

Emotion Detection with Neural Personal Discrimination

Xiabing Zhou¹, Zhongqing Wang^{1*}, Shoushan Li¹,
Guodong Zhou¹, Min Zhang¹

¹School of Computer Science and Technology, Soochow University, China
wangzq.antony@gmail.com,
{zhouxiabing, lishoushan, gdzhou, minzhang}@suda.edu.cn

Abstract

There have been a recent line of works to automatically predict the emotions of posts in social media. Existing approaches consider the posts individually and predict their emotions independently. Different from previous researches, we explore the dependence among relevant posts via the authors' backgrounds, since the authors with similar backgrounds, e.g., *gender*, *location*, tend to express similar emotions. However, such personal attributes are not easy to obtain in most social media websites, and it is hard to capture attributes-aware words to connect similar people. Accordingly, we propose a Neural Personal Discrimination (NPD) approach to address above challenges by determining personal attributes from posts, and connecting relevant posts with similar attributes to jointly learn their emotions. In particular, we employ adversarial discriminators to determine the personal attributes, with attention mechanisms to aggregate attributes-aware words. In this way, social relationship among different posts can be better addressed. Experimental results show the usefulness of personal attributes, and the effectiveness of our proposed NPD approach in capturing such personal attributes with significant gains over the state-of-the-art models.

1 Introduction

The advent of social media and its prosperity enable the creation of massive online user-generated content including opinions and product reviews. Analyzing such user-generated contents allows to detect the users' emotional states, which are useful for various downstream applications.

In the literature, there are a large number of works on emotion detection (Roberts et al., 2012; Abdul-Mageed and Ungar, 2017; Gupta et al., 2017), both discrete and neural models have been

used to predict the emotions of posts in social media. For example, Roberts et al. (2012) used a series of binary SVM classifiers to detect the emotion of a post, while Gupta et al. (2017) used sentiment based semantic embedding and a LSTM model to learn the representation of a post for emotion detection.

Different from previous researches, which consider each post individually, we think that posts in social media are much correlated by the authors' backgrounds. Motivated by the *principle of homophily* (Lazarsfeld et al., 1954), the idea that similarity and connection tend to co-occur, or "birds of a feather flock together", suggests that users, connected by mutual personal backgrounds, may hold similar opinions toward a post (Thelwall, 2010). In the literature, the personal attributes, such as gender, location, age, have been proved useful in personal background construction (Li et al., 2016): people with different attributes tend to express their emotions through different ways. For example, the *happiness* emotion in [E1] is expressed through some femininity sense words, such as "little brother", "handsome", while the *happiness* emotion in [E2] is expressed using a dialectal word "bashi(*comfortable*)", which contains strong characteristic of the Sichuan dialect¹. Therefore, it is necessary to jointly detect the emotion of posts with the personal attributes.

[E1] 祝贺中国男篮夺冠，新疆小哥哥，真是帅。

(*Congratulations to Chinese basketball team on becoming the champion, the little brother of Xinjiang is really handsome.*)

¹Sichuan is a southwest province in China, Sichuan dialect(Sichuanese) is quite specific and different from standard Mandarin.

* corresponding author

[E2] 这种天气吃这个真是巴适!
(*It is just bashi to eat this in such weather!*)

However, the personal attributes are not easy to obtain in most social media websites. On one hand, most websites may not contain useful personal information. On the other hand, people are normally not willing to attach their personal information in social media. Besides, integrating personal attributes into emotion detection is challenging, since it is hard to capture attributes-aware words, such as “little brother” and “bashi”(comfortable), to connect the posts with similar backgrounds. Although there are some related works on either personal attribute extraction (Wang et al., 2014) or emotion detection with personal attributes (Li et al., 2016), none of them address both challenges at the same time.

In this paper, we propose a Neural Personal Discrimination (NPD) model with both adversarial discriminators and attention mechanisms to tackle above challenges. Here, the *Adversarial discriminators* (Goodfellow et al., 2014) are used to determine the personal attributes, e.g., gender or location, of a post, providing the inherent correlation between emotions and personal backgrounds, while the *Attention mechanisms* (Wang et al., 2016) are utilized to aggregate the representation of informative attributes-aware words into a vector for emotion prediction, providing insights into which words contribute to a personal background. Experimental results show the usefulness of personal attributes in emotion detection, and the effectiveness of our proposed NPD model with both adversarial discriminators and attention mechanisms over the state-of-the-art discrete and neural models.

