to the representation of social media content for inferring individual SES. Existing related work mostly leverage original user level attributes without considering relations among attributes. However, attributes are more or less coupled via explicit or implicit relationships (Cao, Ou, & Philip, 2011; Wang et al., 2013). For example, some attributes may share the same statistical information on social media. Therefore, it is natural to hypothesize that the user level attributes are related to each other in some way. To this end, we propose a coupled attribute representation method inspired by Wang et al. (2013) to efficiently capture the important coupling information of user level attributes.

More specifically, as illustrated in Fig. 3, we consider two kinds of interaction relations among platform-based user level attributes: the intra-coupled interaction within an attribute with the correlations between every attribute and its own powers, and the inter-coupled interaction among different attributes with the correlations between each attribute and the powers of other attributes.

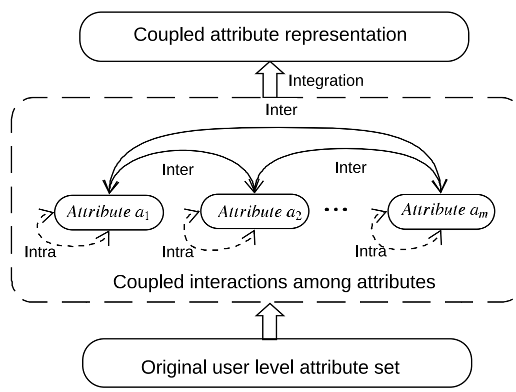


Fig. 3. An overview of coupled user level attribute representation.

Firstly, we map the original attribute space to an expanded space for incorporating linear and nonlinear information through a power expansion as follows:

$$\{\langle a_1\rangle^1, \langle a_1\rangle^2, \dots, \langle a_1\rangle^L, \langle a_2\rangle^1, \langle a_2\rangle^2, \dots, \langle a_2\rangle^L, \dots, \langle a_m\rangle^1, \langle a_m\rangle^2, \dots, \langle a_m\rangle^L\} \tag{7}$$

where $\langle a_j\rangle^p (1 \le p \le L, 1 \le j \le m)$ denotes the pth power of the corresponding value of attribute a_j.

Leveraging the power expansion, the intra-coupled interaction within an attribute a^n_j is defined as an $L \times L$ matrix $\boldsymbol{M}_{Ia}(a_j)$, with considering the correlations between the attribute a_j and its own powers $\langle a_j\rangle^p$.

$$\boldsymbol{M}_{Ia}(a_j) = \begin{pmatrix} \theta_{11}(j) & \theta_{12}(j) & \cdots & \theta_{1L}(j) \\ \theta_{21}(j) & \theta_{22}(j) & \cdots & \theta_{2L}(j) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{L1}(j) & \theta_{L2}(j) & \cdots & \theta_{LL}(j) \end{pmatrix}, \tag{8}$$

where $\theta_{pq}(j)$ denotes the Pearson product–moment correlation coefficient between $\langle a_j\rangle^p$ and $\langle a_j\rangle^q$. Here, we use the revised correlation coefficient by taking account of the p-values for testing the hypothesis of no correlation between attributes, i.e., if p-value is no less than 0.05, the correlation coefficient is set to 0.

Besides, the inter-coupled interaction between numerical attribute a_j and other attributes a_k $(k \neq j)$ is defined as an $L \times L \cdot (m-1)$ matrix $\boldsymbol{M}_{Ie}(a_j|\{a_k\}_{k\neq j})$.

$$\boldsymbol{M}_{Ie}(a_j|\{a_k\}_{k\neq j})$$
$$= \begin{pmatrix} \delta_{11}(j,k_1) & \cdots & \delta_{1L}(j,k_1) & \cdots & \delta_{11}(j,k_{m-1}) & \cdots & \delta_{1L}(j,k_{m-1}) \\ \delta_{21}(j,k_1) & \cdots & \delta_{2L}(j,k_1) & \cdots & \delta_{21}(j,k_{m-1}) & \cdots & \delta_{2L}(j,k_{m-1}) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \delta_{L1}(j,k_1) & \cdots & \delta_{LL}(j,k_1) & \cdots & \delta_{L1}(j,k_{m-1}) & \cdots & \delta_{LL}(j,k_{m-1}) \end{pmatrix}, \tag{9}$$

where $\delta_{pq}(j,k_i)$ denotes the Pearson product–moment correlation coefficient between $\langle a_j\rangle^p$ and $\langle a_{k_i}\rangle^q$, and $\{a_k\}_{k\neq j} = \{a_{k_1},\dots,a_{k_{m-1}}\}$ is the set of attributes other than a_j.

