Neural Gender Prediction from News Browsing Data

Chuhan Wu[1], Fangzhao Wu[2], Tao Qi[1], Yongfeng Huang[1], and Xing Xie[2]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
{wu-ch19,qit16,yfhuang}@mails.tsinghua.edu.cn
[2] Microsoft Research Asia, Beijing 100080, China
{fangzwu,xing.xie}@microsoft.com

Abstract. Online news platforms have attracted massive users to read digital news online. The demographic information of these users such as gender is critical for these platforms to provide personalized services such as news recommendation and targeted advertising. However, the gender information of many users in online news platforms is not available. Fortunately, male and female users usually have different pattern in reading online news. Thus, the news browsing data of users can provide useful clues for inferring their genders. In this paper, we propose a neural gender prediction approach based on the news browsing data of users. Usually a news article has different kinds of information such as title, body and categories. However, the characteristics of these components are very different, and they should be processed differently. Thus, we propose to learn unified user representations for gender prediction by incorporating different components of browsed news as different views of users. In each view, we use a hierarchical framework to first learn news representations and then learn user representations from news representations. In addition, since different words in news titles and bodies usually have different informativeness for learning news representations, we use attention mechanisms to select important words. Besides, since different news articles may also have different informativeness for gender prediction, we use news-level attentions to attend to important news articles for learning informative user representations. Extensive experiments on a real-world dataset validate the effectiveness of our approach.

Keywords: Gender Prediction · News Browsing · Multi-view Learning · Attention Mechanism

1 Introduction

Online news platforms such as Google News have attracted massive users for online news reading [4]. The demographics of users such as gender are very important for these platforms to provide personalized services to their users [3]. For example, with the gender information of users, online news platforms can make more personalized new recommendations to them [23], e.g., recommending

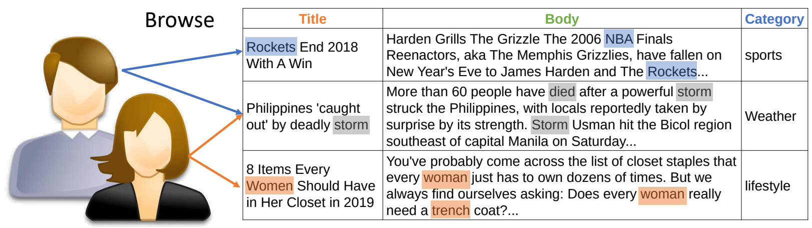


Fig. 1. Several example news articles browsed by a female user and a male user. Important words are highlighted using color bars.

NBA news to male users and fashion news to female users. In addition, advertisers can display their ads to users on news platforms more effectively [22]. For example, advertisers can display ads of dress and makeups to female users, and ads of shaver and tie to male users. Without the gender information of users, advertisers may display the ads of dress to both female and male users, which may be less effective [29]. Thus, the gender information of users is critical for news platforms to provide personalized services [15].

However, the gender information of many users in online news platforms is not available, making it difficult to provide personalized services for them [24]. Luckily, many users click and browse news articles displayed on these platforms, and there are usually some differences in news reading patterns between male and female users. For example, as illustrated in Fig. 1, a female user may browse a news article about dressing fashion styles, while a male user may browse a news article about NBA. Thus, the news browsing data of users can provide useful clues for inferring their genders. In this paper, we explore the problem of gender prediction of users on news platforms based on their browsed news.

Our work is inspired by several observations. First, a news article usually contains different kinds of information such as title, body and topic categories, which are all useful for inferring the gender of users. For example, as shown in Fig. 1, the title of the third news indicates that this news is about fashions, and the body of this news contains the details. Moreover, the topic categories of news articles are also informative for gender prediction. For instance, if a user frequently browses news about lifestyles and fashions but rarely browses sports news, then we can infer this user is probably a female user. Thus, incorporating different fields of news has the potential to learn more accurate user representations for gender prediction. Second, the characteristics of different news fields are very different. For example, news titles are usually short and concise sentences, while news bodies are usually long documents with detailed information, and topic categories are usually tags. Thus, they should be handled differently. Third, different words in the same news usually have different informativeness in learning news representations . For example, in Fig. 1 the word "Woman" is more informative than the word "2019" for learning informative news representations. In addition, different news articles browsed by the same user may also have different informativeness in learning user representations for gender prediction. For example, the third news in Fig. 1 is more informative than the second news in inferring the gender of this user.