2 Related Work

Earlier works on emotion detection are based on discrete models. For example, Yang et al. (2007) built a support vector machine (SVM) model and a conditional random field (CRF) model for the emotion detection. Bhowmick et al. (2009) used a multi-label k NN model to classify a new sentence into multiple emotion categories. Quan et al. (2015) proposed a logistic regression model for social emotion detection. Recently, with the development of artificial intelligence, neural network models have been successfully applied to various NLP tasks (Collobert et al., 2011; Goldberg,

2016). However, few works use neural network models for emotion detection. Abdul-Mageed and Ungar (2017) used a gated recurrent neural network model for emotion detection with a large-scale dataset. Zhang et al. (2018) used an auxiliary and attention based LSTM to detect emotion on a cross-lingual dataset.

Lexicon and social information are very important for emotion detection, and there are many researches focus on this topic. For example, Srapparava and Mihalcea (2008) used WordNet-Affect to compute the sentimental score of a post. More recently, In addition, Hovy (2015) used both the age and gender information of the authors to improve the performance of sentiment analysis. Vosoughi et al. (2016) explored the relationship among locations, date time, authors and sentiments.

Different from previous works which consider each post individually, we think that the posts in social media can be connected through the authors’ backgrounds and should be better addressed. On the basis, we propose a neural personal discrimination model to determine the personal background attributes from each post through adversarial discriminators, and aggregate the representation of informative attributes-aware words through attention mechanisms.

3 Vanilla Model for Emotion Detection

In this section, we propose a vanilla model. In the next section, we show how to utilize the neural personal discrimination model to improve the vanilla model by capturing personal attributes.

3.1 Document Representation

In general, we denote a post as a document d with n words $\{w_1, w_2, \dots, w_n\}$. Given the post, we use a standard Long Short-Term Memory (LSTM) model to learn the shared document representation. Specially, we transform each token w_i into a real-valued vector x_i using the word embedding vector of w_i , obtained by looking up a pre-trained word embedding table D via the skip-gram algorithm to train embeddings (Mikolov et al., 2013). We then employ the LSTM model over d to generate a hidden vector sequence $\{h_1, h_2, \dots, h_n\}$. At each step t , the hidden vector h_t of the LSTM model is computed based on the current vector x_t and the previous vector h_{t-1} with $h_t = \text{LSTM}(x_t, h_{t-1})$. The initial state and all stand LSTM parameters are randomly initialized and

tuned during training.

3.2 Multi-label Emotion Detection

Emotion detection aims to predict the emotion labels of posts. We follow (Wang et al., 2016) which adopts five kinds of emotions² in the study. Since there may be more than one emotion in a post, emotion detection can be considered as a multi-label classification task: we use K emotion-specific binary perceptions ($K = 5$) to predict if the post has the corresponding emotion or not. The advantage of multi-label classification is that it learns and predicts all the emotion labels jointly.

Formally, giving an input vector H , a hidden layer is first used to induce a set of high-level features for each emotion j :

$$\tilde{H}^j = \sigma(W^j H + b^j), \quad (1)$$

and then, \tilde{H}^j is used as inputs to a softmax output layer:

$$\hat{y}^j = \text{softmax}(\tilde{W}^j \tilde{H}^j + \tilde{b}^j) \quad (2)$$

Here, W^j , b^j , \tilde{W}^j , and \tilde{b}^j are model parameters.

3.3 Training

Given the word sequence in a post, our training objective is to minimize the cross-entropy loss over a set of training examples $(x_i, y_i)_{i=1}^N$, with a ℓ_2 -regularization term,

$$J_y(\theta_y) = - \sum_{i=1}^N \sum_{j=1}^K y_i^j \log \hat{y}_i^j + \frac{\lambda}{2} \|\theta_y\|^2 \quad (3)$$

where y_i^j represents the label of the j -th emotion for x_i , θ_y is the set of model parameters and λ is the parameter for ℓ_2 regularization.

In this paper, the model parameters are optimized by AdaGrad (Duchi et al., 2011), and Skip-gram algorithm (Mikolov et al., 2013) is used for word embedding.

4 Neural Personal Discrimination Model

The drawback of above vanilla model is that it does not consider the deep personal relationship among different posts.

