Neural Demographic Prediction in Social Media with Deep Multi-view Multi-task Learning

Yantong Lai[1,2], Yijun Su[3], Cong Xue[2(✉)], and Daren Zha[2]

[1] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[2] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{laiyantong,xuecong,zhadaren}@iie.ac.cn
[3] JD.com, Beijing, China

Abstract. Utilizing the demographic information of social media users is very essential for personalized online services. However, it is difficult to collect such information in most realistic scenarios. Luckily, the reviews posted by users can provide rich clues for inferring their demographics, since users with different demographics such as gender and age usually have differences in their contents and expressing styles. In this paper, we propose a neural approach for demographic prediction based on user reviews. The core of our approach is a deep multi-view multi-task learning model. Our model first learns context representations from reviews using a context encoder, which takes semantics and syntactics into consideration. Meanwhile, we learn sentiment and topic representations from selected sentiment and topic words using a word encoder separately, which consists of a convolutional neural network to capture the local contexts of reviews in word-level. Then, we learn a unified user representation from context, sentiment and topic representations and apply multi-task learning for inferring user's gender and age simultaneously. Experimental results on three real-world datasets validate the effectiveness of our approach. To facilitate future research, we release the codes and datasets at https://github.com/icmpnorequest/DASFAA2021_DMVMT.

Keywords: Demographic prediction · Context · Sentiment and topic views · Multi-task learning

1 Introduction

User demographics have been useful for personalization and recommendation. However, collecting such information in most realistic scenarios is difficult and the collected data might not be real. Thus, how to infer effective user demographics from public available data has attracted both academia and industry.

Luckily, many researchers have studied ways to infer user demographics from social media texts. A common approach relies on lexical features [1,5,15].

C. S. Jensen et al. (Eds.): DASFAA 2021, LNCS 12682, pp. 271–279, 2021.
https://doi.org/10.1007/978-3-030-73197-7_18

For instance, Sap et al. [15] derive gender and age predictive lexica over social media for prediction. Baslie et al. [1] leverage unigrams and characters to identify author's gender and language variety. Gjurković and Šnajder [5] use Linuistic Inquiry and Word Count (LIWC) [14] (a psychological dictionary) to detect user's personality. These hand-crafted features provide good explainability and are of high quality, but they require much manual labor and may ignore rich semantics in text. Recently, deep learning and pre-trained word embeddings have been widely used for demographic prediction. For example, Bayot et al. [2] apply word2vec [12] and convolutional neural network (CNN) [9] to infer user's gender and age. More recently, Wu et al. [19] leverage hierarchical attention mechanism and Tigunova et al. [16] combine attention mechanism with CNN for demographic prediction. Despite these methods perform well, they do not consider the useful sentiment and topic information in text, which have been proved important in demographic prediction [20].

In this paper, we propose a neural approach for demographic prediction based on user reviews. Instead of merging all reviews from the same user into a long text, our approach learns user representations using a deep multi-view multi-task learning model. Our model first learns context representations with a context encoder, to represent rich semantics and syntactics in reviews. Meanwhile, we learn sentiment and topic representations from selected sentiment and topic words with a word encoder respectively. Each word encoder contains a CNN to capture the local contexts in word-level. Then, we obtain a unified user representation integrating from context, sentiment and topic representations. Since gender and age are correlated, we apply multi-task learning for capturing latent influence between them. In the end, we perform experiments on three real-world datasets and the results validate the effectiveness of our model on demographic prediction.

2 Related Work

User demographic prediction with social media text is often regarded as a classification task in natural language processing (NLP) field. Traditional demographic prediction methods mostly rely on hand-crafted linguistic features, such as lexicons [15], unigrams [1] and LIWC [5]. Despite these hand-crafted features based methods preform well, they generally require much manual labor to collect and may ignore rich semantics in text. With the development of word embeddings and deep learning, researchers [2,16] begin to infer demographics using implicit context representations. For instance, Bayot et al. [2] leverage word2vec [12] and CNN [9] to infer gender and age. Further, attention mechanism has been proposed to capture informative contents in text [16,19].

Recently, multi-task learning has been utilized in demographic prediction [17,18]. For example, [17] leverages multi-modal data from Twitter, i.e., users' profiles, following network and tweets, to infer user demographics and location. Additionally, Wang et al. [18] make use of images and user profiles for demographic prediction.

