User Classification with Multiple Textual Perspectives

Dong Zhang, Shoushan Li*, Hongling Wang, Guodong Zhou
Natural Language Processing Lab
School of Computer Science and Technology, Soochow University, China
dzhang@stu.suda.edu.cn, lishoushan@suda.edu.cn
hlwang@suda.edu.cn, gdzhou@suda.edu.cn

Abstract

Textual information is of critical importance for automatic user classification in social media. However, most previous studies model textual features in a single perspective while the text in a user homepage typically possesses different styles of text, such as *original message* and *comment from others*. In this paper, we propose a novel approach, namely ensemble LSTM, to user classification by incorporating multiple textual perspectives. Specifically, our approach first learns a LSTM representation with a LSTM recurrent neural network and then presents a joint learning method to integrating all naturally-divided textual perspectives. Empirical studies on two basic user classification tasks, i.e., gender classification and age classification, demonstrate the effectiveness of the proposed approach to user classification with multiple textual perspectives.

1 Introduction

User attribute classification, also namely user classification for short, is a task which aims to leverage user-generated content to automatically predict user's attributes, such as gender (Wang et al., 2015), age (Rao et al., 2010; Sap et al., 2014) and location (Cheng et al., 2010). Recently, the growth of online social networks provides the opportunity to perform user classification in a broader context (Bollen et al., 2011; Sadilek et al., 2012; Lampos and Cristianini, 2010; Zamal et al., 2012). Basically, user classification is a fundamental task not only in sociolinguistic studies, but also in many real applications, such as recommender systems, and online advertising (O'Connor et al., 2010; Preotiuc-Pietro et al, 2015).

Text style	***User A*** ***Gender: female***	***User B*** ***Gender: male***
Original Message	"*Just bought the lipstick, look beautiful?*"	"*The first day, hard work.*"
Retweeted Message	"*Seaweed mask, it is so remarkably efficient.*"	"*Love her, take her to see the sea.*"
Comment from others	"*Sister, you're so pretty!*"	"*Go to see my latest message*"
Comment to others	"*Thanks.*"	"*Sister, you're so pretty!*"

Table 1: Some examples of different text styles in two users' homepages in a social media

Currently, machine learning approaches have dominated the research on user classification where statistic classifiers are learned with labeled data and various kinds of features, such as textual features,

* Corresponding author

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2112–2121, Osaka, Japan, December 11-17 2016.

behavior features, and social connection features (Preotiuc et al., 2015, Lampos et al., 2016). Among these features, textual features are most popular and they are good clues to infer the user attributes (Zhu et al., 2015; Li et al., 2015). For example, in Table 1, User A publishes a text "*Just bought the lipstick, look beautiful?*" which could be used to infer the user to be a *female* since *females* are more likely to buy a lipstick.

However, user-generated text sometimes possesses different styles, especially in social media. For instance, in Table 1, a homepage in a social media contains at least four kinds of text, namely *Original Message*, *Retweeted Message*, *Comment From Others*, and *Comment To Others*. Almost all previous studies do not distinguish these different styles of text, which might hurt the classification performance. For instance, in Table 1, User A has a *Comment From Others* "*Sister, you're so pretty!*" and User B has the same text but belongs to a different text style, i.e., *Comment To Others*. When the classifier do not carefully differentiate these text styles but merely mix all textual information together, using the sample of User A as training data is more likely to classify User B to be the same gender due to the same text "*Sister, you're so pretty!*". Obviously, this is a wrong prediction because User B is a *male* and the word "sister" is used to call someone else. Therefore, a better way to leverage textual knowledge in social media should be able to distinguish different styles of text.

In this paper, we address the above challenge by proposing a novel approach called ensemble LSTM recurrent neural network. Specifically, we first consider the features from each style of text as a separate textual perspective. Then, we train a Long Short-Term Memory (LSTM) network for each textual perspective respectively. Third, we add a merge layer to combine all LSTM representations by joint learning so as to fuse all textual knowledge. Empirical studies demonstrate that our approach performs much better than many strong baseline approaches.

