

Sarram, Golnaz; Ivey, Stephanie S.

Article

Evaluating the potential of online review data for augmenting traditional transportation planning performance management

Journal of Urban Management

Provided in Cooperation with:

Chinese Association of Urban Management (CAUM), Taipei

Suggested Citation: Sarram, Golnaz; Ivey, Stephanie S. (2022) : Evaluating the potential of online review data for augmenting traditional transportation planning performance management, Journal of Urban Management, ISSN 2226-5856, Elsevier, Amsterdam, Vol. 11, Iss. 1, pp. 123-136, <https://doi.org/10.1016/j.jum.2022.01.001>

This Version is available at:

<https://hdl.handle.net/10419/271450>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

HOSTED BY



ELSEVIER

Contents lists available at ScienceDirect

Journal of Urban Management

journal homepage: www.elsevier.com/locate/jum

Research Article

Evaluating the potential of online review data for augmenting traditional transportation planning performance management

Golnaz Sarram^{*}, Stephanie S. Ivey

Intermodal Freight Transportation Institute (IFTI), Herff College of Engineering, University of Memphis, P.O. Box, Memphis, TN, 38152, USA

ARTICLE INFO

Keywords:

Planning
Social media
Sentiment analysis
Project decision-making
Performance management

ABSTRACT

Synchrony of transportation planning with revealed public preferences and expectations has traditionally been difficult to achieve. In other domains, such as manufacturing and design, frameworks have been developed using data-driven analytic processes to extract customer preferences and expectations to inform product development. This study builds from approaches to understanding customer satisfaction in other domains to investigate the potential of social media review mining for bridging the gap between understanding of neighborhood livability and conventional transportation planning decision-making practices. The novelty of the approach is in interpreting an automated user-defined translation of qualitative measures of livability by evaluating users' satisfaction of the neighborhoods through social media and enhancing the traditional approaches to defining livability planning measures (through a sentiment analysis and visualization package). The provided data-driven neighborhood satisfaction package draws the marketing structure of transportation network elements (analogous to the 'products' in other domains) with residential nodes as the center of the structure. In this study, the FastText algorithm is tested to employ a social review analysis, common neighborhood aspects are annotated, and neighborhood satisfaction is visualized and compared to the associated demographic side information. Through extracting the important aspects of the public concerns, advanced visualization helps transportation planners better understand the implications for community livability and provides insight into an approach for conducting more equitable public outreach. This approach has the potential to capitalize on residents' interests in social media outlets and to increase public engagement in the planning process by encouraging users to participate in online neighborhood satisfaction reporting. This study demonstrates a proof of concept for a social review analysis and specifies the methodology for agencies to augment planning performance management. (word count: 273).

1. Introduction

While transportation planning processes are frequently thought of as being important for helping communities meet the future demand for moving people and goods (such as through input analyses, defining policies and goals, and guiding investments and performance measurement), creating a sustainable transportation system to address the economic, environmental, and social equity issues within a community is also critical. To do this, a comprehensive planning framework is required that includes the study and evaluation

^{*} Corresponding author. Intermodal Freight Transportation Institute (IFTI), Herff College of Engineering, University of Memphis, Memphis, TN, 38152, USA.

E-mail addresses: gsarram@memphis.edu (G. Sarram), ssalyers@memphis.edu (S.S. Ivey).

<https://doi.org/10.1016/j.jum.2022.01.001>

Received 10 November 2020; Received in revised form 12 October 2021; Accepted 13 January 2022

Available online 5 February 2022

2226-5856/© 2022 The Authors. Published by Elsevier B.V. on behalf of Zhejiang University and Chinese Association of Urban Management. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of user satisfaction. Developing a framework that accurately represents revealed versus stated preferences of users, including behavioral and attitudinal data that have been traditionally undervalued in the data collection and modeling processes, is challenging and requires research and development of new planning methodologies and tools. Ultimately, a data-driven transportation planning approach that provides more accurate and updated (real-time) information for transportation planning can improve the decision making process and lead to more livable communities where needs of diverse residents are considered.

Planning agencies have struggled with achieving public involvement goals to appropriately inform decision making processes. The topic of livability creates even greater difficulty in this regard in that perceptions of livability are abstract and difficult to quantitatively assess. Understanding patterns of public perceptions related to livability can provide richer information for planning decisions. By first understanding public quality of life perceptions and expectations, we can design performance measures and project choices for effective and efficient investments. Incorporating state-of-the-art performance measures requires defining measurable livability metrics reflecting stakeholders' needs, wants, and behaviors. Much of the work surrounding livability to date has focused on residential surveys and traditional measures of congestion or other attributes to indicate livability. The connections between livability perceptions and consumer preference theory/demand modeling have only recently entered transportation literature (more discussed in Sarram & Ivey, 2018). Integrating such approaches from other domains in transportation settings, particularly in the case of livability studies, can provide greater insight and can lead to more informed decision-making.

One popular approach in consumer product demand modeling for guiding product design decisions is Discrete Choice Analysis (DCA), adopted from market research and transportation planning (Ben-Akiva & Lerman, 1985). Through establishing the relationships between key customer attributes and engineering attributes (a function of the design options), richer information is integrated into the decision-making process.¹ The manufacturing and the design domains have also developed frameworks of data-driven analytic processes for decision makers to integrate traditional data sources with subjective and objective impact factors (Chien et al., 2016, Lin & Chien CF, 2016).

Meanwhile, the artificial intelligence (AI) domain has provided data mining technique for extracting valuable customer experience and preference information from large market and product data. The approach assists the quantification of some parameters in design models. Tucker and Kim (2008 and 2011) first implemented data mining techniques merged with multilevel optimization to predict the maximum company profit for a particular product architecture, and later, predicted the trend of customer preferences for generating a demand model that reflects changes in the preferences. In the case of mass customization, the technique has been used to balance the heterogeneity of customer preferences and the complexity of products. Clustering analysis and association rule mining have been applied widely to learn the association between customer features and product requirements (Jiao & Zhang, 2005; Morency et al., 2007).

Investigators have extended the study to the domain of product configuration (Moon et al., 2010; Shao et al., 2006; Song & Kusiak, 2009). There are few available data-driven studies associated with spatial configurations (Engin et al., 2019; Long, 2015), however, the synchrony of transportation planning products with public preferences and expectations in this context has been overlooked. Due to the competitiveness of the market, the manufacturing domain for new products has developed beyond use of limited survey information and now includes input from technology and retailer websites (e.g. Amazon, Google and Yelp) through applying text mining to online reviews. Li et al., 2015, performed association rule mining to analyze customer opinions in online reviews and to investigate the product attributes that are desired by specific user types. Zhang, Narayanan, & Choudhary, 2010, presented a feature-based product ranking technique instead of only using the existing automated ranking mechanisms for overall product quality. They identified features within a product category and analyzed their frequencies based on the subjective and comparative sentences in reviews and modeled the relationships between the products. These approaches have not yet been applied in the urban planning domain for informing neighborhood planning decisions. However, there is significant potential to improve public involvement outcomes through attracting residents to provide data in a different and more efficient way that reflects individual choice preferences.

Therefore, this study proposes an approach for planning practices through interpreting a quantitative translation of qualitative measures of livability, defining a marketing structure of transportation network products (neighborhoods) and evaluating users' satisfaction of these products to replace traditional livability and other subjective measures. The approach is focused on translation of consumer preference theory to the transportation planning domain and because the heterogeneity of perceptions between livability performance measures and satisfaction have been overlooked, this study will initiate a framework exploring the subjective livability perceptions of residents that can address the diversity of satisfaction. Therefore, the current study focuses on evaluating the potential for using text mining (discussed more as sentiment analysis in section 2.3) and existing online platforms for informing transportation planning practice.

