Recommendations to the New Cities Foundation Task Force on Connected Commuting

CALIFORNIA PATH PROGRAM, INSTITUTE OF TRANSPORTATION STUDIES
and
CENTER FOR INFORMATION TECHNOLOGY RESEARCH IN THE INTEREST OF SOCIETY
UNIVERSITY OF CALIFORNIA, BERKELEY

Prepared for: The New Cities Foundation

Prepared by:

Greg Merritt
Nathaniel Bailey
California PATH Program
Institute of Transportation Studies
University of California, Berkeley
2105 Bancroft, Suite 300
Berkeley, CA 94720

Professor Alexandre Bayen
Dept. of Civil and Environmental Engineering
Center for Information Technology Research in the Interest of Society (CITRIS)
California PATH Program
University of California, Berkeley
642 Sutardja Dai Hall
Berkeley, CA 94720-1720
This work was performed as part of the Center for Information Technology Research in the Interest of Society (CITRIS) and the California PATH Program of the University of California, in cooperation with the New Cities Foundation, Ericsson, the San Jose Department of Transportation, Orange Labs, Waze and Roadify. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.
# Table of Contents

Acknowledgements .................................................................................................................. 6  
Executive Summary .................................................................................................................. 7  
  Problem Statement .................................................................................................................. 7  
  Background ............................................................................................................................. 7  
  Analysis .................................................................................................................................. 7  
  Recommendations .................................................................................................................. 7  
  Conclusion .............................................................................................................................. 8  
Introduction ............................................................................................................................... 9  
  Background ............................................................................................................................. 10  
  Text Sentiment Analysis ........................................................................................................ 10  
  Text Sentiment Analysis for the Social Web .......................................................................... 11  
  Text Sentiment Analysis for the Social Web in the Transportation Domain ...................... 11  
  Transportation Domain Text Sentiment Analysis for Participatory Commuting Apps ........ 12  
Data Analysis ............................................................................................................................ 13  
  Waze Data Set ....................................................................................................................... 13  
  Type Distribution of Reports and Comments ....................................................................... 16  
  Comment Lexicon by Type ...................................................................................................... 18  
  Sentiment Analysis ................................................................................................................ 21  
  Sentiment by Report Type ...................................................................................................... 22  
  Sentiment by Day of Week .................................................................................................... 23  
  Sentiment by Time of Day ..................................................................................................... 24  
  Roadify Data .......................................................................................................................... 31  
Discussion and Recommendations ............................................................................................. 35  
  Mobile Travel Applications as Pre-Filters for Travel-Related Texts .................................... 35  
  Category Tags and Context as Mechanisms for Refining Domain ...................................... 35  
  Geotagging ............................................................................................................................. 36  
  Utility of Sentiment Analysis ................................................................................................. 36  
  Limitations of Current User Commenting Rates ................................................................... 37  
  Explicitly Encouraging Positive Interactions Among Participatory Commuters .................. 38  
  Tuning Regional Commute Programs to Target Times and Topics Most Relevant to Commuter Sentiment .................................................................................................................. 38  
  Routing for Positive Sentiment and an Improved Commute Experience .......................... 39
Crowdsourced Traveler Data as Travel Network Condition Reporting

Conclusions

References

Appendix A: SentiStrength 2 Training and Use

Optimization of Term Strengths

Cross-Validation of Initial Optimized Emotion Lookup Table

Report of Term Weights for Bad Classifications

Cross-Validation with Fully-Trained Term Lists

Scoring of the Waze Corpus

Figures

Figure 1: Geographic Region of Waze Data Collection

Figure 2: Waze Report and Comment Counts, by Month

Figure 3: Fraction of Waze Reports with Attached Comments, by Month

Figure 4: Waze Report and Comment Data Entry Screens

Figure 5: Distribution of Waze Reports by Type

Figure 6: Distribution of Waze Comments by Report Type

Figure 7: Word Cloud for Waze Report Type "Hazard"

Figure 8: Word Cloud for Waze Report Type "Traffic Jam"

Figure 9: Word Cloud for Waze Report Type "Accident"

Figure 10: Word Cloud for Waze Report Type "Police"

Figure 11: Word Cloud for Waze Report Type "Chit chat"

Figure 12: Waze Comment Positive Valence Classification Distribution by Report Type

Figure 13: Waze Comment Negative Valence Classification Distribution by Report Type

Figure 14: Waze Comment Positive and Negative Valence Classification by Day of Week

Figure 15: Waze Comment Positive and Negative Valence Classification by Day of Week, Report Types Hazard, Traffic jam, Accident, and Police

Figure 16: Positive Valence and Negative Valence by Hour of Day, All Report Types, Weekdays

Figure 17: Positive Valence and Negative Valence by Hour of Day, All Report Types, Weekends

Figure 18: Positive Valence and Negative Valence by Hour of Day, All Report Types, Mondays

Figure 19: Positive Valence and Negative Valence by Hour of Day, All Report Types, Tuesdays

Figure 20: Positive Valence and Negative Valence by Hour of Day, All Report Types, Wednesdays
Tables

Table 1: Waze Data Fields........................................................................................................................................14
Table 2: Waze Comments with Both Strong Negative and Strong Positive Valence Classification.............................................22
Table 3: Classification of Sample Roadify Comments. For the sample shown, classification results are identical when run with either the general, untrained SentiStrength 2 lists or the Waze-trained lists..................................................34
Table 4: Roadify Comments Misclassified When Evaluated with Word Lists Trained for the Waze Domain...........................................34
Table 5: Waze Training. This table shows a sample of the 1,000 randomly-selected user comments and human-assigned valence ratings. This human-classified subset of Waze comments was used as the reference for subsequent calibrations.

Table 6: SentiStrength 2 Term Weight Optimization Command.

Table 7: Examples of EmotionLookupTable.txt Valence Training Results.

Table 8: Ten-Fold Cross-Validation Command. Initial run, after first optimization of emotion lookup table.

Table 9: termWeights Command.

Table 10: Partial List of Waze Terms Added to Emotion Lookup Table.

Table 11: Sample of Waze Slang Term Mappings Added to Slang Lookup Table.

Table 12: Ten-Fold Cross-Validation Command, Fully-Trained Term Lists.

Table 13: Comparison of Initial and Final Ten-Fold Optimizations. The reported values are correlations between the sentiment predictions and the human-coded values of the 1,000-comment training set. "Within 1 Point" accounts for exact matches and for near-matches that differ by +1 or -1 valence rating.
Acknowledgements

Research supported by New Cities Foundation. The support of the Foundation is gratefully acknowledged. We would like to particularly thank Naureen Kabir for her coordination and support of the Task Force.

