CIRRELT

CIRRELT-2023-07

A Methodology to Assess Passengers' Perceptions of Transit Services and Impact of Incidents

Banafsheh Mehri
Martin Trépanier
Yves Goussard

January 2023

Bureau de Montréal
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

A Methodology to Assess Passengers' Perceptions of Transit Services and Impact of Incidents

Banafsheh Mehri[1,2,*], Martin Trépanier[1,3], Yves Goussard[2]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Electrical Engineering, Polytechnique Montréal

[3] Department of Mathematics and Industrial Enginnering, Polytechnique Montréal

Abstract. Public transit agencies have long been concerned with taking into account the users' opinions about the quality of transit services. However, collecting and analyzing customer feedback data is a challenging task that demands extracting and analyzing textual data on a large scale and translating it into useful information. In this study, we focus on gaining insights from public transit official tweets and users' retweets and responses. The abundance of data provided by social media was shown to provide the ability to improve the efficiency of current traffic management systems and transit analytics. This study adds to the body of knowledge by presenting a framework for analysing transit users' opinions about transportation services using Twitter data. We propose a method to collect the tweets and further apply an unsupervised topic modeling machine learning technique on this collected data in order to discover underlying latent topics in users' messages. Then, we perform sentiment analysis per tweets as well as per extracted topics to explore users' opinions and understand the reasons behind their opinions. The results of this study could enable the public transit agencies to measure the quality of their services from users' point of view and to assess the impact of events and incidents on customer satisfaction.

Keywords: Intelligent Transportation Systems (ITSs), public transit, transportation incidents, text mining, Twitter, transit service performance.

Acknowledgements. The authors gratefully acknowledge the financial support provided by the Natural Sciences and Engineering Council of Canada (NSERC), and the Société de transport de Montréal (STM) for data provisioning.

* Corresponding author: banafsheh.mehri@polymtl.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2023

1 Introduction

Twitter is one of the widely used online social media services. The way information is communicated and exchanged has changed dramatically as a result of vast usage of social media platforms [1]. People use the Twitter platform to emanate an increasing flow of data concerning various aspects of human life, such as transit. Users of social media platforms may be considered as *Citizen Sensors* with the potential to record and characterise events by sending messages with spatial footprints, in form of rich semantic and multimedia information [2].

Social media data offers remarkable possibilities for integrating urban transit and technology in a coherent manner. They have the ability to add a frame of reference to the monitoring and evaluation of transit performance. Such data combined with actual transit analytical structures enable agencies to enhance traffic management and operations while also allowing users to have a greater understanding of their local surroundings. More significantly, they will aid in the development of evidence-based and data-driven transit policy and investment decisions. The benefits of leveraging such rich and large amount of data to manage mobility systems have begun to be realized by progressive transit agencies.

For example, the city of Los Angeles collaborated with Esri to create a geographic data visualisation platform. One of their initiatives tracks pedestrian and bike incidents and injuries in traffic accidents in order to discover risk factors and develop preventative techniques [1]. In addition, the city of LA also teamed up with Google Waze to gather data from users of the navigation app to determine the traffic congestion hotspots [3].

In public transit, ridership is often used as a proxy for income. Hence, transit agencies strive to strike a balance between the highest possible ridership and the lowest feasible operational expenses [4]. Service quality such as reliability, comfort, and convenience, service coverage, station accessibility, and user experience at all times and in presence of incidents are all the factors and characteristics that might influence transit ridership[5].

Transit agencies mostly perform Customer Satisfaction Surveys (CSS) among passengers to assess customer experience via passenger comments on several pre-defined areas regarding transit service quality [6]. However, the high cost, small sample size, and inadequate resolution of survey data have been key roadblocks that prevent the full utilization of survey results to influence investment decisions. In addition, compared to pre-defined assessments areas of service quality and incident impacts present in surveys, passengers' real experiences may convey altogether different stories.