In this paper, we propose a neural gender prediction approach to infer the genders of online news platform users from their news browsing data. In our approach, we propose to learn unified user representations via a multi-view learning framework from the titles, bodies and topic categories of their browsed news articles as different views of these users. In each view, we use a hierarchical model to first learn hidden news representations, and then learn user representations from representations of browsed news. In addition, since different words in the same news title or body usually have different informativeness in representing this news, we apply attention mechanism to select important words. Besides, since different news articles may also have different informativeness in learning user representation for gender prediction, we apply news-level attention networks in each view to select important news articles. Extensive experiments are conducted on a real-world dataset, and the results show that our approach can achieve satisfactory performance in predicting the genders of online news platform users and consistently outperform many baseline methods.

2 Related Work

User demographic prediction is an important task in both natural language processing and data mining fields, and has attracted wide attentions in recent years [12]. Existing user demographic prediction methods are mainly based on online behaviors and user generated content such as blogs, forum posts and social media messages [25, 19, 14, 17, 8, 6, 16, 31, 18, 13, 11]. For example, Rosenthal and McKeown [25] proposed to use logistic regression to infer user ages from blogs. They used many kinds of features including blog content features, stylistic features, behavior features and user interest features. Nguyen et al. [17] proposed to use linear regression with Lasso regularization to predict user ages. They used several different kinds of texts, such as blogs, forum posts and transcribed telephone speeches to build user representations. Peersman et al. [19] applied SVM to predict genders of Twitter users based on word n-gram and character n-gram features extracted from their tweets. However, these methods cannot effectively utilize contextual information. In addition, these methods cannot distinguish informative contexts and records from uninformative ones.

In recent years, several deep learning based methods have been proposed for user demographic prediction [31, 28, 5, 27, 5, 26, 1, 2, 29]. For example, Zhang et al. [31] used LSTMs to predict the genders and ages of social media users based on their microblogging messages, retweeted messages, comments from others and the comments to others. Wang et al. [28] proposed a CNN based method to jointly predict the ages and genders of social media users based on their messages in a collaborative manner. Wu et al. [29] proposed a hierarchical user representation model to predict the ages and genders of users in commercial search engines based on their search queries. Different from these methods which rely on blogs, social media data or search queries, our approach predicts the genders of users on news platforms based on their news browsing data. In addition, existing methods for demographic prediction usually aggregate different kinds

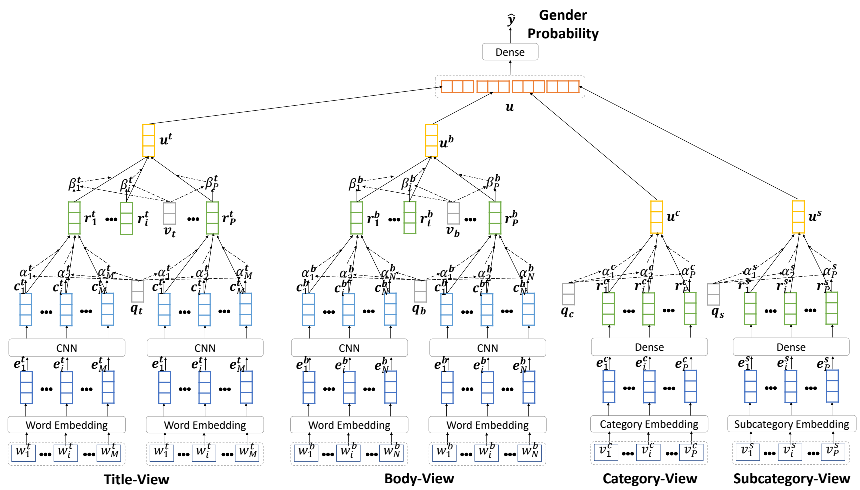


Fig. 2. The framework of our approach.

of user information together and ignore the differences of their characteristics, which may be suboptimal for user demographic prediction.