For user u_i, the attribute values of a_j and its powers are presented as a vector:

$$\widetilde{\boldsymbol{z}}_i(a_j) = [\langle v_{ij}\rangle^1, \langle v_{ij}\rangle^2, \dots, \langle v_{ij}\rangle^L], \tag{10}$$

while the attribute values of other attributes $\{a_k\}_{k\neq j}$ and their powers are denoted as another vector:

$$\widetilde{\boldsymbol{z}}_i(\{a_k\}_{k\neq j}) = [\langle v_{ik_1}\rangle^1, \langle v_{ik_1}\rangle^2, \dots, \langle v_{ik_1}\rangle^L, \dots,$$
$$\langle v_{ik_{m-1}}\rangle^1, \langle v_{ik_{m-1}}\rangle^2, \dots, \langle v_{ik_{m-1}}\rangle^L]. \tag{11}$$

Here, the attribute value of user u_i on attribute a_j is v_{ij}. We incorporate the intra-coupled interaction and the inter-coupled interaction into a new coupled attribute representation, a $1 \times L$ vector $\boldsymbol{r}_i(a_j)$, for user object u_i on the numerical attribute a_j as follows:

$$\boldsymbol{r}_i(a_j) = \widetilde{\boldsymbol{z}}_i(a_j) \odot \boldsymbol{w} \otimes [\boldsymbol{M}^n_{Ia}(a_j)]^T$$
$$+ \widetilde{\boldsymbol{z}}_i(\{a_k\}_{k\neq j}) \odot \underbrace{[\boldsymbol{w}, \boldsymbol{w}, \dots, \boldsymbol{w}]}_{m-1} \otimes [\boldsymbol{M}^n_{Ie}(a_j|\{a_k\}_{k\neq j})]^T, \tag{12}$$

where $\boldsymbol{w} = [1/(1!), 1/(2!), \dots, 1/(L!)]$, $\odot$ denotes the Hadamard product and $\otimes$ indicates the matrix multiplication. After considering all the d_n original numerical attributes, we obtain the final coupled user level attribute representation for the user object u_i as follows:

$$\boldsymbol{r}^a_i = [\boldsymbol{r}_i(a_1), \boldsymbol{r}_i(a_2), \dots, \boldsymbol{r}_i(a_m)] \in \mathbb{R}^{L\cdot m} \tag{13}$$

To further capture the latent relationships between high level features, we link the raw attribute vector $\boldsymbol{r}^a$ to the k-length representation vector $\boldsymbol{u}^a$ in terms of a fully connected network as follows:

$$\boldsymbol{u}^a = \boldsymbol{W}^a \boldsymbol{r}^a + \boldsymbol{b}^a \tag{14}$$

where $\boldsymbol{W}^a$ and $\boldsymbol{b}^a$ are both learnable parameters, encoding the interaction strength over attributes in the fully-connected layer.

Consequently, in the user level, we concatenate the social media text representation and the coupled user level attribute representation to obtain the social media content representation $\boldsymbol{u} = [\boldsymbol{u}^t; \boldsymbol{u}^a]$.

Individual SES Prediction. We employ a fully connected network layer and a softmax layer to project the social media content representation $\boldsymbol{u}$ into SES distribution of C classes as follows:

$$p = softmax(\boldsymbol{W}^u \boldsymbol{u} + \boldsymbol{b}^u). \tag{15}$$

In this model, the cross-entropy error between ground truth SES level distribution and predicted SES level distribution is defined as loss function for optimization when training:

$$L = -\sum_{u\in U}\sum_{c=1}^{C} p^g_c(u) \cdot \log(p_c(u)), \tag{16}$$

where p^g_c denotes the probability of SES label c with ground truth being 1 and others being 0, and U represents the training social media users. p_c is the predicted probability of SES label c.

Fig. 4. A demonstration of user search function in Sina Weibo.

4. Data collection and preprocessing

We create a real data set which contain social media users' contents and convincing SES labels for SES prediction task. This section presents the data collection and preprocessing in details.

4.1. Data collection

In the field of sociology, several studies have shown that socioeconomic index is correlated with occupational status (Blau & Duncan, 1967; Hauser & Warren, 1997; Li, 2005; Treiman, 1977) and provide some simplified mapping between SES and occupations like the Standard Occupation Classification (SOC) hierarchy attached to socioeconomic categorizations in conjunction with the National Statistics Socio-Economic Classification (Elias & Birch, 2010; Rose & Pevalin, 2010).

To create a real dataset that can be used for predicting SES of social media users, according to the China Occupation Classification, for each major occupation we queried Sina Weibo's search API to retrieve a maximum of 500 user accounts whose certificated person card best matched the occupation keywords. For example, as shown in Fig. 4, after searching users with the occupation keyword "CEO", the best matched users whose certified person cards contain the keyword are listed. To remove potential ambiguity in the raw user set, we manually inspected accounts and filtered out those which belong to companies and other occupations. After that, we collected the rest users' microblogs posted from February 2017 to February 2018 and their public platform-based user level attributes. Fig. 5 demonstrates a sample of social media content of a user in Sina Weibo. Finally, we extracted active users who published more than ten microblogs during this given period. In total, about 50% of the accounts were removed after this filtering process. As a result, the final data set consists of 20452 users from 73 occupations and 6,893,746 unique microblogs. For the user privacy, the collected data set and any further research results can only be used for this work.