2. Literature review and research plan

2.1. Approaches for understanding quality of life

As discussed in Sarram & Ivey, 2018, there is limited academic literature in urban planning on developing rich understanding of

¹ The application of DCA for demand estimation is known for addressing unobserved taste variations, unobserved attributes, and modeling deficiencies. The approach considers the preference structure of customers through identifying the key customer attributes for the Kano (a widely-used consumer preference theory model) selected customer utility function of the DCA demand analysis (Wassenaar et al., 2005; Michalek, Feinberg, & Papalambros, 2004).

livability needs and community quality of life. The current studies mostly represent aggregated opportunity proxies or hierarchical classification processes. These methods are limited by the need for direct public involvement for quantifying residential satisfaction and identifying the potential alternatives for improving communities based on users' preferences and differentiated levels of satisfaction.

From an urban economic standpoint, Albouy & Lue, 2015, estimated a behavioral model to measure willingness-to-pay to access local amenities based on combined costs, rent, wage, and commuting differentials for inferring changes in quality of life. Household preference is defined as a utility function based on aggregated amenities in a quality of life index. The model assumes each mobile and informed household chooses place-of-residence and place-of-work combinations (j,k) with similar satisfaction. Therefore, with homogeneous households, all observed combinations (j,k) must drive the same level of utility. The overall study results show that neighborhood quality within metro areas varies substantially due to the artificial amenities created through use of a homogenous approach. Artificial amenities arise from a homogenous approach to assessing 'willingness to pay' for a set of determinants, where restricted feasible options result in the appearance of choice (for example because of higher rent, a person lives further away from desired amenities, making it artificially appear they are willing to accept a longer commute to live in an outlying area). The limitation inherent in this approach may be addressed through considering customer satisfaction patterns in the case of transportation policy considerations, where more information is required about individual preferences.

Similarly, Couture & Handbury, 2017, conducted a within-city analysis on understanding urban revival using 10 years of data from the Consumer Expenditure Survey (CEX) and the National Household Travel Survey (NHTS). The applied location choice model used residential-only data due to the limitation on the age and education group for the workplace location data. The authors discussed the importance of various explanations for the location decisions of the young and college-educated. The results show that varied preferences of these groups for non-tradable services (restaurants, bars, gyms, and personal services) in addition to their travel and expenditure shares can explain the diverging location decisions. The decisions depend on the presence of other young college graduates due to the higher demand of networking and socializing. Considering these results, it can be inferred that the nature of satisfaction is interconnected with the neighborhood quality. This means that the demand for social activities, depending on the individual priorities and preferences, affects residential location choices. As people's choices regarding residential locations are made by best satisfying their priorities rather than finding a perfect place to live, it is necessary to invest in and plan neighborhoods efficiently and thoughtfully through deep understanding of related factors.

In 2012, the Nashville Metropolitan Planning Organization (MPO) conducted a regional household travel survey that included a health component (Nashville Area, 2013). The travel behavior and general health of 6000 households were surveyed including a subset of 1000 households recruited into a GPS and health component of the study. The study compared the summary statistics in a fine geographic resolution, used health status data to identify health priority areas instead of using common demographic approaches alone, and modeled the potential health impacts of active transportation in the region. The results illustrated the relationship between transportation, physical activity and overall health. The study goal was to gather baseline data for the MPO to quantify the effect of health and safety criteria in the evaluation and ranking of future transportation projects on the overall health of Nashville residents. The study results have been used in policy approaches for enhancing the livability of Tennessee communities. Later, Meehan & Whitfeld, 2017, described the multifaceted approach and projects that the Nashville MPO has undertaken to improve the transportation system, quality of life and health through using the aforementioned dataset. The study discussed that the MPO has impacted public and environmental health through the objective of providing mobility for workers and residents. The model results were subject to several limitations such as over-valued benefits in behavioral change and prevention of chronic diseases. However, the study efforts were focused mostly on the transportation effects on public health which overlooks other lifestyle factors (e.g. work stress). While the study is a demonstration of the interplay between the measure of public health and providing mobility for increasing community livability, it only concentrated on analyzing one component of livability (health). A more comprehensive correlation between residential travel behavior and quality of life components is neglected here.

2.2. Assessing a customer centric framework

A primary understanding of the association between the mobility impedances and users' stated satisfaction along with correlation to neighborhood attributes can identify the variables that are more important in order to directly measure neighborhood livability. Therefore, in planning practices, there is a need for a joint metric that considers the trade-off between community livability and network mobility. Caltrans has developed a mobility planning guide² to structure a set of network mobility performance indicators around six smart mobility issues of location efficiency, reliable mobility, health and safety, environmental stewardship, social equity, and robust economy. This guide served as a basis for the framework developed for the current research. For a long time, user satisfaction surveys in transportation planning have considered satisfaction intervals and cardinal scale applications; however, the complexity of users' innate perceptions has been disregarded.

In the marketing domain, Mittal & Kamakura, 2001, presented a conceptual model capturing the relationship among satisfaction ratings, latent satisfaction, repurchase behavior and consumer characteristics. The results showed the relation is highly non-linear and also presents that consumers at the same level of rated satisfaction have different repurchase rates due to the different characteristics they have. Attribute-performance management is also recommended in order to optimize overall satisfaction. The approach maximizes performance based on attributes that have the largest weight in determining overall satisfaction and considers each customer segment individually.

² Smart Mobility 2010

In transit planning, Kim & Chung, 2016, applied the GMM approach to public transportation data to address the heterogeneity of the human complex cognitive system and to evaluate the Likert scale accuracy on this matter. The study presented new implications in analyzing public transportation satisfaction surveys.

2.3. A new approach enabled by social media

The increasing digital social connectiveness of our society provides an opportunity to connect these social ties on the geospatial level and improve consideration of human dynamics on quality of life. The analysis behind such an approach has to consider not only the decision sets of transportation planning but also user preferences. An approach is needed that is able to: 1) reveal the users' preferences, from online review information to direct inquiries of their requirements, 2) interpret them as a complex, through comparing the results with similar residents' satisfactions and 3) integrate the performance of the transportation network surrounding the neighborhood complexes. This type of information about how users make decisions and the ultimate criteria considered in decision-making is very important for planning agencies to deeply understand livability considerations through a customer-centered approach.

Machine learning models maximize predictive accuracy on unseen data, rather than observed data which is used for training. They are mostly non-linear, have different bias-variance tradeoffs and regularization mechanisms capable of controlling model complexity (Kleinberg, Mullainathan, & Raghavan, 2016; Paredes et al., 2017, pp. 780–785). Covariates may also be simultaneously selected with model derivation. These techniques outperform traditional econometric models on prediction, however, this comes at the expense of interpretability. Since discrete choice models place more emphasis on using features that are easily mapped to human behavior and well understood causal relationships and correlations, this study aims to exploit the power of pattern recognition models in prediction accuracy and appraising the potential of interpreting the causal relationship among the livability features through an interactive system. As a breakthrough, users are not required to fill out surveys for manifesting their preferences. Although, the result could also be adopted for use as a pre-survey method to determine expectations, gaps, and services that need to be targeted in actual survey and action planning.

While there is a large body of work related to preferences of residents extracted based on their activities, this is the first to develop a review analysis package based on the users stated preferences in the transportation planning domain. This framework provides an efficient online text classification and visualization approach for automating subjective performance measures used in planning. Developing this data-driven neighborhood satisfaction package draws the marketing structure of transportation network products with residential nodes as the center of the structure.

Hu, Deng, & Zhou, 2019, conducted a semantic and sentiment analysis on online neighborhood reviews. They compared latent Dirichlet allocation (LDA) and multi-grain LDA models to identify the semantic topics (aspect) in Niche neighborhood reviews dataset. The authors extracted the probabilistic distribution of topics, for instance safety, and modeled these as a probabilistic distribution of words, such as safe, crime, and police. In order to quantify the sentiments of the identified aspects, they proposed to leverage an aspect rating analysis (LARA) model to derive aspect-specific ratings from online neighborhood reviews. The purpose has been resolving the requirement for deriving sentiment lexicons from extreme opinions in a certain domain. At the end, they correlated the subjective perceptions extracted from the dataset and the objective socioeconomic attributes of NYC neighborhoods. However, the models have their own assumptions and limitations of LDA ignoring the order of words in texts, and LARA assuming that the overall rating of a reviewer follows a Gaussian distribution with its mean as the weighted sum of the aspect-specific ratings. The correlation results also did not sufficiently reflect the study's objective measures.