There are only a few approaches to predict user demographics based on news browsing data [8, 21]. For example, Phuong et al. [21] proposed to use SVM to predict genders of users based on their news website browsing histories. They used various hand-crafted features such as news categories, topic features extracted by LDA, access time, and sequential features extracted by webpage k-grams. However, these methods cannot effectively utilize the contexts and orders of words in news, which are very important for gender prediction. Different from all the aforementioned methods, our approach predicts the genders of users on news platforms based on their news browsing use a multi-view learning framework to learn unified user representations by regarding different kinds of news information as different views of users. Experiments on benchmark dataset show our approach can outperform these baseline methods.

3 Our Approach

In this section, we introduce our approach in detail. There are two major modules in our model. The first one is *user representation*. which learns representations of users from their browsed news. The second one is *gender classification*, which classifies the genders of users according to their representations. In the *user representation* module, we use a multi-view learning framework to learn unified user representations by incorporating news titles, bodies and topic categories as different views of users. The framework of our approach is shown in Fig. 2.

3.1 User Representations from Title View

The first view in the *news representation* module is *title-based user representation*. It is used to learn representations of users from the titles of their browsed news. In this module, we first use a *word encoder* to learn title representations from words, and then use a *title encoder* to learn title-based user representations from news titles. As shown in Fig. 2, there are three layers in the *word encoder*.

The first one is word embedding. It is used to convert a news title with M words from a word sequence $[w_1^t, w_2^t, ..., w_M^t]$ into a vector sequence $[\mathbf{e}_1^t, \mathbf{e}_2^t, ..., \mathbf{e}_M^t]$.

The second layer is a convolutional neural network (CNN) [9]. Local contexts within a news title are very important for understanding this news. For example, in the title of the first news in Fig. 1, the local contexts of "deadly" such as "storm" is very useful for learning accurate representations of this title. Thus, we employ a CNN network at word-level to learn contextual word representations by capturing local contexts. It takes the aforementioned embedding sequence as input, and outputs a contextual word representation sequence $[\mathbf{c}_1^t, \mathbf{c}_2^t, ..., \mathbf{c}_M^t]$.

The third layer is a word-level attention network. Usually, different words in the same news title may have different informativeness in representing this news. For example, in the title of the third news in Fig. 1 the word "woman" is much more informative than "Items" for learning gender discriminative news representations. Thus, we propose to apply attention mechanism at word-level to select important words for learning informative news representations. The attention weight α_i^t of the i-th word in the same news title is formulated as:

$$a_i^t = \mathbf{q}_t \times \tanh(\mathbf{U}_t \times \mathbf{c}_i^t + \mathbf{u}_t), \quad \alpha_i^t = \frac{\exp(a_i^t)}{\sum_{j=1}^{M} \exp(a_j^t)}, \tag{1}$$

where $\mathbf{V}_t$ and $\mathbf{v}_t$ are linear transformation parameters, $\mathbf{q}_t$ is the attention query vector. The final title representation is the summation of the word contextual representations weighted by their attention weights, i.e., $\mathbf{r}^t = \sum_{j=1}^{M} \alpha_j^t \mathbf{c}_j^t$.

In the *title encoder* module, we learn title-based user representations based on the representations of browsed news titles. The titles of different news usually have different informativeness for learning user representations for gender prediction. For example, the title of the first and third news are very informative for learning gender discriminative user representations, while the second news is uninformative since it is widely browsed by both male and female users. Thus, we propose to use a news-level attention network to select important news for learning informative user representations. Denote the attention weight on the title of the i-th news browsed by a user as β_i^t, which is calculated as follows:

$$b_i^t = \mathbf{v}_t \times \tanh(\mathbf{W}_t \times \mathbf{r}_i^t + \mathbf{w}_t), \quad \beta_i^t = \frac{\exp(b_i^t)}{\sum_{j=1}^{P} \exp(b_j^t)}, \tag{2}$$

where P is the number of browsed news, $\mathbf{W}_t$, $\mathbf{w}_t$ and $\mathbf{v}_t$ are parameters. The final representation of a user in this view is the summation of the title representations weighted by their attention weights, i.e., $\mathbf{u}^t = \sum_{i=1}^{P} \beta_i^t \mathbf{r}_i^t$.

3.2 User Representations from Body View

The second view is *body-based user representation*, which is used to learn user representations from the bodies of browsed news.

