6th International Conference on Computational Social Science IC2S2
July 17-20, 2020, MIT Media Laboratory Boston, MA USA

Using Text Mining to Examine the Role of Topics of Election Promises in Predicting Election Results

Xin Jin
xin.jin@cityu.edu.hk

Department of Media and Communication, City University of Hong Kong
Centre for Communication Research, City University of Hong Kong

Keywords: election promise, policy preference, election prediction, text mining, word2vec

Introduction

As one of the most popular topics within political communication research, elections have been paid a lot of attention to, especially the predictors of election result. Previous literature has proposed and adequately verified the effect of demographics (e.g. age, gender, and education), party effect (e.g. liberal and conservative), personal image (e.g. clothing and smile), political party funding, media strategy, and some other factors of candidates through the process of election. While, as an important tool for voters to evaluate candidates, form prospective beliefs about them, and make a voting decision (Born et al., 2018), the election promise made by candidates through their campaigns and discourses has received less scholarly attention. Generally, voters have their own policy preferences, and candidates who can implement the preference in their election promises will maximize the chances of winning the election (Jasim Alsamydai et al., 2013).

Meanwhile, online political information provider like voter guidance website has improved the openness and transparency in organizing and carrying out election, which can improve the engagement of public (Xenos and Moy, 2007) and make the relationship between the election promise and result more observable. And the online political information is the data source of this study.

To identify different policy focuses of candidates, one feasible way is to extract topics from election promises of each candidate and quantify them. Compared with traditional quantitative methods like content analysis, computational methods such as text mining could save both manpower and time and avoid the coder bias to a certain degree. In this study, the data of candidates of the 2018 local councilor election in Taiwan were used to answer the following questions:1. How can researchers extract and quantify topics computationally from election promises? 2. Do policy preferences exist in the local councilor election in Taiwan? If it exists, which topic of election promises could predict a successful election result?

Topic Extracting and Quantifying

The data (data set: http://dwz1.cc/T9H6PGU) were scraped from *Vote Taiwan* (https://votetaiwan.tw), a voter guidance website with a high popularity in Taiwan where election promises and other political information of each candidate ($N = 1753$) are publicly available. For the topic extracting, previously, supervised methods were commonly used in machine-based political text mining but found to require high-quality training sample and had lack of efficiency (Hillard et al., 2008). Unsupervised methods, such as LSA and LDA were also criticized as they focused exclusively on word co-occurrence without accounting for the word context. Here in this study, we employed word2vec (Mikolov et al., 2013), a context-

awareness method for distributed representation of words and phrases, and K-means clustering to extract topics from election promises (source code: http://dwz1.cc/t8zyllj). Skip-gram model of word2vec was used due to the capacity of accurate representation of infrequent words for a small training sample size. Finally, 33 clusters of words were extracted and identified as topics based on 6,665 distinctive vectorized words with 300 dimensions which are partly shown in Table 1.

To quantify the topic, weighted term frequency of each topic of one certain candidate was calculated according to the election promise. Thus, for each candidate, there would be 33 new variables as the "topic index" and are the form of percentage.

Number of words	*Cluster of words (English translation)*	*Topic*
64	公宅, 租屋, 樂業, 廣建, 居住公寓, 電梯, 住戶, 房屋, 修繕…… (public house, house renting, live in peace, build more, apartment, elevator…)	#housing
33	動物, 收容, 寵物, 疫苗, 毛小孩, 領養, 犬貓, 結紮…… (animal, animal shelter, pet, vaccine, fur-kid, adopt, dog and cat, ligate…)	#animal protection
108	幼教, 托育, 育兒, 托兒所, 幼稚園, 學前, 褓母…… (early edu, daycare, baby care, nursery, kindergarten, preschool, babysitter…)	#child care
59	農業, 行銷, 產銷, 生產, 精緻化, 農民, 農漁 …… (agriculture, sale, produce and sale, produce, delicate agriculture, farmer, fisherman…)	#agricultural product sales

Table 1. The clustering result of election promises

Topics of Election Promises and Policy Preferences

To compare elected (N = 912) and non-elected candidates (N = 841), logistic regression was employed, where independent variables were topic index, the party, demographics, and some other factors, while the dependent variable was the election result. Logistic regression revealed that, successful election results were positively associated with a certain group of topics of election promises, indicating that the policy preference indeed exists in the 2018 Taiwan local councilor election. Specifically, candidates who put more emphasis on #agricultural product sales, #education, #tourism, #smart city, #labor issue, and 17 other topics are more likely to be elected, in addition to previous councillorship (Table 2).

