

Investigating User Ridership Sentiments for Bike Sharing Programs

Subasish Das¹, Xiaoduan Sun¹, Anandi Dutta²

¹Civil Engineering Department, University of Louisiana at Lafayette, Louisiana, USA

²Center for Advanced Computer Studies, University of Louisiana at Lafayette, Louisiana, USA

Email: sxd1684@louisiana.edu, xsun@louisiana.edu, axd1329@louisiana.edu

Received 29 January 2015; accepted 27 March 2015; published 31 March 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Bike sharing is considered a state-of-the-art transportation program. It is ideal for short or medium trips providing riders the ability to pick up a bike at any self-serve bike station and return it to any bike station located within the system's coverage area. The bike sharing programs in the United States are still very young compared to those in European countries. Washington DC was the first jurisdiction to devise a third generation bike sharing system in the US in 2008. To evaluate the popularity of a bike sharing program, a sentiment analysis of the riders' feedback can be performed. Twitter is a great platform to understand people's views instantly. Social media mining is, thus, gaining popularity in many research areas including transportation. Social media mining has two major advantages over conventional attitudinal survey methods—it can easily reach a large audience and it can reflect the true behavior of participants because of the anonymity social media provides. It is known that self-imposed censor is common in responding to conversational attitudinal surveys. This study performed text mining on the tweets related to a case study (Capital Bike share of Washington DC) to perform sentiment analysis or opinion mining. The results of the text mining mostly revealed higher positive sentiments towards the current system.

Keywords

Bike Sharing, Social Media, Twitter Mining, Text Analytic, Sentiment Analysis, Opinion Mining

1. Introduction

Public bicycle programs, also known as bike sharing programs, can be described as a short-term bicycle rental facility for inner-city transportation that provides bikes at self-serve stations. In recent years, these programs have rapidly emerged in major cities all around the world. However, the bike sharing programs in the United States are still very young compared to those in European countries. Of the programs in the US, Washington DC

was the first to implement a third generation bike sharing system in 2008. The system was called SmartBike and was replaced in 2010 by “Capital Bikeshare” [1].

Twitter is one of the most popular micro-blogging platforms in recent years. The users can share their instantaneous thoughts or information on a wide range of topics or interests through short messages known as “tweets”. Sentiment analysis (also known as opinion mining) aims to automatically extract the expressed opinions to understand the sentiments towards a system and develop new perceptions. Sentiment analysis of tweets can expand knowledge and improve intelligence to provide better services and develop inventive prospects for bike sharing programs [2]. This study aims to perform sentiment analysis on the tweets related to Capital Bikeshare (by mentioning or hashtagging) to conduct a sentiment analysis on the current program’s users. The findings of the study will help authorities take strategic actions.

2. Literature Review

A surge of interest in effective computational methods has been seen in recent years, ranging from opinion mining to subjectivity detection to sentiment analysis. These methods usually focus on the identification of public opinions, feelings, sentiments, assessments, beliefs, and conjectures in natural language. While available literature on the use of sentiment analysis has expanded in recent years, there is still a lack of study on transit development through bike sharing programs.

A global view of bike sharing characteristics by analyzing data from 38 systems located in Europe, the Middle East, Asia, Australia and the US was conducted by O’Brien *et al.* [3]. Barbosa and Feng classified the subjectivity of tweets on traditional features with the inclusion of some Twitter specific clues such as retweets, hashtags, links, emoticons, and question marks [4]. For sentiment classification of subjective tweets, the same set of features was also used. The sentiment analysis of tweets was studied by Davidov *et al.* [5]. The authors took a supervised learning approach. Apart from the traditional features, the method also used hashtags, smileys, punctuations, and their frequent arrangements. These features were shown to be quite effective. The work of Gonzalez-Ibanez studied the problem in the context of sentiment analysis using Twitter data to distinguish between sarcastic tweets and non-sarcastic tweets that directly convey positive or negative opinions (indifferent or informative statements were not considered) [6]. Wang proposed a graph-based hashtag approach to classifying tweet sentiments [7]. Sentiment analysis via Twitter mining was introduced in transportation research by Collins *et al.* [8]. The study attempted to devise sentiments of the transit riders by measuring Twitter feeds.