Meanwhile, estimating neighborhood satisfaction of individual users based on their reviews on Twitter platform has been the focus of limited research works: Saeidi et al., 2017 investigated whether sentiment analysis can be used to predict demographic attributes for neighborhoods of London. The correlation results between the two datasets, Yahoo! Answers and Twitter, shows differences in the nature of information. The former contained more encyclopedic information, while the later provided current sociocultural aspects. In another multimedia study, Flaes et al., 2016 analyzed both visual and text content posted on Twitter and Flickr (a multimodal content analysis on multimedia data). The result from microblogging platform Twitter (64 thousand geolocated tweets within a 10-mile radius of the city center of Amsterdam) appeared more suitable for sentiment analysis than the content-sharing website Flickr (64 thousand images taken in and around the city of Amsterdam). The sentiment scores detected from Twitter show significant relationships with the demographics of inhabitants, ethnic background, income and education level. For both platforms the economic indicators (i.e. people living on welfare checks and income) show significant relationship with the sentiment scores, however no relationship with the livability or social index of a neighborhood. The authors however, have not yet identified the adjustment factors to the sentiment scores that are important to the residents of a city. Twitter microblogs have also been used to discover the correlation (0.35) between collective sentiments of tweets and the deprivation index of the involved communities (Quercia et al., 2012).

Gibbons et al., 2019 evaluated Twitter sentiment analysis to predict self-rated local health data from the American Community Survey. They computed the hedonometer grand mean score for each census tract based on the health sentiment lexicons from previous health studies and used a multivariate generalized linear model to identify how neighborhood attributes like aggregated sentiment affect population-level screening behaviors. The major deficiency with lexicon-based sentiment analysis metrics such as Hedonometer is the inability to properly take context into account (Pang & Lee, 2008). The study results shows some limitations in profanity usage (as an indicator of mood) and local inconsistencies due to the aggregation in a smaller geographic scale.

Zivanovic et al., 2018, collected four years of tweets in the area of Bristol between January 2012 to September 2016. The study explored the use of Twitter data in quality-of-life research and the approach was a semi-manual content analysis using computer-assisted qualitative data analysis and GIS software. The study concluded that Twitter data may be useful to indicate the emergence of concerns not identified by traditional quality of life surveys and complement them. On the other hand, the study did not find the Twitter data

representative enough and discussed limitations with using it (e.g. migration and demographic bias) that may render invisible certain segments of the population.

However despite these limitations, the current study will evaluate the use of Twitter datasets rather than Niche, Nextdoor, Yahoo! Answers and other sources of neighborhood reviews due to the following reasons: first, abundant available sources of training datasets that are labeled and have shown reasonable performances in state-of-the-art NLP research, and second, open access to the trending real-time information. These features make Twitter data more accessible and potentially more valuable for use in transportation planning applications for public agencies.

There are numerous reasons for using Twitter as a data source for research purposes. Factors making Twitter appealing for research include its global popularity (and large user set), ease of with which data can be searched for key topics, and the ability to further refine searches through the use of hashtags (Ahmed et al., 2017). Additionally, the Twitter API provides open access for researchers and Twitter is particularly well suited for sentiment analysis due to the consistency of post lengths (constrained to 140 characters).

Thus, the current research implements insights from three disciplines, transportation planning, marketing and Natural Language Processing (NLP), to design a data-driven and efficiency-based planning sentiment analysis (SA) package to present how neighborhood satisfaction can estimate residential quality-of-life, and identify users' preferences and priorities.

3. Methodology

The focus of this research is to:

- Evaluate the potential of Twitter datasets for providing insight useful for transportation planning related to livability assessment in the US and
- Develop a review analysis package based on users' stated preferences in the transportation planning domain.

The intent is for this framework to provide an efficient online text classification and visualization approach for automating extraction of user-defined performance measures that can be used in planning decision-making. Developing this data-driven neighborhood satisfaction package draws the marketing structure of transportation network products with residential nodes as the center of the structure.

Recent NLP works, including the winners of SemEval Aspect-Based Sentiment Analysis (ABSA)³ tasks (Baziotis et al., 2017; Cliché, 2017), have been using deep learning (DL) methods.⁴ For practicality in transportation planning applications, the current study evaluates a fast and efficient language processing approach (including the dataset sources and sentiment analysis approach).⁵

To this aim, neighborhood reviews from the Twitter platform in Memphis, San Francisco and New York City (NYC) were appraised to determine if the social media outlet can be used to indicate public satisfaction of neighborhoods and determine the key user-defined performance measures related to this subjective metric. These three case studies were selected to provide a range of city characteristics in terms of population size, geographic scale, and community type. The methodology for this study included:

- Extraction of neighborhood-relevant data from Twitter for three cities in the US over a six month period
- Testing of a more efficient language processing approach (using fastText) for the planning domain
- Applying the fastText model to develop sentiment profiles for neighborhoods included in case study cities
- Developing sentiment maps to 1) discover existence of possible satisfaction patterns in each city, 2) compare the public expectation and online expression cultures in these three case studies, and 3) examine relationships to Census tract demographic profiles.
- Evaluating the potential of Twitter data for informing transportation planning decisions related to livability.

3.1. Data collection

For the current study, the Twitter API platform was used to extract data for three cities to examine its potential for use in neighborhood satisfaction analysis by transportation planners. The approach to extracting tweets and the selected training dataset are described in the following sections.

3.2. Neighborhood tweets

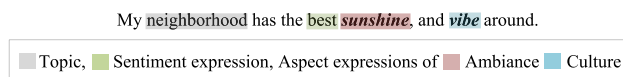
Topic related reviews usually contain at least one of multiple identified possible terms. Separating each search term with spaces and

³ ABSA refers to determining the opinions expressed on different features of entities, e.g., of a transit service, or a neighborhood (Hu & Liu, 2004). An aspect is an attribute of an entity, e.g. the timing of a metro, or the safety of a neighborhood. This level of analysis incurs an additional annotation task to achieve a domain-related training dataset that is also labeled regarding polarity of the aspect (Cataldi et al., 2013).

⁴ Such as Concurrent Neural Network (CNN), and Recurrent Neural Network (RNN).

⁵ The problems with using ABSA-DL methods are first, the ABSA methods needs the availability of domain specific human annotated training dataset, including the identified topics of the text (called the Aspect) and second, the DL algorithms can be very expensive time-wise. Despite having great performances in laboratory practices, deep learning models are very slow to train and test on real-time and large scale datasets.

“OR” using the search_tweet function in R allows extraction of relevant data. For our case, the word “neighborhood” was used to filter and select potentially relevant tweets. The example below represents the determinant elements (topic, sentiment, and aspect) of a review in sentiment analysis:



Text from tweets with neighborhood topics, such as those expressing aspects of a neighborhood and the users' related sentiment (i.e. ‘cars speed through my neighborhood – it is not safe to walk’ or ‘I love walking to the park in my neighborhood’) were collected during a roughly 6 month period, between May 2019 to November 2019, during this study to obtain representative data samples for examining a proof of concept. Moreover, according to the Pew Research statistics of adults using the Twitter services (24% of internet users or 21% of U.S. adults⁶), the data was considered to be relatively representative in each census tract. As collected Tweets span a 10 day time period, the inquiries were continuously submitted every 10 days.⁷

The goal of this data collection activity was to extract a variety of neighborhood characteristics following a word choice strategy. Therefore, the neighborhood Tweets were queried from Twitter public API⁸ including the words neighborhood AND ‘my’ OR ‘me’ as the person pronouns. This experimental approach of extracting the data was identified in order to reduce the number of irrelevant Tweets and real estate advertisements. Entities in our text datasets are locations and aspects are the topics of Tweets. The TwitterR package was used for the open source software R (Schweitzer, 2014).

