Chapter 10

Sarcasm Detection in Twitter Data: A Supervised Approach

Santosh Kumar Bharti
Pandit Deendayal Petroleum University, India

Sathya Babu Korra
National Institute of Technology Rourkela, India

ABSTRACT

Posting sarcastic messages on social media like Twitter, Facebook, WhatsApp, etc., became a new trend to avoid direct negativity. Detecting this indirect negativity in the social media text has become an important task as they influence every business organization. In the presence of sarcasm, detection of actual sentiment on these texts has become the most challenging task. An automated system is required that will be capable of identifying actual sentiment of a given text in the presence of sarcasm. In this chapter, we proposed an automated system for sarcasm detection in social media text using six algorithms that are capable to analyze the various types of sarcasm occurs in Twitter data. These algorithms use lexical, pragmatic, hyperbolic and contextual features of text to identify sarcasm. In the contextual feature, we mainly focus on situation, topical, temporal, and historical context of the text. The experimental results of proposed approach were compared with state-of-the-art techniques.

INTRODUCTION

Sentiment analysis is a procedure to extract the attitude of a speaker or a writer concerning some target (Pang, 2002, PP. 79-86). Social media and social networking have fueled the online space, including Facebook, Amazon, Twitter, etc. as ratings, reviews, comments, etc. are everywhere. Social media content is growing rapidly with every passing day. With the large volumes of information generating daily, identifying clear and most consumer reliable information about their preferences has become very tough task nowadays. To get the correct and trusted information, individuals are showing interest towards analysis of social networking content. For every business, online reviews have become the

DOI: 10.4018/978-1-5225-7186-5.ch010

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
deciding factor which can break or make a product in the marketplace. Sentiment analyzer is used as a tool to understand how consumers are reacting to an event or a new product through auditing social networking posts and comments.

Sarcasm is derived from the French word “Sarcasm” that means “tear flesh” or “grind the teeth”. In simple words, sarcasm is a way to speak bitterly. Macmillan Dictionary defines, “…sarcasm is an activity of saying or writing the opposite of what you mean, or of speaking in a way intended to make someone else feel stupid or show them that you are angry…” (Macmillan, 2007). For example: “it feels great being bored”. In this example, the literal meaning of the sentence is different than what the speaker intends to say using sarcasm. While writing sarcasm in text, people often use only positive words to convey a negative opinion instead of real negative. However, the average human reader face problem in sarcasm detection for social media content such as tweets, reviews, blogs, online forums, etc.

In this paper, the tweets of Twitter are used as the dataset for sarcasm detection. Twitter is a microblogging social networking site where a user can read and post the messages. The Twitter allow posting a message of a limited length of 140 characters. Due to the limitations, the users often use symbolic and figurative texts to express their feelings such as @username, smilies, emoji, exclamation mark and interjection words. Recognition of these symbolic and figurative texts in tweets is the most tedious task in NLP. In the realm of Twitter, the author has observed several types of sarcastic tweets as shown in Table 1 that occurred frequently.

The author has already discussed sarcasm type $T_1$, $T_2$ and $T_3$ in their previous article (Bharti, 2015). The remaining types of sarcastic tweets are as follows:

- **Type $T_4$** depicts the users’ likes and dislikes behavior while posting the tweets on Twitter. To learn the behavioral habit of a particular Twitter user, one can analyze and observe his likes and dislikes habit. Using these likes and dislikes lists of a particular user; one can analyze the sarcastic tweets from the particular user’s account. For example, if any Sachin Tendulkar fan who likes to post the tweet about Sachin and his Twitter account consist a tweet like “I love to see Sachin’s failure in batting” which contradict his like’s habit so, one can be easily identify that given tweet is sarcastic.
- **Type $T_5$** says about tweets contradicting universal facts. For example, if any user has posted a tweet “sun is revolving around the earth” and the fact is earth revolves around the sun. In such cases, user is intentionally negating the universal facts in their tweet, and there is a high probability that the tweet is sarcastic.
- **Type $T_6$** says about tweets that contradict temporal facts, which may change over a period. For example, “India had an amazing win in cricket world cup 2015 final”, while the fact is Australia

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>Sarcasm as a contradiction between positive sentiment and negative situation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_2$</td>
<td>Sarcasm as a contradiction between negative sentiment and positive situation.</td>
</tr>
<tr>
<td>$T_3$</td>
<td>Tweets that starts with interjection word.</td>
</tr>
<tr>
<td>$T_4$</td>
<td>Sarcasm as a contradiction between likes and dislikes.</td>
</tr>
<tr>
<td>$T_5$</td>
<td>Sarcasm as a contradiction between tweet and the universal facts.</td>
</tr>
<tr>
<td>$T_6$</td>
<td>Sarcasm as a contradiction between tweet and its temporal facts.</td>
</tr>
<tr>
<td>$T_7$</td>
<td>Positive tweet that contains a word and its antonym pair.</td>
</tr>
</tbody>
</table>