Sentiment analysis of the tweets is considered rather challenging due to the size of the text. The goal of this study is to classify emerging societal trends based on opinions, natures, attitudes, approaches and views of the users towards bike sharing programs. The contributions of this paper will help the urban transitpolicy makers better anticipate potential impacts of policy measures and better communicate expected profits and concerns.

3. Methodology

3.1. Bike Sharing Riders

The information of the riders was collected from the trip history data of Capital Bike share which is available for public access [9]. When a rental occurs within the system, the software used by Capital Bike share collects basic data about the trip. All private data including member names has been removed from these files. The data from each year is divided into four quarterly datasets. Each file contains 7 columns: duration of trip, start date and time, end date and time, starting station name and number, ending station name and number, identification number of bike used and member type. Members are divided into two groups: subscriber (annual or monthly) and casual (1 to 3/5 days member).

Figure 1 depicts the transaction details of the bike sharing program starting from the 4th quarter of 2010. The yearly highest peak of transactions occurs in the 2nd (April-June) and 3rd (July-September) quarters for both subscriber riders and casual riders. This happens because the average temperature of these two quarters is higher than the other two quarters in a year. Comparing the usage between the 3rd quarter of 2012 and 2013, the usage increased by 28% for subscriber riders and 56% for casual riders in 2013.

3.2. Twitter Mining

Twitter is the most notable social media tool for microblogging service. The user posts, tweets, do not exceed

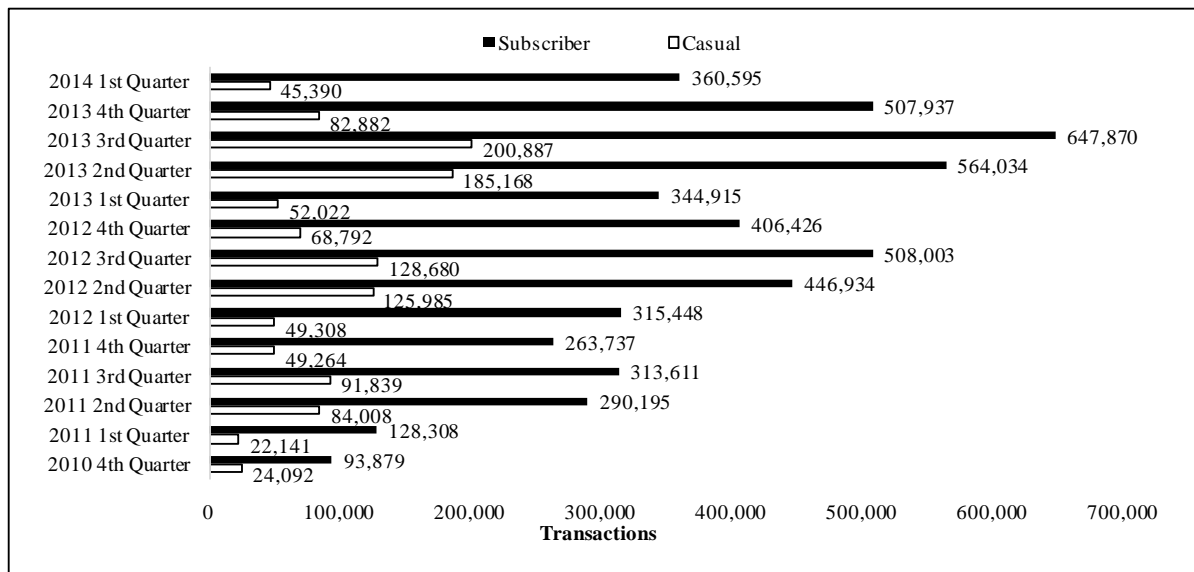


Figure 1. Capital Bikeshare user transactions per quarter.