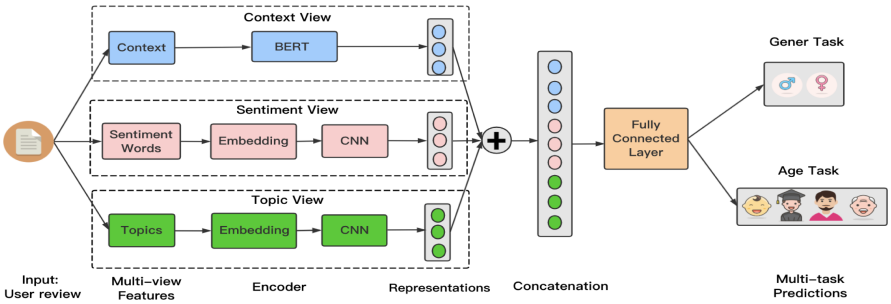

Fig. 1. Our framework for demographic prediction

In this paper, the approach we propose is different from existing methods. First, our model makes better use of context by considering both semantics and syntactics in text. Second, we capture high-order interactions from context, sentiment and topic views. Third, our approach utilizes the correlation between demographics and applies multi-task learning for better performance.

3 Methodology

In this section, we will introduce our approach in details (Fig. 1). The input of our model is a user review $s = \{w_1, w_2, ..., w_n\}$. We define a task set $U = \{u_1, u_2, ..., u_m\}$ and use ϕ_{u_i} to denote the parameters for the i-th task. The outputs of our model are probability distributions for all tasks, i.e. for task u_i is $\Pr(y_{u_i}|s, \phi_{u_i})$, where y_{u_i} represents the class in it.

3.1 Context View

As shown in Fig. 1, context view aims to capture semantics and syntactics information from a user review s, and produces a contextualized latent representation H_c. Thus, the learning process of H_c could be regarded as review-level embedding and we apply a pre-trained language model BERT [4] for encoding. The encoding procedure of BERT contains token embeddings, segment embeddings and position embeddings. Unlike static pre-trained word embeddings (e.g., word2vec [12]), token embeddings could solve polysemy and vary according to their context. Segment embeddings aim to capture inter-sentence syntactics, while position embeddings indicate the position of each words in the review s. The final context embeddings v^c of the review s are the sum of token embeddings, segment embeddings and position embeddings. In addition, the special tokens `[CLS]` and `[SEP]` are used for labeling classification tasks and separating segments respectively. Finally, we output the final hidden state of the first special token `[CLS]` as the context representation H_c in context view.

3.2 Sentiment View

As illustrated in Fig. 1, sentiment view is responsible to learn a hidden representation H_s for sentiment words, which are automatically extracted from a user review s. It mainly contains three steps.

The first step is sentiment words extraction. Sentiment words are indicative for demographic prediction, so we utilize a sentiment dictionary AFINN [13] to automatically extract sentiment words from the review s. If no words appear in the sentiment dictionary, we set "NA" in the sentiment words set S.

The second step is word embedding. Through this step, each sentiment word is mapped to a d-dimensional dense vector $v^s \in \mathbb{R}^{h \times d}$ using a word embedding lookup table $\mathbf{E} \in \mathbb{R}^{V \times d}$, where V is vocabulary size.

The third step is word encoding. We use a CNN as word encoder, to capture semantics from the sentiment words embeddings v^s. It learns the contextual representation through convolutional filters which slide ω-grams per step. We apply N filters to learn semantics from sentiment words embeddings v^s and obtain a feature map $\mathbf{c^s} = [\mathbf{c^s}_1, \mathbf{c^s}_2, ..., \mathbf{c^s}_N]$. The i-th feature map is $\mathbf{c^s}_i = [\mathbf{c^s}_{i_1}, \mathbf{c^s}_{i_2}, ..., \mathbf{c^s}_{i_{h-\omega+1}}]$. After generating feature map $\mathbf{c^s}$, max-over-time pooling operation is performed to capture the most important feature on each dimension of vector by taking the maximum value. For the i-th feature in $\mathbf{c^s}$, the feature after max-over-time pooling is $\hat{c_i^s} = \max\{\mathbf{c^s}_i\}$.

