Personality Traits Prediction Based on Users' Digital Footprints in Social Networks via Attention RNN

Shipeng Wang*, Lizhen Cui†, Lei Liu†, Xudong Lu†, Qingzhong Li†
* † *School of Software, Shandong University, Jinan, China*
† *Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan, China*
**wsp_sdu@163.com,*†*{clz, l.liu, dongxul, lqz}@sdu.edu.cn*
*Correspondence to: Qingzhong Li <**lqz@sdu.edu.cn**>*

***Abstract*—With the increasing popularity of social networks, massive digital footprints of individuals in online service platforms are generated. As a result, an emerging technology namely personality trait analysis has drawn much attention. The prediction and analysis of personality trait is an efficient way to voting prediction, review analysis, decision analysis and marketing. The existing studies generally employ classification models while ignore the temporal property of digital footprints, which may lead to unsatisfactory results. To make an improvement, this paper proposes an effective method to predict the personality traits by taking the temporal factors into account through the use of Attention Recurrent Neural Network (AttRNN). The experimental results based on the dataset of 19000 Facebook volunteers suggest the proposed method is effective for predicting personality traits.**

***Keywords*-Personality Traits Prediction; Digital Footprints; Recurrent Neural Network; Social Networks;**

I. Introduction

Human personality is a lifelong feature that can reflect the nature of human behavior and influence people's behavioral decisions. At the same time, personality is also an important factor influencing people's behavioral tendencies and psychological tendencies, where traits are defined as a cross-situational and temporally stable set of individual attributes. Analysing personality traits can help people understand human behavior and predict the future behavior in both real life and the network of crowd intelligence [1]. In addition, it's beneficial to the emergence of computational social science [2] and new services technologies such as recommended systems [3] and personalized search engines [4]

Traditional measurement methods in psychology research require a lot of manpower, material resources and financial resources, which make it difficult to handle a large number of users. Data-driven behavioral personality prediction can avoid some subjective biases and leverage the rich behavioral data, which can reflect human true personality traits in full perspectives.

With the constant use of the Internet, various online social networks, such as e-commerce platforms and social media platforms, have generated hundreds of millions of user digital footprints. Predicting personality traits based on user digital footprints has always been of great interest to the researchers [5] and made great success on voting prediction analysis, review analysis and personality analysis in *mypersonality* [1].

Recently, many machine learning methods have been employed to predict personality traits, with the main focus on classical classification model, such as Linear Regression [6] and Support vector machine to predict personality traits. In fact, personality traits tend to fluctuate slightly over time. However, the existing research ignores the influence of the temporal factors.

Therefore, this paper proposes to employ Attention Recurrent Neural Network (AttRNN) to predict the personality traits by taking the temporal factors into account. In this paper, we embed the users' discrete behaviors into footprints and introduces attention mechanism into Recurrent Neural Network. Through empirical study, we find that the personality trait "Openness" prediction accuracy (0.48) of the proposed method, measured by the Pearson correlation coefficient, is higher than the prediction accuracy of Openness (0.43) delivered by the state-of-the-art [6].

II. Methodology

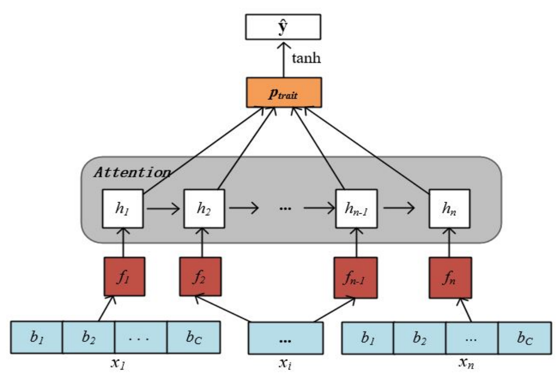

Figure 1. The framework of attention recurrent neural network

In our model, we embed the users' discrete behaviors into footprints and adopt Recurrent Neural Network with the attention mechinism in the hidden layer to predict users'

[1]https://www.psychometrics.cam.ac.uk/productsservices/mypersonality

2474-2473/20/$31.00 ©2020 IEEE
DOI 10.1109/SCC49832.2020.00015

Authorized licensed use limited to: Machakos University College. Downloaded on May 25,2021 at 02:01:14 UTC from IEEE Xplore. Restrictions apply.

personality traits on social networks. Figure 1 shows the framework of Attention Recurrent Neural Network (AttRNN).

A. Sequential Representation of Users Digital Footprints.

For the user digital footprints in social networks, we can represent them by multiple-hots encoding method. Suppose there are **c** (**c** $\in \mathbb{R}$) types of user's network behaviors (such as 'Like' items) in the social networks. Since there is at most one behavior for each user at any moment, in order to describe the user state more accurately, we express the all behaviors of each user as the footprints occurring in **t** (**t** $\in \mathbb{R}$) time intervals. These intervals are equally spaced, such as one day or one week. Suppose there are ***n*** time intervals(***n*** $\leq$ **c**).

