SATP: Sentiment Augmented Topic Popularity Prediction on Social Media

Weizhi Gong, Zuowu Zheng, Xiaofeng Gao(✉), and Guihai Chen

Shanghai Key Laboratory of Scalable Computing and Systems,
Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai 200240, China
{gongwz,waydrow}@sjtu.edu.cn,{gao-xf,gchen}@cs.sjtu.edu.cn

Abstract. In this paper we propose a topic popularity prediction model to quantifiy popularity more accurately with senmantic information and incorporates sentiment ino popularity prediction.

Keywords: Topic popularity prediction · Sentiment analysis · Popularity quantification · Social media

1 Introduction

Most existing topic popularity prediction models measure popularity simply using forwarding or view count, ignore semantic relation between posts and topic, which may lead to inccuracy. Therefore, we propose a model to quantify popularity more accurately with posts' semantic information and incorporate sentiment into popularity prediction.

2 Model Construction

For the input of a series of posts about a certain topic, we aim to output the predicted popularity.

We first quantify the topic popularity using posts' forwarding number weighted by words' relevance WR to the topic. We use the idea of PageRank [1] to calculate word w's relevance to the topic $WR(w)$ in post d, defined in Eqn. (1):

$$WR(w_i) = \frac{1-\theta}{|d|} + \theta \cdot \sum_{j \to i} \frac{\rho(w_i, w_j)}{\sum\limits_{k \to j} \rho(w_k, w_j)} \cdot WR(w_j), \tag{1}$$

where $|d|$ is the length of the post that contains w_i, $\rho(w_i, w_j)$ is distance between word w_i and w_j, which is a linear combination of their semantic distance calculated by *Word2Vec* and lexical distance.

We then use MPQA subjective lexicon [3] to evaluate sentiment intensity of the subjective words and tag 64 most popular emojis with their sentiment

S. Yangui et al. (Eds.): ICSOC 2019, LNCS 11895, pp. 575–577, 2019.
https://doi.org/10.1007/978-3-030-33702-5

intensity. With the sentiment intensity of subjective words and emojis, we create dataset to train the hybrid architecture [2] of Bi-LSTM and convolutional network to calculate sentiment intensity of a certain post. After getting popularity and sentiment time series, we use convolutional network to learn their dependence and learn the history influence information of each time period and use Autoregressive model to predict future popularity.

3 Experiments

We test our model on the Twitter data from Dec. 23, 2017 to Mar. 19, 2018 on different topics. Here we use result of topic Gun Control and Trump as an example to illustrate performance of our model. We evaluate the accuracy of each prediction using *Mean Square Error (MSE)*, the result is shown in Fig. 1 and Tab. 1.

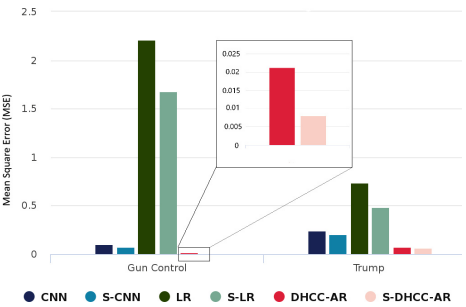

Fig. 1. Prediction performance of each model on each topic, the horizontal ordinate represents two topics, the vertical ordinate represents *MSE*. The "S" prefix means combining with sentiment intensity

Table 1. Improvement of prediction for each network when combining sentiment

S-CNN/CNN	S-LR/LR	S-DHCC-AR/DHCC-AR
21.39%	29.10%	38.71%

4 Conclusions

In this paper we propose a topic popularity prediction model, *SATP*. This model quantifies popularity using semantic information, making the quantification more accurate and explainable. We use the idea of Autoregressive for popularity prediction. To incorporate sentiment into prediction, we use *CNN* to learn sentiment and popularity's data dependence and history influence. We evaluate *SATP* on Twitter dataset by comparing mean square error (*MSE*).

Acknowledgement. This work was supported by the National Key R&D Program of China [2018YFB1004703]; the National Natural Science Foundation of China [61872238, 61672353]; the Shanghai Science and Technology Fund [17510740200]; the Huawei Innovation Research Program [HO2018085286]; the State Key Laboratory of Air Traffic Management System and Technology [SKLAT M20180X]; and the Tencent Social Ads Rhino-Bird Focused Research Program.

References

1. Gao, T., Bao, W., Li, J., Gao, X., Kong, B., Tang, Y., Chen, G., Li, X.: DancingLines: an analytical scheme to depict cross-platform event popularity. In: Hartmann, S., Ma, H., Hameurlain, A., Pernul, G., Wagner, R.R. (eds.) DEXA 2018. LNCS, vol. 11029, pp. 283–299. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98809-2_18
2. Wang, C., Jiang, F., Yang, H.: A hybrid framework for text modeling with convolutional rnn. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2061–2069 (2017)
3. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347–354 (2005)