Finally, we concatenate all the features after max-over-time pooling operations as $H_s = [\hat{c_1^s}, \hat{c_2^s}, ..., \hat{c_N^s}]$. Additionally, H_s is the output for sentiment view.

3.3 Topic View

Topic view learns a latent representation H_t for topics (Fig. 1). The learning procedure mainly consists of three steps.

The first step is topics extraction. We make use of the traditional topic modelling method, latent Dirichlet allocation (LDA) [3], to extract topics automatically. Generally, a user review contains several topics and each topic would be formed by a set of words. Thus, a user review s could be seen as a document consisting of k topics. For each topic z in the document, a distribution φ_z on V_t is sampled from a Dirichlet function, where V_t represents a vocabulary consisting of a set of topic words. Then, LDA estimates the distribution $p(z|w)$ for $z \in s^P, w \in V_t^P$, where P denotes the set of word positions in the user review s. Finally, we get the topics set $T = \{w_1^t, w_2^t, ..., w_r^t\}$ according to the distribution $p(z|w)$, where r is the length of the sequence. Topics embedding and encoding is similar to that in sentiment view, which has been detailed described in Sect. 3.2.

Finally, we output topics latent representation $H_t = [\hat{c_1^t}, \hat{c_2^t}, ..., \hat{c_N^t}]$ for topic view, where $\hat{c_i^t}$ represents the i-th feature after max-over-time pooling.

3.4 Training Procedure

The training procedure of our framework consists of two stages: obtaining a unified user representation and multi-task learning.

Table 1. Statistics of three datasets, including number of total reviews (#Total), number of English reviews with gender and age information (#Extracted), female, male and various age categories distributions of extracted reviews.

Datasets	#Total	#Extracted	Female%	Male%	(0, 18]%	(18, 30)%	[30, 40)%	[40, 99)%
Denmark	646084	826	25.67	74.33	0.3632	8.11	31.60	59.93
US	648784	37060	39.82	60.18	0.13	11.39	29.91	58.58
UK	1424395	129175	40.38	59.62	0.18	8.06	18.59	73.17

The first stage is to obtain a unified user representation for context representation H_c, sentiment representation H_s and topic representation H_t. We apply averaging operation on context representation H_c, to get a good summary of the semantics. Then, we concatenate representations from context, sentiment and topic views as $\mathbf{H} = H_c \oplus H_s \oplus H_t$, where $\oplus$ denotes the concatenation operation.

The second stage is to predict demographics simultaneously using multi-task learning. We pass the concatenated feature $\mathbf{H}$ through a fully-connected layer and output a compressed latent feature vector $\mathbf{X}$. Then, the shared latent vector $\mathbf{X}$ is given to a task-specific layer to calculate the probability distribution $\tilde{y_{u_i}}$:

$$\tilde{y_{u_i}} = \text{softmax}(W_{u_i}^T \mathbf{X} + b_{u_i}) \tag{1}$$

where W_{u_i} and b_{u_i} represent the weights and bias for task u_i respectively.

For each task u_i, we minimize the cross-entropy of the predicted and true distributions as following:

$$L(\phi_{u_i}) = -\sum_{k=1}^{K}\sum_{c=1}^{C} y_k^c log(\tilde{y}_k^c) \tag{2}$$

where L denotes the cross-entropy loss function. y_k^c and $\tilde{y}_k^c$ are the true label and prediction probabilities for task u_i. K represents the total number of training samples and C is the total number of classes.

Finally, the total loss $\mathcal{L}$ of our approach is optimized for global objective function:

$$\mathcal{L} = \lambda_G \mathcal{L}_G + \lambda_A \mathcal{L}_A \tag{3}$$

where λ_G and λ_A are the weights for gender and age task respectively, and $\mathcal{L}_G$ and $\mathcal{L}_A$ are the losses of gender and age tasks.

4 Experiments

4.1 Experimental Setup

In this section, we conduct extensive experiments on three user reviews datasets [7] from TrustPilot (a website for online user reviews) to verify the effectiveness of our framework. The datasets collect users reviews and the reviewers' profiles (e.g., gender and birth year) from Denmark, United Kingdom and the United

States of America. We only extract English reviews with reviewer's gender and birth year information from the three datasets, using a language identifier fastText API[1]. Motivated by [18], we categorize user age into four categories: (0, 18] (18, 30), [30, 40) and [40,99]. The statistics of the three datasets are summarized in Table 1.