For every user, his or her network behavior footprints in ***j***-st time interval (***j*** $\leq$ **t**) can be expressed as $x_j \in \{0,1\}^{\boldsymbol{c}}$. If the user liked ***i***-th item in ***j***-th time interval, x_{ji} = 1, else x_{ji} = 0. At the same time, the user's personality traits can be measured by specific personality traits scale (such as the Big Five Model - International Personality Item Pool), and the personality traits are represented by real valued *trait*, that is, *trait* $\in \mathbb{R}$. For Big Five Model personality traits, trait includes openness, conscientiousness, extroversion, agreeableness and neuroticism. Taking the openness prediction of user as an example, the Users Digital Footprints in the ***j***-st time interval can be expressed as $f_j = x_j$. In AttRNN model, we use the same model framework to predict each of the five personality traits respectively.

B. Personality Trait Attention.

The goal of Attention Recurrent Neural Network is to predict the user's personality traits based on the user footprints $(f_1, f_2, ..., f_n)$. We can obtain the hidden layer state $(h_1, h_2, ..., h_n)$ of user behavior by applying Recurrent Neural Network. In order to reduce the impact of redundant information on personality prediction, this paper uses the attention mechanism to process historical information, and the calculation method is as follows:

First, we calculate the weight α_i at hidden state h_i as,

$$\alpha_i = {W_\alpha}^T h_i + b_\alpha, \tag{1}$$

where W_α and b_α are parameters to be learned. Then apply the softmax function to normalize the weight vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]$ as follows:

$$\boldsymbol{\alpha} = \mathrm{softmax}\left([\alpha_1, \alpha_2, \ldots, \alpha_n]\right). \tag{2}$$

Then we can calculate the user' current personality status P_{trait} in the social network through the weight vector $\boldsymbol{\alpha}$ and historical information $\boldsymbol{h}$. The calculation method is as formula (3):

$$P_{trait} = \sum_{i=1}^{n} \alpha_i \mathrm{h}_i. \tag{3}$$

C. Personality Trait Prediction.

The final personality representation of the user P_{trait} contains the network behavior features of them and it can be used to predict their personality traits. In this paper, *tanh* function is used as the activation function of the output layer. The activation function is as formula (4):

$$\hat{\mathrm{y}} = \beta \tanh\left(W_s * P_{trait} + b_s\right), \tag{4}$$

where W_s and b_s are the parameters to be learned, and β is a fixed real value, which is related to the score range of the scale. For example, if the value of the openness score is [-5,5] , then $\beta_{openness}$ = 5.

For loss function, this model uses cross entropy to measure the difference between prediction results $\hat{\mathrm{y}}$ and real results y, which is shown as follows:

$$\mathrm{cost} = \sum_{i=1}^{N} -y_i \log\left(\hat{y}_i\right) - \left(1 - y_i\right) \log\left(1 - \hat{y}_i\right). \tag{5}$$

III. Experiments

A. Dataset Description

The dataset used in our experiment is obtained from the myPersonality.org database. In our experiment, the user-like dataset [7] contains data corresponding to 110,728 Facebook users with 1,580,284 Likes (digital footprints) and their Five Factor Model [8] personality scores, which is obtained from the International Personality Item Pool (IPIP) [9] questionnaire. And the Facebook users' likes consist of various contents such as pop music, movies, games, famous people, books, proverbs, organizations.

B. Experiments

Table I
Prediction Accuracy Expressed by PCCs

	ope	con	ext	agr	neu
Linear Regression	0.43	0.29	0.4	0.3	0.3
AttRNN	**0.48**	**0.31**	0.35	0.29	**0.31**
BiGRU	0.41	0.23	0.29	0.26	0.26

After data preprocessing according to [7], we complete our experiments based on the user-like dataset composed of 19,700 Facebook users with 8,530 Likes. Then we employ AttRNN and BiGRU to predict personality traits of Facebook users. BiGRU model is also a neural network and it is a two-layer GRU model, one layer is forward propagation and the other is backward propagation, which can better process context information and solve timing problems in text emotion analysis tasks. In our experiments, we convert the Like records of a user to ten sequences and take them as the input of AttRNN and BiGRU to predict Five Factor Model personality traits. The training set contains 18,000 users and the test set contains 1700 users.

Authorized licensed use limited to: Machakos University College. Downloaded on May 25,2021 at 02:01:14 UTC from IEEE Xplore. Restrictions apply.