To obtain a convincing SES label for each user, sociology researchers who have studied the socioeconomic status of occupations in China for a long time were invited to assign a high (level A), middle (level B) or low (level C) SES to each user in our data set according to

Fig. 5. A sample of social media content in Sina Weibo.

the occupation information. Level A mainly contains higher managerial, administrative and professional occupations, such as CEO, judge, mayor, etc. Level B contains many intermediate occupations, such as teacher, actor, programmer, etc. Level C contains routine and manual occupations, such as shop assistant, deliveryman, repairman, etc. The distribution of users across SES levels is 3,974 users with level A, 12,451 users with level B and 4,027 users with level C. Finally, after undersampling for ensuring data set balance, the final distribution of users across classes is 3,974 users with level A, 4,004 users with level B and 4,027 users with level C.

4.2. Data preprocessing

For the collected dataset, We did the following data preprocessing. With regard to the social media text data, we remove punctuation, non-Chinese words, digits, and specific symbols, meanwhile converting all the traditional Chinese words in the social media text into simplified Chinese words. Then we choose to leverage a Chinese Language Technology Platform (LTP) (Che, Li, & Liu, 2010), an integrated Chinese processing platform which includes a suite of high performance natural language processing (NLP) modules and relevant corpora, to segment Chinese text of each microblog into a sequence of Chinese words. Besides, a separate Chinese Wikipedia dataset[2] is used as a reference corpus in order to build the word embedding representations. Regarding the platform-based user level attributes, in this paper, we extract seven user level attributes, i.e., the number of followers, followees, microblogs, proportion of forwarded microblogs, the average number of favorites, forwarded, comments per microblog.

5. Experiments and evaluation

We conduct extensive experiments on our built Sina Weibo dataset to demonstrate the efficiency and robustness of the proposed model on the individual SES prediction task.

5.1. Experimental settings

We employ distributed representation method (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) to build word embedding for social media text. Specifically, we only retain words which appear more than 5 times to build the vocabulary based on our whole textual dataset and a separate Chinese Wikipedia dataset. Then we pre-train the Word2Vec model with the vocabulary in an unsupervised fashion with default parameter settings. Finally, we obtain a 50-dimensional word embedding vector for each word in the dataset. The word embedding is capable of capturing context of a word in a document, semantic similarity and relation with other words. In addition, we leverage the coupled attribute representation for the platform-based user level attribute embedding. To take advantage of information from testing objects, we use both the training and testing objects' attributes in the coupled attribute representation within an unsupervised manner. For the user level attributes, the original attribute dimension is 7. We set the power expansion value $L = 6$ so that the extracted coupled attribute dimension is 42.

In the experiments, we set the dimension of the hidden states in LSTM cell to be 32 so that a combination of forward and backward LSTMs gives us 64 dimensions for microblog and social media content annotations. In order to speed up training, we limit that the maximal length of every microblog is 40 words and a social media user has 50 microblogs at most. We use 80% of the data for training and the remaining 20% for testing. We use Adam (Kingma & Ba, 2015) to update parameters with setting initial learning rate as 0.005 when training. In order to ensure the soundness and robustness of experimental results, the procedure of training data repeats 10 times and we report the averaged prediction performance as final results.

Table 1
SES prediction performance for the baseline models and the proposed model.

Models	Accuracy	Precision	Recall	F1-score
Feature engineering based models				
LR	0.3972	0.3461	0.3973	0.2951
SVM	0.4832	0.4484	0.4825	0.4330
GP	0.5563	0.5544	0.5560	0.5519
Neural network based models				
RNN	0.6215	0.6280	0.6212	0.6168
SRNN	0.6323	0.6471	0.6319	0.6274
att-SRNN	0.6498	0.6777	0.6493	0.6413
CORE	**0.6689**	**0.6880**	**0.6684**	**0.6611**

5.2. Performance evaluation

To evaluate the performance of the proposed model, we compare CORE with two groups of benchmark methods.

The first group consists of feature engineering based models in previous related works (Lampos et al., 2016; Preoţiuc-Pietro, Lampos et al., 2015; Preoţiuc-Pietro, Volkova et al., 2015). These methods first extract several kinds of features, which contains platform-based user level attributes and textual features extracted from social media text (i.e., the frequency of the 1-grams and the frequency distribution across latent topics represented by clusters of 1-grams (Lampos et al., 2016)). Then, existing classic machine learning methods are applied, including logistic regression (**LR**) with Elastic Net regularization, Support Vector Machine (**SVM**), and Gaussian Process (**GP**), for classification.