For their participation in and contributions to the Task Force launch meeting, we would like to thank Naureen Kabir and Mathieu Lefevre of New Cities Foundation; Patrik Cerwall, Monika Byléhn, Nimish Radia, and Kshitiz Singh of Ericsson Research; Manuel Pineda of the San Jose Department of Transportation; Di-Ann Eisnor and Michal Habdank-Kolaczkowski of Waze; Scott Kolber and Ethan Arutunian of Roadify; Gabriel Sidhom of Orange Labs; and Ryan Langon of General Electric.

Naureen Kabir, Nimish Radia, Kshitiz Singh, and Tania Leppert provided valuable input and feedback for the sentiment analysis work that makes up the majority of this report, and overall management of the work of the Task Force.

We would like to thank Waze and Roadify for granting access to their unique data sets for analysis in this report. Waze's Fej Shmuelevitz and Di-Ann Eisnor worked with our team to deliver their data sets for our research work. Similarly, we appreciate the efforts of Daniel Robinson, Ethan Arutunian, and Scott Kolber to establish a Roadify data feed mechanism.

Professor Joan Walker and Leah Anderson of UC Berkeley are acknowledged for their helpful feedback in the areas of data analysis and data presentation.

UC Berkeley's Professor Walker, Romain LeBlanc, and Lisa Hammon are also gratefully acknowledged for providing feedback on the design of the Foundation’s commuter surveys.

Yiguang (Ethan) Xuan of UC Berkeley provided assistance with overlays of data on maps.

Scott Myers of UC Berkeley contributed to development of the Roadify data feed.

Special thanks are due to Professor Michael Thelwall of Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton for use of the SentiStrength 2 algorithm for sentiment strength detection. Professor Thelwall’s assistance with SentiStrength code modifications is greatly appreciated.
Executive Summary

Problem Statement

The New Cities Foundation Task Force on Connected Commuting seeks new ways to improve commuters’ experience of their daily travel. This report explores techniques to ascertain commuter sentiment in order to inform efforts to improve the commute experience.

Background

Text sentiment analysis is a technique for quantifying the emotion associated with written works. This procedure lends itself to bulk evaluation of large quantities of texts for general positive and negative sentiment ratings.

Recent efforts have refined this technique to work well with short, "microblog" texts, and allow for tuning of reference sentiment dictionaries to more accurately classify texts related to a narrow topic range. In this report, this technique is applied to text comments written by users of the commute-related smartphone applications Waze and Roadify.

Analysis

Sentiment dictionaries of the SentiStrength 2 algorithm were trained to reflect the Waze data lexicon and the valence of important words in this domain. We apply this trained algorithm to a set of 15,131 Waze user comments for classification of positive and negative sentiment.

We created an automated feed mechanism to collect daily sets of Roadify user comment data, and evaluate this data for suitability in text sentiment analysis in this application.

We show association of classified sentiment to day of week, time of day, type of report, geographic location, and lexicon.

Recommendations

The following recommendations are made on the basis of the data observed, which is very partial and incomplete. These are a list of tentative recommendations that follow from empirical observations. These are not based on exhaustive surveys, and are thus to be taken with caution, for what they are: qualitative assessments based on the observed data.
1. Commuter comments collected by smartphone applications should be recognized as a high-quality, real-time source of commuter sentiment data; opinions mined from such data sets provide a valuable metric of commuter experience.

2. The value of commuter comment data sets is greatly increased by presenting users with the requirement of assigning categories for comments; commute-related smartphone applications should include this feature.

3. Smartphone applications should geotag comments to increase the utility of sentiment data.

4. Application designers must follow current best practices in areas related to safety and privacy.

5. Sentiment analysis engines must be trained to suit the domain to which they're applied.

6. Higher rates of traveler reporting and commenting greatly improve the utility of the techniques described in this report; efforts should be made to increase these rates.

7. To increase the exchange of positive communication, mobile commute applications should explicitly solicit and share positive user comments.

8. Government and employer commute programs should consider mined sentiment and topics to better focus their efforts on times, locations, and themes of concern to commuters.

9. Travel planning and routing applications should offer itinerary choices that optimize for positive experience based on analyzed historical user comment data sets; these applications should also explicitly collect commuter ratings of their travel experience for future analysis.

10. Transportation network planners and operators should look to sentiment analysis as a metric to gauge the impact of infrastructure changes that is less costly and more timely than traditional surveys.

As a parallel effort to the research presented in this report, the Task Force has undertaken a project to survey San Jose area commuters to better understand the role of smartphone applications in the daily commute. Separate from this report, the UC Berkeley team provided feedback for survey design.

**Conclusion**

Crowdsourced comment data from mobile smartphone applications such as Waze and Roadify is an underutilized source of commuter data. Such data offers great potential for improving the urban commute experience through the application of text sentiment analysis to inform a variety of technologies, products, services, and innovations.
Introduction

The difficulties and stresses related to the daily commute are widely recognized as important factors contributing negatively to the quality of life in urban areas. A study by Kahneman et al. (2006) indicates that commuting is one of the least-enjoyable activities of daily life, rated more negatively than working and housework; another survey (Hewlett_Packard, 2004) suggests that commuters regularly undergo stresses greater than those of fighter pilots.

Traditionally, commuting problems are addressed via infrastructure improvements and coordinated long-range planning, and typically focus on metrics such as increasing capacity and minimizing delay. The New Cities Foundation Task Force on Connected Commuting seeks new ways to improve commuters’ experience of their daily travel.

The increasing use of smartphones and increasingly ubiquitous connectivity introduce the potential for commuters to not just connect socially, but to also inform one another about their commutes in real time and in a one-to-many fashion.

Task Force partners Waze and Roadify have created mobile applications that specifically invite commuters to exchange information about current, developing local travel conditions. This information exchange includes short text messages written by travelers and intended for public consumption by other travelers making similar trips.

This study explores the validity of using the emerging technique of text sentiment analysis of "microblog" texts to process and extract meaning from aggregate collections of these terse, informal communications, and evaluates the utility of this technique as an important element of potential technologies, products, services, and innovations aimed at improving the participatory commuting experience.
Background

The techniques discussed in this report rely on machine learning and text sentiment analysis applied to crowdsourced text messages generated by travelers using smartphone applications. This section reviews general principles and previous work.

Text Sentiment Analysis

Sentiment analysis, sometimes called opinion mining, is the collection and analysis of subjective ratings of text or images, typically using simple scales of parameters such as affective valence (positive to negative emotional response) and arousal (intensity of emotion). Common applications of text sentiment analysis include the evaluation of reviews of consumer products for commercial purposes, surveys of public opinion through analysis of blog and discussion forum postings, and determination of viewer response to films through bulk, automated analysis of reviews and commentary. (Pang, 2008)

Several different approaches have been used to automate classification of text samples for valence. This is generally done by scanning a text sample for the occurrence of words that match words from a reference list. The reference words have pre-assigned valence values; statistical calculations are run on the valence values of matched words.