3.3. Training dataset

According to Bojanowski et al., 2017, fastText exploits character-level similarities between words, therefore the model is supposed to have a good performance modeling infrequent words or even using restricted size datasets. However, in order to reach an accurate result the current study tested fastText with different training datasets including: 1) SemEval-2017 Task 4⁹ (subtask A-sentiment analysis in Twitter) accumulated datasets, which was developed from around 60,000 tweets collected since 2013. This task also provides topic-based labeled datasets (subtask B–C) but the topics were not applicable for the current research; and 2) Sentiment 140¹⁰, developed from a collection of tweets using keyword searches by Stanford University (Go et al., 2009). Among tested datasets for the current study, fastText models were able to provide accurate results only using the large training dataset of Sentiment 140.¹¹

3.4. Shallow neural network

This study implemented fastText, an NLP library for learning of word embeddings and text classification created by Facebook AI Research (FAIR) lab.¹² This open-source library also works on generic standard hardware and is able to fit on portable devices. It requires remarkably lower training time and has better scalability than deep learning models.¹³

3.5. Preprocessing, reverse geocoding and spatial aggregation

Sentiment 140 contains 1,600,000 tweets in three categories of Negative, Neutral and Positive. However, fastText is able to only open around 1,048,576 in csv format. Due to its binary prediction design regarding the sentiment categories, Neutral groups were removed from the Sentiment 140 dataset for the current study (discussed more in the result and discussion section). Approximately 10%

⁶ <https://www.pewresearch.org/internet/2016/11/11/social-media-update-2016/>.

⁷ The Twitter API platform offers three tiers of search APIs: standard, premium, and enterprise, which provide free and paid access to either the last 30 days of the Tweets or access to Tweets as early as 2006 with full data fidelity. However, the private company shuts searches down in cases of detecting individuals running the queries and overloading the system. According to the Twitter guidance, the analysts and developers are allowed to accommodate different searches of 1500 tweets and Twitter does not archive the searches for more than three to seven days' query on the API, depending on their numbers (Twitter, 2018).

⁸ <https://developer.twitter.com/en/use-cases/academic-researchers>.

⁹ <http://alt.qcri.org/semeval2017/task4/>.

¹⁰ <http://help.sentiment140.com/for-students>.

¹¹ The Sentiment 140 dataset has been used in other NLP research and was created automatically as opposed to manually annotated tweets (Rosenthal et al., 2017). The approach assumed that any tweet with positive emoticons were positive and vice versa.

¹² The softmax layer in deep learning processes is replaced with a hierarchical softmax and each node represents a label. Therefore, the model only predicts the probabilities for a limited number of labels and parameters which reduces the training time. FastText also considers the imbalanced nature of the classes through using Huffman algorithm to represent categories by building the category tree. The depth in the tree is smaller and leads to further computational efficiency without sacrificing accuracy (Bojanowski et al., 2017).

¹³ This makes an efficient tool to build NLP models for the classification of neighborhood reviews and generating real-time predictions. To enable the application of these models in the planning decision-making process, a simple pre-processing step is introduced, as described in the following section.

of the dataset was used as a validation set to compare the accuracy of the trained model. Therefore, in total around 948,576 samples were used as training and 100,000 tweets as validation sets.

A manual inspection revealed that both training and validation datasets were evenly balanced between positive and negative labels, thus the upsampling function was not required. FastText removes hashtags, URLs, HTMLs, punctuation, numbers, emojis and special characters as they are not included in the sentiment analysis; converts the text into lowercase; and applies the same cleaning process on the validation dataset as well. It also allows recognition even for misspelled words or concatenation of words. During the training process, the Tweets will be converted to tokens (split text into pieces) which will be used as an input to the model.¹⁴ A sequence of characters or contained unit of text is called a token. Depending on how a token is created, it may contain words, emotions, hashtags, links, or even individual characters. FastText works at the word level and uses word unigrams, or tokens, as the inputs.

3.5.1. Location categorization and spatial unit of analysis

Sometimes users discuss several characteristics of different neighborhoods in one Tweet, therefore the same Tweet was assigned for every mentioned neighborhood. Each of the datasets from the case study cities was manually (using the excel search function and looking for the trending neighborhood name abbreviations) reviewed to examine how residents were indicating neighborhood names (location) in the Tweets. In two of the case studies, Memphis and San Francisco, neighborhood reviews are more precisely articulated and addressed the neighborhood names (NNs) directly. A manual aspect annotation process was applied only to meaningful samples of data from Memphis and San Francisco because the samples contained non-geocoded tweets that mentioned the neighborhood names. This phase was performed based on the SemEval 2016 task 5 (ABSA-16) annotation guideline. The annotator was familiar with the problem definition and the task.

While the appropriate focus of analysis for planning applications in a livability context may be the neighborhood unit, this study collects neighborhood-related Tweets and evaluates the results at the zipcode level, due to the availability of census demographic data at this level. This is intentional since the current study is focused on evaluation of Twitter data for transportation planning applications, and it is important to assess demographic trends found in the study datasets.

3.5.2. Aspects

Based on ABSA-16 annotation guidelines (SemEval 2016; Task 5),¹⁵ the annotation process included three steps:

- 1) Identifying the entity E and the attribute A pair (E#A) towards which an opinion is expressed. An initial inventory of entity types (e.g. safety, mobility, environmental) was defined and lists of their related attribute labels (e.g. crime, bikeability, park) are extracted. These entity categories are defined based on livability performance measures from Sarram and Ivey, 2017. For the purpose of the current study, the Sarram and Ivey (2017) livability performance measures were categorized into five types: environmental, socioeconomic, safety, mobility and amenity, as shown in Table 1. The measures are based on a feature selection analysis (Ranker method) developed on a scored livability survey dataset. The algorithm contains a search approach and an evaluation measure that ranks all possible subsets of the features. The final subset is the most determinant in the accuracy of machine learning classifiers for predicting the neighborhood livability scores. The priority was not considered for the current research and the term ‘amenity’ was defined based upon amenity types defined by Couture & Handbury, 2017.
- 2) Assigning polarity to each E#A pair of a sentence, from a modified set of P = {positive, negative}. The neutral label has not been considered in the current study due to the increased performance of the fastText classifier (discussed more in the result and discussion section).
- 3) Extracting the opinion target expression (OTE) as an explicit reference (mention) to the reviewed entity E of the E#A pair. OTE can be a named entity, a common noun or a multi-word term. It is uniquely identified by its starting and ending offsets.

The aim was to evaluate if the social media review data are applicable for planning project decision-making purposes while comparing the results with the extracted features from the previous study (Sarram & Ivey, 2017), integrating the entity types list, and developing their paired attribute labels for future studies.

4. Result and discussion

Twitter data were collected for 6 months and the total number of relevant Tweets extracted for each case study area was: Memphis, X; San Francisco, X, and NYC, X. However, for the data to be useful for livability applications it is necessary that Tweets be associated with a specific geolocation. When just the geocoded Tweets were extracted, the total number remaining in each case study was: Memphis, 110; San Francisco 650; NYC 1800. The aggregated datasets were used separately as a test set during the fastText training. The stored dataset in each case study includes several types of user information, retweets, place names, latitudes and longitudes. Note that even though a substantial number of meaningful tweets were collected in each case study, only the geocoded tweets could be considered as applicable data for visualization and decision-making purposes. Therefore, the focus of current study was on using the location-based information and evaluating its potential for use in transportation planning studies, where agencies could access much larger datasets using paid options within the Twitter API. Because of the relative size of the case study datasets, only that for NYC was used in visualization

¹⁴ <https://fasttext.cc/docs/en/supervised-tutorial.html>.

¹⁵ http://alt.qcri.org/semeval2016/task5/data/uploads/absa2016_annotationguidelines.pdf.

Table 1

List of extracted features in final training data set (Sarram & Ivey, 2017).

Priority	Top Ranked Features in Order	Performance measure
1	Experiencing negative environmental issues	Environmental
2	Living in an economically thriving neighborhood	Socioeconomic
3	Knowing my neighbors	Safety
4	Minimal road congestion	Mobility
5	Quality affordable housing	Socioeconomic
6	Having a park in my neighborhood	Amenity
7	How often stuck in traffic due to trains	Mobility
8	Feeling safe in my neighborhood	Safety
9	Experiencing presence of freight or heavy trucks traffic	Mobility
10	Living close to school/work	Mobility
11	Having alternative transportation options	Mobility

evaluations. The other two case studies were used to inform the aspect extraction.

