Chapter 21
Opinion Mining of Twitter Events using Supervised Learning

Nida Hakak
Maharshi Dayanand University, Haryana, India

Mahira Kirmani
Maharshi Dayanand University, Haryana, India

ABSTRACT

Micro-blogs are a powerful tool to express an opinion. Twitter is one of the fastest growing micro-blogs and has more than 900 million users. Twitter is a rich source of opinion as users share their daily experience of life and respond to specific events using tweets on twitter. In this article, an automatic opinion classifier capable of automatically classifying tweets into different opinions expressed by them is developed. Also, a manually annotated corpus for opinion mining to be used by supervised learning algorithms is designed. An opinion classifier uses semantic, lexical, domain dependent, and context features for classification. Results obtained confirm competitive performance and the robustness of the system. Classifier accuracy is more than 75.05%, which is higher than the baseline accuracy.

1. INTRODUCTION

Opinion mining is used to refer to the task of automatically determining the opinion expressed in text, phrases, sentences or any piece of writing. However, more generally it is used to determine one's attitude towards a particular event or reaction to an event. Here, attitude means qualitative opinion, feeling or reaction to some situation that triggered the event. Opinion measurement and sentiment analysis for Twitter data are attracting much research both in academia and industry. Millions of users of Twitter are posting tweets about their daily life, write opinions on a variety of issues. Twitter is easily accessible through smartphones and other web services and thus is a preferred media for communication. Hence internet users are shifting from traditional communication tools like blogs and mailing lists to twitter.

DOI: 10.4018/978-1-6684-6303-1.ch021

More specifically, as Twitter users are expressing their opinions about several issues including religious matters, political views, and reviews of e-services, products or even movie reviews twitter has become a viable source for opinion measurement and detection. Such data is extremely potent for the industry for feedback and political parties to frame their policies and strategies.

Dey, Babo, Ashour, Bhatnagar & Bouhlel (2018) presented a detailed implementation of strategies and challenges to social network intelligence in their study. Twitter posts are 240 characters long messages called tweets. Java, Song, Finin & Tseng (2007) in their study suggest that users use Twitter for following reasons (1) for information source (2) to be in touch with family and friends and (3) seeking information about trends happing worldwide. Opinion mining involves data mining and Natural Language Processing (NLP) techniques to uncover hidden information and opinions from social web's substantial textual sources. Opinion mining from the text written in natural languages is challenging as it requires a deep understanding of explicit and implicit, regular and irregular, syntactic and semantic rules of language. Therefore, opinion mining is a challenge for NLP researchers for taking utilizing tools of NLP for efficient and effective opinion mining systems and thus leading to substantial practical impact.

Most of the companies are using opinion mining systems to create and automatically maintain reviews written by their customers for their popular products. Opinion mining can also be used by companies to improve customer relationship. Opinion mining also finds its applications in the recommender systems used by e-commerce sites. Moreover, opinion mining can play an essential part in the anti-spam policy drafting. Political parties also use opinion mining to know if the people support their decisions and programs or not and use feedback to frame their policies and strategies for future.

Opinion mining involves tracking users perception towards the brand or an event so as to capture mood of the public towards the brand or some political movement. Automatic opinion mining uses the NLP tools and machine learning techniques to effectively extract sentiments in text. Several studies exist in literature which have used machine learning to automatically detect the opinion like (Mohammad, Zhu, Kiritchenko & Martin, 2015; Yan, Turtle & Liddy, 2016; Gore, Diallo & Padilla, 2015). Pak & Paroubek (2010) used multinomial Naıve Bayes classifier with linguistic features to perform opinion analysis of collected twitter corpus. Grigori Sidorov (Sidorov, Miranda-Jiménez, Viveros-Jiménez, Gelbukh, Castro-Sánchez, Velásquez, & Gordon, 2012) used several machine learning algorithms for automatic detection of opinions in a Spanish language Tweet corpus.

In this paper, we discuss opinion mining of Twitter event done used state-of-art NLP techniques and machine learning tools. We show how to use Twitter for the corpus of opinion detection system. We have collected a corpus of 4,928,436 tweets about event namely Kashmir Unrest 2016 downloaded from the twitter from 10, July 2016 to 31, December 2016. We then preprocess tweets cleaning noise and performing other text transformations. Then the state-of-art linguistic analysis is performed using NLP techniques and then built an opinion classifier using supervised learning algorithms. We have also developed a rich opinion corpus using crowd-sourcing.

1.1. Contribution

The contributions of our research are:

1. We present a method for efficient preprocessing of text for removing and replacing slangs and to correct misspelled words.
2. We have built a vibrant opinion mining corpus for supervised learning algorithms.