The valence values of words on the reference list are numerical ratings of emotional response assigned by human subjects; for convenience of calculation, terms that elicit positive emotion are assigned positive valence values, and terms that elicit negative emotion are assigned negative valence values.

For longer texts, approaches such as that of SO-CAL (Taboada, 2011) calculate and compare total positive and negative sentiment scores for all list matches for the complete text, and, using a weighting factor, generate a net positive or net negative classification.

For shorter texts, a simple form of valence classification is the net valence. This is calculated by summing together the maximum (positive) valence score to be found from any word in the text with the minimum (negative) valence score to be found from any word in the text.

For some applications of analysis of short texts, positive and negative scores are evaluated independently, recognizing that even a single, short text may simultaneously express both strong positive and strong negative sentiment.
While rated lists of common words can be used to carry out sentiment analysis of any general text, best results are obtained if reference word lists are "trained" to better reflect the valence of the words as they’re typically used in the domain of study. For example, while the term "backup" might have positive valence in consumer reviews of personal computing products, it tends to have a negative association in the domain of automobile commuting.

**Text Sentiment Analysis for the Social Web**

In recent years, social Web applications have become a source of increasingly large quantities of written texts of very short length. Twitter is the best-known example of such a "microblogging" service. Twitter texts (tweets) are limited to 140 characters each, and the service’s several hundred million users worldwide (Wasserman, 2012) currently post about 400 million tweets per day (Farber, 2012).

Large aggregations of microblog posts cannot be read and analyzed by people, particularly not in real time. Text sentiment analysis offers a method for extracting measures of emotion from these large data sets. (Pak, 2010)

In practice, tweets and other short, informal messages offer special challenges to text sentiment classification when compared to traditional, longer text samples of reasonably-well-crafted prose. For example, abbreviations are often used in order to fit many words within a sparse character limit and to speed up data entry while using the physically small, cumbersome data entry interfaces of mobile devices. Misspellings are common in this variety of informal, non-vetted writing, and people often incorporate non-alpha characters and non-word strings.

The text sentiment analysis algorithm used in the this study, SentiStrength 2, addresses these special characteristics of texts from the social Web; see below and Appendix A for additional information.

**Text Sentiment Analysis for the Social Web in the Transportation Domain**

While microblogs have been mined for measurement of general sentiment, gauging of political opinion, and marketing, little work has been done to mine for sentiment in the transportation domain. Most notably, Collins et al. (2012) used Twitter posts as a source of transit riders’ sentiment as a performance metric of the Chicago Transit Agency’s rail service.

While the Ukkusuri study demonstrated that analysis of tweets offers the possibility of quickly and inexpensively aggregating opinion for performance metrics, significant challenges were found selecting relevant tweets about the rail system from the "fire hose" of available tweets, most of which do not relate to travel.
Similarly, colleagues of the authors of this report have examined the Twitter feed for travel data associated with a southern San Francisco peninsula location, and have reported similar "signal to noise" challenges.

**Transportation Domain Text Sentiment Analysis for Participatory Commuting Mobile Apps**

As the use of smartphones increases, commuters have increasing access to mobile applications designed to facilitate travel. Such applications may provide various combinations of static information (e.g., published train schedules) and real-time information (e.g., bus arrival times or developing highway congestion). They may also permit users to share timely transportation network information with other travelers (e.g., availability of parking or location of a recent vehicle collision), and may also invite users to comment on their observations and experience.

These crowdsourced comments solicited from travelers using mobile smartphone applications are effectively "pre filtered" to be predominantly related to the particular domains of travel and transportation associated with the mobile application used to collect the comments. Such pre-filtered data sets are much better-suited to studies of sentiment in the realm of participatory commuting than data sourced from general, non-topic-specific social media microblogging services.

Mobile applications can also prompt users to categorize or contextualize their comments, providing finer domain distinction than that provided simply by the use of the particular application.

Furthermore, most smartphones are equipped with geolocation features that enable applications to report GPS coordinates. Geolocation data from smartphones was first shown to be useful for traffic estimation by Bayen et al. and algorithms have been developed to infer vehicle trajectory from sparse geolocation data. (Work, 2010) Participatory commuting smartphone applications offer the potential to combine sentiment mined from rich, commute-related microblog posts with trajectory and other geolocation data.

This study explores the use of sentiment analysis of travel-related microblog posts from smartphone application users, and suggests ways that these techniques might be used to improve the daily travel experience of urban commuters.
Data Analysis

Waze Data Set

Waze contributed anonymized traveler reporting data for this study. This data set consists of 114,256 user reports spanning twenty-six contiguous months, from January 1, 2010 through February 29, 2012, within a geographic bounding box described by a rectangle with corners at the following coordinates:

37.43 latitude, -122.09 longitude
37.22 latitude, -121.74 longitude

This region corresponds approximately to the City of San Jose, California, and immediate surroundings; see map, Figure 1.

![Figure 1: Geographic Region of Waze Data Collection. The Waze data set analyzed in this study came from user reports submitted from reported locations in the region shown. Map data ©2012 Waze](image)

The Waze data set includes the fields shown in Table 1.
### Table 1: Waze Data Fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Sample Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Date</strong></td>
<td>Time user initiated report within Waze app: Year-month-day hour:minute:second (GMT)</td>
<td>2011-08-27 00:28:21</td>
</tr>
<tr>
<td><strong>Alert Type</strong></td>
<td>User-selected alert category: 0=Chit Chats 1=Police 2=Accidents 3=Jams 5=Hazards 6,7=Other</td>
<td>3</td>
</tr>
<tr>
<td><strong>Text</strong></td>
<td>User comment (optional)</td>
<td><em>I love living so close to the coast but the beach traffic on my commute home is obscene.</em></td>
</tr>
<tr>
<td><strong>X</strong></td>
<td>Longitude, degrees</td>
<td>-121.98489</td>
</tr>
<tr>
<td><strong>Y</strong></td>
<td>Latitude, degrees</td>
<td>37.220044</td>
</tr>
</tbody>
</table>

This set of 114,256 reports includes 15,131 reports with text in the optional comment field. For the 26-month data set overall, this represents a comment rate of just over 13% of reports.