The other group is composed of neural network based methods, which are widely leveraged in recent text classification related work. There have been a variety neural networks proposed for text-based classification. In this work, we mainly focus on the coupling methods used in these works not the neural network itself. Hence, we choose the following baseline methods:

- **RNN** represents each word with the word embedding vector and feeds each user's word embedding vectors into the Recurrent Neural Network (RNN) (Zhang, Li, Wang, & Zhou, 2016). Afterwards, the hidden vectors of RNN are averaged to obtain social media text representation for individual SES prediction.
- **SRNN** considers the hierarchical structure of social media text following the Hierarchical Attention Network (HAN) (Yang et al., 2016). We first likewise construct a user level social media text

[2] https://dumps.wikimedia.org/zhwiki/

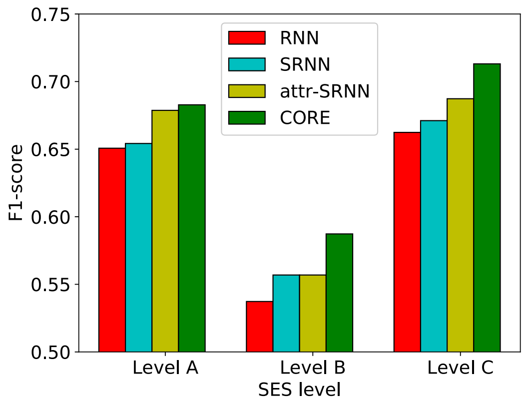


Fig. 6. Performance comparison for each SES level.

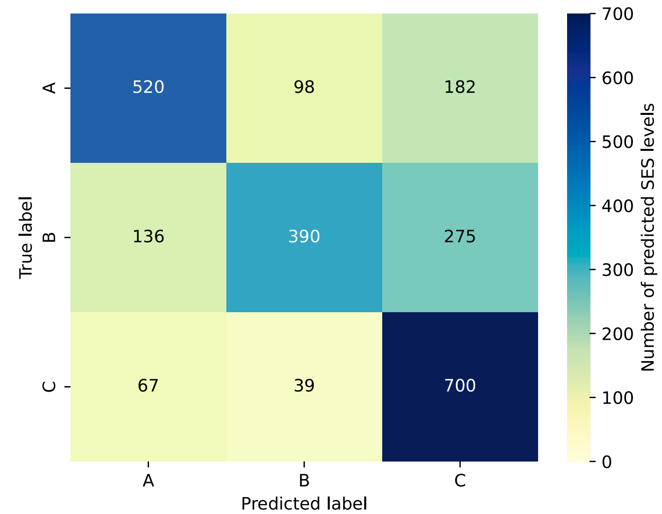


Fig. 7. The confusion matrix for SES prediction.

Table 2
Performance of the proposed model for each SES level.

	Level A	Level B	Level C
Precision	0.7192	0.7400	0.6050
Recall	0.6500	0.4868	0.8684
F1-score	0.6828	0.5873	0.7131

representation by first building representation of microblogs with word embedding and then aggregating those into a user-level representation.

- **attr-SRNN** leverages SRNN to represent social media text, and then combines platform-based user level attributes without considering coupled attribute representation to represent social media content for the individual SES prediction task.

To show the robustness of our experimental results, we further employ BiLSTM in the above baseline methods. The hyperparameters of BiLSTM in the baseline models are same as our proposed model. We report experimental results of all methods in terms of accuracy, precision, recall and F1-score. Particularly, accuracy is calculated as the number of correctly predicted testing samples divided by the total number of testing samples. For each SES level, *precision* is defined as the fraction of correctly predicted positive observations over the total predicted positive observations. *Recall* is calculated as the number of correctly predicted positive observations divided by the number of the all observations in actual class. *F1-score* is the harmonic mean of Precision and Recall, which is calculated as:

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

For the 3-way classification, precision, recall and F1-score are macro-averaged, which take into account the skewed class label distributions by weighting each class uniformly.

As illustrated in Table 1, we can observe that **CORE** greatly outperforms the baseline models in terms of all metrics. Compared with the neural network based methods, three machine learning based baseline methods have much lower performance, which indicates that the extracted user level features and textual features cannot represent social media content very well. This is because the traditional feature engineering methods are unable to capture some important coupling information of social media content, such as the order of text sequence, the hierarchy of social media text and latent interaction relationships among user level attributes. Although only considering social media text representation, **RNN** significantly outperforms these machine learning based methods with about 6%–13%, 7%–28%, 7%–23% and 6%–32% higher performance scores in terms of accuracy, precision, recall and F1-score respectively. This implies that **RNN** can learn social media text representation much better compared with predefined textual features owing to considering the order of text sequence. Due to considering the hierarchical structure of social media text, **SRNN** achieves higher performance than **RNN**. In addition, **attr-SRNN** enhances the prediction performance with about 1.4%–3% higher performance score, which proves that appropriately fusing user level attributes can improve the ability of representing social media content. Furthermore, through considering the linear and nonlinear relationships among platform-based user level attributes, **CORE** can take into account the couplings in the social media text as well as various couplings among user level attributes, which significantly improves the prediction performance compared with baseline models.