In this study, we think that the posts in social media can be connected through the authors' backgrounds. Therefore, we propose a Neural Personal Discrimination (NPD) model to connect people and learn their emotions collectively. We use *adversarial discriminators* to determine the personal attributes to construct the personal profiles, and employ *attention mechanisms* to aggregate attributes-aware words. In this way, the social relationship between different posts can be well addressed.

Figure 1 illustrates our proposed neural personal discrimination model for emotion detection. In particular, we first learn the representation of each post using a LSTM model, same as the vanilla joint model. Then, we use adversarial discriminators to determine the personal attributes of each post. Finally, we employ attention mechanisms to aggregate the representation of informative attributes-aware words into a vector for the emotion prediction. In the following of this section, we illustrate the details of the infrastructure one by one.

4.1 Personal Adversarial Discriminators

A straightforward way to jointly detect personal attributes and the emotion of a post is to treat emotion detection as a multi-label classification. However, such model may not be able to separate the posts from different attributes directly, and thus fail to learn the relationship between the emotion of a post and the personal backgrounds of the authors. To address this issue, we utilize adversarial discriminators to determine the personal attributes of the authors, and to learn the emotion and the attributes of the authors collectively. Adversarial networks have achieved much success in various studies, especially in image and text generation (Goodfellow et al., 2014; Wang and Wan, 2018; Fedus et al., 2018). In this part, we propose two adversarial discriminators, i.e., a gender discriminator and a location discriminator, to determine the personal attributes of each post.

Gender Discriminator. The gender discriminator is employed to determine the author's gender of each post. Let $g_i \in [0, 1]$ represents the probability of the gender label (female or male) for the gender discriminator, and f is the function parameterized by θ_f which maps an embedding vector to a hidden representation h_i^g from the post x_i . Here, the gender discriminator $G(h_i^g; \theta_g) \rightarrow \hat{g}_i$ param-

²The five emotions include *happiness*, *sadness*, *anger*, *fear*, and *surprise*, please refer to "Experiments Section" of the paper for more details.

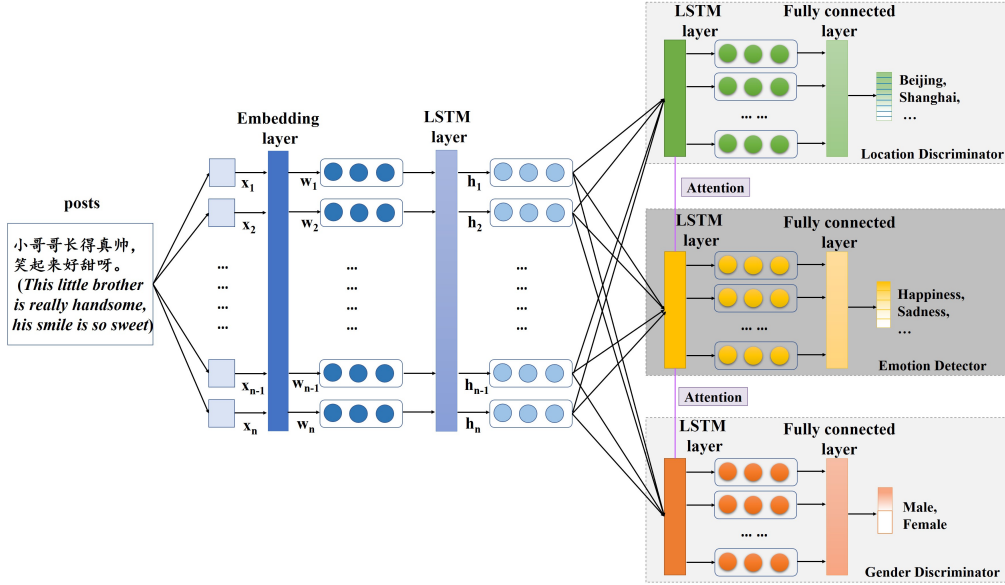


Figure 1: Overview of the neural personal discrimination model.

eterized by θ_g maps a hidden representation vector h_i^g to a predicted gender label \hat{g}_i with the loss function is:

$$J_{gend} = \sum_{i=1}^N \{g_i \log \hat{g}_i + (1 - g_i) \log(1 - \hat{g}_i)\}, \quad (4)$$

where $\hat{g}_i = G(f(x_i))$.

In this study, the gender discriminator is trained towards a saddle point of the loss function through maximizing the loss over θ_g while minimizing the loss over θ_f (Ganin et al., 2017).