Figure 2 shows month-to-month report and comment activity for the San Jose Waze data set. (Note that months may have 28, 29, 30, or 31 days; given that Waze usage is lower on weekends than weekdays, and that the number of weekends in a given month can vary, calculating daily rates within a given month was not considered to be more meaningful than using simple per-month counts for comparing month-to-month trends.)
These monthly report and comment rates generally increase through the period of study, with large increases in reporting rates in later months. This correlates well with Waze's self-reported increase of application users in this time period: a July 2012 Waze blog entry indicates "an impressive 2012 for us so far as we've doubled our Wazer population from 10 Million to 20 Million in only 6 months." (Feldman, 2012)

Notably, the fraction of reports that includes user-contributed comments has decreased over time; see Figure 3. Through 2010, comments were included with one third to one half of all reports each month; by early 2012, comment rates had fallen below 10% for this data set.
However, this decrease in comment rate has been outstripped by the overall rate of increase in reporting in the San Jose area, yielding a general trend of increasing numbers of comments month-to-month, particularly through 2011 and into 2012.

An examination of reports from a period spanning mid May through late June of 2012 (a period outside the 26-month period used for this report) indicates that these trends continue for the San Jose area: reporting rates continue to increase, and the comment faction continues to decrease slightly, but not enough to prevent a net increase of number of comments per month.

These trends are discussed in the Discussion and Recommendations section of this report.

**Type Distribution of Reports and Comments**

When submitting a report, Waze users first select a report type; see the Waze report type selection screen in Figure 4.

The data set for this study includes reports of type Traffic jam, Police, Accident, Hazard, and Chit Chat; see . A small fraction of reports (0.4%) from other categories are also included.
Figure 5 shows the distribution of report types for the 26-month San Jose data set. Nearly half of the reports have been assigned by users to the "Traffic jam" category. "Hazard" and "Police" each account for just under one fifth of all reports, with the remainder approximately split between "Accident" and "Chit chat."

When only reports with comments are considered, the distribution of report types is distinctly different. The "Chit chat" category contributes to just over half of the commented reports. This is followed by Jams, Hazards, Police, Accidents, and Other in order of decreasing fraction; see Figure 6.
Analysis of word occurrence statistics of Waze comments from this data set shows high differentiation of vocabulary across report type. Each report type was found to have a distinctly different set of most-frequently-occurring words.

We present Waze user lexicon by report type as word clouds in Figure 7 - Figure 11. A word cloud displays each unique word from a set of words, with rendered font size a function of frequency of occurrence; more-frequently-occurring words are larger, and less-frequently-occurring words are smaller. In these particular renderings, commonly-occurring English words (such as "the," "a," and "to") have been excluded. For each set, only the 300 most-frequently-occurring words are included.

These word clouds reveal that within each report type, comment lexicon distribution is distinct. That is, statistically, when Waze users add a comment to a contributed report, the range of subjects discussed is related to the user-selected report type.
Figure 7: Word Cloud for Waze Report Type "Hazard"

Figure 8: Word Cloud for Waze Report Type "Traffic Jam"
Figure 9: Word Cloud for Waze Report Type "Accident"

Figure 10: Word Cloud for Waze Report Type "Police"
Sentiment Analysis

Sentiment classification of the user comments in this data set was carried out using the SentiStrength 2 algorithm. See Appendix A for documentation of SentiStrength 2 training for the specific domain of Waze user comments.

For some applications, net sentiment – classification of a text as predominantly positive or predominantly negative – can be useful. For this data set, even with its short-length texts, many examples include both strongly-negative and strongly-positive sentiment; to classify a text with such high-response words as net neutral would mask the strong sentiment expressed in the comment. Each comment was classified separately for both positive and negative sentiment with the trained, domain-specific SentiStrength 2 emotion and slang dictionaries.

In this report, the positive valence classification scale ranges from 1 (neutral) to 5 (strongest positive), in integer steps; the negative valence classification scale ranges from -1 (neutral) to -5 (strongest negative). Every comment receives a positive classification and a negative classification, although one or both may be neutral. Error! Reference source not found. shows examples of Waze comments that each feature both strong positive and strong negative classification.
Table 2: Waze Comments with Both Strong Negative and Strong Positive Valence Classification

<table>
<thead>
<tr>
<th>Positive Classification</th>
<th>Negative Classification</th>
<th>Comment Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-4</td>
<td>Loves the carpool lane when he uses it but hates everyone in it when he's driving by myself that's fair right?</td>
</tr>
<tr>
<td>3</td>
<td>-3</td>
<td>Why is traffic so bad these days? Better economy? lol</td>
</tr>
<tr>
<td>3</td>
<td>-4</td>
<td>Fantastic weekend despite the rain... and terrible traffic ... and the big box retailers...</td>
</tr>
<tr>
<td>4</td>
<td>-5</td>
<td>Love the rain hate the traffic</td>
</tr>
<tr>
<td>4</td>
<td>-3</td>
<td>I love living so close to the coast but the beach traffic on my commute home is obscene.</td>
</tr>
<tr>
<td>3</td>
<td>-4</td>
<td>Dear traffic. Have I told u lately how much I hate u with a passion.. Thank god for my car being Hella small and fast ! Lol I’m movin thru</td>
</tr>
</tbody>
</table>

**Sentiment by Report Type**

Just as lexicon varies by report type, so does distribution of positive and negative sentiment classification. Figure 12 and Figure 13 show these distributions.

![Figure 12: Waze Comment Positive Valence Classification Distribution by Report Type](image-url)
Figure 13: Waze Comment Negative Valence Classification Distribution by Report Type

Sentiment by Day of Week

In Figure 14 and Figure 15, we present positive and negative sentiment distributions by day of week. Figure 14 source data includes all Waze user comments, regardless of report type. For comparison, this is repeated in Figure 15 with source data is restricted to the report types Hazard, Traffic jam, Accident, and Police; report types Chit chat and Other are omitted.

Figure 14: Waze Comment Positive and Negative Valence Classification by Day of Week. This heat map presents daily positive valence distribution and negative valence distribution, which are independent. Heat values represent fraction of total comments (of the respective polarity) with the indicated classification value. Valence values of +1 and -1 were included in distribution calculations, but are omitted from display since they correspond to neutral classification and overshadow the stronger classifications on this scale.
Figure 15: Waze Comment Positive and Negative Valence Classification by Day of Week, Report Types Hazard, Traffic jam, Accident, and Police. This heat map presents daily positive valence distribution and negative valence distribution, which are independent, for aggregated report types of Hazard, Traffic jam, Accident, and Police. Heat values represent fraction of total comments (of the respective polarity) with the indicated classification value. Valence values of +1 and -1 were included in distribution calculations, but are omitted from display since they correspond to neutral classification and overshadow the stronger classifications on this scale.

