Political opinion mining in a Mexican Twitter corpus

Gemma Bel-Enguix, Ricardo Jiménez, Gerardo Sierra

Instituto de Ingeniería

Universidad Nacional Autónoma de México

ABSTRACT

Studying social media is a key research topic in the last years for several disciplines: complex systems, social science, sociology, linguistics, and artificial intelligence. Opinion mining, that involves the use of data mining and machine learning, is an arising area of natural language processing that aims at automatically analyze the feelings and opinions of people behind a text. This paper focuses on the task of performing opinion mining in Twitter. The corpus we worked on contains more than 12 million tweets, generated from October 2017 to June 2019 in Mexican territory and they are mostly in Spanish. Our goal is to implement an analytical system of public opinion which is able to monitor networks in real-time. Up to now, the main interest of our research is political opinion. The method is based on linguistic analysis of the tweets, identification of keywords and hashtags, and polarity classification. We show how we could be able to easily predict the results of the Mexican Presidential Elections of 2018 using only the parameters above mentioned. We plan to extend the methodology and test it in other elections with a twofold objective: being a constant reflect of the flux of political opinion and testing if Twitter has predictive capabilities, this is, if the opinions in social media are a good projection of the society.

Keywords: opinion mining; social media; natural language processing; political opinion; Twitter