Performance of each SES level. In addition, we plot Fig. 6 to demonstrate the prediction performance of these models for each SES level in terms of F1-score. We can observe that the proposed **CORE** has the highest F1-score on each SES level prediction task, which further validates the effect of these three types of key coupling information of social media content on the individual SES prediction.

Table 2 shows the detailed prediction performance of **CORE** for each SES level in terms of precision, recall and F1-score. We can observe that relatively more users are wrongly assigned as Level C and about half Level B users are assigned as the other SES levels. In terms of F1-score, we can observe that it is more difficult to correctly classify users from the Level B class (lowest F1 score). Fig. 7 illustrates the confusion matrix for the SES prediction results of the proposed model. Intuitively, we can also observe that Level B users are more likely to be predicted as Level C, which may be because there exist some similar platform behaviors between some Level B user and Level C users. In addition, another possible reason why the prediction performance is limited is that significant differences exist between the reality and what some users disclose on social media. In the future, we will investigate ways to further improve the performance by better understanding the nature of these errors in the model.

Coupled attribute representation analysis. To validate the performance and robustness of the coupled attribute representation, we evaluate the performance of **CORE** by varying the power expansion value L from 2 to 10 and compare it with **attr-SRNN** which does not consider the coupled attribute representation. Note that L is a hyperparameter in the coupled attribute representation. In Fig. 8, we present the performance of **CORE** and **attr-SRNN** over different power expansion values in terms of accuracy, precision, recall and F1-score.

For all the evaluation metrics, the proposed **CORE**, considering the coupled user level attribute representation, outperforms **attr-SRNN** no

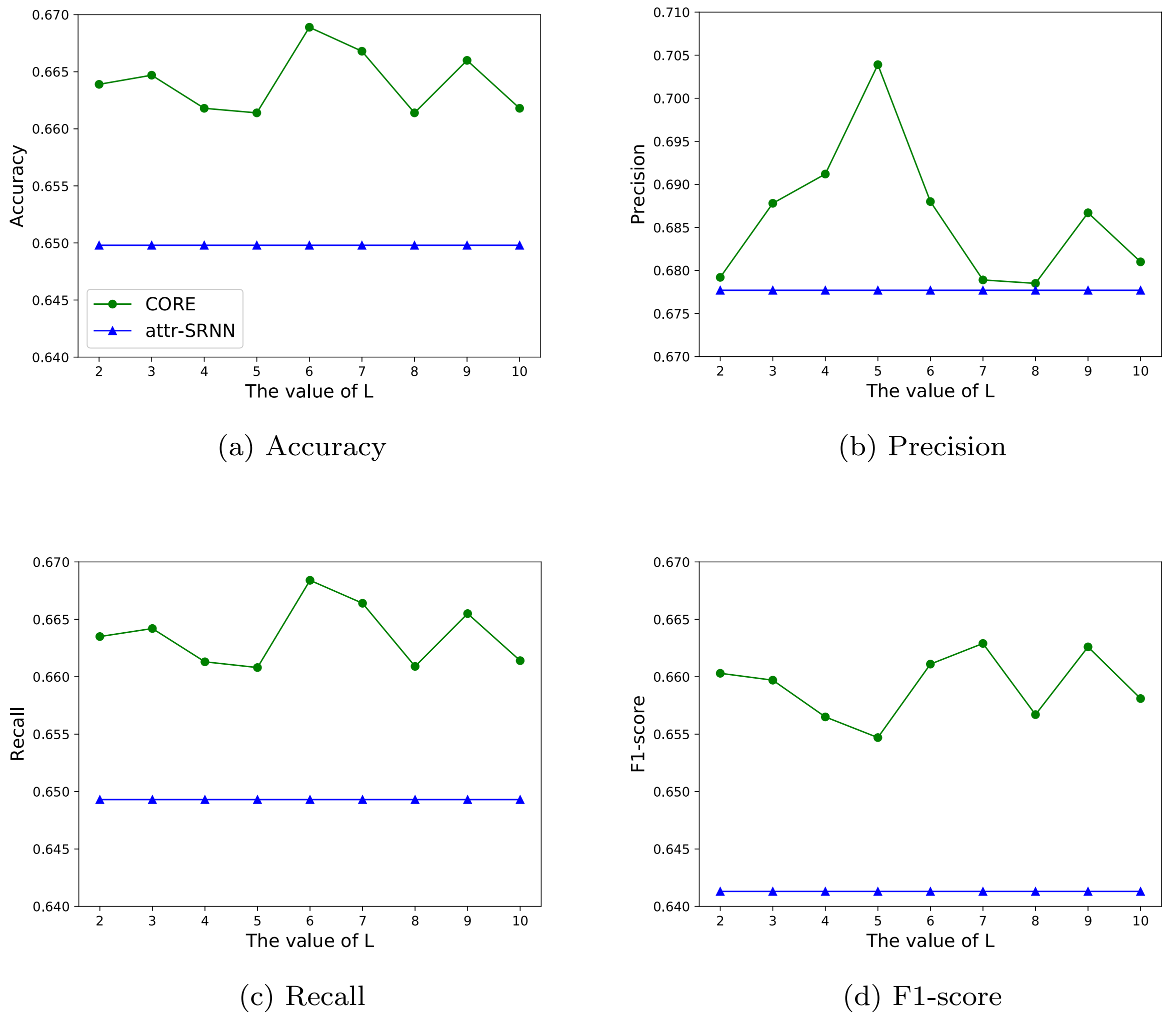


Fig. 8. Performance over various power expansion value L.