**Sentiment by Time of Day**

Figure 16 and Figure 17 present sentiment by hour of the day for aggregations of weekdays and weekends, respectively; Figure 18 - Figure 24 present hourly sentiment for each day of the week. This set is repeated in Figure 25 - Figure 33 with source data restricted to the report types Hazard, Traffic jam, Accident, and Police; report types Chit chat and Other are omitted. Time of day is local time, converted from source data GMT and corrected for daylight savings.

Note that in most cases, commenting rates are quite low during the overnight hours. Because of this, statistical averaging is poor for valence values shown for late night and early morning. Larger data sets would be required if there were a desire to better determine traveler sentiment during these times; see Discussion and Recommendations for more on data quantity.

As we might expect, these figures depict stronger negative sentiment distributions during weekday morning and evening commute times than during mid-day.
Figure 16: Positive Valence and Negative Valence by Hour of Day, All Report Types, Weekdays

Figure 17: Positive Valence and Negative Valence by Hour of Day, All Report Types, Weekends

Figure 18: Positive Valence and Negative Valence by Hour of Day, All Report Types, Mondays

Figure 19: Positive Valence and Negative Valence by Hour of Day, All Report Types, Tuesdays
Figure 20: Positive Valence and Negative Valence by Hour of Day, All Report Types, Wednesdays

Figure 21: Positive Valence and Negative Valence by Hour of Day, All Report Types, Thursdays

Figure 22: Positive Valence and Negative Valence by Hour of Day, All Report Types, Fridays

Figure 23: Positive Valence and Negative Valence by Hour of Day, All Report Types, Saturdays
Figure 24: Positive Valence and Negative Valence by Hour of Day, All Report Types, Sundays

Figure 25: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Weekdays

Figure 26: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Weekends

Figure 27: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Mondays
Figure 28: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Tuesdays

Figure 29: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Wednesdays

Figure 30: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Thursdays
Figure 31: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Fridays

Figure 32: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Saturdays

Figure 33: Positive Valence and Negative Valence by Hour of Day, Select Report Types (Traffic jam, Hazard, Accident, Police), Sundays

Figure 34 - Figure 37 present geographic distribution and lexicon information for positive and negative sentiment comments for weekday morning and afternoon commute times. The geographic emotional valence maps for the aggregated 26-month Waze data set show concentrations of reports that are distinct across polarity and time of day. The associated word clouds indicate distinct subject areas for each combination of polarity and time of day.
Figure 34: Weekday Morning Positive Valence Map and Word Cloud. Map points indicate reported GPS location of Waze reports with positive valence classification between 2 and 5 submitted between the times of 05:00 and 10:00, Monday-Friday. Larger points indicate classifications with stronger positive valence classification. Word cloud represents the same data set. Map data ©2012 Waze

Figure 35: Weekday Afternoon Positive Valence Map and Word Cloud. Map points indicate reported GPS location of Waze reports with positive valence classification between 2 and 5 submitted between the times of 15:00 and 20:00, Monday-Friday. Larger points indicate classifications with stronger positive valence classification. Word cloud represents the same data set. Map data ©2012 Waze
Figure 36: Weekday Morning Negative Valence Map and Word Cloud. Map points indicate reported GPS location of Waze reports with negative valence classification between -2 and -5 submitted between the times of 05:00 and 10:00, Monday-Friday. Larger points indicate classifications with stronger negative valence classification. Word cloud represents the same data set. Map data ©2012 Waze

Figure 37: Weekday Afternoon Negative Valence Map and Word Cloud. Map points indicate reported GPS location of Waze reports with negative valence classification -2 and -5 submitted between the times of 15:00 and 20:00, Monday-Friday. Larger points indicate classifications with stronger negative valence classification. Word cloud represents the same data set. Map data ©2012 Waze

Roadify Data

Roadify contributed a data set of user comments collected via their mobile smartphone application for use in this study. As part of the work of this study, we built an automated data feed mechanism in cooperation with Roadify to collect anonymous user data aggregated daily by Roadify.

Roadify’s application invites travelers to submit short comments, each associated with a specific transit line; see screenshot in Figure 38. The size of the contributed Roadify data set, while large, was, at the time of data collection, just below the threshold for executing statistically meaningful domain training and sentiment analysis akin to that done for the contributed Waze data set. However, we are able
to present an initial evaluation of comment data from this transit-oriented mobile smartphone application.

Figure 38: Roadify Comment Data Entry Screen

We note that the Roadify user comments are topically centered around travel on specific transit service providers and the individual lines they operate. This data source is therefore highly filtered compared to comments harvested from general social microblogs.

Figure 39 and Figure 40 present aggregate word clouds of the full Waze and Roadify data sets, aside from the omission of common English words such as "the," "to," "and." These word clouds clearly show that user comments collected via each application are generally "on topic" in the context of the particular application, and, notably, distinctly different from one another.
We classified Roadify user comments with SentiStrength 2 using general, untrained lists; samples appear in Table 3.
Table 3: Classification of Sample Roadify Comments. For the sample shown, classification results are identical when run with either the general, untrained SentiStrength 2 lists or the Waze-trained lists.

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
<th>Comment Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-1</td>
<td>Good morning everyone 1 Train On Time and it's little chilly but nice, beautiful sunny day. Have a great day everyone.</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>I love this bus ride</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>I love the Q train!</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>Wow on time, am loving this app</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>Absolutely gorgeous dusk-nighttime passenger ride!!</td>
</tr>
<tr>
<td>3</td>
<td>-3</td>
<td>1 Train On Time great day nice and sunny - Weekend nightmare Note that won't be any trains from 242nd to 168th street only shuttle buses.</td>
</tr>
<tr>
<td>3</td>
<td>-2</td>
<td>Great service but WAY too expensive. Connect it with the MTA!</td>
</tr>
<tr>
<td>1</td>
<td>-4</td>
<td>I hate this f&amp;&amp;$: bus,</td>
</tr>
<tr>
<td>1</td>
<td>-4</td>
<td>Just got on the N on bay parkway. I hate this train.</td>
</tr>
</tbody>
</table>

For the particular samples shown in Table 3, classification results are identical when evaluated by SentiStrength 2 using the Waze-trained lists or the untrained lists. This is not the general case, however; using lists trained for one domain to classify texts from another domain can be expected to give poor classification results in many cases. These differences are accounted for by terms whose valence classifications differ between the Waze context and general use. One example is the word "stop" as seen in Table 4. In the context of automobile travel, "stop" (including the variation "stopped" as in "traffic is stopped") was rated with a valence of -3 when trained for the Waze domain. In the context of transit, the term "stop" is used frequently, but often in a sentiment-neutral way to indicate a routine action of a bus or train (when used as a verb) or the location of a bus or rail station (when used as a noun).