Note that the motivation of employing LSTM as our single-perspective learning approach is that LSTM equips with a special gating mechanism that controls access to memory cells and it is powerful and effective at capturing long-term dependencies (Bengio et al., 1994). This advantage is helpful for modeling text and thus this approach has been successfully applied to a variety of NLP tasks, such as machine translation (Bahdanau et al., 2015), sentiment analysis (Tang et al., 2015), and sequence labeling (Chen et al., 2015).

The remainder of this paper is organized as follows. Section 2 overviews related work on user classification. Section 3 introduces data collection. Section 4 proposes our multi-perspective ensemble LSTM approach with multiple textual perspectives for user classification. Section 5 evaluates our approach with a benchmark dataset. Finally, Section 6 gives the conclusion and future work.

2 Related Work

Over the last decade, many previous studies have been devoted to the research on user classification with multiple attributes, such as user gender and user age.

User gender classification has been extensively studied in several domains, such as Blog (Peersman et al., 2011; Gianfortoni et al., 2011), E-mail (Mohanmad et al., 2011), YouTube (Filippova, 2012) and Micro-blog (Liu et al., 2013). More recently, some studies focus on some specific application scenarios on gender classification, such as multi-lingual gender classification (Ciot et al., 2013; Alowibdi et al., 2013), inferring gender by crowd (Nguyen et al., 2014) and interactive gender classification (Li et al., 2015).

User age classification has been studied in two main domains, i.e., blog (Burger and Hender son, 2006) and social media (Machinnon and Warren, 2006). In the blog domain, Schler et al. (2006) focus on textual features extracted from the blog text, such as word context features and POS stylistic features. Burger and Henderson (2006) explore some social features, such as location, time, and friend features, related to blogger age. Other studies, such as Rosenthal and McKeown (2011) and Goswami et al. (2009) explore both the textual and social features in automatic age classification. In the social media domain, Mackinnon and Warren (2006) explore some kind of social features, i.e., the relationship between users to predict a user's age and country of residence in a social network. Peersman et al. (2011) apply a text categorization approach to age classification with textual features only, i.e., word unigrams and bigrams. More recently, Marquardt et al. (2014) propose a multi-label classification approach to predict both the gender and age of authors from texts. Specifically, besides the word features, they also adopt some sentiment and emotion features in their approach.

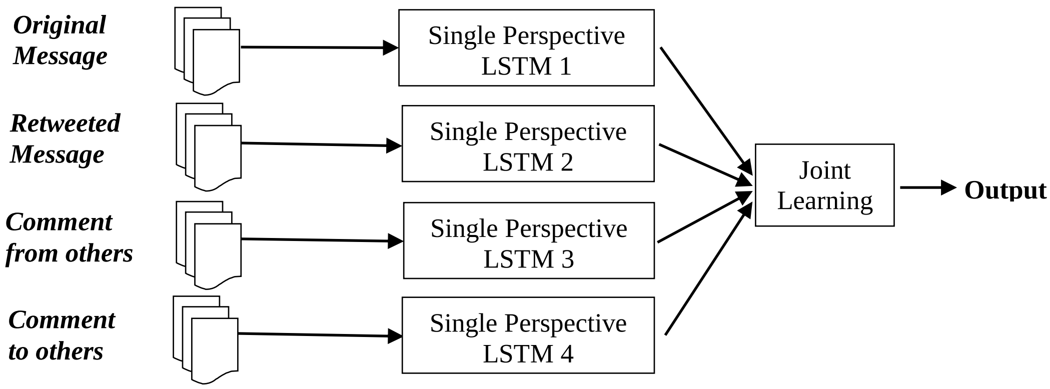

Figure 1: The framework of multi-perspective ensemble LSTM neural network

Some other user attributes, such as user location (Cheng et al., 2010), political orientation (Rao et al., 2010) and user occupational class prediction (Preotiuc-Pietro et al, 2015) are also popularly studied in recent years. Unlike all previous studies, this paper employs a deep learning approach to user classification and different styles of textual features are treated separately.