Location Discriminator. The location discriminator is employed to determine the authors' location of a post³. Let $\ell_i^j \in [0, 1]$ represents probability of the j -th location information of the i -th post and $j \in \{1, 2, \dots, m\}$, where, m is the number of provinces. The loss function is:

$$J_{loc} = \sum_{i=1}^N \sum_{j=1}^m \ell_i^j \log \hat{\ell}_i^j, \quad (5)$$

where, $\hat{\ell}_i^j = L(f(x_i))$ and $L(h_i^\ell; \theta_\ell)$ are the parameters of the location discriminator, $h_i^\ell = f(x_i)$ is the hidden represent from the post x_i

From the optimization of both discriminators, we can find that both h^g and h^ℓ represent the latent feature representation of posts, which integrate the discrimination of various personal information. With the goal at $G(\theta_g)$ and $L(\theta_\ell)$ try best

³In this study, location discriminator is used to detect the author's province of a post.

to determine the gender and location of the authors. The adversarial network makes use of min-max optimization.

4.2 Personal Attention Mechanisms

In emotion detection, not all words contribute equally to the representation of emotions and personal attributes. Hence, we employ attention mechanisms to extract the words that are important to the personal backgrounds of posts, and to aggregate the representations of those informative attributes-aware words. With regard to the two adversarial discriminators, we propose attention mechanisms to build two representation (v^g and v^ℓ) from the gender and location discriminators respectively, and then concatenate them together to construct the overall personal representation through the informative attributes-aware words .

Gender Attention. We use an attention function to aggregate the gender-aware representation of the salient words to formulate the gender attention vector v^g . Here, the gender attention model outputs a continuous vector $v^g \in \mathbb{R}^{d \times 1}$ recurrently by feeding the hidden representation vectors $\{h_1^g, h_2^g, \dots, h_{n_t}^g\}$ as inputs. Specifically, v^g is computed as a weighted sum of h_i^g ($0 \leq i \leq n_t$), namely

$$v^g = \sum_i^{n_t} \alpha_i h_i^g \quad (6)$$

where n_t is the hidden variable size, $\alpha_i \in [0, 1]$ is the weight of h_i^g , and $\sum_i \alpha_i = 1$. For each

piece of hidden state h_i^g , the scoring function is calculated as follows:

$$v_i^g = \tanh(W_g h_i^g + b^g) \quad (7)$$

$$\alpha_i = \frac{\exp(v_i^g)}{\sum_j \exp(v_j^g)} \quad (8)$$

Location Attention. Similar with the gender attention mechanism, the location attention model outputs a continuous vector $v^\ell \in \mathbb{R}^{d \times 1}$ recurrently by feeding the hidden representation vectors $\{h_1^\ell, h_2^\ell, \dots, h_{n_k}^\ell\}$ as inputs. Specifically, v^ℓ is computed as a weighted sum of h_i^ℓ ($0 \leq i \leq n_t$), namely

$$v^\ell = \sum_i^{n_k} \beta_i h_i^\ell \quad (9)$$

$$v_i^\ell = \tanh(W_\ell h_i^\ell + b^\ell) \quad (10)$$

$$\beta_i = \frac{\exp(v_i^\ell)}{\sum_j \exp(v_j^\ell)} \quad (11)$$

where, n_k is the location hidden variable size, and β_i is the same setting as α_i .

Finally, we concatenate v^g and v^ℓ to capture the overall personal representation through all the personal attributes discriminators.

$$V = v^g \oplus v^\ell \quad (12)$$

4.3 Adversarial Training with Neural Personal Discrimination

The proposed NPD model can be trained in a end-to-end manner once we obtain the loss function of the emotion detector and the attribute discriminators. Our ultimate training goal is to minimize the loss function with parameters $\theta = \{\theta_f, \theta_y, \theta_g, \theta_\ell\}$ as follow:

$$J(\theta) = \lambda_1 J_y + \lambda_2 J_{gend} + \lambda_3 J_{loc} \quad (13)$$

Where λ_1 , λ_2 and λ_3 are the weight parameters to balance the importance of losses between the emotion detection and the two personal attribute discriminators.