Predictor	*Coefficient*	*Standard error*
Topic index		
#finance	.25*	.13
#gender equality	.06	.12
#anti-drug and violence	.19	.12
#public security	.15	.10
#entrepreneurship	.13	.10
#tourism	.32**	.10
#taxation	0.16	.09
#disadvantaged group	.49**	.19
#smart city	.40**	.13
#animal protection	.24*	.10
#administrative transparency	.26**	.09
#internationalization	.02	.10
#serving the people	.62**	.20
#streamline administration	.75**	.25
#land planning	.20	.12
#technology industry	.34**	.11
#agricultural product sales	.54***	.16
#resource allocation	.54**	.18
#transportation	.20*	.09
#public facility	.39**	.13
#housing	.31*	.12
#elderly care	.24*	.11
#culture and art	.21*	.10
#environment protection	.28*	.12
#subsidy	.21	.13
#medical	.19*	.10
#legal advice	.22*	.09
#partisanship	-.23	.15
#education	.33***	.10
#national affairs	.22	.12
#child care	.24*	.11
#aborigines	.41	.22
#labor issue	.29**	.10
Personal features		
Age	-.52***	.08
Gender	-.07	.07
Highest education	.13	.07
Previous councillorship	1.15***	.09
Number of proposals	-.03	.09
Smile index	.12	.07
Community support	-.15*	.07
Parties		
KMT	.52	\
DPP	-.10	\
IP	-.23	\
Other	-.28	\

Missing values of the highest education and age were replaced by mode value and mean value respectively; Constant = .15, Pseudo R^2 = .29, -1 × Log Likelihood = 863.18; *p < .05; **p < .01; ***p < .001.

Table 2. The result of logistic regression

Discussion and Conclusion

Findings resonate with the literature (Yang & Chen, 2017). First, the topic distribution of one candidate can present his or her policy focuses, and if the policy focus is consistent with the policy preference of voters, the candidate is more likely to be elected. Second, based on the topic, what voters care about the most could be observed. Different from the presidential election of Taiwan whose candidates focus more on international and military topics, candidates of the local councilor election focus more on topics closely related to the daily life of local residents. Third, the "home style" of local election could be found. On one hand, candidates adopts different discourse strategies to address their election promises which are the reflection of their intention to appeal to target voters based on their knowledge of the local public opinion. On the other hand, the election promise information in this study is supposed to be accessed mainly through website, which might facilitate the selective exposure, memory and sharing procedures of voters, and thereby lead to voting decisions identical to their initial concerns on public issues. Finally, this study verified the applicability and the feasibility of the innovative idea for extracting topics from text, "word2vec and K-means clustering" (Guan, Zhang, & Zhu, 2016), and applied it to the Chinese natural language processing.

References

Born, A., van Eck, P., & Johannesson, M. (2018). An Experimental Investigation of Election Promises. *Political Psychology*, *39*(3), 685–705. https://doi.org/10.1111/pops.12429

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, *69*, 214–224. https://doi.org/10.1016/j.eswa.2016.10.043

Giraudy, M. E. (2015). *Conservative Popular Appeals: The Electoral Strategies of Latin America's Right Parties*. *UC Berkeley Electronic Theses and Dissertations*. University of California, Berkeley. Retrieved from https://escholarship.org/uc/item/95r0q1dd

Guan, L., Zhang, Y., & Zhu, J. (2016). Social Media Processing. In L. Yuming, G. Xiang, H. Lin, & M. Wang (Eds.), *Social Media Processing : 5th National Conference, SMP 2016* (pp. 85–97). Nanchang, China: Singapore : Springer Singapore : Imprint: Springer. https://doi.org/10.1007/978-981-10-6805-8

Hillard, D., Purpura, S., & Wilkerson, J. (2008). Computer-assisted topic classification for mixed-methods social science research. *Journal of Information Technology and Politics*, *4*(4), 31–46. https://doi.org/10.1080/19331680801975367

Jasim Alsamydai, M., & Hamdi Al Khasawneh, M. (2013). Basic Criteria for the Success of the Electoral Candidates and their Influence on Voters' Selection Decision. *Advances in Management & Applied Economics*, *3*(3), 1792–7552.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space, 1–12. Retrieved from http://arxiv.org/abs/1301.3781

Ryoo, J., & Bendle, N. (2017). Understanding the social media strategies of U.S. primary candidates. *Journal of Political Marketing*, *16*(3–4), 244–266. https://doi.org/10.1080/15377857.2017.1338207