

LSTM Based Personality Prediction from Social Media

Thahira M, Shafna K P

Dept. of Computer Science & Engineering
MGM College of Engineering and
Pharmaceutical Sciences, Valanchery,
Kerala, India

tharasaid@gmail.com, shafnasaid1@gmail.com

Abstract- The usage of social media networks are increasing rapidly. Which are used as a platform to share People's feelings, emotions and experiences etc. This Proposed method predict the person's behaviour from social media networks. Personality prediction from social media networks are become a challenging task. This method follow the big-five-factor model (OCEAN) namely, openness to experience (O), conscientiousness (C), extraversion (E), agreeableness (A), neuroticism (N) for defining personality. Here the mathematical regular expression operation is used for the data preprocessing. There are three feature selection algorithms such as Pearson correlation coefficient (PCC), information gain(IG) and chi-squared (CHI) method are used for the relevant features. Deep learning neural network such as LSTM is used as a classifier for predicting personality of a person. This Proposed method improves the accuracy of personality prediction compare to other existing system.

Keywords- Lstm, regular expression operation, chi-squared method, Pearson correlation coefficient, information gain(IG)

I. INTRODUCTION

Nowadays, the usage of social media networks are very high. These social media networks are used to build a social networks or social relationships with peoples. The most famous social media platforms are Facebook, Twitter, Google+, and Instagram etc.

The popularity of these sites are higher day by day because of the ease access throughout the world and it have user-friendly interfaces. By using these social media platforms we can start communication with others within a short period of time. While using these sites, the users are facilitated by many activities, such as posting statuses, sharing others' posts, liking others' posts, commenting on others' posts, chatting directly with the friends, and playing online games with the friends etc. In this project, predict the personality behaviour of a person from these social media platform. Predicting personalities of a person from this digital media is a computationally challenging problem.

Here different kind of methods are used for the personality prediction. For defining personality, we follow the widely used big-five-factor model (OCEAN)[1] namely, openness to experience (O), conscientiousness (C), extraversion (E), agreeableness (A), neuroticism (N). Data preprocessing, feature extraction and feature selection methods are used for find out the relevant features. In this method, finding the most

relevant feature is one of the most challenging tasks. Those relevant features are trained to the classification model, and then testing is performed. Here Deep learning neural network is used as a classifier such as Long Short term Memory (LSTM). The paper categorized as follows. Section II describe the literature survey of the previous methods include data preprocessing, feature extraction, feature selection, classification. Section III explain the proposed method. Finally, the Section IV gives the conclusion

II. LITERATURE SURVEY

There have been several studies about the prediction of behaviour from the social media. In this section, we are explain about the previous methods that is feature extraction, feature selection and classification.

1. Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media

A.

In this paper[1] big-five-factor models (BFFM) are used for defining the personality of a person, there are four positive personality traits namely, openness to experience, conscientiousness, extraversion, agreeableness, and one personality negative trait that is neuroticism. Different feature selection algorithms, such as the chi-squared (CHI) method, Pearson coefficient correlation, information gain (IG), correlation-based feature subset

(CFS)-based subset evaluation, and symmetrical uncertainty attribute evaluation, are used to predict the big-five personality traits. Linguistic Inquiry and Word Count (LIWC) tool is used for the feature extraction it is the one of the relevant method for extraction. In this paper, propose five different classifiers, namely, the naive Bayes(NB), decision tree (DT), random forest (RF), simple logistic regression (SLR), and support vector machine (SVM) for the training and testing of classification model. These five classifiers are machine learning classifiers. By using these classifiers the accuracy of prediction will be less compare to Deep learning method.

2. Predicting Facebook-Users' Personality based on Status and Linguistic Features via Flexible Regression Analysis Techniques

In this paper[2], it address that the problem of predicting the personality of Facebook users by using SVR and LR analysis respectively. Linguistic Inquiry and Word Count (LIWC) tool and Latent Dirichlet Allocation (LDA) are used to extract facebook user's statuses and posts. By using this LDA method the performance of regression techniques increases. The topics extracted by LDA have been converted to suitable matrices that have been then used to predict the BIG5 personality scores. In this paper, it include a pre-processing step and feature extraction step. In preprocessing it cleaned the unwanted data in order to eliminate characters, names, spaces, stop words from this statuses and bag of words have been created.

Then applied to the following regression techniques into this dataset: (i) LR, (ii) SVR with linear kernel , (iii) SVR with polynomial kernel, and (iv) SVR with RBF kernel. SVR is a widely used regression technique which is based on Support Vector Machine.

3. Personality Prediction System from Facebook Users

In this paper [3], Big five model personality method is used for predicting the personality of a person from facebook user information. Comparing the other literature method it introduce deep learning neural network classification process. It gives high accuracy for the personality prediction. But it is possibly used to small number of dataset for this study. It is the disadvantages of this paper. For the classification process, traditional machine learning and deep learning algorithms are used here. Traditional machine learning algorithms included Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Gradient Boosting, and Linear Discriminant Analysis (LDA). Deep learning method is MLP (Multi-Layer Perceptron). MLP consists of input, hidden, and output layers which is using a basic algorithm for training,

known as backpropagation. In the Features Selection method it is used for filter or remove the features and to get low correlation for the personality prediction. Chi-square method is used for the feature selection and the correlation values are calculated by using this algorithm. In preprocessing NLTK library is used for removing names, stop words and stemming etc.