Specifically, Eq. 13 is defined by finding a saddle point $\hat{\theta}_y, \hat{\theta}_f, \hat{\theta}_g, \hat{\theta}_\ell$ such that

$$(\hat{\theta}_f, \hat{\theta}_y) = \arg \min_{\theta_f, \theta_y} J(\theta_f, \theta_y, \hat{\theta}_g, \hat{\theta}_\ell) \quad (14)$$

$$\hat{\theta}_g = \arg \max_{\theta_g} J(\hat{\theta}_f, \hat{\theta}_y, \theta_g, \hat{\theta}_\ell) \quad (15)$$

$$\hat{\theta}_\ell = \arg \max_{\theta_\ell} J(\hat{\theta}_f, \hat{\theta}_y, \hat{\theta}_g, \theta_\ell) \quad (16)$$

As suggested previously, a saddle point is defined by Eq. 14–Eq. 16, and can be achieved as a stationary point the gradient updates:

$$\theta_f \leftarrow \theta_f - \mu \left(\lambda_1 \frac{\partial J_y^i}{\partial \theta_f} - \lambda_2 \frac{\partial J_{gend}^i}{\partial \theta_f} - \lambda_3 \frac{\partial J_{loc}^i}{\partial \theta_f} \right) \quad (17)$$

$$\theta_y \leftarrow \theta_y - \mu \lambda_1 \frac{\partial J_y^i}{\partial \theta_y} \quad (18)$$

$$\theta_g \leftarrow \theta_g - \mu \lambda_2 \frac{\partial J_{gend}^i}{\partial \theta_g} \quad (19)$$

$$\theta_\ell \leftarrow \theta_\ell - \mu \lambda_3 \frac{\partial J_{loc}^i}{\partial \theta_\ell} \quad (20)$$

where μ is the learning rate.

5 Experimentation

5.1 Experimental Settings

We collect the data from *Weibo.com*, one of the most popular SNS websites in China. We crawl all the posts and corresponding personal profiles from the website. The dataset contains 11,157 microblog posts from 839 users. We employed six graduated students to annotate the corpus with a well-defined annotation guideline. Every two annotators annotate a same part of corpus, if they have disagreement on some posts, we ask another annotator to vote with them. The annotation guideline is based on Lee et al. (2013). Five basic emotions are annotated, namely *happiness*, *sadness*, *fear*, *anger* and *surprise* (Lee et al., 2013; Wang et al., 2017). Table 1 illustrates the statistics of each emotion. From the table, we find that the frequency of *happiness* and *sadness* are similar. Moreover, the frequency of *fear* and *anger* is much less than other three emotions.

We randomly select 70% posts as the training data, and remaining 30% posts as the test data. For evaluation, F1-measure is used to evaluate the performance of proposed model in each emotion. Average F1-measure is used to evaluate the overall performance of all emotions.

The setting of hyper-parameters is: vocabulary size is 2000, batch size is set to 32, dropout ratio is 0.2, learning rate μ is set 0.0001, and $\lambda_1 : \lambda_2 : \lambda_3 = 1 : 1 : 1$

Table 2: Experimental results of different models.

Method	F1.					
	Happiness	Sadness	Anger	Surprise	Fear	Average
SVM	0.628	0.462	0.390	0.117	0.091	0.338
Abdul17	0.656	0.492	0.429	0.103	0.111	0.358
Vaswani17	0.644	0.494	0.392	0.113	0.121	0.353
NPD	0.657	0.510	0.459	0.135	0.127	0.378

Table 1: The statistics of emotion distribution in the dataset

Emtion	Post Number
Happiness	2,915
Sadness	2,454
Fear	359
Anger	153
Surprise	601
None	4,675

5.2 Experimental Results

We compare the proposed Neural Personal Discrimination (NPD) model with several representative baselines models in Table 2, where, 1) **SVM** is a widely used baseline to predict the emotion of a post in social media (Yang et al., 2007). 2) **Abdul17** is a standard LSTM model which consist of a LSTM layer and a fully connected layer, and it is modified from the model in Abdul-Mageed and Ungar (2017). The LSTM model yields the state-of-the-art performance on emotion detection in recent researches. 3) **Vaswani17** is an improved LSTM model with a self-attention mechanism. The self-attention mechanism is used to capture the structural information and has been successfully applied in various natural language processing tasks recently (Cheng et al., 2016; Vaswani et al., 2017)

From Table 2, we find that all of the neural models outperform SVM significantly. This indicates that neural models are much more effective than discrete models in emotion detection. In addition, our proposed NPD model outperforms both the standard LSTM model (Abdul17) and the improved LSTM model with self-attention (Vaswani17) significantly. This shows the effectiveness of our proposed NPD model with both adversarial discriminators and attention mechanisms. This also shows the usefulness of personal attributes for emotion detection. Moreover, we find that the performance of Vaswani17 is even lower than the standard L-

STM model. This shows that simply integrating a self-attention mechanism may not be able to well capture informative words for emotion detection in social media.

