Polarity Classification of Arabic Sentiments

Mohammed N. Al-Kabi, Information Technology Faculty, Zarqa University, Zarqa, Jordan
Heider A. Wahsheh, College of Computer Science, King Khaled University, Abha, Saudi Arabia
Izzat M. Alsmadi, University of New Haven, West Haven, CT, USA

ABSTRACT

Sentiment Analysis/Opinion Mining is associated with social media and usually aims to automatically identify the polarities of different points of views of the users of the social media about different aspects of life. The polarity of a sentiment reflects the point view of its author about a certain issue. This study aims to present a new method to identify the polarity of Arabic reviews and comments whether they are written in Modern Standard Arabic (MSA), or one of the Arabic Dialects, and/or include Emoticons. The proposed method is called Detection of Arabic Sentiment Analysis Polarity (DASAP). A modest dataset of Arabic comments, posts, and reviews is collected from Online social network websites (i.e. Facebook, Blogs, YouTube, and Twitter). This dataset is used to evaluate the effectiveness of the proposed method (DASAP). Receiver Operating Characteristic (ROC) prediction quality measurements are used to evaluate the effectiveness of DASAP based on the collected dataset.

KEYWORDS

Arabic Multi-Media Classification, Arabic Text Analysis, Arabic Text Classification, Big Data, Document-Level, Opinion Mining, Sentiment Analysis, Social Network

1. INTRODUCTION

Millions of Arab Internet users enjoy the use of Online Social Networks (OSNs). Arab users of social media generate millions of interactions and a huge amount of opinionated data about different topics each day. This wealth of opinionated Arabic data includes beside the text comments, images, audios, and videos that are posted through user accounts to display their sentiments and opinions about a wide range of topics.

Sentiment analysis (SA) is used to automatically analyse the opinionated data about different aspects of life like products, sportsmen, actors, actresses, politicians, governments, etc. Opinionated data includes reviews and comments posted by social media users. Nowadays, many companies analyse these posts to enhance their services and products. Therefore, there is no need to conduct surveys and opinion polls anymore by those parties. SA is an essential tool that used to make the right decision by companies, political campaigns, governments, politicians, candidates, economists, movie stars, sportsmen, etc. In business, for example, customers usually read posts of the consumers of a certain product before deciding to buy it, and the point of views of different owners highly affect their decision to buy a product or to look for different alternatives.

SA is not a straightforward task since the polarity of a review is not always fixed, but it depends on the domain it belongs to. Let us consider a simple example for the word “cheap” that is considered in economics domain as a positive word, while in the politics domain, it is considered to be negative. SA is highly dependent on the context and the domain. Furthermore, the authors of these posts (reviews) use slang, emoticons, lengthened words (Words with repeated letters), and their posts are
characterized by a high percentage of spelling mistakes relative to those found in newspaper articles due to the nature of the devices used to generate these posts and their authors. Previous experiences of specialists in this field reveal that human understandings of comments and reviews are different, and, therefore, different conclusions are drawn by different persons. Therefore, it is harder for software or website tools to identify precisely the polarities of these comments and reviews.

Most of the studies of SA are about English reviews and comments, and there are relatively few SA studies in other languages such as French and Arabic (Ghorbel, & Jacot, 2011). Arabic sentiment analysis is the core topic of this study, and it is used to identify the polarity of the collected Arabic text sentiments. The scope of this study does include image, audio, and video posts generated by social media users, besides an analysis to Arabic text posts. The task of determining the polarity of video posts is much harder than text posts. Therefore, to enhance the accuracy of video posts annotations, for example, the body language can be included. Supervised classification of the collected Arabic dataset is presented in this study. Our paper proposes a new method called Detection of Arabic Sentiment Analysis Polarity (DASAP) using a modest dataset of Arabic comments, posts, and reviews of different types: text, images, videos, and audios. This dataset is collected from social network websites (i.e. Facebook, Blogs, YouTube and Twitter). Receiver Operating Characteristic (ROC) prediction quality measurements are used to evaluate the effectiveness of the proposed approach.

The rest of this paper is organized as it follows: Section 2 presents the related work to this study, and section 3 presents the methodology followed to accomplish this study. Section 4 presents the data collection. Section 5 shows text and images lexicons while section 6 presents opinion extraction methods. Section 7 shows the experiments and results. Section 8 presents the conclusion and future work.

2. RELATED WORK

This section introduces studies that are closely related to our study. A review of studies of Arabic sentiment analysis is presented in this section, with a discussion of their differences from our study. A study that is closely related to our study is conducted by (Morency, Mihalcea, & Doshi, 2011). It is an interesting study since it combines multi-modal to conduct sentiment analysis. Visual, audio, and text features are used together to identify the polarity of sentiments under consideration. They adopt statistical analysis to identify the most effective multimodal features that help to identify the polarities of different video opinions they collect from YouTube. Statistical analysis conducted by (Morency, Mihalcea, & Doshi, 2011) indicates that polarized words, smiles, gazes, pauses, and voice pitches are the most influential features to identify the polarities of the video opinions. Furthermore, they collected from YouTube a dataset of video opinions that are annotated to be used by researchers. Those authors conclude that joint use of multiple modalities helps to improve the effectiveness of classifiers used to identify the polarities of video opinions.

One of the earliest studies to identify the subjectivity in Arabic is conducted by (Abdul-Mageed, & Korayem, 2010), and it is based on a portion of Penn Arabic Treebank that consists of 1552 Arabic sentences that are distributed among 200 documents. The two authors annotated manually each of the 1552 Arabic sentences by labelling each one as either subjective or objective, where 815 Arabic sentences were labelled as objective and the other 737 Arabic sentences were labelled as subjective. They attempt in their study to discover the challenges that morphologically rich languages (MRLs) pose to text classification and identifying subjectivity. The tokenization and determining part of
Related Content

A Reference Ontology Based Approach for Service Oriented Semantic Interoperability
www.igi-global.com/chapter/reference-ontology-based-approach-service/72950?camid=4v1a

Relationships among Information Technology, Service Quality, and Overall Satisfaction of the Customers in Life Insurance Corporation of India
www.igi-global.com/chapter/relationships-among-information-technology-service-quality-and-overall-satisfaction-of-the-customers-in-life-insurance-corporation-of-india/140860?camid=4v1a
A Layered Communication Architecture for Power Routing in the Smart Grid
[www.igi-global.com/article/a-layered-communication-architecture-for-power-routing-in-the-smart-grid/113321?camid=4v1a](www.igi-global.com/article/a-layered-communication-architecture-for-power-routing-in-the-smart-grid/113321?camid=4v1a)

Web Technology Systems Integration Using SOA and Web Services
[www.igi-global.com/chapter/web-technology-systems-integration-using-soa-and-web-services/203415?camid=4v1a](www.igi-global.com/chapter/web-technology-systems-integration-using-soa-and-web-services/203415?camid=4v1a)