The vast amount of user-generated textual data on Twitter can be collected and used as a supplement to enhance the data collected via traditional approaches. This augmented data can potentially be used to better characterize customer opinion about the quality of public transit services and customer

[1] https://ladotlivablestreets.org/

reaction in the event of an incident on the transit networks. In this study, we propose an approach to systematically collect and analyze the tweets published on Twitter by the Société de transport de Montréal (STM) and its users over six years from 2015 until 2020. We then apply a topic modeling technique to this data to identify the abstract themes from tweet observations.
Furthermore, we analyze the sentiments of customers of STM via their retweets and their responses according to the identified topics. Finally, we crosscheck the results of the analysis of tweets with the STM metro incident dataset in order to determine the potential impact of such incidents on customer satisfaction. This strategy is far less expensive and time-consuming compared to traditional approaches that can help transit agencies to plan and perform transit services more effectively with regard to incorporating consumer feedback into the process.

2 Literature review

Users may discuss their everyday activities and provide their thoughts on any topic on social networks, which have become one of the most important communication tools in modern culture. These platforms have grown in popularity in recent years. In 2020, over 3.6 billion people were using social media worldwide, this number projected to increase to almost 4.41 billion in 2025 [2]. Several studies have sought to use social media to conduct transit research in several major categories such as monitoring of traffic conditions [7] [8] [9] [10] [11], estimation of travel demand [12] [13] [14], assessment of mobility behaviour [15] [16] [17] and incident and disaster modeling [18] [19] [20] [21] [22].

Twitter data and text processing techniques such as sentiment analysis and topic modelling have been employed in recent studies to examine public opinion and consumer satisfaction of transportation, and transit in particular. Developing a sentiment detection system, researchers [23] used tweets in Chicago that had keywords of train name to gauge transit riders' contentment. Their findings show that transit users are more likely to express negative sentiments about a scenario such as *power outage* than positive sentiments. Schweitzer [24] also utilised tweets to gauge how people felt about public transit. She discovered that Twitter users show more positive sentiments about public transit than they do about other government services such as police departments.
Furthermore, she found that transit agencies that reply directly to their users' inquiries and comments have higher levels of positive sentiment.
Steiger et al. [25] analyzed public transit data and identified important transit hubs in the city of London using social media data from Twitter, Foursquare, Instagram, and Flicker. They utilized a density-based spatial clustering to find clusters of points that were densely packed together. They discovered that the observed clusters were physically oriented along London's rail segments. The

[2] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

findings were confirmed using an OpenStreetMap [3] overlay of the major rail and public transit networks.

Luong and Houston [26] also used Twitter data to analyze public opinions of light rail service. They used the names of seven light rail lines in Los Angeles as keywords to collect data from the Search Twitter API. Then they used sentiment analysis to look at how people on Twitter felt about light rail in Los Angeles.

Gal-Tzur et al. [27] worked on identifying the criteria for gathering transit-related knowledge from social media. They designed their study based on the understanding of the nature of social media content with the goal of harvesting it for transport planning and management. Their findings showed the potential of social media data in offering to help transit policy makers with useful insights for creating transit policies.

El-Diraby, et al. [28] focused on using social and semantic network analysis to support the modeling the customer satisfaction of transit systems. They used data from the Vancouver Transit Agency and they monitored the TransLink [4] Twitter account for eleven months. Their findings resulted in the discovery of negative sentiments toward the transit system, resulting in identifying the services that the agency can improve to increase customer satisfaction.

In addition, Ali et al. [29] used Twitter, Facebook, and TripAdvisor[5] to analyze users' opinions. They used sentiment classification to study traffic control and management systems of intelligent transportation systems. They reported that their method achieves an accuracy of 93%, which shows that their proposed approach can be effective for sentiment classification.

Then Korfiatis et al. [30] used topic modeling on airline passengers' online reviews to recognize customer satisfaction and service quality to extract the quality features of airlines from review text. Their results showed the importance of the cost aspect of airlines amongst all other service quality dimensions, thus explaining the success of low-cost carriers in the airline market.

Qi et al. [31] focused on exploring the public perceptions towards the transit-related events. They presented a framework to facilitate public sentiment detection and analysis of transit services from Twitter in the Miami-Dade County transit [6] by using topic modeling. The framework can be used as a tool for stakeholders to enable a holistic understanding of public opinions on transit services and increase the degree of public participation in transit management.