5.3 Analysis and Discussion

In this subsection, we analyze the influence of different factors in the proposed NPD model, and give some statistics and examples to illustrate the effectiveness of the proposed NPD model with different personal attributes.

5.3.1 Influence of Personal Attributes

We illustrate the influence of personal attributes in the proposed NPD model in Table 3, where, 1) **LSTM-attributes** is a LSTM based multi-label classification model, which predicts both the emotion and the attribute labels of each post collectively. 2) **NPD-gender** ablates the location attribute, i.e., only considering the gender attributes in the NPD model. 3) **NPD-location** ablates the gender attribute, i.e., only considering the location attributes in the NPD model.

From the table, we can find that the performance of LSTM-attributes is much lower than LSTM model. This indicates that simple multi-label classification setting is not effective for integrating personal attributes. This may be due to the fact that basic multi-label setting fails to learn the relationship between the emotions of posts and the personal attributes of the authors well. In addition, both the NPD-gender and the NPD-location perform better than the LSTM model respectively. This shows the effectiveness of the proposed NPD model with both adversarial and attention networks, and the usefulness of both gender and location attributes. Finally, the proposed NPD model with all the attributes significantly outperforms all the other models. This suggests that we should integrate all the personal attributes for emotion detection in social media.

Table 3: Comparison of various models with different personal attributes.

Method	Average F1.
LSTM	0.358
LSTM-attributes	0.341
NPD-gender	0.360
NPD-location	0.363
NPD	0.378

Table 4: Influence of network structures.

Method	Average F1.
LSTM	0.358
LSTM-attention	0.363
LSTM-adversarial	0.375
NPD	0.378

5.3.2 Influence of Network Structures

After analyzing the influence of different attributes, we analyze the influence of network structures in Table 4, where **LSTM-attention** ablates the adversarial discriminators, and only utilizes attention mechanisms with a multi-label classification setting, and **LSTM-adversarial** ablates the attention mechanisms and only utilizes adversarial discriminators for emotion detection.

From Table 4, we can see that both the attention mechanisms (LSTM-attention) and the adversarial discriminators (LSTM-adversarial) are effective in emotion detection. Moreover, the adversarial discriminators are much more effective than the attention mechanisms. This shows that the personal attribute discriminator is more important than learning informative attributes-aware words. Moreover, the NPD model obtains the best results by integrating both adversarial discriminators and attention mechanisms.

5.3.3 Statistics

We give the statistics of gender and location to explore the correlation between emotion and personal attributes.

Distribution of Gender. Figure 2 illustrates the distribution of emotions between genders. Here, the Y-axis is conditional probability of each emotion given gender. From the figure, we can find that women tend to express the sadness emotion, while men tend to express anger emotion. This may be due to the fact that the different personality has different emotion expressions, i.e., sentimentality of the female and the impulsion of the

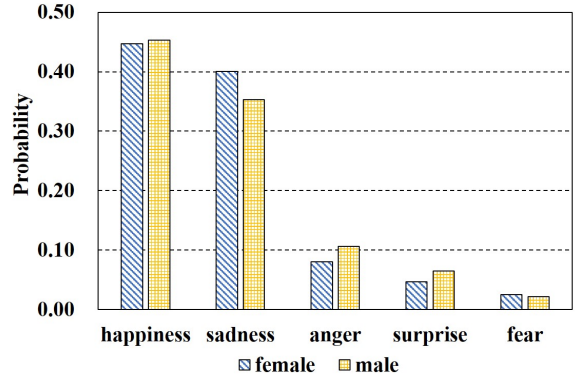


Figure 2: Distribution of gender.

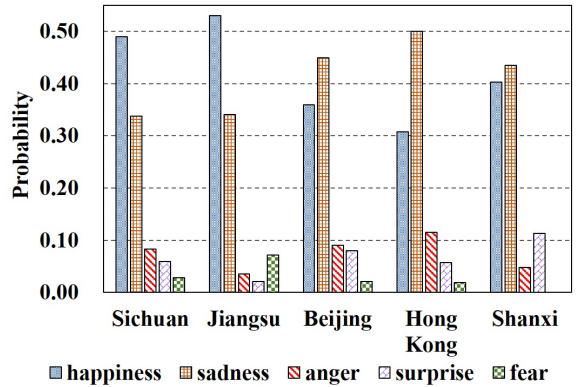


Figure 3: Distribution of location.

male.