140 characters and no privacy conditions can be imposed. The reflection of opinions and information is performed in real-time. Twitter generates a large chunk of textual content daily. The textual content can be explored by means of text mining, natural language processing, information retrieval, and other methods. With 107.7 million accounts created before January 1st, 2012, the US now merely represents 28.1% of all Twitter users. In December of 2011 alone, about 5.6 million new accounts were created within the US. Based on the data from March of 2012, Twitter has 140 million active users. One can argue whether Twitter stratifies the necessarily representative sample data of the outside world. However, a contextualization of the social media data with the appropriate mechanism may provide important insights. The keys to successful Twitter mining depend on several factors, such as data exploration technique, appropriate algorithm, target specification, and responsiveness of the post. The Twitter terms are defined here in brief for easy interpretation:

Tweet: A short message or post from an account holder on Twitter. The account holder's identity is known as the Twitter handle. The text spans a maximum of 140 characters.

Retweet: Retweet forwards a tweet from users to their followers which is almost similar to e-mail forwarding.

Hashtag: It is denoted by a word with a preceding “#” symbol (e.g., #bikeshare). It is generally used before a relevant keyword or phrase, with no spaces, in tweets to categorize those tweets and help them show more easily in a Twitter Search.

Mention: Mention acknowledges a user with the symbolic “@” sign without using the “reply” feature.

The Twitter handle of Capital Bikeshare is “*bikeshare*”. The tweets from the Capital Bikeshare timeline were collected for nine months (October 2013-June 2014). The Twitter handle was created in July of 2010. Nearly 6200 official tweets were made from this account between July of 2010 and February of 2015. The account has twelve thousand followers. It is important to note that the one-time tweet extraction limit from a Twitter handle is limited to 3200. Popular data mining “R” packages “twitter” and “tm” were used in this study to extract tweets for analysis [10] [11]. Twitter currently implements two forms of authentication in the new model, both still leveraging open standard for authorization (OAuth). These two forms are: 1) Application-user authentication which is the most common form of resource authentication in Twitter's OAuth 1.0A implementation to date. 2) Application-only which is a form of authentication where user application makes Application Programming Interface (API) requests on its own behalf, without a user context [12]. The collected tweets (count = 591), by using newly implemented Twitter OAuth, are distributed by hour of the day in **Figure 2**. Most of the tweets were made from 11 AM to 11 PM.

3.3. Sentiment Analysis

Sentiments are central to almost all human activities and are key influencers of our behaviors. Most of the hu-

man beliefs and perceptions of humankind is based on how others see and evaluate the world. For this reason, people often seek out the sentiments of others in order to make a better decision. This is not only true for individuals but also true for various programs and organizations. Opinions and related concepts such as sentiments, evaluations, attitudes, and emotions are the subjects of study of sentiment analysis. **Figure 3** illustrates the flowchart of the sentiment analysis procedure conducted in this study.

It is important to note that the sentiment lexicons have domain-specific sentiment values, the sentiment classification performance of a given text may vary according to the calculation process of the sentiment for that text. Various sentiment lexicons with different format and research focus have been developed to aid the classification of positive and negative annotations in the mining-ready texts. Both similarity and diversification are noticed while comparing the listed words and their ratings. Constructing a domain-specific sentiment lexicon is essential to tackle the classification problem of sentiment analysis. The researchers of this study are currently developing a sentiment lexicon appropriate for transportation related tweets. This work remains a prospective topic for future research. A list of positive and negative sentiment words in English (around 6800 words) was used to perform the sentiment analysis on the tweets. This list was compiled by Hu and Liu in 2004 [13]. After comparing with different sentiment lexicons, this list has been considered as the most relevant lexicon.

It will be interesting to mine the Twitter data related to “@bikeshare” and “#bikeshare” to understand the sentiment of the bike users. A function named “score. sentiment”, introduced by Breen, was used to produce the score count of each tweet [14]. This function will mine each tweet by using the positive and negative word lexicons and produce a positive, negative or zero score. A tweet with a “+2” score means that this particular tweet has two positive words by mentioning or hashtagging “bikeshare”. A tweet with a negative score indicates the negative words used in a particular tweet. **Table 1** lists eighteen tweets associated with bike share. A general inspection on the first tweet associates three words representing positive sentiment: open, available and veracity.