Although the studies described above offer useful insight into the usage of social media data in public transit analysis, the majority of them merely utilized a simple keyword search to filter out transit tweets. However, based on our preliminary analysis, the majority of these tweets may not accurately reflect users' comments on transit service quality. In addition, since data collection from social media platforms is challenging due to the limitation of APIs, most of these studies are limited to only few months of historical data. Furthermore,

[3] https://www.openstreetmap.org/map=3/71.34/-96.82
[4] https://www.translink.ca/
[5] https://www.tripadvisor.ca/
[6] https://www.miamidade.gov/global/transportation/home.page

up to the present, only few studies has focused at cross joining operational transit data and social media data. The volume of data generated by the transit network grows over the time in lockstep with the volume of data generated by social media overtime. Therefore, in order to infer customer satisfaction, public transit research could indeed benefit from further pairing of operational transit data sources with social media data.
As such, this research aims to propose a systematic approach to collect Twitter data over several years back in time, adjusting and applying modern text processing techniques to examine the content of the data collected in order to determine common themes in the domain of transit and to determine the overall polarity of the themes. Then, by using the STM's operational data which contains information on incidents, we characterize their potential impact by analyzing the tweets associated with such incidents, validating assumptions of patterns within social media data.

3 Methodology

Figure 1 shows the overall diagram of the proposed methodology. In this section we present a framework to collect and pre-process Twitter data. Then, we assess transit performance using transit users' opinions expressed on tweets. In this study, we use a topic modeling technique to sift tweets that are relevant to the actual user experience of the transit system. Then, using sentiment analysis techniques we determine whether the sentiment expressed in customer posts is positive, negative, neutral, or mixed, which helps to determine the overall polarity of the topics and the polarity of customer opinions. Finally, by using the STM's operational data which contains information on incidents, we characterize their potential impact by analyzing the tweets associated with such incidents.

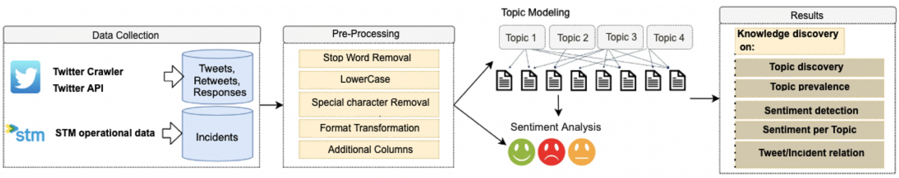

Fig. 1 Overview of the framework

The proposed framework consists of three major components. The first step is the data collection methodology and data preprocessing. The next step consists of text processing which consists of Topic modeling and Sentiment Analysis. Then, the third component includes knowledge discovery by After cross-joining the Twitter data with STM operational data, generating and analysing the results of text processing and in addition validating the results with STM in an iterative manner to customize the general methodology to the specific characteristic of transit system. This section describes these three components in further detail.

3.1 Study Area

Greater Montreal[7] is the most populous metropolitan area in Quebec and the second most populous in Canada. Statistics Canada identified Montreal's Census Metropolitan Area as 4,258.31 square kilometres [8] with a population of 4,027,100 [9].
The low-density municipalities are located on the fringe of Metropolitan Montreal. Most of these cities and towns are semi-rural. The main part is composed of densely populated municipalities located in close proximity to Downtown Montreal. It includes the entire Island of Montreal, Laval, and the Urban Agglomeration of Longueuil. The Montreal Metro network is operated by the Société de transport de Montréal (STM). It is a rubber-tired underground rapid transit system serving Greater Montreal, Quebec, Canada. It has expanded since the 1960s from 26 stations on three separate lines to 68 stations on four lines totalling 69.2 kilometres[10] in length, as shown in Figure 2, serving the north, east and centre of the Island of Montreal with connections to Longueuil, via the Yellow Line, as well as the suburb of Laval, via the Orange Line. Montreal Metro delivers an average of 1,500,000 daily unlinked passenger trips per weekday[11]. According to the STM, the Metro system had transported over 7 billion passengers as of 2010.