matter what the L value is, which validate the efficiency and robustness of the coupled attribute representation method. To be more specific, the proposed **CORE** improves by 1.1%–1.9% (Accuracy), 0.2%–2.6% (Precision), 1.2%–1.9% (Recall), and 1.3%–2.2% (F1-score) for the individual SES prediction. That is to say, fusing the coupled user level attribute representation can assist in enhancing the performance of individual SES prediction.

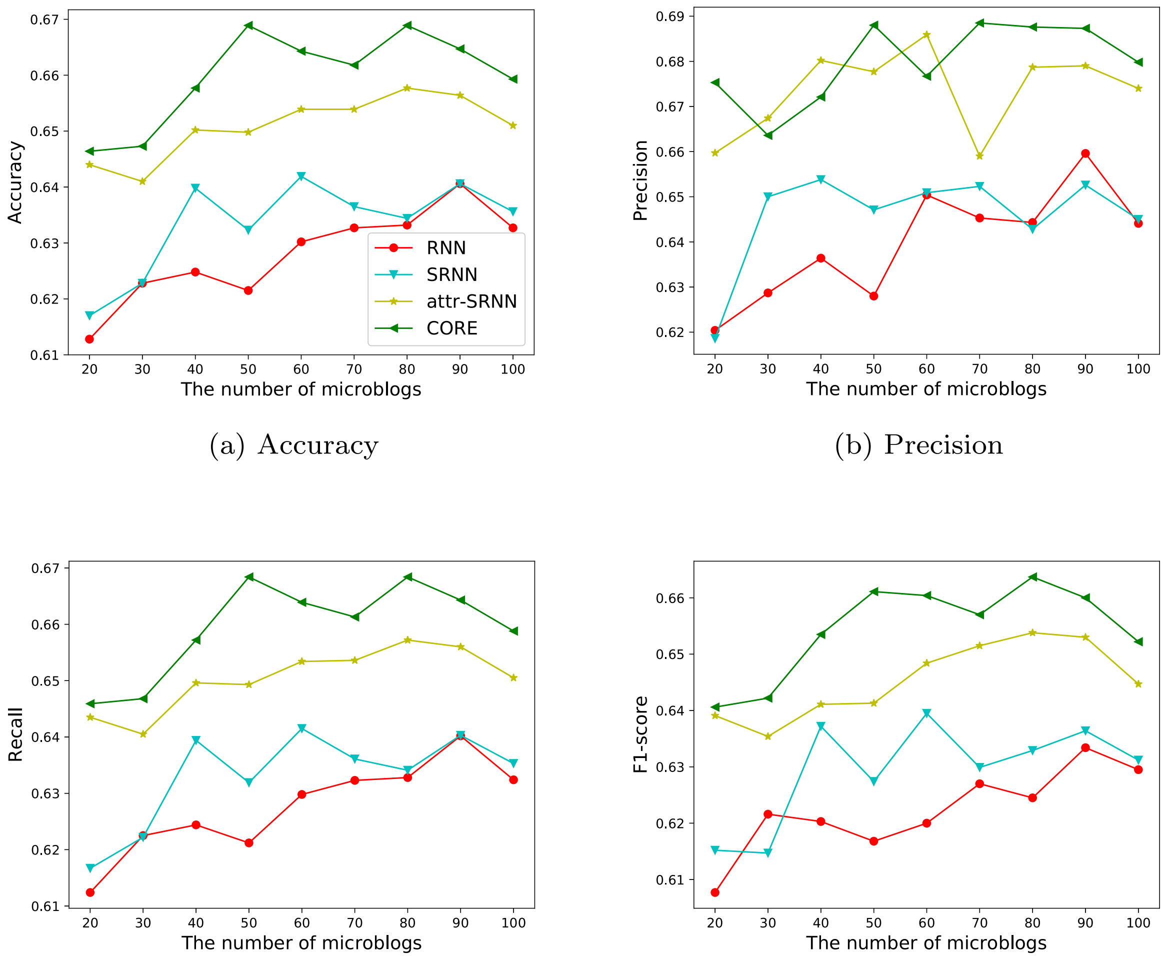
Performance comparison over microblog numbers. To further investigate the performance and robustness of the proposed model over social media content with various microblog numbers, Fig. 9 compares the performance of the proposed **CORE** and other three neural network based baseline models **RNN**, **SRNN** and **attr-SRNN** under different microblog number settings (i.e., maximum microblog number parameter).