Similar with the title view, there are also two major components in this view, i.e., a *word encoder* to learn body representations from words, and a *body encoder* to learn user representations from news bodies. The *word encoder* also has three layers, i.e., a shared word embedding layer to convert the word sequence (denoted as $[w_1^b, w_2^b, ..., w_N^b]$, where N is the number of words) of a news body into a vector sequence, a word-level CNN network to learn contextual word representations (denoted as $[\mathbf{c}_1^t, \mathbf{c}_2^t, ..., \mathbf{c}_M^t]$) by capturing local contexts in news bodies, and a word-level attention network to select important words for learning informative representations of news bodies. Denote the bodies of the news browsed by users as $[b_1, b_2, ..., b_P]$. We apply the *word encoder* in this view to each body to obtain the hidden representation sequence $[\mathbf{r}_1^b, \mathbf{r}_2^b, ..., \mathbf{r}_P^b]$ of news bodies.

The *body encoder* is used to learn representations of users from the body representations of their browsed news. Different news bodies may also have different informativeness for learning user representations for gender prediction. For example, the body of the first news in Fig. 1 is more informative than the second one in inferring the gender of a user. Thus, we apply a news-level attention network to select important news based on their body representations. Denote the attention weight of the i-th body as β_i^b, which is computed as:

$$b_i^b = \mathbf{v}_b \times \tanh(\mathbf{W}_b \times \mathbf{r}_i^b + \mathbf{w}_b), \quad \beta_i^b = \frac{\exp(b_i^b)}{\sum_{j=1}^{P} \exp(b_j^b)}, \tag{3}$$

where $\mathbf{v}_b$, $\mathbf{W}_b$ and $\mathbf{w}_b$ are parameters. The final news body based representation of a user is the summation of the contextual body representations weighted by their attention weights, i.e., $\mathbf{u}^b = \sum_{j=1}^{P} \beta_j^b \mathbf{r}_j^b$.

3.3 User Representations from Category/Subcategory View

The third view is *category-based user representations*, which is used to learn user representations from the topic categories of browsed news. On many online news platforms such as MSN News, a news article is classified into a general topic category (e.g., "sports") and a finer-grained subcategory (e.g., "basketball_nba") to target user interests more effectively. Usually, news categories are important clues for gender prediction. For example, news articles about fashions are more likely to be clicked by female users than male users. Thus, we propose to incorporate the categories and subcategories of news to learn gender discriminative user representations. There are three layers in this module.

The first one is category/subcategory embedding. Denote the input category ID sequence as $[v_1^c, v_2^c, ..., v_P^c]$ and the subcategory ID sequence as $[v_1^s, v_2^s, ..., v_P^s]$. We use a category/subcategory embedding layer to convert the sequences of discrete category and subcategory IDs into sequences of low-dimensional vectors, which are respectively denoted as $[\mathbf{e}_1^c, \mathbf{e}_2^c, ..., \mathbf{e}_P^c]$ and $[\mathbf{e}_1^s, \mathbf{e}_2^s, ..., \mathbf{e}_P^s]$.

The second one is a dense layer. It is used to learn hidden representations of categories and subcategories as follows:

$$\mathbf{r}_i^c = \mathrm{ReLU}(\mathbf{W}_c \times \mathbf{e}_i^c + \mathbf{w}_c), \quad \mathbf{r}_i^s = \mathrm{ReLU}(\mathbf{W}_s \times \mathbf{e}_i^s + \mathbf{w}_s), \tag{4}$$

where $\mathbf{W}_c$, $\mathbf{w}_c$, $\mathbf{W}_s$ and $\mathbf{w}_s$ are linear transformation parameters.

