

Arabic Lexicon Learning to Analyze Sentiment in Microblogs

Authors : Mahmoud B. Rokaya

Abstract : The study of opinion mining and sentiment analysis includes analysis of opinions, sentiments, evaluations, attitudes, and emotions. The rapid growth of social media, social networks, reviews, forum discussions, microblogs, and Twitter, leads to a parallel growth in the field of sentiment analysis. The field of sentiment analysis tries to develop effective tools to make it possible to capture the trends of people. There are two approaches in the field, lexicon-based and corpus-based methods. A lexicon-based method uses a sentiment lexicon which includes sentiment words and phrases with assigned numeric scores. These scores reveal if sentiment phrases are positive or negative, their intensity, and/or their emotional orientations. Creation of manual lexicons is hard. This brings the need for adaptive automated methods for generating a lexicon. The proposed method generates dynamic lexicons based on the corpus and then classifies text using these lexicons. In the proposed method, different approaches are combined to generate lexicons from text. The proposed method classifies the tweets into 5 classes instead of +ve or -ve classes. The sentiment classification problem is written as an optimization problem, finding optimum sentiment lexicons are the goal of the optimization process. The solution was produced based on mathematical programming approaches to find the best lexicon to classify texts. A genetic algorithm was written to find the optimal lexicon. Then, extraction of a meta-level feature was done based on the optimal lexicon. The experiments were conducted on several datasets. Results, in terms of accuracy, recall and F measure, outperformed the state-of-the-art methods proposed in the literature in some of the datasets. A better understanding of the Arabic language and culture of Arab Twitter users and sentiment orientation of words in different contexts can be achieved based on the sentiment lexicons proposed by the algorithm.

Keywords : social media, Twitter sentiment, sentiment analysis, lexicon, genetic algorithm, evolutionary computation

Conference Title : ICCICN 2019 : International Conference on Computational Intelligence and Communication Networks

Conference Location : Dubai, United Arab Emirates

Conference Dates : January 30-31, 2019