Distribution of Location. Figure 3 is an example of the distribution of emotions between locations. As discussed in the above section, we use the province of authors as location attributes. Here, the Y-axis represents the conditional probability of each emotion given location. From the figure, we can find that the authors' location can may influence their emotions in many aspects. For example, people tend to express the positive (i.e., happiness) emotion than the negative emotion in Jiangsu. One of the most comfortable and developed regions in China. Due to air pollution and populations, people tend to express the negative emotion than the positive emotion in Beijing. In addition, people in Hong Kong always feel crowding and tend to express the sadness emotion. Finally, people always feel happy and comfortable in Sichuan, well known as "country of paradise" in China.

From the results of statistics, we can see that the personal attributes, like gender, location, can affect the emotion detection. Also, the experimen-

Table 5: Examples predicted labels of LSTM and NPD

Post	LSTM	NPD
[E3]@嬷嬷俊,连九份芋圆都会做,巴适惨了,冰冰凉Q弹水果甜品. (@Momo Jun, this guy can even make Taro Balls from Jiufen of Taiwan. It is so bash. This ice fruit dessert is so soft and chewy.)	Sadness	Happiness
[E4]下期节目,不容错过,有我的男神出场. (Next tv show is not to be missed, because my Mr. Mcdreamy will be coming.)	None	Happiness
[E5]我这身体不敢恭维啊,吃了好多的冰,结果痛经了吧,赶紧多喝点红糖姜茶. (I can't compliment my body. Eating too much ice leads to my dysmenorrhea. I must to drink more brown sugar ginger tea)	None	Sadness

tal results of personal attribute influence shows the same conclusion.

5.3.4 Case Study

We select three examples from the test set to evaluate the effectiveness of the proposed NPD model for better comparison with the LSTM model in Table 5. In [E3], through “bashi(*conformable*)” is a strong cue word of the *happiness* emotion, it is treated as a general word without indicating any location information from the LSTM model. At the same time, the location information (Sichuan Province) determined by the location discriminator and the attention mechanism makes “bashi(*conformable*)” critical in the NPD model, and these enable the NPD model determines the correct *happiness* emotion. In [E4] and [E5], both posts contain implicit gender information. For example, “Mr. Mcdreamy” implies the female’s affection for the handsome male, and “dysmenorrhea” is a female physiological disease. Without the background of such gender information, it is impossible to infer any potential emotional information from these two words. This explains why the LSTM model fails to detect any emotion from the examples (*None* of emotion). However, our NPD model successfully determines the gender of these two posts by the gender discriminator, and improves the weight of these two words for emotion detection by the attention mechanisms. In this way, the emotions of the three examples are correctly detected through the proposed NPD model.

6 Conclusion

Most of previous studies consider each post individually in emotion detection, one of the most important tasks in sentiment analysis. However, since the posts in social media are generated by

users, it is natural that these posts can be connected through authors’ personal background attributes. In this paper, we propose a neural personal discrimination model with both adversarial discriminators and attention mechanisms to connect posts with similar personal attributes. Here, the adversarial discriminators are used to determine the personal attributes, and the attention mechanisms are employed to aggregate attributes-aware words. In this way the social correlations between different posts can be captured. The experimental results show the usefulness of the personal attributes and the effectiveness of our proposed neural personal discrimination model in modeling such personal attributes with significant performance improvement over the state-of-the-art baselines.

Acknowledgment

This work was support by National Natural Science Foundation of China (No. 61702518, No.61806137, No.61525205) and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.. This work was also supported by the joint research project of Alibaba and Soochow University. Finally, we would like to thank the anonymous reviewers for their insightful comments and suggestions.

References

- Muhammad Abdul-Mageed and Lyle H. Ungar. 2017. [Emonet: Fine-grained emotion detection with gated recurrent neural networks](#). In *ACL*, pages 718–728.
- Plaban Kumar Bhowmick, Anupam Basu, Pabitra Mitra, and Abhishek Prasad. 2009. [Multi-label text classification approach for sentence level news emotion analysis](#). In *ICPRAI*, pages 261–266.