The histogram of Fig. 1 visualizes the relative frequencies of Tweets in NYC neighborhoods. Percentage of tweets (ratio in %) in each neighborhood is calculated based on the total amount of collected tweets in NYC. In all case studies the majority of neighborhoods have less than 100 related Tweets. Saeidi et al., 2017 reported similar results in London, UK. This suggests that there is a lack of available direct residential expression in random online neighborhood reviews, indicating outreach activities will be needed by planning agencies to encourage residents to use Twitter to express their neighborhood satisfaction. Potential strategies for addressing this limitation will be further discussed in next section.

4.1. Aspect category annotation

During the training procedure, it was noticed that the majority of neighborhood Tweets are either critical or complementary and a neutral opinion was infrequently observed. This observation triggered the idea of reconsidering the SemEval 2016 annotation guideline and removing the opinion polarity of Neutral from the training labels. This helped notably in developing more accurate fastText models and to provide more accurate sentiment classification results. A similar conclusion was reached by Shah et al., 2018, who determined that varying characteristics of datasets will require adjustment of annotation guidelines.

A summary of the results of aspect category annotation for meaningful neighborhood Tweets from San Francisco areas are presented in Table 2. A example set of Tweets is also provided in Table 3. A review of the aspect entities and attributes for the total dataset indicates:

1. Besides the *Crime* and *Crash* attributes in San Francisco, *Verbal Assault* is a third mentioned **Safety** concern.

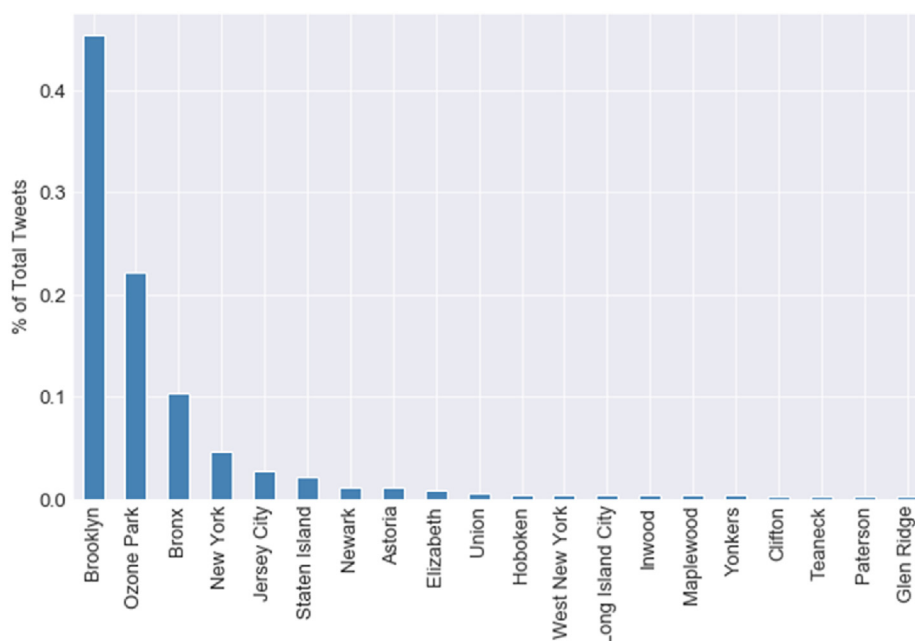


Fig. 1. Histogram of percentage of Tweets per zipcodes in New York.

Table 2
Summary of extracted neighborhood Entities and Attributes.

Entity	Attributes					
	Memphis			San Francisco		
Safety	Crash	Crime	Crash	Crime	Verbal Assault	
Mobility	Walkability	Traffic	Bikeability	Walkability	Transit	Traffic
Cultural	Community	Racism	Community	Racism		
Environmental	Ambiance	Noise	Ambiance	Sustainability		
Amenities	Quality		Quality			
Socioeconomic			Affordability			

2. *Bikeability* and *Transit* are indicated besides *Walkability* in the neighborhood **Mobility** aspect category of the San Francisco area.
3. *Affordability* is not mentioned frequently as a **Socioeconomic** concern in Memphis.
4. In San Francisco, *Parking availability* is Tweeted as a trending Traffic issue effecting neighborhood **Mobility**, however in Memphis the effect of a public place and related *Events* on the Traffic and the neighborhood quality are more important.
5. *Ambiance* in both cities is considered as a key **Environmental** attribute (e.g. having a park, great view) while *Noise* and *Sustainability* are reported in Memphis and San Francisco respectively.
6. Having a supportive and active neighborhood *Community* are addressed as important **Cultural** vibes while *Racism* is also reflected in the language of both cities' Tweets.
7. Proxy to the neighborhood **amenities** (e.g. restaurants and groceries) was important in the residential satisfaction.

The aspect-based results are compatible with the extracted features from Sarram and Ivey (2017), while adding a new entity of neighborhood **Culture or life style**. This new neighborhood attribute has not been previously suggested by the neighborhood livability survey results which could also relate to the socioeconomic characteristics of a neighborhood as well.

4.2. fastText sentiment analysis

Like most machine learning algorithms, performance of a fastText model varies depending on the preprocessing of data and the choices of hyperparameters. The optimal values of these parameters tend to vary depending on the dataset or task. While fastText has recently developed an autotuning hyperparameter feature,¹⁶ the current study searched for the best set manually to be able to directly monitor the relationship between performance of the text classifiers and the choices of training datasets. The optimality of the training parameters is dependent on the quantity of the dataset.¹⁷ The fastText default values for epoch and lr are 5 times and 0.05.¹⁸

An overview of the results of fastText models compared between two different training datasets is shown in Table 4. The predicted results from each model were tested manually in order to confirm the reported accuracy of fastText models for transportation planning/livability contexts.

In all final models, the same hyperparameters lr, epoch, and wordNgrams are interplayed while the dimension parameter is changed between 20 and 100. This suggests that the extended training size in this study yields an improvement in the fastText validation process, therefore the model needs less dimension in order to train.

In order to evaluate the performance of the learned classifier, fastText requires splitting the data into training and validation sets and similar to other classification algorithms, the accuracy of the model is calculated based on the fraction of correctly predicted Tweets in each process. Here, two validation sets are used to test the learned classifier (see Table 4). The fastText model developed on the Sentiment 140 dataset in both training and validation processes provided a result on the neighborhood Tweets test set that outperforms other tried datasets including SemEval 17. This means, for the smaller size dataset (SemEval 17), fastText validation results are slightly worse than those of the Sentiment 140 model (0.65 vs 0.83). This is while retaining a short training time in each scenario (less than 5 min), far less than performance time of the state-of-the-art deep learning algorithms.

¹⁶ <https://fasttext.cc/docs/en/autotune.html>.

¹⁷ The model components consist of epoch, learning rate (lr), word n-grams and dimension (dim), with the standard ranges respectively between 5 and 50, 0.1–1.0, 1–5 and 100–300. The most important hyperparameters of the model are the range of size for the subwords and its dimension. Word n-grams relates to sequencing of the tokens in the text. As mentioned in the preprocessing section, word vectors capture hidden information about a language (e.g. word analogies and semantic) and the subwords are all the substrings contained in a word between the minimum (minn) and the maximum (maxn) size. Subword-level information is in particular interesting to build vectors for unknown words. Meanwhile, the dimension controls the size of these vectors. The larger the dimension, the more information can be captured. However, it requires learning more data and it is also harder and slower to train. The epoch parameter determines the number of times a model will loop over the data. Adding more epoch would increase the training time, therefore in case of massive datasets, it is recommended to loop over it less often. On the other hand, the higher the learning rate hyperparameter, the faster the model converges to a solution. This is while the risk of overfitting to the dataset would increase.

¹⁸ <https://fasttext.cc/docs/en/unsupervised-tutorial.html>.