3 Data Collection

Our data are collected from Sina Micro-blog[1], a famous Micro-blogging platform in China. From the website, we crawl each user's homepage which contains user information (e.g., *name*, *age*, *gender*, *verified type*), and their posted messages. The data collection process starts from some randomly selected users, and iteratively gets the data of both their user attributes including gender and age. Different styles of text in each user's homepage are collected and they are:

1) ***Original message***: the messsages which are originally published by the user;
2) ***Retweeted message***: the messages which are retweeted by the user;
3) ***Comment from others***: the comments which are written by other users;
4) ***Comment to others***: the comments which are written by the user.

For gender classification, we randomly select 3000 *male* and 3000 *female* users for our empirical study and for age classification, we randomly focus on two age categories: *80s* (birthday between 1980 and 1989), *90s* (birthday between 1990 and 1999), each of which contains 3000 samples.

Table 2 shows the statistics about the average number of messages each user possessed in his/her homepage. From this table, we can see that each style of text has a decent number of messages or comments where *original message* and *comment to others* have more messages or comments than the other two styles.

	Gender		Age	
	Male	*Female*	*80s*	*90s*
Original message	148	158	154	153
Retweeted message	84	95	86	90
Comment from others	140	189	175	189
Comment to others	83	121	105	128

Table 2: Statistics about the average number of messages each user processed in his/her homepage

4 Our Approach

We treat the four styles of text as four textual perspectives for user classification and learn a multi-perspective ensemble LSTM recurrent neural network to make full use of all these perspectives. In general, our approach consists of two main components: (1) learning a new representation via a single-

[1] http://weibo.com/

perspective LSTM recurrent neural network of one type of user perspective. (2) employing a merge layer via joint learning to combine four different types of user perspectives. Figure 1 shows the framework overview of our approach and the two main components, i.e., single-perspective LSTM and multi-perspective ensemble LSTM via joint learning, will be discussed in detail.

4.1 Single perspective LSTM

In this study, we apply the implementation used by (Graves, 2013). The LSTM units at each time step t are defined to be a collection of vectors in $\mathbb{R}^d$: an input gate i_t, a forget gate f_t, an output gate o_t, a memory cell c_t and a hidden state h_t.

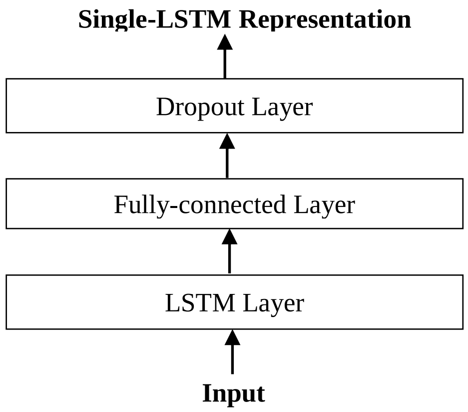

Figure 2: The framework of single perspective LSTM

Figure 2 illustrates the model architecture of our single-perspective LSTM where only one single LSTM layer is used. The input contains the representation of one type of textual perspective. According to the transition mode above, the input propagates through LSTM layer, Fully-connected layer and Dropout layer. The computing functions are given as following:

$$h^* = \phi\left(\omega^T h + b\right) \quad (1)$$

$$g = h^* \cdot D\left(p\right) \quad (2)$$

Where ϕ is the non-linear activation function, employed "*relu*" in our model and h is the output from LSTM layer. D denotes the dropout operator and p denotes a tune-able hyperparameter (the probability of retaining a hidden unit in the network).

4.2 Multi-perspective Ensemble LSTM via Joint Learning

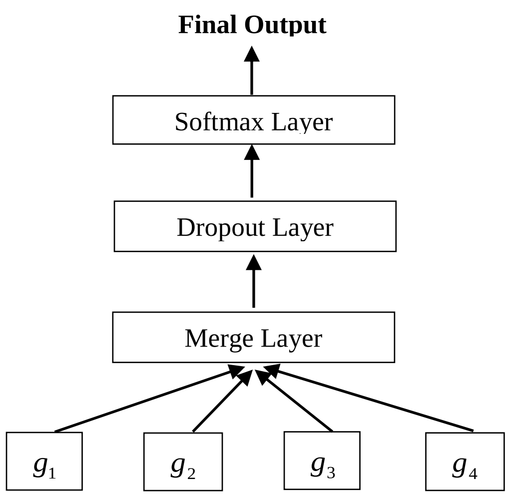

Figure 3: The framework of our multi-perspective ensemble LSTM approach

In order to distinguish the four types of textual perspectives and make full use of them legitimately, we propose a multi-perspective ensemble LSTM via joint learning to incorporate classification knowledge in *original message*, *retweeted message*, *comment from others* and *comment to others* separately. Figure 3 shows the framework of our multi-perspective ensemble LSTM approach where g_1, g_2, g_3 and g_4 are four LSTM representations learned from four single-perspective LSTM neural networks with four styles of textual perspectives.