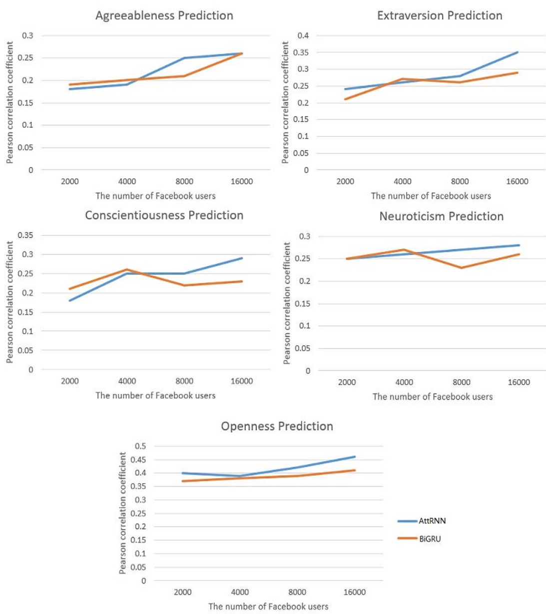

Figure 2. The results of personality traits prediction

C. Results of Personality Traits Prediction

The accuracy of the personality traits prediction on each dimension of FFM is expressed by the Pearson correlation coefficient. Table I shows the prediction accuracy of linear regression model [6], BiGRU model and AttRNN. AttRNN overall outperforms linear regression and BiGRU model with significantly better performance on predicting openness, conscientiousness and neuroticism. The result of BiGRU model is poor and there may be two reasons. Firstly, the BiGRU model is unsuitable for personality trait prediction based on Facebook Likes because the content of the likes contain too little information. It is too difficult for the BiGRU model to learn some knowledge. Secondly, the performance of BiGRU model is limited by small data volume. In comparison, our method performs better.

Figure 2 shows the results of personality traits. Compared to four other personality traits, the accuracy of openness (0.48) is pretty higher than baselines, because openness is largely expressed through personal interests and preferences and the Facebook user-likes dataset reflects the users' interests appropriately. It is also confirmed that people tend to show their real personalities in an anonymous environment.

In the meantime, as the scale of data increases, the accuracy of the AttRNN model is gradually increasing, which can be verified by the results in Figure 2. And all the results show AttRNN method is effective in massive data.

IV. Conclusion and Future Work

Massive digital footprints of online social networks makes it possible to predict personality traits by means of data-driven methods. In this paper, we propose to use Attention Recurrent Neural Network (AttRNN) to deal with the personality traits prediction problem, which takes the sequence of digital footprints into account and introduces attention mechanism into traditional RNN model. Experimental results proved the effectiveness of AttRNN model for personality traits prediction task.

As future work, we will consider the implicit meaning of the users' digital footprints, such as deep semantic features, and further analyze and predict personality traits in combination with the attention mechanism to improve the accuracy of the model.

Acknowledgment

This work is partially supported by National Key R & D Program No.2017YFB1400100, Shandong Provincial Major Scientific and Technological Innovation Project 2019JZZY020505, MSTIP 2019JZZY010109, TaiShan Industrial Experts Programme of Shandong Province tscy20160404.

References

[1] S. Wang, L. Cui, L. Liu, X. Lu, and Q. Li, "Projecting real world into crowdintell network: a methodology," *International Journal of Crowd Science*, 2019.

[2] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann *et al.*, "Social science. computational social science." *Science (New York, NY)*, vol. 323, no. 5915, pp. 721–723, 2009.

[3] M. Elahi, M. Braunhofer, F. Ricci, and M. Tkalcic, "Personality-based active learning for collaborative filtering recommender systems," in *Congress of the Italian Association for Artificial Intelligence*. Springer, 2013, pp. 360–371.

[4] W. T. Leung, D. L. Lee, and W. C. Lee, "Pmse: A personalized mobile search engine," *IEEE Transactions on Knowledge & Data Engineering*, vol. 25, no. 4, pp. 820–834, 2013.

[5] V. Kaushal and M. Patwardhan, "Emerging trends in personality identification using online social networks—a literature survey," *ACM Transactions on Knowledge Discovery From Data*, vol. 12, no. 2, p. 15, 2018.

[6] K. Michal, S. David, and G. Thore, "Private traits and attributes are predictable from digital records of human behavior," *Pnas*, vol. 110, no. 15, pp. 5802–5805, 2013.

[7] M. Kosinski, Y. Wang, H. Lakkaraju, and J. Leskovec, "Mining big data to extract patterns and predict real-life outcomes." *Psychological Methods*, vol. 21, no. 4, pp. 493–506, 2016.

[8] R. R. Mccrae and O. P. John, "An introduction to the five-factor model and its applications," *J Pers*, vol. 60, no. 2, pp. 175–215, 2010.

[9] L. R. Goldberg, J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough, "The international personality item pool and the future of public-domain personality measures ," *Journal of Research in Personality*, vol. 40, no. 1, pp. 84–96, 2006.

Authorized licensed use limited to: Machakos University College. Downloaded on May 25,2021 at 02:01:14 UTC from IEEE Xplore. Restrictions apply.