Table 3

Example Tweets from San Francisco with annotated aspect categories. Subjective phrases are italicized. Topics are **bold** and aspect phrases are **bold italic**.

Tweets	{Entity# Attribute, “OTE”, Polarity}
Another Mission restaurant closes due to skyrocketing rents . Another sad loss for my neighborhood	1. {SOCIOECONOMIC#AFFORDABILITY, “rents”, negative}
Just spent a speedy lunchtime in the Castro — my old neighborhood! — and am reminded of how this place has the best sunshine , the best movie palace , and the best vibe around.	1. {AMENITIES#QUALITY, “place”, positive} 2. {ENVIRONMENTAL#AMBIANCE, “vibe”, positive}
How lucky I was to call this place my last stop every day *rainbow emoji* This is about 3 blocks from where I live and right by where my wife walks every weekday. SF has been incredibly tone deaf on the safety concerns of the residents of this neighborhood. I can't understand how the judge released this man after he violently assaults this woman.	1. {SAFETY# VERBALASSAULT, “assaults”, negative}
Big news for Oceanview . My food desert of a neighborhood finally opened a coffee shop . No longer walking two miles for Arco gas station coffee. @ Blue House Café	1. {AMENITIES#QUALITY, “coffee shop”, positive} 2. {MOBILITY#WALKABILITY, “walking”, positive}
Neptune Gardens is a neighborhood on the West end of Alameda that combines the accessibility of shopping and restaurants with the advantages of being largely residential .	1. {MOBILITY#WALKABILITY, “accessibility”, positive} 2. {CULTURAL#COMMUNITY, “residential”, positive}
@JesseArreguin @DavidChiu @davidying @WatsonLadd My former affordable housing boss: “I live in NB and the neighborhood has already changed. The diversity is disappearing. Prioritize housing over parking. More bikes . Shuttles and buses . We need 100% affordable” #berkmtg	1. {MOBILITY#BIKEABILITY, “bikes”, negative} 2. {MOBILITY#TRANSIT, “buses”, negative} 3. {CULTURAL#COMMUNITY, “diversity”, negative} 4. {SOCIOECONOMIC#AFFORDABILITY, “affordable”, negative}
@prinzrob @akronisticloror My own neighborhood is my least favorite area there's no bike lanes ; everyone drives like they're on a rural back road (i.e., carelessly). Though biking up San Pablo through west Berkeley has been the most terrifying of all my bike routes ; the closest I've come to being hit .	1. {MOBILITY#BIKEABILITY, “bike”, negative} 2. {SAFETY#CRASH, “hit”, negative}

Table 4

Training accuracy - comparison between two labeled Twitter datasets.

model data		hyperparameters				accuracy	
training	validation	lr	epoch	wordNgrams	dim	training	validation
Sent. 140	<i>SemEval 17</i>	0.05	15	2	20	0.943476	0.65021
		0.05	15	2	100	0.941563	0.618878
Sent. 140	<i>Sent. 140</i>	0.05	15	2	20	0.94473	0.83298

4.3. Sentiment maps

An online Jupiter visualization package was developed to translate the neighborhood Twitter reviews from qualitative to quantitative data (refer to the code availability section). This visualization tool receives the extracted neighborhood Tweets, preprocesses the data, builds the fastText models and visualizes the aggregated classified sentiments at the zipcode level. The goal is to gain new insights about the neighborhood satisfaction to facilitate the project decision-making process.

To consider more reliable sentiments, only neighborhoods with more than one third of the average Tweets were taken into account. Next, the binary predictions were replaced with the related labels, 0-negative and 1-positive, to calculate the total neighborhood sentiments using the polar aggregation. The research results in Fig. 2 reveals available satisfaction maps of Twitter neighborhood data in NYC for a short data collection span of six months. In the New York region the neighborhoods of Brooklyn (zipcode 11226), Ozone Park (zipcode 11417) and Bronx (zipcode 10461) have the most active inhabitants on the Twitter platform related to neighborhood perceptions.

These limited detected sentiments are compatible with the national Twitter penetration rate in the U.S. According to a new Pew Research Center survey¹⁹ conducted in Feb. 2019, around one-in-five U.S. adults (22%) say they use Twitter. Among this population, in a late 2018 survey²⁰, it is also revealed that the Twitter users tend to be younger than 50 (mainly between 30 and 49), have at least a college degree, and are wealthier than the general public (earn at least \$75,000 a year). Additionally, 13% of the adult users in U.S. keep their accounts private. The major portion of users who opt not to post tweets publicly are women, who are also more active when it comes to creating or favoriting tweets. Thus, it can be expected that a large proportion of the tweets would be shared without a specific reference of location (either in the user profile or added to the tweet). Mechanisms to address these limitations are needed to make Twitter more valuable for the transportation planning domain. If steps are taken to create more balanced expression via Twitter, then

¹⁹ <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>.

²⁰ <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/#fn-22366-1>.

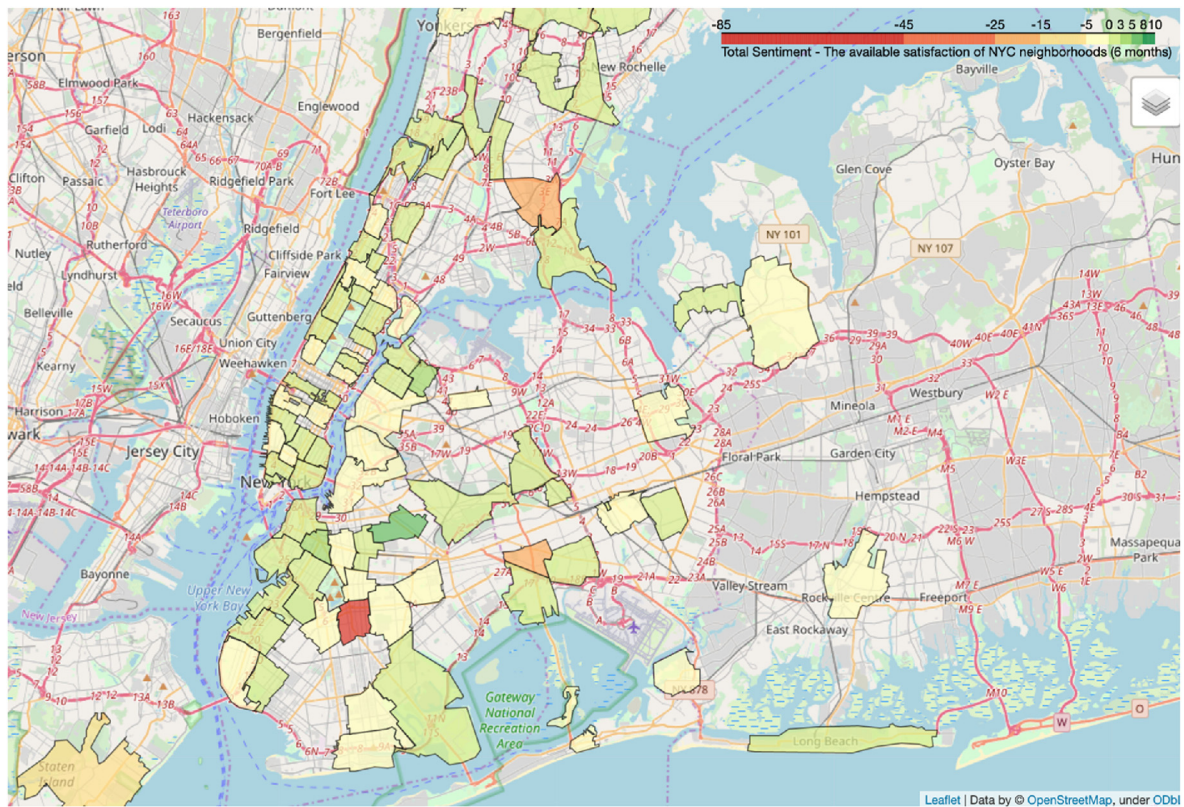


Fig. 2. Sentiment-derived satisfaction of neighborhoods in New York.

sentiment analysis and visualization can provide a very valuable tool for transportation planners to quickly identify areas (census tracts) where further investigation and investment may be warranted, as demonstrated in Fig. 2 (where red areas indicate predominantly negative sentiment while green areas indicate positive). Besides extracting the important aspects of public concerns, the visualized map can help transportation planners to develop a more equitable approach to public outreach by ensuring residents in areas with more negative sentiment are aware of opportunities to engage in the planning process and have access to do so. The sentiment visualization can also quickly indicate problem areas and allow planners to overlay other relevant data. When coupled with aspect information, this would provide better understanding of factors contributing to livability, particularly in areas where residents demonstrate greater dissatisfaction. The neighborhood sentiments in Fig. 2 range from green to red, representing more satisfactory to dissatisfactory. The scale of the colors is adjustable depending on the relative frequencies of data between the neighborhoods.