The third one is a news-level attention network. Different categories and subcategories usually have different importance for gender prediction. For example, the news in the "sports" category is more informative than the news in the "weather" category, since the latter one is usually browsed by both male and female users. Thus, we apply a news-level attention network to select important news according to their topic categories. Denote the attention weight of the i-th topic category as α_i^c, which is calculated as:

$$a_i^c = \mathbf{q}_c \times \tanh(\mathbf{V}_c \times \mathbf{r}_i^c + \mathbf{v}_c), \quad \alpha_i^c = \frac{\exp(a_i^c)}{\sum_{j=1}^{P} \exp(a_j^c)}, \tag{5}$$

where $\mathbf{q}_c$, $\mathbf{V}_c$ and $\mathbf{v}_c$ are parameters. The category based user representations are the summation of the hidden category representations weighted by their attention weights, i.e., $\mathbf{u}^c = \sum_{j=1}^{P} \alpha_j^b \mathbf{r}_j^c$. The subcategory based user representations $\mathbf{u}^s$ can be computed in a similar way. The final unified user representations is the concatenation of the user representations learned from different views, i.e., $\mathbf{u} = [\mathbf{r}^c; \mathbf{r}^{sc}; \mathbf{r}^t; \mathbf{r}^b]$.

3.4 Gender Classification

The *gender classifier* is used to predict the probability $\hat{\mathbf{y}}$ of a user in different gender categories from his/her representations, which is computed by: $\hat{\mathbf{y}} = softmax(\mathbf{W}_y \times \mathbf{u} + \mathbf{w}_y)$, where $\mathbf{W}_y$ and $\mathbf{w}_y$ are parameters. The loss function we use is cross entropy, which is formulated as follows:

$$\mathcal{L} = -\frac{1}{N_g} \sum_{i=1}^{N_g} \sum_{k=1}^{K_g} y_{i,k} \log(\hat{y}_{i,k}), \tag{6}$$

where N_g is the number of user with gender labels, K_g is the number of gender categories, $y_{i,k}$ and $\hat{y}_{i,k}$ respectively denote the ground-truth and predicted probability of the i-th user in the k-th gender category.

4 Experiments

4.1 Datasets and Experimental Settings

Since there is no off-the-shelf dataset for news based gender prediction study, we built one by crawling 10,000 users from MSN News[3]. Among them, the gender information of 4,228 users is available which were used for our experiments. We also collected their news browsing histories in a month, i.e., from Dec. 13, 2018 to Jan. 12, 2019. The statistics of our dataset are shown in Table 1. We randomly sampled 80% of users for training, 10% for validation, and 10% for test.

[3] https://www.msn.com/en-us/news

# male users	2,484	**avg. # words per news title**	11.29
# female users	1,744	**avg. # words per news body**	730.72
# categories	15	**# subcategories**	284

Table 1. Statistics of our dataset.

In our experiments, we used the 300-dimensional pre-trained Glove embedding [20] to initialize the word embeddings. The CNN networks had 400 filters, and their window size was 3. The dimensions of attention query vectors were 200. The optimizer we used was Adam [10]. We added 20% dropout to each layer. The batch size was set to 50. These hyperparameters were tuned on the validation set. We independently repeated each experiment 10 times and reported the average results in terms of accuracy and macro F1-score.

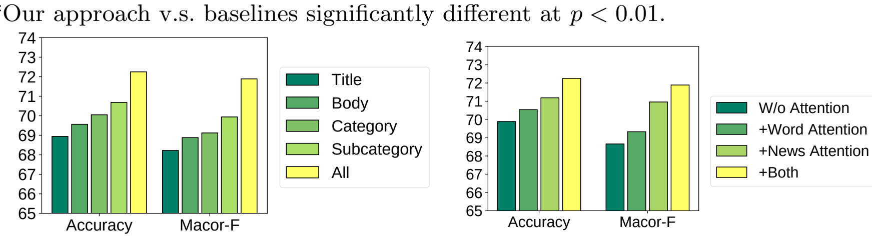
4.2 Performance Evaluation

We evaluate the performance of our approach by comparing it with several baseline methods, including: (1) *LinReg* [17], linear regression with Lasso regularization. (2) *SVM* [19], support vector machine; (3) *LR* [25], logistic regression; (4) *CNN* [9], convolutional neural network. (5) *LSTM* [7], long short-term memory network. (6) *CNN-Att*, CNN with attention mechanism. (7) *LSTM-Att*, LSTM with attention mechanism. (8) *HAN* [30], a hierarchical attention network for document classification. (9) *HURA* [29], a hierarchical attention network for document classification. (10) *Ours*, our neural gender prediction approach with news browsing data. In traditional methods (1-3), we used features including news category/subcategory IDs and TF-IDF features extracted from news titles and bodies as the input. In baseline methods based on neural networks (4-9), we used the combination of news titles, bodies, categories and subcategories by aggregating them into a long document. The results of these methods are reported in Table 2. According to the results, we have several observations.