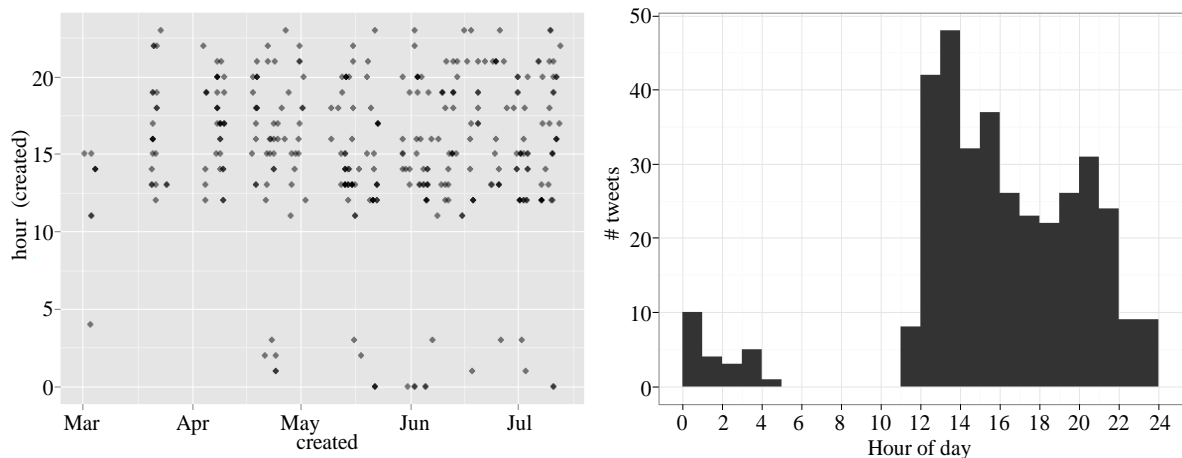


Figure 2. Distribution of the tweets by hour of day.

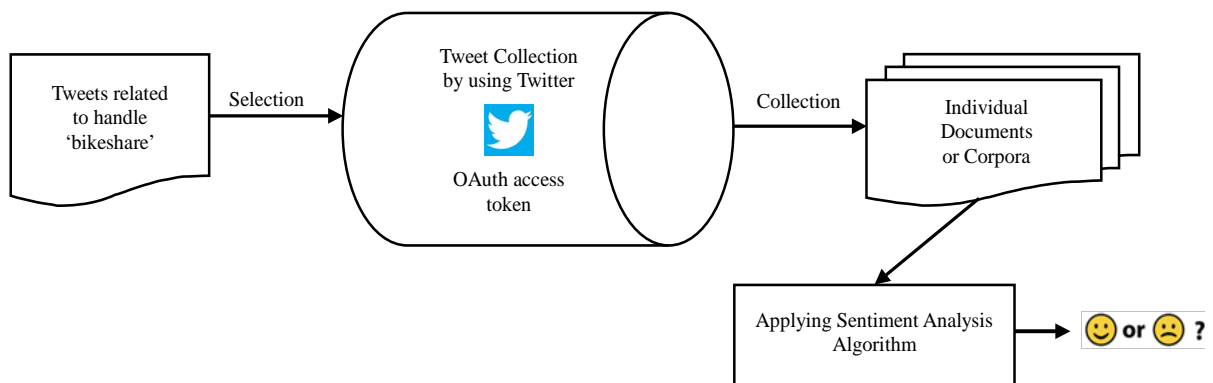


Figure 3. Flowchart of sentiment analysis.

Table 1. Sentiment score of the tweets.

