# Public opinions of light rail service in Los Angeles, an analysis using Twitter data

Thuy T.B. Luong, University of California, Irvine
Douglas Houston, University of California, Irvine

**Abstract**
Understanding commuters' perceptions, attitudes, and behavior is an important component of transportation planning and management. Collecting such information using traditional survey or interview methods is costly and burdensome, but mining attitudinal data from social networking media could potentially provide insights into the temporal alignment of public opinion with transportation system dynamics. We demonstrate this potential by examining facets of public posts on Twitter about light rail transit services in Los Angeles in terms of sentiment analysis, topic modeling, and the interaction between posters and retweeters. Results provide new insights into how transit users present themselves and their opinions, engage with government agencies, react to events/policies, and share information with others on social media. We also demonstrate an interactive online interface that transit service providers could use to display and monitor real-time feedback and sentiment along different lines in the area's light rail system.
**Keywords:** public opinion, sentiment analysis, public transit, social network, communication
**Citation**: Luong, T.B.T., Houston, D. (2015). Public opinions of light rail service in Los Angeles, an analysis using Twitter data. In *iConference 2015 Proceedings.*

**Contact**: luongt1@uci.edu

## 1 Introduction

With the explosive growth of social web services and mobile devices, individuals and organizations are increasingly using information from social networking media for decision-making. Twitter is a form of blogging that allows users to send brief text updates or micromedia such as photographs or audio clips. Through the service, users provide information on geo-location (GPS coordinates from smartphones), timing of where and when they provided information, and their opinions/reviews. Several social sectors (e.g. politics, entertainment and business) have been gathering and analyzing information from Twitter (Mittal & Goel, 2012; Tumasjan, Sprenger, Sandner, & Welpe, 2010). Using information from this data source could reduce the need to conduct surveys, opinion polls, and focus groups because there is such abundance of information is already publicly available. However, finding opinions on Twitter and distilling the information into meaningful patterns is a time consuming task. The average human reader will have difficulty identifying relevant sites and extracting and summarizing the opinions from large numbers of daily Twitter posts. We developed an analysis system based on text mining techniques to address this need.

Though using Twitter data for opinion mining is popular in many fields, its use in transportation planning and management sector remains still limited. In 2013, Iteris (Mai & Hranac, 2013) compared incident records from the California Highway Patrol with Twitter messages related to roadway events over the same time period. In another study, Twitter data were obtained from the riders of the rapid transit system of the Chicago Transit Authority which allows members of Twitter to monitor sentiments of transit users (Collins, Hasan, & Ukkusuri, 2012). However, these studies are still far from sufficient public opinion mining. This paper extends these previous efforts by introducing ways to understand opinion data in Twitter in term of sentiment analysis, topic modeling, and the interaction between posters and retweeters, which could contribute to improved situational understanding, crisis communications, enhancing and delivering transport policy goals. We conduct detailed analysis for each line in 7-line rail transit system in Los Angeles.

## 2 Methodology

### 2.1 Data collection

We started collecting tweets from May 1st 2014 to October 1st 2014 using the Search Twitter API, with radius 50 miles around Los Angeles using name/account's name of 7 LR lines. After cleaning data, there are 8,515 tweets left for analysis, including 4,131 tweets were retweeted.

## 2.2 Data analysis

- **Exploratory analysis of retweet relationships**

Nearly half of the collected tweets were retweeted. By retweeting a tweet, all of users' connections are able to see the message, therefore reaching more people. Thus, we used exploratory analysis to identify salient themes in the tweeted data. First, we classified the accounts. The type of tweet posters includes: government, service, school, firm, light rail agency, transportation agency (non-light rail), and person. Then a network graph of relationships between tweet posters and retweeters was created, in which each edge connects two accounts that appear in the same tweets, one from poster, and the other from retweeter.

Using our collected tweets, we found that the commuters retweeted from other individuals the most, and then transit agencies, while schools and firms did not have strong retweet connections (Figure 1). It means the focus of most retweets was the transfer of transit-related information from person to person. Therefore, in order to spread information or promote public transit, the agencies could utilize Twitter as a social marketing tool because of its incredibly powerful virality.