Table 4: Roadify Comments Misclassified When Evaluated with Word Lists Trained for the Waze Domain

<table>
<thead>
<tr>
<th>Negative Classification (untrained)</th>
<th>Negative Classification (Waze)</th>
<th>Comment Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-3</td>
<td>Love the 38L. Stops at about every fourth stop and saves 10-15 minutes easy every time I get on it versus the 38.</td>
</tr>
<tr>
<td>-1</td>
<td>-3</td>
<td>Does it stop at 59 st ?</td>
</tr>
<tr>
<td>-1</td>
<td>-3</td>
<td>Train is right on time at flushing ave stop</td>
</tr>
</tbody>
</table>

For the particular samples shown in Table 4, classification results are identical when evaluated by SentiStrength 2 using the Waze-trained lists or the untrained lists.
Discussion and Recommendations

The recommendations in this report are made on the basis of the data observed, which is very partial and incomplete. These tentative recommendations follow from empirical observations. These are not based on exhaustive surveys, and are thus to be taken with caution, for what they are: qualitative assessments based on the observed data.

Mobile Travel Applications as Pre-Filters for Travel-Related Texts

While general social media applications such as Twitter may generate large quantities of user-contributed text, they invite users to chat about any topic. This can present challenges for analysis when the area of study requires the data stream to be filtered for a specific domain of interest.

In the domain of transportation and commuting, mobile smartphone applications are now widely used by travelers to assist with routing and deliver reports about current condition of the transportation network. Microblog texts and user comments solicited via transportation-oriented mobile applications can provide a pre-filtered data set to the researcher or practitioner interested in mining opinion in the transportation domain.

While some additional filtering may further increase the quality of the data set, user comment data sets from transportation-specific mobile applications do not necessarily require first-order filtering to select for transportation-related data, as is generally required when using data streams from general-use social applications.

*We recommend the use of transportation-related smartphone applications as high-quality sources of classifiable texts in the commuting/transportation domain.*

Category Tags and Context as Mechanisms for Refining Domain

While the Waze and Roadify mobile applications solicit user comments within a domain of interest – automobile travel and transit travel, respectively – they also categorize or contextualize comments in ways that facilitate analysis.

As shown in Figure 4, Waze users must first categorize each new report as belonging to one of several categories. Word clouds from comments in each category (Figure 7 - Figure 11) show clear, finer distinctions of subtopics within the domain of automobile travel for each report category.
Similarly, the Roadify data set features contextual subcategorization by association of comments with specific transit lines; see Figure 38.

We recommend that transportation-related smartphone applications be designed to prompt users to assign each submitted comment to an appropriate category, and to associate comments with relevant contextual travel information. This requires little additional action on behalf of the user, and allows the researcher and practitioner to carry out more finely-grained analysis of submitted texts.

**Geotagging**

Most modern smartphones include geolocation functionality, so location data is generally available for inclusion with mobile travel application data. By associating mined sentiment and topics with specific locations, additional insight may be gained about travel in a specific region of interest. If combined with additional trajectory data points for commute trips, sentiment analysis may become applicable to specific routes, origins, and destinations.

We recommend that mobile application user comments be geotagged in the interest of more accurately associating them with specific locations and transportation network segments.

While geotagging may present privacy concerns, these concerns are not unique to this genre of application; see, for example, Hoh (2008). As this area is currently evolving, we recommend that mobile application engineers follow current best practices in this area.

**Utility of Sentiment Analysis**

This report has demonstrated the viability and usefulness of sentiment analysis in the domain of commute-specific microblogging of automobile travelers. Through bulk processing of thousands of comments, patterns of positive and negative valence have been shown to vary distinctly as functions of:

- Location
- Time of day
- Day of week
- User-selected category

Emotional valence training of a domain-specific emotion dictionary (and related supporting dictionaries) increased the statistical accuracy of valence classification of crowdsourced Waze traveler reports; see Appendix A for more information about
training for the data set used in this study, and discussion of Table 4 for an example of the importance and value of using appropriately-trained dictionary lists.

*We recommend that sentiment analysis be considered as a viable metric for evaluating travelers’ experience of their commutes.*

*We strongly recommend the training of emotion and other supporting dictionaries to improve the accuracy of classification of texts from any coherently-defined data set.*

**Limitations of Current User Commenting Rates**

Sentiment analysis is a "data hungry" application. Indeed, it is best-suited to evaluating large quantities of texts: even a set of several hundred microblog entries can be evaluated for a range of subtle sentiment in very sophisticated way in a short time by a human reader.

While the availability of texts from Waze users in the San Jose area has been seen to generally increase over the 26-month period of study, comments must be aggregated over a range of one or several parameters such as time, space, or report type in order to constitute a set large enough to be analyzed for sentiment. While current reporting rates in the San Jose area are shown in this report to allow for meaningful analysis, these rates are clearly on the threshold of utility for real-time, localized, category-specific evaluation.

In the interest of leveraging sentiment analysis as a useful tool in the domain of commuter travel, we recommend the promotion of the use of mobile travel applications as well as the encouragement of submission of comments by application users.

We note, as shown in Figure 3, that among San Jose area Waze users, comment rates as a fraction of total report rates appear to be decreasing slightly but steadily from late 2011 through the present; even though comment rates have increased through this time, they have been outpaced by the growth in uncommented reports. We propose several possible contributions to this trend. Waze "early adopters" may be interested in using more features of the application; latecomers may be more casual users, less interested in contributing reports. Also, awareness over distracted driving has been increasing; users may be increasingly less likely to attempt to behave in this risky behavior. *We recommend that application designers encourage users to contribute comments, while following current best practices and adhering strictly to all applicable laws and regulations regarding use of mobile devices.* New technologies that promise hands-free text communication may result in new possibilities for traveler reporting and commenting.
Explicitly Encouraging Positive Interactions Among Participatory Commuters

Commuting is generally associated with negative sentiment, as reflected in this and other studies. The Connected Commuting Task Force seeks to understand how to improve commuters’ daily travel experience. A simple – but, we think, profound – result of this study is that when offered the comment category of "Chit chat," participating travelers contributed texts with predominantly positive sentiment; see Figure 12 and Figure 13. This suggests that including general social interaction alongside traffic and travel commentary – encouraging commuters to simply "chat" with one another as part of their daily travel – may increase positive sentiment and improve the commute experience.

We recommend that participatory commuting apps encourage commuters to communicate socially in addition to offering mechanisms for sharing transportation network status reports, which are predominantly negative in sentiment.