Score	Tweets
1	“Carole jolly is next speaker. #bikeshare”
-1	“#Local group worried about locations of #bikeshare stations ”
-1	“Fixing problems we’ve created for ourselves #BikeYVR #HelmetChoice #BikeShare”
2	“Nice work! RT @poochica: Ha! I beat @jwetz to 18th and RIA. @bikeshare”
2	“I love riding a #bikeshare in #dc and I love #bikeshare”
2	“RT @GRPBS: Thanks to everyone who visted! It’s great to have your support #bikesharE”
4	“#Bikeshare #dc doing awesome jobs. Nice bikes. Kisoks.”
1	“The last week of #BikeMonth calls for a new nice #bikeshare bag. Have you seen @po_campo’s?”
1	“Just add rider: http://t.co/8Ev3dqhGke @timbuk2seattle #Free #Bikeshare”
2	“Happy best month of the year aka #bikeshare BIKE MONTH!”
2	“Nice work! 18th and RIA. #bikeshare #dc”
2	“I love riding a #bikeshare in #manhattan and I love #bikeshare @CitibikeNYC”
2	“Any folks in #frederick interested in #bikeshare? well here’s your lucky day!”
4	“#Bikeshare #dc #Capitlbikeshare riding awesome. Location incredible. Quick. Feeling great.”

As the used lexicon is not completely domain specific, we find the sentiment score +1 in place of +3. Inspection on the second tweet finds only one positive term: happy. In this case, the score accurately shows the value as positive one.

People use the hashtag symbol “#” before a relevant keyword or phrase (no spaces) in their tweets to categorize those tweets and help them show more easily in Twitter Search. The “hashtag bikeshare” (#bikeshare) generates more tweets than “mention bikeshare” (@bikeshare). The frequencies of the sentiment score for both words are shown in **Figure 4** by using R package “ggplot2” [15]. The plot clearly shows the higher positive sentiments towards the bike sharing program.

The generated tweet dataset for the hashtags was further divided into two other groups: very positive and very negative. **Figure 5** reveals the boxplot of the tweets with positive, negative, very positive, very negative and indifferent/neutral/informative scores. The frequency of positive and very positive tweets is higher than that of the negative and very negative tweets. The highest frequencies are visible at indifferent tweets with no sentiment.

Another contribution of this research is that the findings would be helpful in developing a sentiment lexicon specially developed for transportation-related terms. This paper does not develop any real-time Twitter sentiment analysis tools which is considered as a prospective topic for future research. A real-time Twitter analytic tool would be a good tool for the policy makers to adopt on-time decisions. The concepts developed in this paper would be helpful to perform the sentiment analysis on the auto-generated, real-time tweets.

4. Conclusions

This paper presents a preliminary investigation on how text mining techniques can be used to extract knowledge from tweets associated with a bike sharing program. This paper contributed in performing sentiment analysis on the tweets of the users related to a bike sharing program. This study performs text categorization according to affective positive/negative valence annotation in order to gain subjective information. The findings reveal that the positive responses towards the current system were higher in frequency than the negative responses. Exploration on the neutral and negative tweets will be helpful for the authorities to understand what is lacking in the current system. The discovery in the terms will be used in the future to develop a sentiment lexicon specially designed for transportation-related terms.

Social media text mining works, as performed in this study, by examining and analyzing the posted text and links that reveal a person or group’s uncensored subjectivity and attitudinal polarity (positive or negative). Two