The merge layer is designed to combine four types of user representation with a standard concatenation operation, i.e.:

$$g^* = [g_1; g_2; g_3; g_4] \tag{3}$$

Finally, a softmax output layer is used for classification. The model's prediction $label_{pred}$ is the class whose probability is maximal, specifically:

$$label_{pred} = \arg\max_i P(Y = i | x, W, U, V) \tag{4}$$

In our joint learning, the training objective is the penalized cross-entropy error, i.e.:

$$J = -\sum_{i=1}^{n_c} t_i \log y_i + \lambda \sum_{i=1}^{m} \left(\sum_{\varepsilon \in \omega} \left\| W_i^\varepsilon \right\|_F^2 + \sum_{\varepsilon \in \mu} \left\| U_i^\varepsilon \right\|_F^2 + \sum_{\varepsilon \in \nu} \left\| V_i^\varepsilon \right\|_F^2 \right) \tag{5}$$

Where $t \in \mathbb{R}^{n_c}$ is the one-hot represented ground truth and $y \in \mathbb{R}^{n_c}$ is the estimated probability for each class by softmax. (n_c is the number of target classes; m is the number of textual perspectives). In addition, W, U and V represent the corresponding weight matrices connecting them to the gates. $\| \cdot \|_F$ denotes the Frobenius norm of a matrix. $\omega = \{i, f, o, c\}$, $\mu = \{i, f, o, c\}$ and $\nu = \{i, f, o\}$ are the set of different gates (for W's, U's and V's, respectively). λ is a hyperparameter that specifies the magnitude of penalty on weights.

To train our ensemble LSTM, we use Stochastic Gradient Descent with mini-batches. The set of parameters to learn is the set $\theta = \{W, U, V\}$ in each single LSTM RNN of user perspective. The gradients $\partial J / \partial \theta$ are achieved through the back propagation algorithm (a special case of the chain-rule of derivation). Specifically, in terms of W_i^ε, the update equation is given by:

$$W_i^\varepsilon := W_i^\varepsilon + \frac{\partial J}{\partial g^*} \cdot \frac{\partial g^*}{\partial g_i} \cdot \frac{\partial g_i}{\partial h_i^*} \cdot \frac{\partial h_i^*}{\partial h_i} \cdot \frac{\partial h_i}{\partial W_i^\varepsilon} \tag{6}$$

Where $\frac{\partial h_i}{\partial W_i^\varepsilon}$ in LSTM unit will be computed via back propagation though time (BPTT). In the same spirit, U_i^ε and V_i^ε could be obtained as following:

$$U_i^\varepsilon := U_i^\varepsilon + \frac{\partial J}{\partial g^*} \cdot \frac{\partial g^*}{\partial g_i} \cdot \frac{\partial g_i}{\partial h_i^*} \cdot \frac{\partial h_i^*}{\partial h_i} \cdot \frac{\partial h_i}{\partial U_i^\varepsilon} \tag{7}$$

$$V_i^\varepsilon := V_i^\varepsilon + \frac{\partial J}{\partial g^*} \cdot \frac{\partial g^*}{\partial g_i} \cdot \frac{\partial g_i}{\partial h_i^*} \cdot \frac{\partial h_i^*}{\partial h_i} \cdot \frac{\partial h_i}{\partial V_i^\varepsilon} \tag{8}$$

5 Experiments

In this section, we empirically evaluate the performance of our approach to user classification in social media.

