

Overview of the 1st Classification of Spanish Election Tweets Task at IberEval 2017

Maitte Giménez¹, Tomás Baviera², Germán Llorca³, José Gámir³, Dafne Calvo⁴, Paolo Rosso¹, and Francisco Rangel^{1,5}

¹ Pattern Recognition and Human Language Technology (PRHLT) Research Center,
Universitat Politècnica de València

{mgimenez,proso}@dsic.upv.es

² Mediaflows Research Group, Valencian International University
tomas.baviera@campusviu.es

³ Mediaflows Research Group, Universitat de València
{german.llorca, jose.gamir}@uv.es

⁴ Mediaflows Research Group, Universidad de Valladolid
dafne.calvo@uva.es

⁵ Autoritas Consulting, S.A., Spain
francisco.rangel@autoritas.es

Abstract. This paper summarises the COSET shared task organised as part of the IberEval workshop. The aim of this task is to classify the topic discussed in a tweet into one of five topics related to the Spanish 2015 electoral cycle. A new dataset was curated for this task and hand-labelled by experts on the task. Moreover, the results of the 17 participants of the task and a review of their proposed systems are presented. In a second phase evaluation, we provided the participants with 15.8 millions tweets in order to test the scalability of their systems.

Keywords: Topic Classification, Twitter, Elections

1 Introduction

Nowadays, politics has upended by the usage of social media. A political campaign cannot be strategised using only the traditional media. During the election cycle, both politicians and voters engage in conversations about different topics. Politicians and their campaign staff share their policy approaches and bits of the candidates' personal lives. Characterising the influence processes in the public space is one of the most interesting topics in political communication research. Political parties, media and citizens send messages through a complicated media network, where knowing who has the power of agenda setting becomes critical. In this sense, the social media logic has boosted a more active user participation in delivering political messages, accessing more sources, and mobilising for political action. The analysis of this complex media network requires innovative research tools capable of evaluating the different elements in the political information flow [5].

33. De la Peña Sarracén, G.L.: Ensembles of methods for Tweet Topic Classification. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
34. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45(4), 427–437 (2009)
35. Spark-Jones, K.: Report on the need for and provision of an ‘ideal’ information retrieval test collection. *Computer Laboratory* (1975)
36. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. pp. 841–842. ACM (2010)
37. Sudhahar, S., Veltri, G.A., Cristianini, N.: Automated analysis of the us presidential elections using big data and network analysis. *Big Data & Society* 2(1) (2015)
38. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. *Computational linguistics* 37(2), 267–307 (2011)
39. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welppe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM* 10(1), 178–185 (2010)
40. Villar Lafuente, C., Garcés Díaz-Munío, G.: Several approaches for tweet topic classification in COSET - IberEval 2017. In: Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017). CEUR Workshop Proceedings. CEUR-WS.org, Murcia (Spain) (September 19 2017)
41. Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S.: A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In: Proceedings of the ACL 2012 System Demonstrations. pp. 115–120. Association for Computational Linguistics (2012)
42. Zirn, C., Glavaš, G., Nanni, F., Eichorts, J., Stuckenschmidt, H.: Classifying topics and detecting topic shifts in political manifestos. *PolText* (2016)