As previously mentioned in the data collection section, while a portion of Tweets are geotagged or mention the location, there are certain limitations with online reviews that are important to consider. First, as pointed out by Zivanovic et al., 2018, there are risks of ‘migration bias’ and representability with Twitter data. This means that the statement about a specific location could be shared from a completely different location and different time; in addition to the problem that the use of Twitter is very uneven among the population (e.g. by age group, by income group, and languages they use). In another study in 2017, Blank and Lutz investigated the representativeness of different social media platforms and concluded that Twitter users in UK are greatly different from the total population only in terms of age and income. Second, the main challenge of online review data (Twitter, Niche, and Flickr) in comparison with predefined sources of extracting neighborhood information (e.g. surveys, focus groups comments, public meetings) is that they unfortunately do not provide the demographic information of the users who wrote the reviews. When the demographic information are available along with the online reviews, it is possible to develop customer profiling and translate the derived sentiments more accurately into the decision-making components/measures. As people can have different expectations of their living environment, depending on their cultures, backgrounds, and socioeconomic statuses (Hu et al., 2019). Therefore, in order to detect these limitations in the current study, a comparative analysis was necessary at the neighborhood level.

While there is an uneven representation of neighborhoods, the visual content contains some complementary information to the texts making it valuable to employ text and visualization jointly. One popular metadata for explaining these two sets of information is residential demographics that has been commonly used to translate relational information into project decision making components. Thus, this study assumes the differences in neighborhood satisfaction patterns are driven by demographic differences and describes trends that are most likely to shape these patterns, starting with appraising the related Census demographic trends and concluding with possible impacts of those trends on users’ satisfaction and expectations.

Demographic data from US Census Bureau was used as an objective dataset with which the neighborhood residential characteristics were compared. Three demographic characteristics were more distinctive among the most reported neighborhoods: age, mean household income, and educational attainment. According to the Census Demographic information²¹ and from aligning it with sentiment maps in Fig. 2 and the zipcodes, it is inferred that in NYC:

1. Residents younger than 50 and mainly in their 30s are more likely to share Tweets about the quality of their neighborhood.
2. Most of the residents in the active neighborhood areas have pursued higher education at least at the bachelor level.
3. The mean household income in the active neighborhood areas is not the highest level and mainly reflects middle class households (between \$50–70K).

While this study found some significant limitations with Twitter data currently, there is much potential for it to enhance transportation planning strategies. First, the nature of Tweets provides more rich information that may not be obtained through traditional mechanisms because it is not biased by a survey developer. In the example presented in this study, an additional feature influencing livability perceptions was identified through the Twitter analysis that was not uncovered using traditional approaches. Second, Twitter provides an opportunity to identify factors influencing livability that are highly positive or negative (and therefore may be more important to address) as shown in the lack of neutral tweets in the case study datasets. With large amounts of data, high frequency issues can be identified and addressed. Given the current limitations with Twitter data, it is recommended that it be used in combination with other sources of datasets and that data be collected in a longer time span. Transportation agencies can do this more rapidly through use of the premium platform privileges via paid accounts, and can also track changes in sentiment over time. Another suggestion for the future work is instead of collecting the tweets with related keywords or geo tags, test the other available types of twitter APIs. For instance User Timeline API is able to target a group of active Twitter users and Streaming API can collect any tweet within the study area. With these approaches, other neighborhood filtering strategies might be effective on the data after the collection is complete.²² Therefore, the resulting sample size might be larger.

This study also demonstrated the utility of fastText and the Sentiment 140 training dataset for transportation planning contexts. The combination provides an accurate and efficient tool that shows promise for enhancing decision-making. The approach demonstrated in this study can be applied to online data alongside or before developing customer surveys, for example, for the purpose of identifying additional features that need to be examined in transportation planning surveys.

There are numerous strategies that transportation agencies may employ to increase the utility of Twitter for sentiment analysis for livability or other purposes. First, targeted outreach efforts can enhance participation and sharing of meaningful data, without the need for people to share geocoded location data. For example, agencies can develop hashtags and public outreach campaigns to encourage people to share thoughts related to specific topics (such as #TDOTlivability or hashtags incorporating specific neighborhood names and keywords). Intentionally developing Twitter as a means of communication requires not only outreach but also sharing by the agency of how these comments will be used to enhance planning practice and investments, and developing two-way communication.

Additionally, transportation agencies struggle to get input from a variety of populations, including younger adults. Twitter already captures a younger audience, and because adults in their 20s–30s are digital natives, they are much more likely to be motivated to engage via a social media platform rather than traditional venues like public meetings. With carefully planned communication strategies, Twitter can be leveraged to balance participation and increase engagement and perceived relevance of transportation agencies. Low income populations are another demographic traditionally difficult to engage in the transportation planning process. Many communities have programs specifically designed to provide internet access and mobile devices to disadvantaged populations. According to a survey of teen social media use in 2018 by Pew Research Center,²³ young adults from low income communities are using social media at higher rates (70%) than their higher income counterparts (36%). This also points to the opportunity to create education and outreach programs designed to encourage input via Twitter from this population to create more balanced engagement. Expression via social media may also be more attractive to these groups, who may be less likely to express opinions in public forums where they may feel hesitant because of the presence of people from higher income or education demographics who may dominate discussions (Jacobs et al., 2009). And, in fact, social media may be a more convenient and less threatening venue for expressing opinions about neighborhood quality for many people, as evident from this Tweet from a user in San Francisco:

‘after being inspired I found a site that allows me to report any issues I see in my neighborhood and Oakland in general check it out and see if it is available for your area’

The main contribution of this study is in preparing the path for transportation planners to be able to actively identify project performance measures and their impacts on residential quality of life while collectively detecting key decision making components for residents. In this age of increasing IoT and digital interconnectedness, agencies who want to engage stakeholders must begin thinking about engagement strategies in new ways. The use of social media for text mining and sentiment analysis provides a new approach that can not only enhance understanding of factors affecting livability or other transportation considerations, but also has the potential to engage stakeholders that previously have not participated in planning decisions.

²¹ <https://statisticalatlas.com/neighborhood/New-York/New-York/Downtown/Overview>.

²² <https://developer.twitter.com/en/docs/twitter-api>.

²³ <https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/>.

5. Conclusion

This study has provided a review analysis package based on users stated preferences in social media for transportation planning decision-making purposes. The Github instruction page provides guidance for agencies in developing an efficient online text classification tool combined with a visualization approach for automating the performance measure extraction.

In addition to examining the application of fastText algorithms in this domain, a guided annotation process was established towards the extraction of user-defined neighborhood expectations, and a Jupyter notebook approach was developed. The Twitter dataset reveals meaningful information and the results show promise for engaging the public in expressing their neighborhood satisfaction and expectations via a targeted social media strategy. The aspect-based annotated results from this study are compatible with the extracted features from a previous study that used traditional survey techniques. In the current study, an additional neighborhood aspect (Cultural) has been extracted and strategies are suggested for promoting the active use of social media by neighborhood residents in order to achieve more efficient input to the planning process. This approach shows particular potential for livability studies, however; the methodology could be targeted for numerous planning-related interests of transportation agencies.