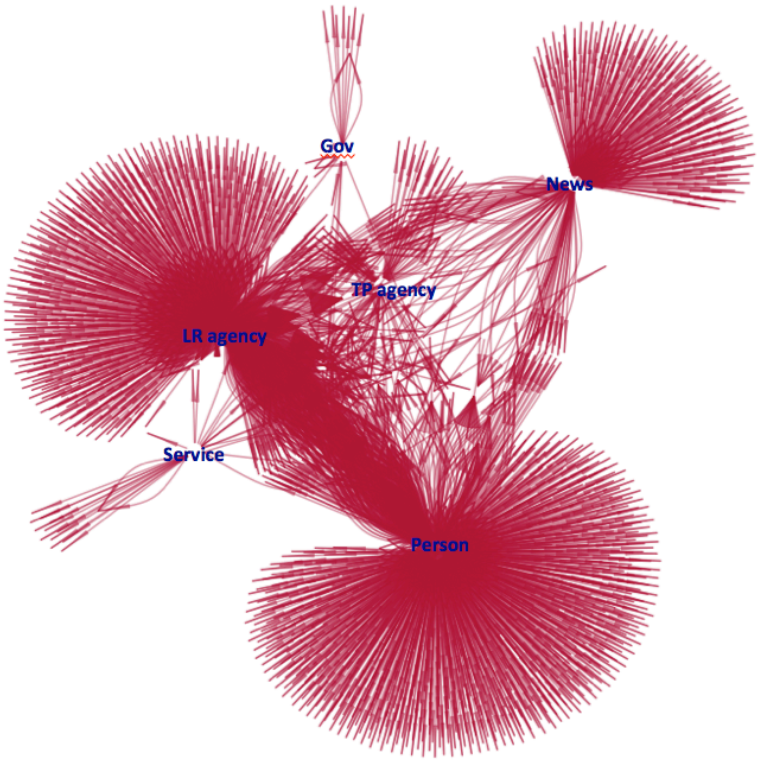


Figure 1: Who retweets whom transit-related information

- **Sentiment analysis**

Sentiment analysis is a form of data mining used to interpret textual and verbal information. By parsing human verbal communication through natural language processing, applying computational linguistics, and deploying text analytics, subjective information (i.e., attitudes, opinions, and intentions) in source materials can be identified and extracted for more intense processing and scrutiny. To extract the sentiment of these tweets automatically, we used an opinion lexicon in English, which is provided by Hu and Liu (Hu & Liu, 2004) after translating emotion icons, emojis to words, and pre-processing raw data. Each tweet was assigned an overall score based on the total score of each of its words. The average sentiment value for each rail transit tweet line is shown in Figure 2. This type of information can transit providers and government and policy makers understand the condition or status of each rail transit line from the commuters' perspective. For example, in the period of analysis (5 months), we found that the Red Line was associated with the most positive tweet sentiments while the Blue Line has the most negative tweet sentiments.

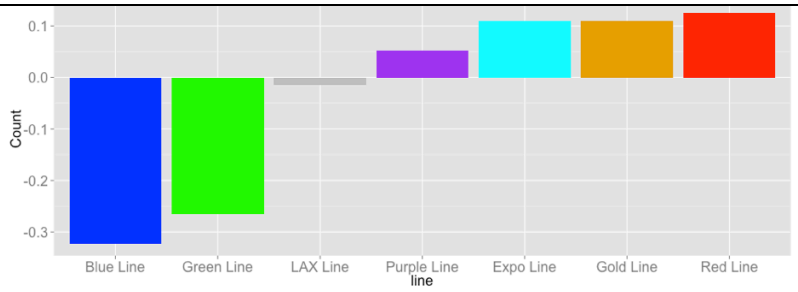


Figure 2: Average sentiment score by rail transit line

Also, for each rail transit line, we conducted sentiment analysis on temporal dimension, to identify when riders had the most negative or positive opinions about the service. Because we built an interactive web-application for this study, we can generate comparisons about sentiments among rail transit lines. For example, for the Blue line, while Mondays had more positive tweets, weekends had more negative tweets (Figure 3).