Table 2. Performance comparison of accuracy and macro-averaged F1 score on the Denmark, US and UK datasets for gender and age inference tasks. Here, gender task weight λ_G and age task weight λ_A are set to 1.

Methods	Denmark				US				UK			
	Gender		Age		Gender		Age		Gender		Age	
	Acc	Fscore	Acc	Fscore	Acc	Fscore	Acc	Fscore	Acc	Fscore	Acc	Fscore
Majority	74.33	–	59.93	–	60.18	–	58.58	–	59.62	–	73.17	–
CNN	71.95	41.84	60.98	24.74	65.52	63.62	56.93	32.92	62.92	61.37	73.06	28.43
BiLSTM	73.17	42.25	64.63	26.17	66.08	64.65	57.93	34.70	63.61	61.59	72.10	26.71
FastText	70.73	41.43	59.76	23.62	66.86	66.11	55.34	27.48	65.36	64.20	76.70	21.70
BERT	73.17	42.25	56.10	19.97	67.59	65.74	58.61	23.11	64.92	61.81	77.04	22.57
RoBERTa	67.58	35.63	56.10	23.96	60.50	37.69	57.85	18.32	60.04	37.51	77.07-	21.76
$\text{HAM}_{\text{CNN}-\text{attn}}$	68.07	45.27	52.77	24.49	63.68	60.15	57.03	24.76	52.12	49.65	67.89	24.85
HURA	70.84	44.70	51.57	25.61	66.42	57.56	54.11	22.42	58.41	41.18	75.69	22.32
Ours	**78.05**	**46.31**	**69.51**	**27.34**	**72.02**	**70.77**	**59.61**	**36.13**	**67.22**	**65.91**	**77.39**	**30.84**

We compared our proposed framework with the following methods: (1) Majority, a majority class based approach; (2) CNN [2], a CNN based model in demographics prediction; (3) BiLSTM [6], a bidirectional Long Short-Term Memory network; (4) FastText [8], a method proposed for efficient text classification; (5) BERT [4], the state-of-art method in various NLP tasks; (6) RoBERTa [11], a state-of-art robust model modified from BERT; (7) $\text{HAM}_{\text{CNN}-\text{attn}}$ [16], a state-of-art method using CNN and attention mechanism in demographic prediction; (8) HURA [19], a hierarchical demographic prediction model based on CNN and attention mechanism. As most works in user demographic prediction, we use accuracy and macro-averaged F1 score as evaluation metrics and report the average results. In all experiments, we preform 10-fold cross-validation, where 8 folds are used for training, 1 for validating and 1 for testing.

For all baselines, we adopt the optimal parameters configurations reported in their works. In our model, we use a BERT-base encoder and set max input length as 300. For sentiment and topic views, we embed sentiment words and topics into a 200-dimension vector. The filter number of word encoder is 128 and the window sizes are 2, 3, 4 and 5. We leverage Adam [10] optimizer with a dropout rate of 0.5 and set learning rate as 1e-3. Additionally, we apply an L2 weight decay 1e-3 to the loss and train our model in 20 epochs with a batch size of 64. For multi-task learning, we use equal gender task weight λ_G and age task weight λ_A as 1.

[1] https://fasttext.cc/docs/en/language-identification.html.

4.2 Experimental Results

Table 2 summarizes the comparison results between our model and baselines on three datasets. We have the following observations: (1) Our framework outperforms the statistics based baseline Majority, because our framework leverages rich information (e.g., semantics, sentiment words and topics) in text other than major class distribution of dataset; (2) Our framework also achieves better performance than neural network based methods, for we could capture context information and local semantics in text; (3) Compared with the Transformers based methods, our framework achieves better performance, especially on the macro-averaged F1 score; (4) As for the state-of-art methods $\text{HAM}_{\text{CNN}-\text{attn}}$ [16] and HURA [19], our model still gets better results by focusing on the effects of high-order interactions among context, sentiment and topics in text.