First, the methods based on neural networks outperform traditional methods such as *SVM*, *LR* and *Linreg*. This is probably because neural network based methods can learn more accurate user representations for gender prediction. In addition, traditional methods usually rely on bag-of-words features, while the contexts and orders of words cannot be fully captured. Second, the methods using attention mechanism usually outperform the methods without mechanism. This is probably because different words and news usually have different informativeness for learning user representations, and selecting important contexts in news can benefit user representation learning. Third, the methods based on hierarchical models (*HAN*, *HURA* and *Ours*) outperform methods based on flatten models (e.g., *CNN-Att* and *LSTM-Att*). This may be because learning user representations in a hierarchical manner can utilize the document structures of news, which can benefit news and user representation learning. Fourth, our approach can outperform other baselines such as *HAN* and *HURA*. This is probably because different kinds of news information have different characteristics, and aggregating different news fields is usually not optimal. In our approach we use a multi-view learning framework to incorporate titles, bodies and categories as different views of news, which can learn better user representations.

Method	25%		50%		100%	
	Accuracy	Fscore	Accuracy	Fscore	Accuracy	Fscore
Linreg [17]	63.89±1.14	62.54±1.19	65.16±1.21	63.86±1.22	66.22±1.22	65.37±1.24
SVM [19]	66.10±0.78	64.98±0.79	66.85±0.82	66.23±0.84	67.26±0.88	66.89±0.90
LR [25]	65.26±0.61	63.89±0.63	66.43±0.66	65.67±0.67	67.33±0.71	66.84±0.72
CNN [9]	66.22±0.87	65.33±0.89	67.35±0.77	66.78±0.77	68.12±0.66	67.68±0.68
LSTM [7]	66.11±0.74	65.05±0.75	67.23±0.61	66.71±0.62	67.99±0.54	67.60±0.55
CNN-Att	66.84±0.68	65.76±0.70	68.01±0.62	67.26±0.62	68.79±0.53	68.32±0.55
LSTM-Att	66.57±0.73	65.44±0.77	67.67±0.64	67.01±0.65	68.45±0.56	68.03±0.57
HAN [30]	67.22±0.58	65.89±0.60	68.54±0.47	68.03±0.49	69.88±0.42	69.67±0.44
HURA [29]	67.95±0.64	66.66±0.66	69.77±0.52	68.89±0.55	70.45±0.41	70.01±0.43
Ours*	**70.03**±0.53	**68.89**±0.55	**71.28**±0.41	**70.78**±0.42	**72.25**±0.37	**71.89**±0.39

Table 2. The performance of different methods under different ratios of training data. *Our approach v.s. baselines significantly different at $p < 0.01$.



(a) Effectiveness of multi-view learning. (b) Effectiveness of attention networks.

Fig. 3. Effectiveness of the multi-view learning framework and attention networks.

4.3 Model Effectiveness

In this section, we will explore the effectiveness of different components of our approach. First, we want to verify the effectiveness of multi-view learning framework in our approach. We compare the performance of our approach and its variants with different combinations of news views, and the results are shown in Fig. 3(a). According to Fig. 3(a), we have several observations. First, we find the category view or subcategory view are very important for our approach. This is intuitive because the topics of news browsed by male and female users usually have huge differences. Thus, incorporating the information of news topic categories can learn more gender discriminative user representations. In addition, we find the model with the subcategory view slightly outperform that with the category view. This may be because subcategories usually contain finer-grained topic information than categories. Second, the model with the body view can achieve better performance than that with the title view only. This is probably because news bodies usually contain the details of news, which can provide richer information than news titles for learning informative news representations. Third, combining all four views can further improve the performance of our approach. These results validate the effectiveness of our multi-view learning framework.