Tuning Regional Commute Programs to Target Times and Topics Most Relevant to Commuter Sentiment

Comparing morning and afternoon weekday word clouds for positive sentiment, Figure 34 and Figure 35, "friday" emerges as a prevalent morning positive word, even among this weekday (Monday through Friday) aggregation of terms. It is quite notable that the terms "friday" and "happy" fade from morning to evening. We’re tempted to infer that when San Jose Waze comment contributors express positive sentiment, one of the most-commonly associated morning themes is "Friday;" this might bear further study. Regional programs that ask commuters to make an extra effort to travel differently (different routes, times, modes, or telecommute) might take advantage of this possible positive association with Friday mornings, while efforts designed to improve the commute experience may do well to focus on days and times other than Friday mornings.

Interestingly, "traffic," a little-used term in the morning positive sentiment word cloud, makes a significant appearance in the evening cloud. Similarly, the prevalence of the term "traffic" increases when going from morning to evening in the negative sentiment word clouds. These comparisons indicate that "traffic" may have a greater mindshare in the evening than in the morning. This suggests the possibility that regional commuter programs designed to ameliorate the experience of congestion may do better to focus particularly on the evening commute rather than the morning commute.

Comparing negative sentiment weekday word clouds, Figure 36 and Figure 37, suggests some variation in themes that commuters report in morning vs. evening comments. Notably, the words "construction" and "closed" feature prominently
among morning weekday negative sentiment terms, but are not dominant in the evening cloud. Meanwhile, "accident" appears in both clouds, but is much more significant in the evening lexicon. While these may simply indicate the prevalence of construction activity during the morning commute and a greater impact of accidents during the evening commute, it is worth noting this differentiation in subjects related to negative sentiment at different times of day. **Efforts to improve the experience of the commuter may do well to address specific subjects that relate to negative sentiment, as revealed from analysis of traveler comments; these factors can likely be associated with specific times and locations for additional refinement.**

**Routing for Positive Sentiment and an Improved Commute Experience**

Trip planning and routing applications often offer options such as shortest travel time, lowest fees (with respect to auto tolls and transit fares), and avoidance of highways. This study reveals that negative and positive sentiment can be localized to specific times and places. **Trip planning and routing applications may be able to look to sentiment analysis of traveler comments to be able to offer lower-stress routing options to commuters seeking a more pleasant commute experience.**

Similarly, geolocalization of reports in the Waze report category "Hazard" demonstrate the potential to crowd-source the identification of areas of the travel network that travelers perceive to be dangerous or unsafe. To complement this category, mobile travel application engineers may consider introducing positive categories such as "Nice road," "Scenic train route," or "Pleasant drive." **We recommend that application designers solicit reports and comments directly related to commuters’ experience of the network, and that this data be used to improve traveler routing.**

**Crowdsourced Traveler Data as Travel Network Condition Reporting**

As shown in this report, word clouds and word count statistics can be used to mine the lexicons used under each Waze rubric to identify the environmental sources of strong sentiment related to each category. This data set also offers the opportunity to compare reported perceived hazards ("Hazard" reports) with external data such as safety statistics and accident rates. Similarly, "Traffic jam" reporting, and its strong negative sentiment (Figure 13), could be compared to incident history to determine the extent to which negative "jam" reporting may relate to actual delay versus perceived unusual delay. **Mobile application data should be considered as a resource to inform network planners and operators – possibly in near real time – about aspects of the network that may require maintenance or modification to perform optimally, and as an additional metric to gauge the impact of infrastructure changes.** This offers the promise of traveler feedback
which, although less detailed, is less costly and much more timely than traditional surveys.
Conclusions

We demonstrate that traveler comment data collected from mobile smartphone applications is well-suited to text sentiment analysis. Such data sets offer texts with highly-domain-specific themes that make them very high-value compared to general social media sources. Training a sentiment analysis algorithm for sensitivity to word usage within a given domain optimizes classification results for texts in that domain.

Classified sentiment is shown in this report to have strong and distinct association with day of week, time of day, type of report, geographic location, and lexicon. The greater the reporting and commenting rates, the greater the possibilities for deeper insight into the commute experience. Contemporary changes in commenting and reporting rates are changing, and may currently be influenced by a number of competing factors.

Classification of commuter comments, particularly when combined with user-selected report categories, travel contexts, and geolocation data, offer many possibilities for improving the urban commute experience by informing a variety of technologies, products, services, and innovations. The authors look forward to the evolution of these technologies in the interest of improving the commute experience in cities worldwide.
References


Appendix A: SentiStrength 2 Training and Use

This Appendix is not meant to stand alone or to serve as a user manual of SentiStrength 2. Rather, the purpose of this Appendix is to document the training and use of SentiStrength 2 so that another SentiStrength 2 user could repeat similar training and analysis.

Complete information about SentiStrength 2 can be found on the pages and linked resources of this Web site:

http://sentistrength.wlv.ac.uk/

For this study, SentiStrength 2 classifications were evaluated in the default mode of separate positive and negative classifications for each sample text.

Optimization of Term Strengths

To begin the training process, we selected 1,000 Waze user comments at random from the 15,131 used for this study and coded them for positive and negative sentiment as a reference sample for training. See Table 5 for sample Waze comments and their human-coded valence values.

Table 5: Waze Training. This table shows a sample of the 1,000 randomly-selected user comments and human-assigned valence ratings. This human-classified subset of Waze comments was used as the reference for subsequent calibrations.

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-2</td>
<td>train is stopped.....beautiful cold morning!</td>
</tr>
<tr>
<td>3</td>
<td>-2</td>
<td>AAA towing car</td>
</tr>
<tr>
<td>1</td>
<td>-3</td>
<td>Freeway closed tonight. 11:30 pm 4;30 am</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>it's not raining. woot!</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>right shoulder</td>
</tr>
<tr>
<td>1</td>
<td>-3</td>
<td>big bag in center of intersection</td>
</tr>
<tr>
<td>1</td>
<td>-3</td>
<td>On ramp to 237 stopped</td>
</tr>
<tr>
<td>1</td>
<td>-4</td>
<td>truck hit cop car on white and quimby lots of cops</td>
</tr>
<tr>
<td>2</td>
<td>-5</td>
<td>multi car accident on shoulder now</td>
</tr>
<tr>
<td>1</td>
<td>-3</td>
<td>ramp from 85s to 87 closed</td>
</tr>
<tr>
<td>1</td>
<td>-4</td>
<td>Happy Monday :-(</td>
</tr>
<tr>
<td>1</td>
<td>-3</td>
<td>all lanes stopped</td>
</tr>
</tbody>
</table>

A new, optimized emotion lookup table (EmotionLookupTable.txt) was created by using the "optimise"[sic] action; command shown in Table 6.
Table 6: SentiStrength 2 Term Weight Optimization Command.