4.Feature Select: a software for feature selection based on machine learning approaches

Feature selection, is a preprocessing step, this feature selection method is used in various sciences such as biology, engineering, computer science, and other fields. It has been developed in the MATLAB programming Language. In this paper [4], introduce a software application called FeatureSelect. In this software based on filter method it have relatively lower performance than the wrapper method. This FeatureSelect software also consists of optimisation algorithms and three types of learners. Here SVM, ANN and DT are used as a learner that can be applied to classification and regression datasets. By using this optimization algorithm that can increase the performance.SVM have a library called LIBSVM and that provides many options for classification and regression problems, and develop FeatureSelect based on this library. It provides a user-friendly and straightforward method of feature selection and that can be use any kind of research, and can easily be applied to any type of balanced and unbalanced data. It gives more accuracy, sensitivity, specificity, etc

5. Recognising Personality Traits Using Facebook Status Updates

In this paper [5], include automatic recognition of personality traits based on Facebook users information such as status updates, network properties and time factors etc. Personality traits are described using five factors, namely extraversion, neuroticism (the opposite of emotional stability), agreeableness, conscientiousness, and openness to experience(Big five model). Here use three kinds of numeric features such as LIWC features, Social Network features, Time-related features. For the training ,here using three types of machine learning algorithms such as Support Vector Machine with a linear kernel (SVM), Nearest Neighbor with k=1 (kNN) and Naive Bayes (NB)

6.Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution

In this paper[6]it is based on the feature selection and feature selection is a preprocessing step, it include removing irrelevant data, increasing learning accuracy, and improving result comprehensibility etc. This paper propose a fast filter method, a feature selection algorithm

that is Fast Correlation Based Filter for Feature Selection (FCBF) is implemented for the relevant feature selection as well as redundancy. By using this FCBF filter method, without using any learning algorithm for training. The efficiency and effectiveness of this method is very high.

III. PROPOSED SYSTEM

We have discussed dissimilar methods for predicting personality from social media that is machine learning based method such as naive Bayes (NB), decision tree (DT), random forest (RF), simple logistic regression (SLR), and support vector machine(SVM). By using this method the accuracy is very less. In this paper, we have to applied the classic classification methods to evaluate the performance of the proposed experimental method. Deep learning neural network such as LSTM [7] are implemented for the experiment process. This LSTM is used as a classifier and these are used for training and testing to this classification model. By using this LSTM neural Network we can improve the accuracy of this prediction. In the data preprocessing step, regular expression operation is used for the removal of names, space, characters etc. In these system LIWC method is used for the feature extraction and three algorithms such as Pearson correlation coefficient (PCC), information gain(IG), and chi-squared (CHI) method are used for the feature selection.

IV. CONCLUSION

In this paper, we are applied a Deep learning based neural network such as Long Short Term Memory (LSTM) neural network as a classifier for the personality prediction of person from social media. This LSTM method is one of the most important technique for predicting personality. It gives more accuracy in prediction than using the machine learning methods. This proposed system solves all the problems of existing system such as prediction of large scale datasets, prediction of high level features etc.

REFERENCES

- [1] Ahmed Al Marouf , Md. Kamrul Hasan, and Hasan Mahmud “Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media,” in Proc. IEEE Int Transactions on Computational Social Systems (Volume: 7, Issue: 3, June 2020).
- [2] P. Howlader, K. K. Pal, A. Cuzzocrea, and S. D. M. Kumar, “Predicting Facebook-users’ personality based on status and linguistic features via flexible regression analysis techniques,” in Proc. 33rd Annu. ACM Symp.Appl. Comput. (SAC), Pau, France, Apr. 2018, pp. 339–345.
- [3] T. Tandra, D. Suhartono, R. Wongso and Y. L. Prasetyo, “Personality prediction system from Facebook users,” in Proc. 2nd Int. Conf. Comput. Sci. Comput. Intell. (ICCSCI), Bali, Indonesia, Oct. 2017, pp. 604–611.
- [4] Y. Masoudi-Sobhanzadeh, H. Motieghader, and A. Masoudi-Nejad, “FeatureSelect: A software for feature selection based on machine learning approaches,” BMC Bioinf., vol. 20, no. 1, p. 170, 2019,doi: 10.1186/s12859-019-2754-0.
- [5] G. Farnadi, S. Zoghbi, M. Moens, and M. De Cock, “Recognising personality traits using Facebook status updates,” in Proc. WCPR, 2013, pp. 14–18.
- [6] L. Yu and H. Liu, “Feature selection for high dimensional data: A fast correlation-based filter solution,” Proc. 20th Int. Conf. Mach. Learn., 2003, pp. 856–863.
- [7] J. W. Pennebaker, M. E. Francis, and R. J. Booth. (2001). *Linguistic Inquiry and Word Count: LIWC2001*. Erlbaum, Mahwah, NJ, USA.[Online]. Available: <https://www.erlbaum.com>

About Authors

Thahira M received the B.Tech degree in Computer Science and Engineering from University Calicut, Kerala, India, in 2016. Presently, she is doing her Post Graduation in the Department of Computer Science and Engineering, at MGM College of Engineering and Pharmaceutical Sciences, Valanchery, Kerala. Her latest research include Data mining, Natural Language Processing, Artificial intelligence, Deep learning and Machine Learning.

Shafna KP obtained her Bachelor’s degree in Computer Science and Engineering from MEA Engineering college perinthalmanna, Kerala, India and Master’s degree in Network Engineering, Information and Technology from Gov. College of Engineering, Idukki, Kerala, India. Currently, she is working as Assistant Professor in the department of Computer Science and Engineering MGM College of Engineering and Pharmaceutical Sciences, Valanchery, Kerala, India. Her current research include Artificial Intelligence, Machine learning, Natural language Processing and Data mining.