From Fig. 9, we can observe the changing performance of four models over different microblog number in terms of accuracy, precision, recall and F1-score. Particularly, the proposed **CORE** consistently outperforms other baseline models for all microblog numbers in terms of accuracy, recall and F1-score. For the precision metric, **CORE** mostly has better performance than other models. It indicates the robustness and flexibility of our model on dataset of different scales. The performance changes over various microblog number mostly show some fluctuations when the number of microblogs ranges from 50 to 100. There might be several reasons resulting in these fluctuations. On one hand, most of Sina Weibo users did not publish more than 50 microblogs during that period, which would incur bias to these users with many microblogs. On the other hand, although considering more microblogs, there might exist many duplicated or noisy microblogs, which would affect the prediction performance.

6. Conclusion and future work

In this paper, we propose a novel coupled social media content representation model called CORE for the individual SES prediction, which can consider three important coupling information of social media content. On one hand, it presents a structure-aware social media text representation method based on recurrent neural network to incorporate the order and the hierarchical structure of social media text. On the other hand, it devices a coupled attribute representation method to take into account intra-coupled and inter-coupled interaction relationships among platform-based user level attributes. From extensive experiments on the built Sina Weibo dataset, we validate the efficiency and robustness of the proposed model by comparing other benchmark models.

In the future, there are many potential research directions of this work. First, we will explore to incorporate microblog level attributes and coupling information between attributes and social media text to improve our model. Second, in fact, most attributes contain categorical and numerical attributes. Therefore, in the next step, we plan to study the embedding representation of categorical attributes and the



(a) Accuracy (b) Precision

(c) Recall (d) F1-score

Fig. 9. Performance over various microblog number.

method of capturing the couplings between categorical and numerical attributes. Third, to reduce the effect of sampling bias in the collected data set, we plan to consider the relationships between users through exploiting their following relations and other implicit relations. Fourth, we will expand our experiments to further analyze the effects of hyper-parameters in our model. Finally, we plan to apply the proposed model to different social media datasets to further verify its efficiency.

CRediT authorship contribution statement

Tao Zhao: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Investigation, Writing – original draft. **Lu Tang:** Visualization, Investigation, Writing – reviewing & editing. **Jinfeng Huang:** Methodology, Investigation, Supervision, Project administration, Writing – reviewing & editing. **Xiaoming Fu:** Methodology, Supervision, Writing – reviewing & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 824019. We would like to thank Prof. Jar-der Luo and Prof. Jingming Liu and their graduate students from the Department of Sociology, Tsinghua University, for their assistance in understanding sociological theories and help in obtaining data labels. We also thank anonymous reviewers for their critical comments on early drafts of this paper.

References

Abitbol, L., & Morales, A. (2021). Socioeconomic patterns of twitter user activity. *Entropy*, *23*, 780.

Aletras, N., & Chamberlain, B. P. (2018). Predicting twitter user socioeconomic attributes with network and language information. In *Proceedings of the 29th on hypertext and social media* (pp. 20–24). ACM.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*, 1137–1155.

Blau, P. M., & Duncan, O. D. (1967). The american occupational structure.. *American Journal of Sociology*, *33*(2), 296.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Borges, G. R., Almeida, J. M., Pappa, G. L., et al. (2014). Inferring user social class in online social networks. In *Proceedings of the 8th workshop on social network mining and analysis* (p. 10). ACM.

Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating gender on twitter. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1301–1309). Association for Computational Linguistics.

Cao, L., Ou, Y., & Philip, S. Y. (2011). Coupled behavior analysis with applications. *IEEE Transactions on Knowledge and Data Engineering*, *24*(8), 1378–1392.

Carlsson-Kanyama, A., & Linden, A.-L. (1999). Travel patterns and environmental effects now and in the future:: implications of differences in energy consumption among socio-economic groups. *Ecological Economics*, *30*(3), 405–417.

Che, W., Li, Z., & Liu, T. (2010). Ltp: a chinese language technology platform. *Journal of Chinese Information Processing*, *2*(6), 13–16.

Culotta, A. (2010). Towards detecting influenza epidemics by analyzing twitter messages. In *Proceedings of the first workshop on social media analytics* (pp. 115–122). ACM.

Ding, S., Gao, X., Dong, Y., Tong, Y., & Fu, X. (2021). Estimating multiple socioeconomic attributes via home location – A case study in China. *Journal of Social Computing*, *2*(1), 71–88.

Ding, S., Huang, H., Zhao, T., & Fu, X. (2019). Estimating socioeconomic status via temporal-spatial mobility analysis – A case study of smart card data. In *Proceedings of the 28th IEEE international conference on computer communication and networks (ICCCN 2019)*.

Elias, P., & Birch, M. (2010). SOC2010: revision of the standard occupational classification. *Economic & Labour Market Review*, *4*(7), 48–55.