Table 3. Ablation study of multiple views in our framework on gender and age attributes

	Denmark				US				UK			
	Gender		Age		Gender		Age		Gender		Age	
	Acc	Fscore	Acc	Fscore	Acc	Fscore	Acc	Fscore	Acc	Fscore	Acc	Fscore
w/o Context view	69.51	41.01	58.54	18.46	65.62	61.45	57.85	29.53	60.04	58.61	76.63	21.69
w/o Sentiment view	73.17	42.25	62.20	25.56	69.13	67.81	57.69	21.34	63.61	61.37	76.70	23.11
w/o Topic view	74.39	43.45	63.54	25.25	69.24	66.55	58.61	25.45	65.48	61.98	76.84	23.49
Ours	**78.05**	**46.31**	**69.51**	**27.34**	**72.02**	**70.77**	**59.61**	**36.13**	**67.22**	**65.91**	**77.39**	**30.84**

To investigate which part contributes more to the performance, we perform ablation study. From Table 3, we could observe that: (1) Our approach performs best when leveraging context, sentiment and topic views; (2) Without context information, the performances on both gender and age tasks decrease sharply for not taking rich semantics into consideration; (3) Incorporating topics with context information performs better than using sentiment words and context. This is because topics may vary from nouns (e.g., shop and price) to adjectives (e.g., quick) and adverbs (e.g., smoothly), while sentiment words mainly consist of adjectives (e.g., helpful). Thus, we can conclude that various types of words contribute more to semantics.

5 Conclusion

In this paper, we study demographic prediction based on user reviews and propose a deep multi-view multi-task learning model. Our model first learns context representations from reviews by considering semantics and syntactics in text. At the same time, we learn sentiment and topic representations separately to capture the local contexts of reviews in word-level. Then our model integrates representations from context, sentiment and topic views and leverages the correlation between demographics to predict. Experimental results show that our model

outperforms many baseline methods on gender and age predictions on three real-world datasets. Further, we perform ablation study to investigate which view contributes more to performance. In our future work, we plan to adapt our study to multilingual user reviews and explore the transferability of our model on more user attributes.

References

1. Basile, A., Dwyer, G., Medvedeva, M., Rawee, J., Haagsma, H., Nissim, M.: N-gram: new groningen author-profiling model–notebook for pan at clef 2017. In: CEUR Workshop Proceedings, vol. 1866 (2017)
2. Bayot, R.K., Gonçalves, T.: Age and gender classification of tweets using convolutional neural networks. In: Nicosia, G., Pardalos, P., Giuffrida, G., Umeton, R. (eds.) MOD 2017. LNCS, vol. 10710, pp. 337–348. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-72926-8_28
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**(1), 993–1022 (2003)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2019)
5. Gjurkovic, M., Šnajder, J.: Reddit: a gold mine for personality prediction. In: Second Workshop on Computational Modeling of Peoples Opinions (2018)
6. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
7. Hovy, D., Johannsen, A., Søgaard, A.: User review sites as a resource for large-scale sociolinguistic studies. In: Proceedings of the 24th International Conference on World Wide Web, pp. 452–461 (2015)
8. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)
9. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
10. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR abs/1412.6980 (2014)
11. Liu, Y., et al.: Roberta: a robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
12. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
13. Årup Nielsen, F.: A new anew: evaluation of a word list for sentiment analysis in microblogs (2011)
14. Pennebaker Francis, J.W.: Linguistic Inquiry and Word Count. Lawrence Erlbaum Associates Mahwah Nj (2012)
15. Sap, M., Park, G., Eichstaedt, J.C., Kern, M.L., Schwartz, A.H.: Developing age and gender predictive lexica over social media. In: Conference on Empirical Methods in Natural Language Processing (2014)
16. Tigunova, A., Yates, A., Mirza, P., Weikum, G.: Listening between the lines: learning personal attributes from conversations. In: The World Wide Web Conference, pp. 1818–1828 (2019)

17. Vijayaraghavan, P., Vosoughi, S., Roy, D.: Twitter demographic classification using deep multi-modal multi-task learning. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 478–483 (2017)
18. Wang, Z., et al.: Demographic inference and representative population estimates from multilingual social media data. In: The World Wide Web Conference, pp. 2056–2067 (2019)
19. Wu, C.: Neural demographic prediction using search query. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (2019)
20. Wu, C., Wu, F., Qi, T., Liu, J., Huang, Y., Xie, X.: Neural gender prediction in microblogging with emotion-aware user representation. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2401–2404 (2019)