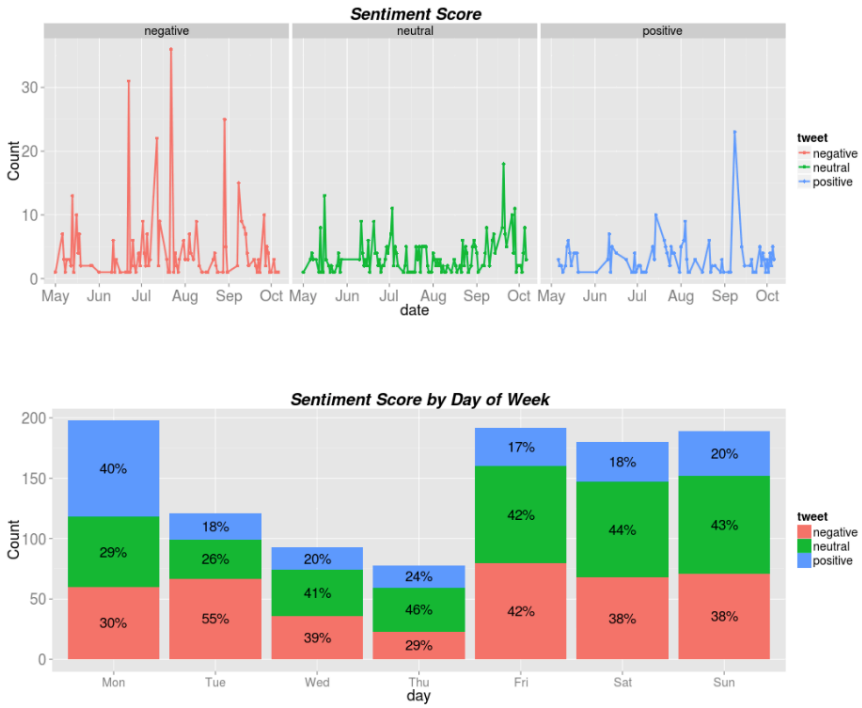


Figure 3: Sentiment score by day of week for the Blue Line

- **Word clustering analysis**

We analyzed the topics of tweets for particular lines to help to explain why a particular line had a negative or positive sentiment score. The technique can help clarify why people like/ dislike a particular segment of service.

To identify the topics of tweets, we used unigram (n-gram of size 1) with k-medoids clustering, to group and partition the frequent words in tweets. An n-gram model is a type of probabilistic language model for predicting the next item in such a sequence in the form of a (n - 1)–order Markov model. This approach allowed us to resolve the number of connections a word have with other words and we graphed these patterns to visualize how words and tweets are connected as a network of terms. The colors and connections colors reflect the number of connections (degrees) each term has with other terms. Thus, we see definite terms that occur together (more connections-degrees) in tweets with larger labels and darker edges (lines). For example, Figure 4 shows that with the Blue Line, the frequency of words such as “delay”, “disable”, “dies”, “fatal”, “beating” are higher than the other words.

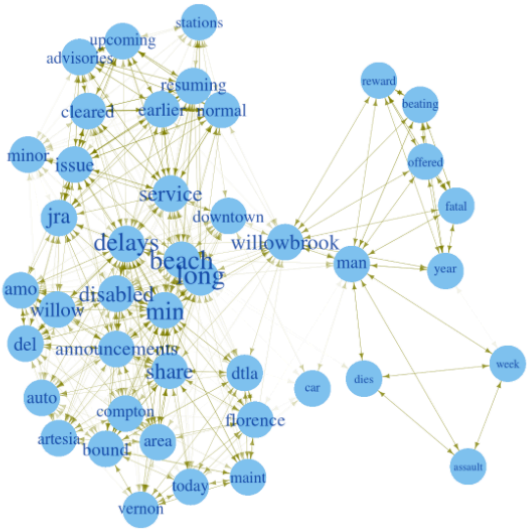


Figure 4: Network analysis of term for the Blue Line

## 3 Conclusion

Building public transit has become an important strategy for reducing vehicle travel and associated emissions. Consistent with this trend, Los Angeles County's main transit agency has committed funds for several new light rail lines and segments scheduled to open by 2019. This research demonstrates ways that transportation planning and management agencies can take advantage of public social media to not only spread information or promote public transit, but also understand more about rail transit users' feedbacks and opinions, especially when dealing with emergencies and incidences. Using text-mining techniques, we provide transit users as well as transit providers differentiated insights about comments, criticisms, responses, and shared information about LR in Los Angles. The methods we developed in this study could also be applied to analysis from other areas of service operations, which could in turn help further improve the public transit policy and management.

## References

Collins, C., Hasan, S., & Ukkusuri, S. (2012). A Novel Transit Riders' Satisfaction Metric: Riders' Sentiments Measured from Online Social Media Data. Proceedings of the 2013 Annual Meeting, Transportation Research Board

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews, 168–177.

Mai, E., & Hranac, R. (2013). Twitter Interactions as a Data Source for Transportation Incidents. Proceedings of the 2013 Annual Meeting Transportation Research Record.

Mittal, A., & Goel, A. (2012). Stock prediction using Twitter sentiment analysis. Standford University,

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. ICWSM, 10, 178–185.

## Table of Figures

Figure 1: Who retweets whom transit-related information ..... 2
Figure 2: Average sentiment score by rail transit line ..... 3
Figure 3: Sentiment score by day of week for the Blue Line ..... 3
Figure 4: Network analysis of term for the Blue Line ..... 4