[7] *Grand Montréal* in French
[8] 1,644.14 sq mi
[9] https://tinyurl.com/adpxrvya
[10] 43.0 mi
[11] Statistics of the end of year 2019

Fig. 2 Overview of the Montreal Metro

3.2 Data Collection

3.2.1 Tweets

For this research we first need to gather tweets of STM metro lines and information about gathered tweets from different sources. In order to do so, a systematic data gathering model has been developed to provide accurate and consistent information for further analysis.
Various types of data fields, including tweet content, retweet numbers, and user profiles can be accessed via the Twitter API. However, the Twitter Official

API has several restrictions on the download and data amount. In addition, the data collection process will be inefficient if the time span is long (e.g., in this research over years from 2015 to 2020) due to the limitation of time constraints set by the API. Nonetheless, Twitter Advanced Search[12] can be used to search through tweets since March 21st 2006 until today in specific areas and can provide filtering options such time limit that can help find specific contents and all the data are open to the public.

Via data extraction tools such as the web crawler we can automatically fetch and extract information from specific websites. Since the results of Twitter Advanced Search are provided in a browser, in order to retrieve historical tweets, we have used a web scraping method.

For this research, in order to automatically download tweets from the Twitter Advanced Search website we developed a web crawler using a Python library called GetOldTweets3[13] . This approach can help to download tweets filtered by location avoiding the issues of rates and time limits set by the Twitter API. When scraping data from Twitter advanced search results, there are few fields available for tweets such as ID, Date (posted on Twitter), Text (body of the tweet) and the Location.

To carry out our study we needed more information about tweets such as retweets, responses, comments, etc. Since we have already collected tweet IDs corresponding to our search criteria in the previous step, we were able to use the Twitter API and retrieve all the available fields with regard to each and every tweets. In order to do so, we developed another Python script that uses the Twitter API to obtain information about the collected tweets in the previous step using tweet IDs. In fact, we used the list of tweets' IDs from the collected tweets that we stored in a database as the reference to call the Twitter API and fetch the complete data available for tweets. To respect another Twitter API limitation on the volume of fetching data, we employed a Lambda function[14] that needs to be called with 450 IDs per hour in which we pass a list of tweets' IDs and Lambda function returns complete information about tweets.

STM publishes a series of tweets on different Twitter accounts, each corresponding to one of the four lines of the subway network providing real-time information during interruptions. Table 1 summarizes STM accounts. Through the aforementioned process, we scraped data from the web (using Twitter advanced search) to collect STM official tweets and then leverage the Twitter API to enrich the tweets by more fields of data (e.g. replies, retweets, responses, etc.) and aggregated all the available information for each and every tweets. A total of 22039 tweets were collected in the time period of 2015-01-01 to 2020-12-17.

[12] https://twitter.com/search-advanced
[13] https://pypi.org/project/GetOldTweets3/
[14] https://aws.amazon.com/lambda/

Table 1 STM official Twitter accounts

Account Name	Link	Subway Line
stm_Verte	https://twitter.com/stm_Verte	Green line
stm_Bleue	https://twitter.com/stm_Bleue	Blue line
stm_Orange	https://twitter.com/stm_Orange	Orange line
stm_Jaune	https://twitter.com/stm_Jaune	Yellow line 9[2]

3.2.2 STM Operational Data

We also had access to a collection of STM data including incidents, events and ridership reports. In particular, this dataset contained 5731 distinct incidents that occurred on metro lines: green, orange, blue, and yellow since 2015-01-01 until 2020-09-30 which consisted of 22 fields such as incident id, incident date, incident start hour, line, symptoms, cause, etc.

3.2.3 Joined Data

In addition, to find Tweets per incidents, as shown in the Figure 3, using the two datasets of incidents and tweets, a new dataset was created which contained tweets that were posted between 15 minutes before the start time of the incident and 1 hour and 45 minutes after the ending time of each and every incident. This joined dataset contains 14302 records corresponding to 1318 incidents with associated tweets. The objective of creating this joined dataset was to get a better understanding of the interrelationships between incidents, STM tweets and users' tweets and responses.

