



MINING OPINION FEATURES IN USERS' REVIEWS

by

MOSTAFA ALI MAHMOUD MOHAMMED

B.Sc Information System Assiut University

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

MASTER OF COMPUTERS AND INFORMATION

Department of Information System

Faculty of Computers and Information

UNIVERSITY OF ASSIUT

ASSIUT, EGYPT.

2013

Supervised by:

Prof. Magdy Mofid Doss

Dr. Taysir Hassan Abdel Hamid

Dr. Abdelrahman Heydar

Examined by:

Prof. Ahmad Sharaf

Prof. Hosny Mohammed Ibrahim

Prof. Magdy Mofid Doss

Dr. Taysir Hassan Abdel Hamid

Abstract

Textual information in the world can be classified into two main categories, facts and opinions. Opinions (Words of mouth) are subjective statements that reflect people's sentiments or perceptions about the entities and events. With the Web, especially with the explosive growth of the user generated content on the Web, the world has changed. Opinion mining, which is called sentiment analysis, can be viewed as a classification process that aims to determine whether a certain document or text is written to express a positive or a negative opinion about a certain object (e.g., a topic, product, or person). One can post reviews of products at E-commerce sites and make comments on almost anything in Internet forums, discussion groups, and blogs, which are collectively called the user generated content. Now, if one wants to buy a product, it is no longer necessary to ask one's friends and families because there are plentiful of product reviews on the Web which give the opinions of the existing users of the product.

As e-commerce is increasingly becoming popular, the number of customer reviews that a product receives grows rapidly. However, for popular products, many online product reviews exist but for other reviews product reviews are very few. These online discussions about particular products may help other online users to make a decision in buying/ not buying those products, like in [amazon.com](http://www.amazon.com)¹ and [ebay.com](http://www.ebay.com)². Since an enormous number of unstructured and ungrammatical reviews on a product exist, opinion mining is getting a crucial research area for better decision making of buying products. However, there is limited interest from Arab customers to leave these reviews. Arab people are interested in using social network as a communication way especially in their revolutions since the Tunisian revolution started in December 2010. Social networks have affected the way the new generation think all over the world, specifically in the Middle East. Their effectiveness appears in the revolutions of Tunisia, Egypt, and Syria. Therefore, social networks opinion mining for Arabic slang language has become essential since it is widely used between the youth generation. Arabic slang language suffers from two main problems, which are the new expressive (opinion) words and

¹ <http://www.amazon.com>

² <http://www.ebay.com>

idioms as well as the unstructured format. Mining Arabic slang language requires efficient techniques to extract youth opinions on various issues, such as news websites.

In this thesis, we explain two opinion mining approaches to mine different kinds of reviews. First, is to summarize the unstructured and ungrammatical users' reviews, based on Support Vector Machine (SVM). Two levels of classification are applied: 1) Features classification and 2) Polarity classification for every feature class. Because the customers' reviews are unstructured and ungrammatical (free text), the average precision is 93.15 and the average recall is 92.41. The second is applied on Arabic comments on social network websites and news portals; we propose a SVM-based classifier for Arabic slang language, applying sentiment analysis, to classify youth news' comments on Facebook. This classifier consists of three main phases: 1) Arabic comments data preparation, 2) Data preprocessing, and 3) data classification. In addition, a Slang Sentimental Words and Idioms Lexicon (SSWIL) of opinion words is built. The lexicon is used by Arab youth in their comments on news topics, Facebook³ posts and comments, tweets in Twitter⁴. SSWIL enhances the classification task to be 86.86% of classified comments instead of 75.35% when using classical opinion words lexicon with precision 88.63 and recall 78 instead of 82.4 and 59.33 respectively.

³ www.Facebook.com

⁴ www.Twitter.com