5.1 Experimental Settings

Dataset: (1) **Gender classification**: the dataset contains 3000 *male* and 3000 *female* users and each user has four styles of text: *original message*, *retweeted message*, *comment from others* and *comment to others*. We randomly select 4200 (70%) users as training data, 600 (10%) users as development data and use the remaining 1200 (20%) users as test data. (2) **Age classification**: the data set contains 3000 *80s* (between 1980 and 1989) users and *90s* (between 1990 and 1999) users and each user has four

Parameter and Description	Value
Size of total unigram features	30000
Dimension of the LSTM layer output	128
Dimension of the fully-connected layer output	64
Dropout rate	0.5
Epochs of iteration	10

Table 3: Parameters setting in LSTM RNN

	ME	CNN	Parallel CNN	LSTM
Original message	0.843	0.843	0.849	0.863
Retweeted message	0.784	0.793	0.788	0.791
Comment from others	0.825	0.798	0.818	0.823
Comment to others	0.736	0.743	0.754	0.776
Average	*0.797*	*0.794*	*0.802*	***0.813***

Table 4: Performance comparison of different approaches with single textual perspective (**Gender Classification**)

styles of text: *original message*, *retweeted message*, *comment from others* and *comment to others*. We randomly select 4200 (70%) users as training data, 600 (10%) users as development data and use the remaining 1200 (20%) users as test data.

Representations: Each message text is treated as a bag-of-features and transformed into binary vectors encoding the presence or absence of each feature. The features include word unigrams, and two kinds of complex features, i.e., F-measure and POS sequence pattern features, which yield the state-of-the-art performance in user classification (Mukherjee and Liu, 2010).

Classification algorithms: (1) The maximum entropy (ME) classifier implemented with the public tool, Mallet Toolkits[2]. (2) The random forest classifier and adaboost classifier implemented with the public tool, scikit-learn[3]. (3) The CNN classifier implemented with the help of the tool Keras[4]. (4) The LSTM classifier implemented with the help of the tool Keras.

Parameters Setting: (1) The most important parameter of RF and ABC is *estimators*, which is set 500 via fine-tuning. (2) The parameters of LSTM are set as shown in Table 3.

Evaluation Measurement: The performance is evaluated using the standard accuracy measurement.

5.2 Experimental Results

Experimental Results on Single Textual Perspective

For thorough comparison, four approaches with single perspective are implemented:

- **ME**: the maximum entropy classifier with all the parameters default.
- **CNN**: the basic bow-CNN is proposed in (Johnson and Zhang, 2014).
- **Parallel CNN**: the extension of bow-CNN, which has two or more convolution layers in parallel to learn multiple types of embedding of small text regions, proposed in (Johnson and Zhang, 2014).
- **LSTM**: the single perspective LSTM introduced in Section 4.1.

Table 4 shows the performance comparison of four approaches to gender classification. From this table, we can see that the text style of *original message* performs best among all four styles of text no matter what classification approach is used. On average, CNN and Parallel CNN performs better than ME. Among the four approaches, LSTM perform best. Significance test shows that our LSTM approach significantly outperforms the other four approaches (p-value<0.05).

Table 5 shows the performance comparison of four approaches to age classification. From the table, we can see that the text style of *original message* performs best among all four styles of text no matter

[2] http://mallet.cs.umass.edu/
[3] http://scikit-learn.org/stable/
[4] https://github.com/fchollet/keras

	ME	CNN	Parallel CNN	LSTM
Original message	0.793	0.775	0.763	0.794
Retweeted message	0.707	0.699	0.733	0.745
Comment from others	0.736	0.761	0.757	0.759
Comment to others	0.745	0.751	0.744	0.760
Average	*0.745*	*0.747*	*0.749*	***0.765***

Table 5: Performance comparison of different approaches with single textual perspective (**Age Classification**)

Approach	RandomForest	Adaboost	Voting LSTM	Weighted_Sum LSTM	***Ensemble LSTM (Ours)***
Accuracy	0.791	0.803	0.853	0.885	***0.908***

Table 6: Performance comparison of five approaches with multiple textual perspective (**Gender Classification**)

Approach	RandomForest	Adaboost	Voting LSTM	Weighted_Sum LSTM	***Ensemble LSTM (Ours)***
Accuracy	0.763	0.744	0.801	0.816	***0.823***

Table 7: Performance comparison of five approaches with multiple textual perspective (**Age Classification**)

what classification approach is used. Similar to the results in gender classification, LSTM still perform best in age classification. Significance test shows that our LSTM approach significantly outperforms the other three approaches (p-value<0.05).