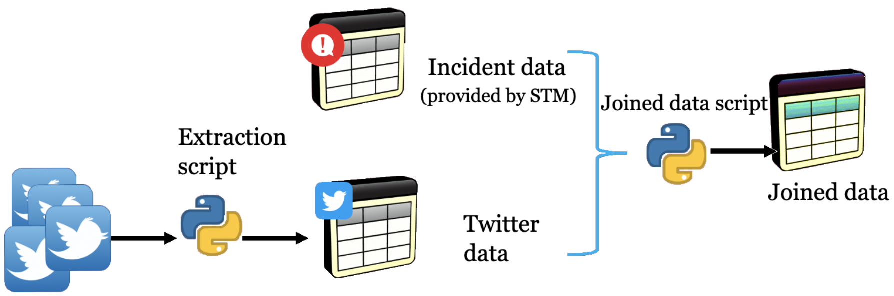

Fig. 3 Joined data schema

3.3 Pre-processing

Collected tweets were in JSON format, but we needed the data in relational tables in order to be able to join and perform unified search across all data. Therefore, we created a relational data store (AWS Relational Data Service

(RDS)[15]) which provides an interface to store and retrieve data in SQL-based format. In order to store the data in a proper format for further analysis, the following steps were taken:

- Removing URLs and hyperlinks such as words starting with "http" or "https".
- Removing special characters such as #, /, or – characters and punctuation signs.
- Removing stop words in both French and English Tweets.
- Removing emoticons as they represent only 0.2% of our data.
- Tokenization and n-grams[16] extraction to break Tweet's text into tokens and accompanying words in n-grams. It has to be mentioned that tokenization of social-media data is relatively more difficult than tokenization of the usual plain text documents. This is due to several specific attributes such as embedded links, icons and/or images that create spaces.
- Transforming and unifying date format to a proper format for tweets as well as for STM operational data.
- String manipulation such as converting all strings to lowercase.
- Text normalization (stemming and lemmatization) for replacing words with their roots. This helps to reduce the dimensionality of the representation matrix when different words, e.g. "write", "writer" and "writing" are mapped into one word "write" and are counted together.
- Assigning lines of the subway to each tweet regarding the STM account that publishes the tweet or retweets and replies to the STM accounts' tweets. For example, the line assigned to tweets published from STM_Verte is all Green and the line assigned to all retweets of such tweets is also Green. The line assigned to all replies to such tweets and retweets is also Green. The same line assignment goes for all four STM subway lines.
- Flatten JSON objects, as each tweet is represented as a nested JSON object with variety of data structure types such as list, dictionary, numbers and strings. We flattened all the tweets with a recursive Python solution to provide a robust way to analyze tweets. In other words, we try to provide consistency in our database and enhance its maintenance.

3.4 Topic Modeling

Topic modeling [32] [33] [34] as a category of text analytics contributes to the process of converting unstructured text into meaningful data for analysis to support fact-based decision-making.

Topic Modeling is used to organize a corpus of documents into "topics" which is a grouping based on a statistical distribution of words within the documents.

[15] https://aws.amazon.com/rds/
[16] an n-gram is a contiguous sequence of n items(words) from a given sample of text

The process of topic modeling is to infer hidden variables such as word distribution for all topics and topic mixture distribution for each document by observing the collection of documents. In other words, the backbone idea of topic modeling is that each topic is a distribution of words and each document is a mixture of topics across a set of documents. For example, a collection of documents that contains frequent occurrences of words such as "traffic", "car", "metro", "collision" or "brake" are likely to share a topic on "transportation". If another collection of documents shares words such as "workout", "match", "gym", or "hockey" it is likely that they are discussing a topic on "sport". The figure that follows shows the relationships among words, topics, and documents.

