‘How Happy Are We?’

Analysing Sentiment Within Twitter Data

Dissertation

James Michael

May 2017

Swansea University
Prifysgol Abertawe

Department of Computer Science

Swansea University

1 Introduction

Social media platforms such as Twitter have fast become popular over recent years. As of June 30, 2016, Twitter reports an approximate of "313 million monthly active users", with "79% of accounts outside the U.S." and "82% active users on mobile" [1]. Liu et al. observe the "conversational" nature of Twitter, with "users mentioning other users in over 50% of tweets today" [2]. A notion shared by Perreault and Ruths, who suggest that using Twitter from a mobile device affects "the way users are generating and consuming microblogging content" [3]. More personal content from mobile devices is perhaps because tweets are being formed in a similar way to text messages, and in part due to the length restrictions on tweets.

This project seeks to answer three questions. How can we determine the happiness of our friends? How can we determine it for people we have a general interest in? How can it be done in the modern world?

Many existing projects have looked at analysing sentiment within tweets, and several are presented in the Related Work section. However, many of these projects gather tweets and data from search queries, with little focus on accounts or people who the user personally finds interesting.

The result is a mobile application which aims to address this issue. It analyses sentiment within Twitter data - however, contrary to many existing works, the application analyses only those accounts followed by the user. This can consist of friends or people the user has a general interest in, meaning the project provides a much more personal experience.

The application allows for identifying the happiness of a friend on a given day – allowing the user to behave accordingly in person. An example scenario could be if a decline in the happiness of a friend is identified. This can be verified by manually checking the detected negative terms across all collected tweets for a user. If the decline is verified, it might prompt the user to offer support or help their friend. Likewise, if an increase in happiness is identified, the user knows their friend has been posting more positive content.

Each user who is analysed by the application is assigned to one of five groups, which ultimately make up a sentiment scale ranging from Very Negative to Very Positive. This first requires each user be given an overall happiness score, which is computed by implementing an equation discussed in Section 2.7, 'The Happiness Equation'. These happiness scores are computed relative to other users on a day, and as such a tweet is only deemed positive if it is positive within the context of all other users on the day.

The application's interface consists of five primary windows, presented in Section 4.10, 'User Interface'. These allow the user to view and select a group, identify the users within a group, browse analysed tweets posted by a user, and identify sentiment terms found within a given tweet.

The project's aims are:

1. Quantify positivity and negativity of all Tweets collected.
2. Quantify positivity and negativity of a given Tweet.
3. Assign users to one of five groups on a sentiment scale.
4. Visualise these groups and the collection of users and tweets within.
5. Real-time functionality.
6. Evaluate the user interface.

Only one of these aims has not been met, real-time functionality. All others have been implemented or completed.