Experimental Results on Multiple Textual Perspectives

For thorough comparison, several ensemble learning approaches with multiple perspectives are implemented:

- **RandomForest**: a popular ensemble learning approach proposed by Strobl et al. (2007). In our implementation, we train multiple decision tree classifiers and employ random forest algorithm to combine them.
- **Adaboost**: a popular ensemble learning approach proposed by (Zhu et al., 2009). In our implementation, we mixture the data of all perspective and use each word feature to form a weak classifier and then combine all feature classifier with adaboost algorithm.
- **Voting LSTM**: we first use each single textual perspective to train a LSTM classifier and then use the voting rule (Kuncheva and Rodriguez, 2014) to combine the obtained label outputs from all single-perspective LSTM classifiers.
- **Weighted_Sum LSTM**: we first use each single textual perspective to train a LSTM classifier and then use weighted sum rule (Marler and Arora, 2010) to combine the obtained probability outputs from all single-perspective LSTM classifiers.
- **Ensemble LSTM (Our approach)**: our joint learning approach as introduced in Section 4.2.

Table 6 shows the performance comparison of all approaches to gender classification when multiple textual perspectives are used. From this table, we can see that, using multiple textual perspectives does not always outperform the best performed approach with a single textual perspective. For instance, when RandomForest and Adaboost are used, the performance of using multiple textual perspective are 0.791 and 0.803 respectively, which are worse than that of using the *Original message* perspective with LSTM classifier, i.e., 0.863. Our ensemble LSTM approach performs best and it performs much better than both the best-performed single perspective LSTM (as shown in Table 4) and other strong ensemble strategies with multiple textual perspectives, such as Voting LSTM and Weighted_Sum LSTM. Significance test shows that our ensemble LSTM approach significantly outperforms other approaches when multiple textual perspectives are used (p-value<0.05).

Table 7 shows the performance comparison of all approaches to age classification when multiple textual perspectives are used. From this table, we can see that our ensemble LSTM approach performs best and it is also performs better than other strong ensemble strategies with multiple textual perspectives, such as Voting LSTM and Weighted_Sum LSTM. Significance test shows that our ensemble LSTM approach significantly outperforms other approaches when multiple textual perspectives are used (p-value<0.05).

5.3 Effectiveness Analysis and Case Study

In order to further illustrate the superiority of our approach, we give a case study as following. Table 8 shows the selected features sorted by the feature selection method of information gain (IG) (Li et al., 2009) when the task of gender classification is considered. We extract the features from the *original message* text and the *retweeted message* text separately.

This table shows the top-10 IG features from the *original message* text and their ranks in the *retweeted message* text. N denotes the sequence number of the feature in the selected features. $|F_f|$ denotes the feature frequency in all samples of female. $|F_m|$ denotes the feature frequency in all samples of male. For instance, the sequence number of emoticon "rabbit" in *original message* is the first, the feature frequency in all samples of female is 5871, and the feature frequency in all samples of male is 1872. It is observed that this feature is usually used by a woman. From the table, we can see that many '*good*' features in *original message*, such as *emoticon [rabbit]*, *亲亲* (*kiss*) and *讨厌* (*hate*), are not ranked top in *retweeted message*. If we merely merge all styles of text, some '*good*' features in one textual perspective would not be as effective as in the scenario when they are separately treated.