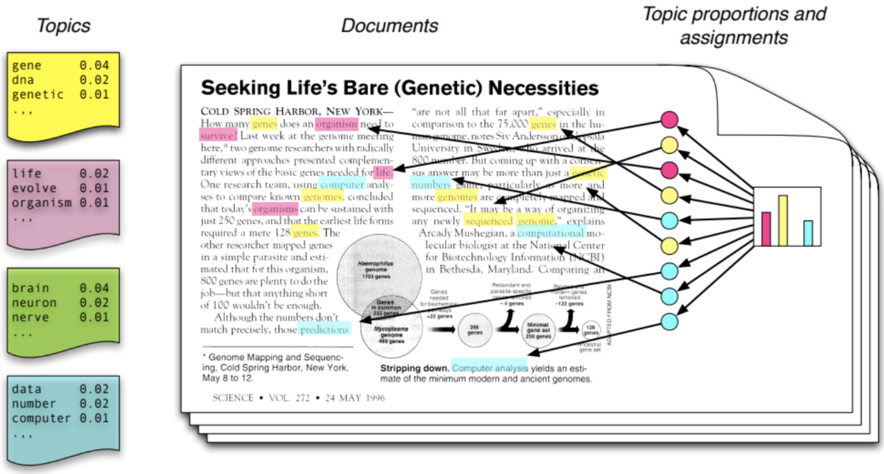

Fig. 4 Source: Blei, D.M. (2010) Probabilistic topic models communications of the ACM, 55(4) 77-84

Each "Topic" is a distribution over words. Each "Document" is a mixture of corpus-wide topics. Each "Word" is drawn from one of those topics.

Topic models are a classical example of probabilistic graphical models that involve challenging posterior inference problems. The Neural Topic Model is a topic modeling method based on Neural Variational Inference [35] [36] and can be used in cases where a fine control of the training, optimization, and/or hosting of a topic model is required, such as training models on text corpus of particular writing style or domain, or hosting topic models as part of a web application. In addition, Neural Topic Model AWS implementation [17] that we used for this research provides us with the flexibility to modify the network architecture as well as hyperparameters to accommodate the idiosyncrasies of

[17] https://docs.aws.amazon.com/sagemaker/latest/dg/ntm.html

the data sets as well as to tune the trade-off between a multitude of metrics such as document modeling accuracy, human interpretability and granularity of the learned topics, based on the application.
In Neural Topic Model the difficult inference problem is framed as an optimization problem solved by scalable methods such as stochastic gradient descent. Compared to conventional inference schemes, the neural-network implementation allows for scalable model training as well as low-latency inference.
Neural Topic Model takes the high-dimensional word count vectors in documents as inputs, maps them into lower-dimensional hidden representations, and reconstructs the original input back from the hidden representations. The hidden representation learned by the model corresponds to the mixture weights of the topics associated with the document. The semantic meaning of the topics can be determined by the top-ranking words in each topic as learned by the reconstruction layer. The training objective of this topic modeling is to minimize the reconstruction error and Kullback–Leibler divergence [37], the sum of which corresponds to an upper-bound on the negative log-likelihood of the data.
As an unsupervised generative model, the main indicator of model training progress is the training loss, which corresponds to the negative log-likelihood of data. To evaluate how well the trained model generalize to unseen data, when we train the model we always supply a validation data set so that the model training progress can be properly assessed and early stopping can be in effect to avoid overfitting. In other words, similar to all unsupervised learning methods, we do not have an accuracy or error metric to compare model training progress to the established prior expectations.
In addition to training loss, which measures how well the model describes and reconstructs data, for topic modeling on text, the top-N words representing each topic should be semantically meaningful, and thus human-interpretable. A sample of the output of the topic model inference is shown in the Figure 6.


```
{
    "precictions": [
        {"topic_weights": [0.02, 0.1, 0, ...]},
        {"topic_weights": [0.25, 0.067, 0, ...]}
    ]
}
```


Fig. 5 Model Inference Output Sample

The probabilities of topic predictions for two separate documents are shown above. The list of non-negative numbers represent the weights for each of the topics assigned to the document. These "topic_weights" represent the strength of topics in each document.

3.5 Sentiment Analysis

To infer sentiment in texts of users' messages, an automated machine learning sentiment analysis based on Amazon Comprehend [18] was performed, that helped to discover user opinion and the pattern of regular dissatisfaction time within a transit system, and also enabled the detection of unusual events occurring within the transit network such as train delays or failure of services.


```
{
"SentimentScore": {
        "Mixed": 0.030585512690246105,
        "Positive": 0.94992071056365967,
        "Neutral": 0.0141543131828308,
        "Negative": 0.00893945890665054
    },
    "Sentiment": "POSITIVE",
    "LanguageCode": "fr"
}
```