Farnadi, G., Tang, J., De Cock, M., & Moens, M.-F. (2018). User profiling through deep multimodal fusion. In *Proceedings of the eleventh ACM international conference on web search and data mining* (pp. 171–179). ACM.

Frege, G. (2003). On sense and reference. *oversatt av Max Black, i J. Guitérrez-Rexach (red.): Semantics: Crictical Concepts in Linguistics*, *1*, 7–25.

Ganu, G., Elhadad, N., & Marian, A. (2009). Beyond the stars: improving rating predictions using review text content. In *Proceedings of the 12th international workshop on the web and databases*.

Gao, J., Zhang, Y.-C., & Zhou, T. (2019). Computational socioeconomics. *Physics Reports*, *817*, 1–104.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 253–262). ACM.

Hasanuzzaman, M., Kamila, S., Kaur, M., Saha, S., & Ekbal, A. (2017). Temporal orientation of tweets for predicting income of users. In *Proceedings of the 55th annual meeting of the association for computational linguistics* (pp. 659–665).

Hauser, R. M., & Warren, J. R. (1997). Socioeconomic indexes for occupations: A review, update, and critique. *Sociological Methodology*, *27*(1), 177–298.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Huang, Q., & Wong, D. W. (2016). Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us?. *International Journal of Geographical Information Science*, *30*(9), 1873–1898.

Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on multimedia* (pp. 795–816). ACM.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd international conference on learning representations*.

Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, *50*, 723–762.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lampos, V., Aletras, N., Geyti, J. K., Zou, B., & Cox, I. J. (2016). Inferring the socioeconomic status of social media users based on behaviour and language. In *European conference on information retrieval* (pp. 689–695). Springer.

Lampos, V., Miller, A. C., Crossan, S., & Stefansen, C. (2015). Advances in nowcasting influenza-like illness rates using search query logs. *Scientific Reports*, *5*, 12760.

Lampos, V., Preoţiuc-Pietro, D., & Cohn, T. (2013). A user-centric model of voting intention from social media. In *Proceedings of the 51st annual meeting of the association for computational linguistics* (pp. 993–1003).

Li, C. (2005). Prestige stratification in the contemporary China: occupational prestige measures and socio-economic index. *Sociological Studies*, *2*, 74–102.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Preoţiuc-Pietro, D., Lampos, V., & Aletras, N. (2015). An analysis of the user occupational class through twitter content. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing* (pp. 1754–1764).

Preoţiuc-Pietro, D., Volkova, S., Lampos, V., Bachrach, Y., & Aletras, N. (2015). Studying user income through language, behaviour and affect in social media. *PloS One*, *10*(9).

Propper, C., Damiani, M., Leckie, G., & Dixon, J. (2007). Impact of patients' socioeconomic status on the distance travelled for hospital admission in the english national health service. *Journal of Health Services Research & Policy*, *12*(3), 153–159.

Qu, L., Ifrim, G., & Weikum, G. (2010). The bag-of-opinions method for review rating prediction from sparse text patterns. In *Proceedings of the 23rd international conference on computational linguistics* (pp. 913–921). Association for Computational Linguistics.

Rao, D., Yarowsky, D., Shreevats, A., & Gupta, M. (2010). Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on search and mining user-generated contents* (pp. 37–44). ACM.

Rose, D., & Pevalin, D. (2010). *Re-basing the NS-SEC on SOC2010: A report to ONS*: *Techincal Report*, University of Essex.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417–453.

Soto, V., Frias-Martinez, V., Virseda, J., & Frias-Martinez, E. (2011). Prediction of socioeconomic levels using cell phone records. *User Modeling, Adaption and Personalization*, 377–388.

Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1422–1432).

Tchokni, S. E., Séaghdha, D. O., & Quercia, D. (2014). Emoticons and phrases: status symbols in social media. In *Proceedings of the eighth international conference on weblogs and social media* (pp. 485–494).

Treiman, D. (1977). Index - occupational prestige in comparative perspective. *American Journal of Sociology*, *85*, 511–514.

Wang, J., Gao, J., Liu, J.-H., Yang, D., & Zhou, T. (2019). Regional economic status inference from information flow and talent mobility. *EPL (Europhysics Letters)*, *125*(6).

Wang, S., & Manning, C. D. (2012). Baselines and bigrams: simple, good sentiment and topic classification. In *Proceedings of the 50th annual meeting of the association for computational linguistics* (pp. 90–94). Association for Computational Linguistics.

Wang, C., She, Z., & Cao, L. (2013). Coupled attribute analysis on numerical data. In *Twenty-third international joint conference on artificial intelligence*.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: human language technologies* (pp. 1480–1489).

Zadeh, A., Chen, M., Poria, S., Cambria, E., & Morency, L.-P. (2017). Tensor fusion network for multimodal sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 1103–1114). Association for Computational Linguistics.

Zhang, D., Li, S., Wang, H., & Zhou, G. (2016). User classification with multiple textual perspectives. In *Proceedings of the 26th international conference on computational linguistics: technical papers* (pp. 2112–2121).