Consequently, the outcome of this work demonstrates the potential of Twitter to significantly increase agencies' public involvement activity and to obtain enhanced information about the efficiency of the system and the areas that need improvements. The package provides a framework for developing a data-driven tool for connecting the condition and importance of transportation products (networks) to public preferences and expectations. Through understanding the revealed aspects of public (consumer) concerns, the visualized sentiment analysis can help transportation planners better understand how residents interact with the transportation network in their community, assess network performance, develop more equitable public outreach and improve communication about planning projects. The sentiment visualization, specifically coupled with aspect extraction, can quickly illustrate problem areas. This in turn can enable project decisions that are more cost effective as they are better aligned with the reality of public expectations and utility rates. Developing strategies that leverage this tool may provide planners the ability to decide more readily on the projects and investment distributions to best enhance the communities they serve.

Code availability

Jupyter notebooks for fastText training, validation and testing, as well as descriptive and sentiment visualization codes are customized in one package towards developing a data-driven planning decision-making tool. To view this package visit [NextUrban](#).

Declaration of competing interest

All the authors have no conflict of interest with the funding entity and any organization mentioned in this article in the past three years that may have influenced the conduct of this research and the findings.

Acknowledgement

This research was funded in part by the University of Memphis Intermodal Freight Transportation Institute, Department of Civil engineering, and the Tennessee Department of Transportation (TDOT).

References

- Ahmed, W., Bath, P. A., & Demartini, G. (2017). Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. In U. K. Bingley (Ed.), *Advances in research ethics and integrity* (pp. 79–107). Emerald, Dec.
- Albouy, D., & Lue, B. (2015). Driving to opportunity: Local rents, wages, commuting costs and sub-metropolitan quality of life. *Journal of Urban Economics*, 89, 74–92.
- Baziotis, C., Pelekis, N., & Doukeridis, C. (2017). *DataStories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis*.
- Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *TACL*, 5, 135–146.
- Cataldi, M., Ballatore, A., Tiddi, I., & Aufaure, M. A. (2013). Good location, terrible food: Detecting feature sentiment in user-generated reviews. *Social Network Analysis and Mining*, 3(4), 1149–1163. CiteSeerX 10.1.1.396.9313.
- Chien, C.-F., Kerh, R., & Lin, K. Y. (2016). Data-driven innovation to capture user-experience product design: An empirical study for notebook visual aesthetics design. *Computers & Industrial Engineering*, 99, 162–173.
- Cliché, M. (2017). *BB twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs*.
- Couture, V., & Handbury, J. (2017). *Urban revival in America, 2000 to 2010*.
- Engin, Z., Dijk, J. V., Lan, T., Longley, P. A., Treleaven, P., Batty, M., & Penn, A. (2019). Data-driven urban management : Mapping the landscape. *Journal of Urban Management*, (May), 1. <https://doi.org/10.1016/j.jum.2019.12.001>
- Flaes, J. B., Rudinac, S., & Worring, M. (2016). What multimedia sentiment analysis says about city liveability. In *Ecir '16* (pp. 824–829).
- Gibbons, J., Malouf, R., Spitzberg, B., Martinez, L., Appleyard, B., Thompson, C., Nara, A., & Tsou, M. H. (2019). Twitter-based measures of neighborhood sentiment as predictors of residential population health. *PLoS One*, 14(7), Article e0219550.
- Hu, Y., Deng, C., & Zhou, Z. (2019). *A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments* (pp. 1–21). Annals of the American Association of Geographers.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of KDD*, 2004.
- Jacobs, L. R., Cook, F. L., & Delli Carpini, M. X. (2009). *Talking together*. Chicago: University of Chicago Press.
- Jiao, J., & Zhang, Y. (2005). Product portfolio identification based on association rule mining. *Computer-Aided Design*, 37, 149–172.
- Kim, S., & Chung, J. (2016). Reinterpretation of the Likert scale for public transportation user satisfaction: Pattern recognition approach. *Transportation Research Record. Journal of the Transportation Research Board*, 2541, 90–99.
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). *Inherent trade-offs in the fair determination of risk scores*.

- Li, S., Nahar, K., & Fung, B. C. M. (2015). Product customization of tablet computers based on the information of online reviews by customers. *Journal of Intelligent Manufacturing*, 26, 97–110.
- Lin, K. Y., & Chien Cf, K. R. (2016). UNISON framework of datadriven innovation for extracting user experience of product design of wearable devices. *Computers & Industrial Engineering*, 99, 487–502.
- Long, Y. (2015). Big/open data for urban management. *Journal Urban Management*, 4, 73.
- Meehan, L. A., & Whitfield, G. P. (2017). Integrating health and transportation in Nashville, Tennessee, USA: From policy to projects. *Journal of Transport & Health*, 4, 325–333.
- Michalek, J. J., Feinberg, F., & Papalambros, P. (2004). Linking marketing and engineering product design decisions via analytical target cascading. *Journal of Production 814 Innovation Management: Special Issues Marketing Design New Product Development*, 22, 42–62.
- Mittal, V., & Kamakura, W. A. (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research*.
- Moon, S. K., Simpson, T. W., & Kumara, S. R. T. (2010). Amethodology for knowledge discovery to support product family design. *Annals of Operations Research*, 174, 201–218.
- Morency, C., Trépanier, M., & Agard, B. (2007). *Measuring transit performance using smart card data*. San Francisco, USA: World Conference on Transport Research. June 24–28.
- Nashville Area, M. P. O. (2013). *Middle Tennessee transportation and health study final report*. Nashville TN: Nashville Area MPO.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Found Trends Information Retrieval*, 2, 1–135.
- Paredes, M., Hemberg, E., O'Reilly, U., & Zegras, C. (2017). Machine learning or discrete choice models for car ownership demand estimation and prediction? 2017. In *5th IEEE International Conference on models and technologies for intelligent transportation systems (MT-ITS)* (pp. 780–785). Naples.
- Quercia, D., Ellis, J., Capra, L., & Crowcroft, J. (2012). Tracking gross community happiness from tweets. In *Proceedings of the international conference on computer supported cooperative work and social computing*.
- Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in twitter. *SemEval@ACL, 2017*, 502–518.
- Saeidi, M., Venerandi, A., Capra, L., & Riedel, S. (2017). Community question answering platforms vs. *Twitter for Predicting Characteristics of Urban Neighborhoods*.
- Sarram, G., & Ivey, S. S. (2017). *Evaluating a survey of public livability perceptions and quality-of-life indicators: Considering freight-traffic impact*. ASCE International Conference on Sustainable Infrastructure 2017.
- Sarram, G., & Ivey, S. S. (2018). *Investigating customer satisfaction patterns in a community livability context: An efficiency-oriented decision-making approach*. ASCE International Conference on Transportation and Development.
- Schweitzer, L. (2014). Planning and social media: A case study of public transit. *Journal of the American Planning Association*, 80(3), 218–238.
- Shah, F., Sirts, K., & Pfahl, D. (2018). *The impact of annotation guidelines and annotated data on extracting app features from app reviews*.
- Shao, X., Wang, Z., Li, P., & Feng, C. J. (2006). Integrating data mining and rough set for customer group-based discovery of product configuration rules. *International Journal of Production Research*, 44, 2789–2811.
- Song, Z., & Kusiak, A. (2009). Optimizing product configurations with a data-mining approach. *International Journal of Production Research*, 47, 1733–1751.
- Tucker, C., & Kim, H. M. (2008). Optimal product portfolio formulation by merging predictive data mining with multilevel optimization. *Journal of Mechanical Design*, 130, Article 041103.
- Tucker, C., & Kim, H. M. (2011). Trend mining for predictive product design. *Journal of Mechanical Design*, 133, 111008.
- Wassenaar, H., Chen, W., Cheng, J., & Sudjianto, A. (2005). Enhancing discrete choice demand modeling for decision-based design. *ASME Journal 811 Mechanical Design*, 127(4), 514–523.
- Zhang, K., Narayanan, R., & Choudhary, A. (2010). *Voice of the customers: Mining online customer reviews for product feature-based ranking*. WOSN.
- Zivanovic, S., Martinez, J., & Verplanke, J. (2018). Capturing and mapping quality of life using Twitter data. *Geojournal*, 85, 237–255, 2020.