Feature	Original message			Retweeted message		
	N	$\|F_f\|$	$\|F_m\|$	N	$\|F_f\|$	$\|F_m\|$
表情符-兔子 (emoticon [rabbit])	1	5871	1872	154	1553	960
亲亲 (kiss)	2	3700	978	104	1186	606
闺蜜 (ladybro)	3	588	103	1	1186	313
NBA	4	53	467	10	120	500
足球 (football)	5	169	1144	5	328	1561
球队 (team)	6	31	378	3	135	810
讨厌 (hate)	7	1854	773	797	1470	943
进球 (goal)	8	23	296	6	96	607
委屈 (grievance)	9	2358	802	--	--	--
男神 (dream guy)	10	1163	331	26	843	334

Table 8: The top-10 IG features from the *original message* text and their ranks in the *retweeted message* text

6 Conclusion

In this study, we propose a novel approach, namely ensemble LSTM, to user classification, which jointly learns textual features from different textual perspectives. Our contributions lie in two main aspects: First, the proposed LSTM approach with a single textual perspective performs much better than traditional approaches, such as ME and CNN, for user classification. Second, the proposed ensemble LSTM approach significantly outperforms both the approaches which use only one single textual perspective and several other ensemble approaches.

In our future work, we attempt to apply bidirectional LSTM in user classification to utilize both the bi-directional contexts. Moreover, in addition to the textual features, we would like to merge social features to further improve performance. What's more, we will apply our proposed multi-perspective ensemble LSTM model in some other tasks of user classification, such as user occupation classification and so on.

Acknowledgements

This research work has been partially supported by four NSFC grants, No.61273320, No.61375073, No.61331011, and No. 61402314.

Reference

Jalal S. Alowibdi, Ugo A. Buy and PHilip Yu. 2013. Language Independent Gender Classification on Twitter. In *Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 739-743.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of ICLR*.

Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 1994. Learning Long-term Dependencies with Gradient Descent is Difficult. *Neural Networks, IEEE Transactions on*, 5(2):157–166.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2(1): 1-8.

John D. Burger and John C. Henderson. 2006. An Exploration of Observable Features Related to Blogger Age. In *Proceedings of AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*, pages 15-20.

Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. In *Proceedings of CIKM*, pages 759-768.

Xinchi Chen, Xipeng Qiu, Chenxi Zhu, Pengfei Liu, and Xuanjing Huang. 2015. Long Short-term Memory Neural Networks for Chinese Word Segmentation. In *Proceedings of EMNLP*, pages 1197-1206.

Morgane Ciot, Morgan Sonderegger and Derek Ruths. 2013. Gender Inference of Twitter Users in Non-English Contexts. In *Proceedings of EMNLP*, pages 1136–1145.

Katja Filippova. 2012. User Demographics and Language in an Implicit Socail Network. In *Proceedings of EMNLP*, pages 1478-1488.

Philip Gianfortoni, David Adamson and Carolyn P. Rosé. 2011. Modeling of Stylistic Variation in Social Media with Stretchy Patterns. In *Proceedings of EMNLP*, pages 49–59.

Sumit Goswami, Sudeshna Sarkar and Mayur Rustagi. 2009. Stylo-metric Analysis of Bloggers' Age and Gender. In *Proceedings of AAAI Conference on Weblogs and Social Media*, pages 214-217.

Alex Graves. 2013. Generating Sequences With Recurrent Neural Networks. *arXiv preprint arXiv:1308.0850*.

Rie Johnson, Tong Zhang. 2014. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. *Eprint Arxiv*.

Ludmila I. Kunchev, Juan J. Rodríguez. 2014. A Weighted Voting Framework for Classifiers Ensembles. *Knowledge and Information Systems*, 38(2): 259-275.

Vasileios Lampos, Nikolaos Aletras, Jens K. Geyti, Bin Zou and Ingemar J. Cox. 2016. Inferring the Socioeconomic Status of Social Media Users based on Behaviour and Language. *European Conference on Information Retrieval. Springer International Publishing*, pages 689-695.

Vasileios Lampos and Nello Cristianini. 2010. Tracking the Flu Pandemic by Monitoring the Social Web. In *Proceedings of the 2nd International Workshop on Cognitive Information Processing*, pages 411-416.

Shoushan Li, Rui Xia, Chengqing Zong and Chu-Ren Huang. 2009. A Framework of Feature Selection Methods for Text Categorization. In *Proceedings of ACL*, pages 692-700.

Shoushan Li, Jingjing Wang, Guodong Zhou, and Hanxiao Shi. 2015. Interactive Gender Inference with Integer Linear Programming. In *Proceedings of IJCAI*, pages 2341-2347.