Then, we conducted experiments to validate the influence of the word-level and news-level attention mechanism on our approach. The performance of our approach and its variant with different combinations of attention networks is

User	Category		Subcategory			Title
Male	autos		autossports			The 2020 Toyota Supra Is Here and It Looks Glorious
	sports		football_nfl			Steelers - Patriots NFL Week 15 : What You Need To Know
	weather		weathertopstories			Fast - moving snowstorm to hit Maine , NH at rush hour
	food&drink		tips&tricks			12 Steak Marinades Every Carnivore Needs
	weather		weathertopstories			Heavy rains cause river and street flooding in mountains and foothills
Female	lifestyle		lifestylefashion			Amy Schumer 's new clothing line is about helping women feel comfy and confident
	lifestyle		lifestylefashion			31 Outfit Ideas to Start the New Year in Style
	health		healthosteoporosis			40 health concerns women should watch out for after 40
	food&drink		recipes			31 of the most beautiful pie crusts and tips to make them
	tv		tvoscars			Kevin Hart Confirms He 's Hosting the 2019 Academy Awards

Fig. 4. Visualization of the word- and news-level attention weights on the titles, categories and subcategories of several news browsed by a randomly selected male and female users. Darker colors represent higher attention weights.

shown in Fig. 3(b). From Fig. 3(b), we have several observations. First, the news-level attention network has an important impact on our approach. This is probably because different news usually have different informativeness for learning user representations, and selecting important news can benefit user representation learning. Second, the word-level attention networks are also important for our approach. This is probably because different words usually have different informativeness for learning informative news and user representations for gender prediction, and our approach can attend to important words in news titles and bodies. Third, combining both kinds of attention networks can further improve the performance of our approach. These results validate the effectiveness of attention mechanism in our approach.

4.4 Visualization of Attention Weights

In this section, we will visually explore the effectiveness of the word-level and news-level attention mechanisms in our approach. The visualization results of the attention weights on news titles, categories and subcategories are shown in Fig. 4. From Fig. 4, we have several observations. First, we find the attention network can effectively recognize important words within news titles. For example, the words “Toyota” and “NFL” are assigned high attention weights, since these words are very informative for learning gender discriminative news and user representations, while the words such as “2020” and “Week” are assigned low attention weights since they are less informative. These results show that our approach can learn informative news representations by selecting important words. Second, we find our approach can effectively recognize important news according to their categories and subcategories. For example, the news in the “autos” and “lifestyles” categories are gained more attentions than those in the “weather” and “food&drink” categories. These results show that our approach can learn informative user representations by modeling the importance of different news topics for gender prediction. Third, we find approach can help learn more informative user representations by selecting news according to their title representations. For example, the news of “40 health concerns ...” is very informative for inferring the gender of users. However, the topic categories of this news are not very gender discriminative, and the category/subcategory views fail to recognize this news. Fortunately, our approach can still highlight this news by utilizing the content of news title, which is useful for learning more informative user representations. These results validate the effectiveness of the word- and news-level attention networks in our approach.

5 Conclusion

In this paper, we propose a neural gender prediction approach to infer the genders of online news platform users from their news browsing data. In our approach, we use a multi-view learning framework to incorporate different kinds of new information such as title, body, category and subcategory as different views of users to learn accurate user representations for gender prediction. In addition, since different words and news usually have different informativeness for gender prediction, we apply attention mechanism at both word and news levels to select important words and news articles for user representation learning. Extensive experiments on a real-world dataset validate the effectiveness of our approach.

Acknowledgments

The authors would like to thank Microsoft News for providing technical support and data in the experiments, and Jiun-Hung Chen (Microsoft News) and Ying Qiao (Microsoft News) for their support and discussions. This work was supported by the National Key Research and Development Program of China under Grant number 2018YFC1604002, the National Natural Science Foundation of China under Grant numbers U1836204, U1705261, U1636113, U1536201, and U1536207, and the Tsinghua University Initiative Scientific Research Program.