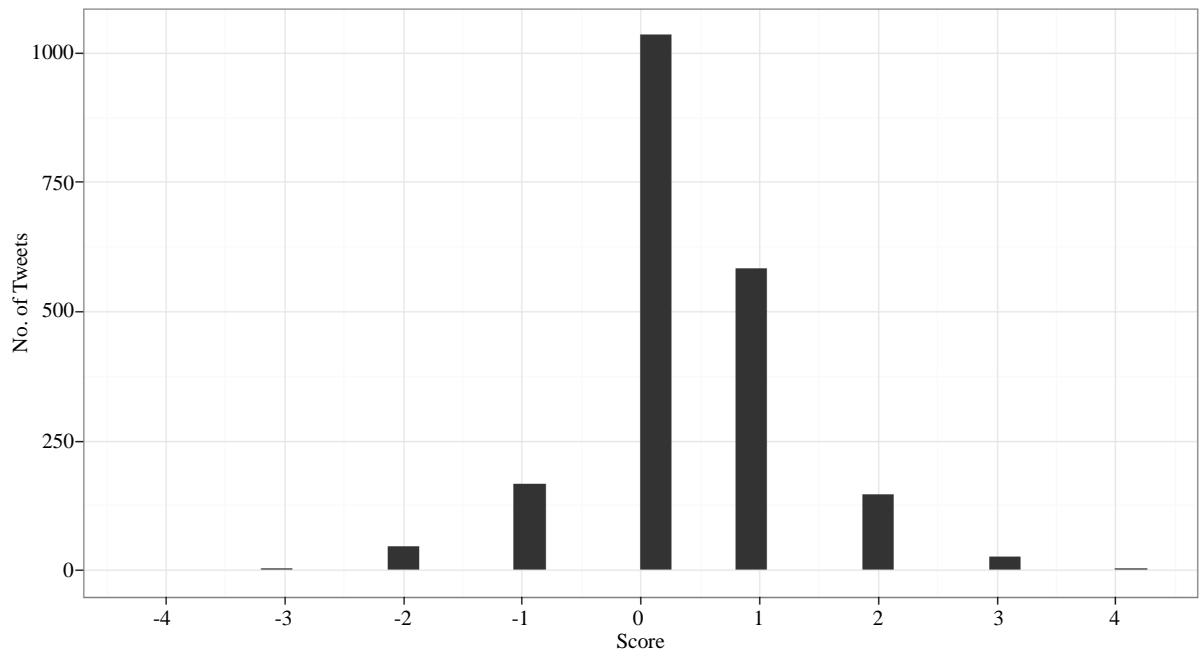


Figure 4. Sentiment score of the tweets.