<table>
<thead>
<tr>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ java -jar SentiStrength.jar sentidata dictionaries/ input training-1000.txt optimise dictionaries/EmotionLookupTable.txt</td>
</tr>
<tr>
<td>Number of texts in corpus: 1000</td>
</tr>
<tr>
<td>Saved optimised term weights to dictionaries/EmotionLookupTable.txt</td>
</tr>
</tbody>
</table>

The emotional valence scores of twenty-four terms from the 2,546-term standard, original emotion lookup table were adjusted by the SentiStrength 2 engine to more accurately reflect the valence of these words as used and rated in the human-scored random sample of 1,000 Waze comments. A sample of several corrected terms is shown in Table 7.

Table 7: Examples of EmotionLookupTable.txt Valence Training Results.

<table>
<thead>
<tr>
<th>Word</th>
<th>Original Valence</th>
<th>Corrected Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>overturn</td>
<td>-2</td>
<td>-4</td>
</tr>
<tr>
<td>accident</td>
<td>-2</td>
<td>-3</td>
</tr>
<tr>
<td>love</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>tire</td>
<td>-2</td>
<td>-3</td>
</tr>
</tbody>
</table>

Cross-Validation of Initial Optimized Emotion Lookup Table

As a reference for evaluation of final training, we ran a ten-fold cross-validation with this corrected emotion lookup table as shown in Table 8.

Table 8: Ten-Fold Cross-Validation Command. Initial run, after first optimization of emotion lookup table.

<table>
<thead>
<tr>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ java -jar SentiStrength.jar sentidata Dictionaries/ input training-1000.txt iterations 30</td>
</tr>
<tr>
<td>Number of texts in corpus: 1000</td>
</tr>
<tr>
<td>Before training, positive Before training, positive: 630 63.0% negative 602</td>
</tr>
<tr>
<td>60.2% Positive corr: 0.49656942 negative 0.42315385 out of 1000</td>
</tr>
<tr>
<td>Running 30 iteration(s) for standard or selected options on file training-1000.txt; results in training-10000_out.txt</td>
</tr>
<tr>
<td>Set of 30 10-fold cross validations finished</td>
</tr>
<tr>
<td>Finished! Results in: training-10000_out.txt</td>
</tr>
</tbody>
</table>

These initial validation results are compared with final validation results later in this Appendix.

Report of Term Weights for Bad Classifications

We ran the termWeights action of SentiStrength to discover terms from the Waze data set that should be added to the emotion lookup table; see command in Table 9.
Using the documented SentiStrength 2 guidelines for addition of terms to the emotion lookup table, 104 new terms from the Waze data set were added; see sample added terms in Table 10.

Table 10: Partial List of Waze Terms Added to Emotion Lookup Table.

<table>
<thead>
<tr>
<th>Sample Added Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>cheers</td>
</tr>
<tr>
<td>clear</td>
</tr>
<tr>
<td>closed</td>
</tr>
<tr>
<td>construction</td>
</tr>
<tr>
<td>crash</td>
</tr>
<tr>
<td>flipped</td>
</tr>
<tr>
<td>heavy</td>
</tr>
<tr>
<td>meter</td>
</tr>
<tr>
<td>moving</td>
</tr>
<tr>
<td>pulled</td>
</tr>
<tr>
<td>radar</td>
</tr>
<tr>
<td>standstill</td>
</tr>
<tr>
<td>stop</td>
</tr>
<tr>
<td>stuck</td>
</tr>
<tr>
<td>traffic</td>
</tr>
<tr>
<td>undercover</td>
</tr>
<tr>
<td>upsidedown</td>
</tr>
<tr>
<td>weekend</td>
</tr>
<tr>
<td>woot</td>
</tr>
<tr>
<td>work</td>
</tr>
</tbody>
</table>

The termWeights output also revealed twenty-nine slang terms that could be added to the slang lookup table; see examples in Table 11.
Table 11: Sample of Waze Slang Term Mappings Added to Slang Lookup Table.

<table>
<thead>
<tr>
<th>Slang Term</th>
<th>Mapped Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>beeyotch</td>
<td>bitch</td>
</tr>
<tr>
<td>chp</td>
<td>police</td>
</tr>
<tr>
<td>fml</td>
<td>fuck my life</td>
</tr>
<tr>
<td>hola</td>
<td>hello</td>
</tr>
<tr>
<td>howdy</td>
<td>hello</td>
</tr>
<tr>
<td>hwy</td>
<td>highway</td>
</tr>
<tr>
<td>popo</td>
<td>police</td>
</tr>
<tr>
<td>tgif</td>
<td>thank god it's Friday</td>
</tr>
</tbody>
</table>

After adding the new, unranked terms to the emotion lookup table and the new slang mappings to the slang lookup table, the optimization procedure (Table 6) was run on the expanded tables to assign valence weights to the newly-introduced terms, further training SentiStrength 2 for the Waze domain.

Cross-Validation with Fully-Trained Term Lists

To evaluate our training of SentiStrength 2 for the Waze domain, we ran a final ten-fold cross-validation, as shown in Table 12.

Table 12: Ten-Fold Cross-Validation Command, Fully-Trained Term Lists.

```
$ java -jar SentiStrength.jar sentidata dictionaries/ input training-1000.txt iterations 30
Number of texts in corpus: 1000
Before training, positive: 805 80.5% negative 698 69.8% Positive corr: 0.79624474 negative 0.66066116 out of 1000
Running 30 iteration(s) for standard or selected options on file training-1000.txt; results in training-10000_out.txt
Set of 30 10-fold cross validations finished
Finished! Results in: training-10000_out.txt
```

These final validation results are compared with initial validation results in Table 13.

Table 13: Comparison of Initial and Final Ten-Fold Optimizations. The reported values are correlations between the sentiment predictions and the human-coded values of the 1,000-comment training set. "Within 1 Point" accounts for exact matches and for near-matches that differ by +1 or -1 valence rating.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Match Exactly</td>
<td>Match Within 1 Point</td>
</tr>
<tr>
<td>After Initial Term Strength Optimization</td>
<td>63.0%</td>
<td>74.1%</td>
</tr>
<tr>
<td>After Full Training</td>
<td>80.5%</td>
<td>91.7%</td>
</tr>
</tbody>
</table>
**Scoring of the Waze Corpus**

We scored the full 15,131-text corpus of the Waze data set used for this study with the fully-trained SentiStrength 2 term lists. The resulting positive and negative classifications of the texts of this corpus are the classifications presented and analyzed in this report.