References

1. Buraya, K., Farseev, A., Filchenkov, A.: Multi-view personality profiling based on longitudinal data. In: CLEF. pp. 15–27. Springer (2018)
2. Ciccone, G., Sultan, A., Laporte, L., Egyed-Zsigmond, E., Alhamzeh, A., Granitzer, M.: Stacked gender prediction from tweet texts and images notebook for pan at clef 2018. In: CLEF. p. 11p (2018)
3. Culotta, A., Kumar, N.R., Cutler, J.: Predicting the demographics of twitter users from website traffic data. In: AAAI. pp. 72–78 (2015)
4. Das, A.S., Datar, M., Garg, A., Rajaram, S.: Google news personalization: scalable online collaborative filtering. In: WWW. pp. 271–280. ACM (2007)
5. Farnadi, G., Tang, J., De Cock, M., Moens, M.F.: User profiling through deep multimodal fusion. In: WSDM. pp. 171–179 (2018)
6. Filippova, K.: User demographics and language in an implicit social network. In: EMNLP. pp. 1478–1488 (2012)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
8. Hu, J., Zeng, H.J., Li, H., Niu, C., Chen, Z.: Demographic prediction based on user's browsing behavior. In: WWW. pp. 151–160 (2007)

9. Kim, Y.: Convolutional neural networks for sentence classification. In: EMNLP. pp. 1746–1751 (2014)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Li, W., Dickinson, M.: Gender prediction for chinese social media data. In: RANLP. pp. 438–445 (2017)
12. Mac Kim, S., Xu, Q., Qu, L., Wan, S., Paris, C.: Demographic inference on twitter using recursive neural networks. In: ACL. vol. 2, pp. 471–477 (2017)
13. Malmi, E., Weber, I.: You are what apps you use: Demographic prediction based on user's apps. In: ICWSM. pp. 635–638 (2016)
14. Mislove, A., Lehmann, S., Ahn, Y.Y., Onnela, J.P., Rosenquist, J.N.: Understanding the demographics of twitter users. ICWSM **11**(5th), 25 (2011)
15. Mukherjee, S., Bala, P.K.: Gender classification of microblog text based on authorial style. Information Systems and e-Business Management **15**(1), 117–138 (2017)
16. Nguyen, D., Gravel, R., Trieschnigg, D., Meder, T.: " how old do you think i am?" a study of language and age in twitter. In: ICWSM. pp. 439–448 (2013)
17. Nguyen, D., Smith, N.A., Rosé, C.P.: Author age prediction from text using linear regression. In: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. pp. 115–123 (2011)
18. Nguyen, D., Trieschnigg, D., Doğruöz, A.S., Gravel, R., Theune, M., Meder, T., De Jong, F.: Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In: COLING. pp. 1950–1961 (2014)
19. Peersman, C., Daelemans, W., Van Vaerenbergh, L.: Predicting age and gender in online social networks. In: SMUC. pp. 37–44 (2011)
20. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: EMNLP. pp. 1532–1543 (2014)
21. Phuong, T.M., et al.: Gender prediction using browsing history. In: Knowledge and Systems Engineering, pp. 271–283. Springer (2014)
22. Rangel, F., Rosso, P., Montes-y Gómez, M., Potthast, M., Stein, B.: Overview of the 6th author profiling task at pan 2018: Multimodal gender identification in twitter. Working Notes Papers of the CLEF (2018)
23. Rangel Pardo, F.M., Celli, F., Rosso, P., Potthast, M., Stein, B., Daelemans, W.: Overview of the 3rd author profiling task at pan 2015. In: CLEF. pp. 1–8 (2015)
24. Reddy, T.R., Vardhan, B.V., Reddy, P.V.: N-gram approach for gender prediction. In: IACC. pp. 860–865. IEEE (2017)
25. Rosenthal, S., McKeown, K.: Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations. In: ACL. pp. 763–772 (2011)
26. Sezerer, E., Polatbilek, O., Sevgili, Ö., Tekir, S.: Gender prediction from tweets with convolutional neural networks: Notebook for pan at clef 2018. In: CLEF (2018)
27. Wang, J., Li, S., Zhou, G.: Joint learning on relevant user attributes in micro-blog. In: IJCAI. pp. 4130–4136 (2017)
28. Wang, L., Li, Q., Chen, X., Li, S.: Multi-task learning for gender and age prediction on chinese microblog. In: NLPCC. pp. 189–200. Springer (2016)
29. Wu, C., Wu, F., Liu, J., He, S., Huang, Y., Xie, X.: Neural demographic prediction using search query. In: WSDM. pp. 654–662. ACM (2019)
30. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In: NAACL. pp. 1480–1489 (2016)
31. Zhang, D., Li, S., Wang, H., Zhou, G.: User classification with multiple textual perspectives. In: COLING. pp. 2112–2121 (2016)