- Jianpeng Cheng, Li Dong, and Mirella Lapata. 2016. [Long short-term memory-networks for machine reading](#). pages 551–561.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. [Natural language processing \(almost\) from scratch](#). *Journal of Machine Learning Research*, 12:2493–2537.
- Niko Colneriç and Janez Demsar. 2018. Emotion recognition on twitter: Comparative study and training a unison model. *IEEE Transactions on Affective Computing*, PP(99):1–14.
- John C. Duchi, Elad Hazan, and Yoram Singer. 2011. [Adaptive subgradient methods for online learning and stochastic optimization](#). *Journal of Machine Learning Research*, 12:2121–2159.
- William Fedus, Ian J. Goodfellow, and Andrew M. Dai. 2018. Maskgan: Better text generation via filling in the ----- . In *ICLR*, pages 1–16.
- Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. 2017. [Domain-adversarial training of neural networks](#). pages 189–209.
- Yoav Goldberg. 2016. [A primer on neural network models for natural language processing](#). *J. Artif. Intell. Res.*, 57:345–420.
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. 2014. [Generative adversarial nets](#). In *NIPS*, pages 2672–2680.
- Umang Gupta, Ankush Chatterjee, Radhakrishnan Srikanth, and Puneet Agrawal. 2017. [A sentiment-and-semantics-based approach for emotion detection in textual conversations](#). *CoRR*, abs/1707.06996. Version 4.
- Dirk Hovy. 2015. [Demographic factors improve classification performance](#). In *ACL*, pages 752–762.
- Paul F Lazarsfeld, Robert K Merton, et al. 1954. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66.
- Sophia Yat Mei Lee, Ying Chen, Chu-Ren Huang, and Shoushan Li. 2013. [Detecting emotion causes with a linguistic rule-based approach](#). *Computational Intelligence*, 29(3):390–416.
- Junjie Li, Haitong Yang, and Chengqing Zong. 2016. [Sentiment classification of social media text considering user attributes](#). In *NLPCC/ICCPOL*, pages 583–594.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). In *NIPS*, pages 3111–3119.
- Xiaojun Quan, Qifan Wang, Ying Zhang, Luo Si, and Wenyin Liu. 2015. [Latent discriminative models for social emotion detection with emotional dependency](#). *ACM Trans. Inf. Syst.*, 34(1):2:1–2:19.
- Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, and Sanda M. Harabagiu. 2012. [Empatweet: Annotating and detecting emotions on twitter](#). In *LREC*, pages 3806–3813.
- Carlo Strapparava and Rada Mihalcea. 2008. [Learning to identify emotions in text](#). In *SAC*, pages 1556–1560.
- Mike Thelwall. 2010. [Emotion homophily in social network site messages](#). *First Monday*, 15(4):1–8.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *NIPS*, pages 6000–6010.
- Soroush Vosoughi, Helen Zhou, and Deb Roy. 2016. [Enhanced twitter sentiment classification using contextual information](#). *CoRR*, abs/1605.05195. Version 1.
- Ke Wang and Xiaojun Wan. 2018. [Sentigan: Generating sentimental texts via mixture adversarial networks](#). In *IJCAI*, pages 4446–4452.
- Xuren Wang and Qihui Zheng. 2013. Text emotion classification research based on improved latent semantic analysis algorithm. In *ICCSEE*, pages 210–213.
- Zhongqing Wang, Sophia Yat Mei Lee, Shoushan Li, and Guodong Zhou. 2017. [Emotion analysis in code-switching text with joint factor graph model](#). *IEEE/ACM Trans. Audio, Speech & Language Processing*, 25(3):469–480.
- Zhongqing Wang, Shoushan Li, Hanxiao Shi, and Guodong Zhou. 2014. [Skill inference with personal and skill connections](#). In *COLING*, pages 520–529.
- Zhongqing Wang, Yue Zhang, Sophia Yat Mei Lee, Shoushan Li, and Guodong Zhou. 2016. [A bilingual attention network for code-switched emotion prediction](#). In *COLING*, pages 1624–1634.
- Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. 2007. [Emotion classification using web blog corpora](#). In *WI-IAT*, pages 275–278.
- Lu Zhang, Liangqing Wu, Shoushan Li, Zhongqing Wang, and Guodong Zhou. 2018. [Cross-lingual emotion classification with auxiliary and attention neural networks](#). In *NLPCC*, pages 429–441.