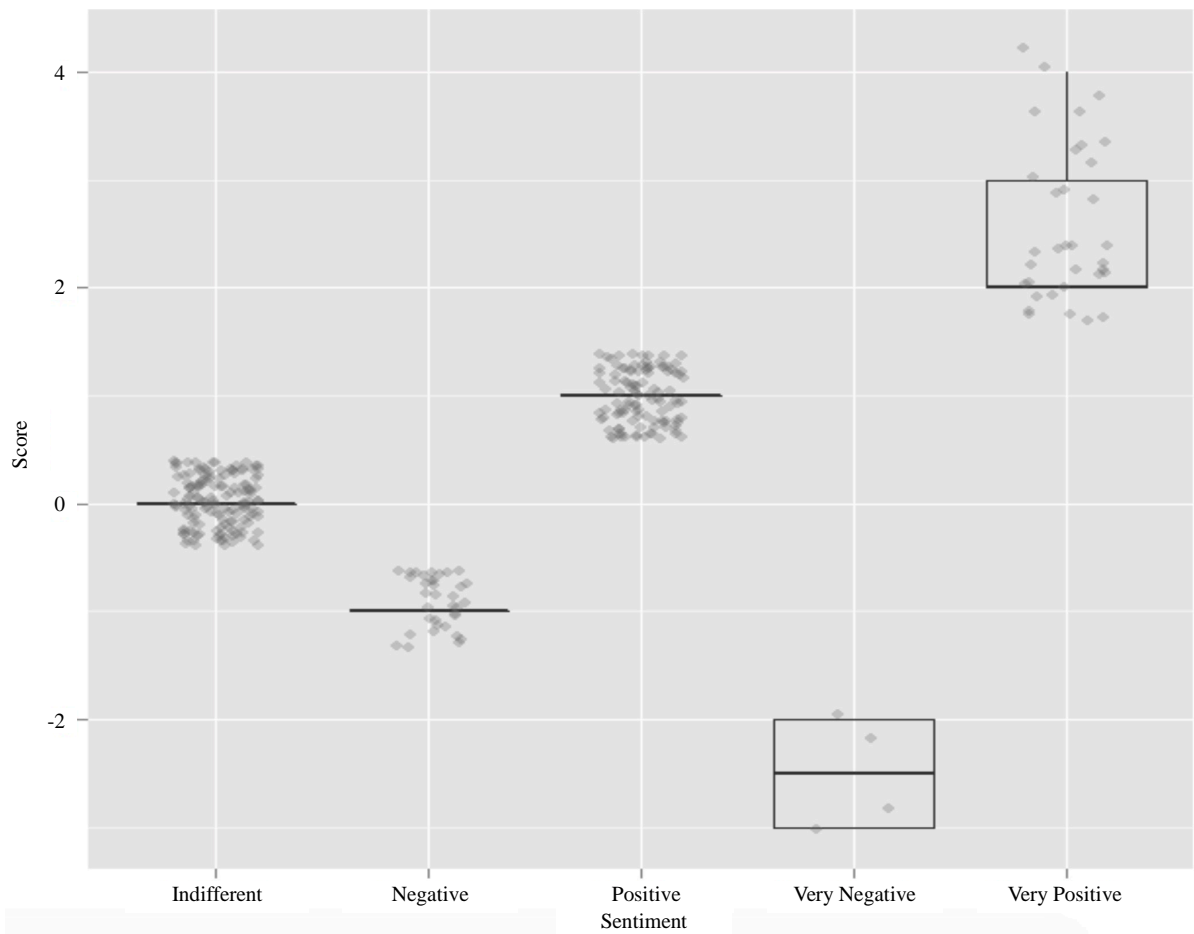


Figure 5. Box-plot of the sentiment scores.

major works were developed in this method; designing a data collection framework and developing a mining tool which can extract the social media information related to the bike sharing program for analysis. This study demonstrates a cheaper survey tool which has a major advantage over conventional attitudinal survey methods; it can easily reach a large audience and it can reflect the true behavior of participants instantly with almost no cost. It is expected that the results from this study will be used in future investigations of text mining of transportation-related countermeasures and services.

References

- [1] Federal Highway Administration (2012) Bike Sharing in the United States: State of the Practice and Guide to Implementation. US Department of Transportation, Washington, DC.
- [2] Tromp, E. (2012) Multilingual Sentiment Analysis on Social Media: An Extensive Study on Multilingual Sentiment Analysis Performed on Three Different Social Media. Lap Lambert Academic Publishing, Germany.
- [3] O'Brien, O., Cheshire, J. and Batty, M. (2014) Mining Bicycle Sharing Data for Generating Insights into Sustainable Transport Systems. *Journal of Transport Geography*, **34**, 262-273.
- [4] Barbosa, L. and Feng, J. (2010) Robust Sentiment Detection on Twitter from Biased and Noisy Data. *Proceedings of the International Conference on Computational Linguistics (Coling'10)*.
- [5] Davidov, D., Tsur, O. and Rappoport, A. (2010) Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. *Proceedings of the International Conference on Computational Linguistics (Coling'10)*.
- [6] González-Ibáñez, R., Muresan, S. and Wacholder, N. (2011) Identifying Sarcasm in Twitter: A Closer Look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: short papers (ACL-2011)*.
- [7] Wang, X., Wei, F., Liu, X., Zhou, M. and Zhang, M. (2011) Topic Sentiment Analysis in Twitter: A Graph-Based Hashtag Sentiment Classification Approach. *Proceeding of the ACM Conference on Information and Knowledge Management (CIKM-2011)*
- [8] Collins, C., Hasan, S. and Ukkusuri, S. (2013) A Novel Transit Rider Satisfaction Metric: Rider Sentiments Measured from Online Social Media Data. *Journal of Public Transportation*, **16**.
- [9] Trip Distribution Data of Capital Bikeshare. <http://www.capitalbikeshare.com/trip-history-data>
- [10] Gentry, J. (2014) twitteR: R Based Twitter Client. <http://cran.r-project.org/web/packages/twitteR/twitteR.pdf>
- [11] Feinerer, I. and Hornik, K. (2013) tm: Text Mining Package. R Package Version 0.5-0.1. <http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf>
- [12] Developer Website for Twitter. <https://dev.twitter.com/docs/auth/oauth>
- [13] Liu, B. (2012) Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies). Morgan & Claypool Publishers, Vermont, Australia.
- [14] Breen, J.R. (2014) Tutorial on Twitter Text Mining. <https://github.com/jeffreymbreen/twitter-sentiment-analysis-tutorial-201107>
- [15] Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer, New York.