Nan Liu, Yanxiang He, Qiang Chen, Min Peng and Ye Tian. 2013. A New Method for Micro-blog Platform Users Classification Based on Infinitesimal-time. *Journal of Information & Computantional Science*. pages 2569-2579.

Ian Mackinnon and Robert Warren. 2006. Age and Geo-graphic Inferences of the LiveJournal Social Net-work. In *Proceedings of ICML*, pages 176-178.

R. Timothy Marler, Jasbir S. Arora. 2010. The Weighted Sum Method for Multi-objective Optimization: New Insights. *Structural and multidisciplinary optimization*, 41(6): 853-862.

James Marquardt, Golnoosh Farnadi, Gayathri Vasudevan, Marie-Francine Moens, Sergio Davalos, Ankur Teredesai and Martine De Cock. 2014. Age and Gender Identification in Social Media. In *Proceedings of CLEF 2014 Evaluation Labs*, pages 1129-1136.

Saif M. Mohammad, and Tony Yang. 2011. Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 70-79.

Arjun Mukherjee and Bing Liu. 2010. Improving Gender Classification of Blog Authors. In *Proceedings of EMNLP*, pages 207-217.

Dong Nguye, Dolf Trieschnigg, A. Seza Doğruöz, Rilana Gravel, Mariet Theune, Theo Meder and Franciska de Jong. 2014. Why Gender and Age Prediction from Tweets is Hard: Lessons from a Crowdsourcing Experiment. *Association for Computational Linguistics*.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of* ICWSM, pages 122–129.

Claudia Peersman, Walter Daelemans, Leona Vaerenbergh. 2011. Predicting Age and Gender in Online Social Networks. In *Proceedings of SMUC*, pages 37-44.

Daniel Preotiuc-Pietro, Vasileios Lampos and Nikolaos Aletras. 2015. An Analysis of the User Occupational Class through Twitter Content. In *Proceedings of ACL*, pages 1754-1764.

Daniel Preoţiuc-Pietro, Svitlana Volkova, Vasileios Lampos, Yoram Bachrach and Nikolaos Aletras. 2015. Studying User Income through Language, Behaviour and Affect in Social Media. *PloS one*, 10(9): e0138717.

Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying Latent User Attributes in Twitter. In *Proceedings of the 2nd International Workshop on Search and Mining Usergenerated Contents*, pages 37–44.

Sara Rosenthal and Kathleen McKeown. 2011. Age Prediction in Blogs: A Study of Style, Content, and Online Behavior in Pre- and Post-Social Media Genera-tions. In *Proceedings of ACL*, pages 763-772.

Adam Sadilek, Henry Kautz, and Vincent Silenzio. 2012. Modeling Spread of Disease from Social Interactions. In *Proceedings of ICWSM*, pages 322-329.

Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and H Andrew Schwartz. 2014. Developing Age and Gender Predictive Lexica over Social Media. In *Proceedings of EMNLP*, pages 1146–1151.

Jonathan Schler, Moshe Koppel, Shlomo Argamon and James Pennebaker. 2006. Effects of Age and Gender on Blogging. In *Proceedings of AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*, pages 199-205.

Carolin Strobl, Anne-Laure Boulestei, Thomas Augustin. 2007. Unbiased Split Selection for Classification Trees Based on the Gini Index. *Computational Statistics & Data Analysis*, 52(1): 483-501.

Duyu Tang, Bing Qin, Ting Liu. 2015. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. In *Proceedings of EMNLP*, pages 1422-1432.

Jingjing Wang, Yunxia Xue, Shoushan Li and Guodong Zhou. 2015. Leveraging Interactive Knowledge and Unlabeled Data in Gender Classification with Co-training. Database Systems for Advanced Applications. *Springer International Publishing*, pages 246-251.

Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors. In *Proceedings of ICWSM*, pages 387-390.

Zhu Zhu, Jingjing Wang, Shoushan Li and Guodong Zhou. 2015. Interactive Gender Inference in Social Media. *Database Systems for Advanced Applications. Springer International Publishing*, pages 252-258.

Ji Zhu, Hui Zou, Saharon Rosset and Trevor Hastie. 2009. Multi-class Adaboost. *Statistics and its Interface*, 2(3): 349-360.