Fig. 6 Sentiment Analysis Output Sample

The sentiment analysis operation returns an object that contains the detected sentiment and a Sentiment Score object. The score represents the likelihood that the sentiment was correctly detected. The "Sentiment" is the most likely sentiment for the text and the numbers are the scores for each of the sentiment categories. For example, in the Figure 6 it is 95 percent likely that the text has a Positive sentiment. There is a less than 1 percent likelihood that the text has a Negative sentiment.
However, our results have been manually checked and corrected by subject matter experts at STM and we used the Sentiment Score to determine if the accuracy of the detection meets the needs of our application. Therefore, based on the STM feedback if the sentiment is detected as Neutral and the neutral score is over 0.5 and the negative score is greater than positive score we detect the sentiment as Negative. In addition, if the sentiment is detected as Neutral and the positive score is greater than 0.6 we detect the sentiment as Positive. As a result, 46% of tweets originally detected as Neutral became Negative and 2% fell under Positive category. All the tweets originally detected as Negative, Positive and the 48% of the rest of Neutral tweets remained under their original detected categories.

4 Experimental Results

Figure 7 (a), (b) and (c) depict the trend of monthly tweets published by STM, users' reaction and the number of distinct users per year, respectively.

[18] https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html

In total STM published 8233 tweets that invoked 13806 user reaction from 4096 distinct users over years from 2015 to 2020. In November 2019 we see a peak in user reaction where STM also communicated the most with users as seen in the Figure 7. It's worth mentioning that there were several severe incidents reported on STM networks on November 2019.

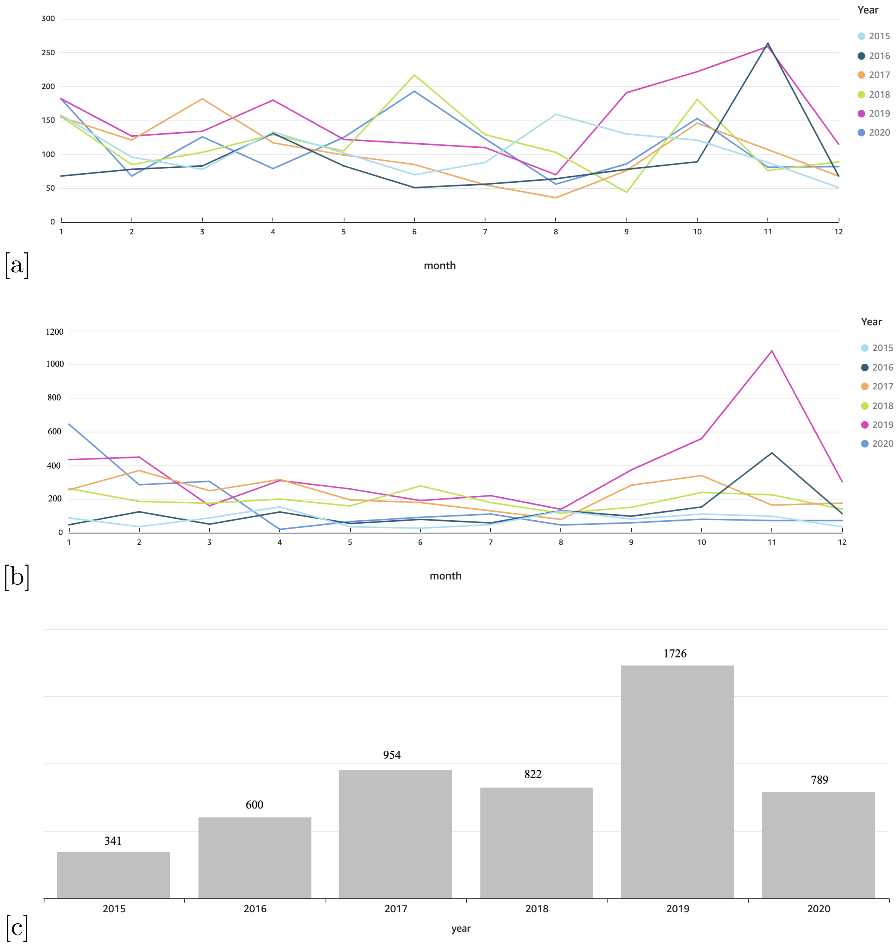

Fig. 7 (a) Number of monthly published tweets by STM. (b) Number of monthly users' reaction to STM tweets. (c) Number of distinct users per year interacting with STM on Twitter.

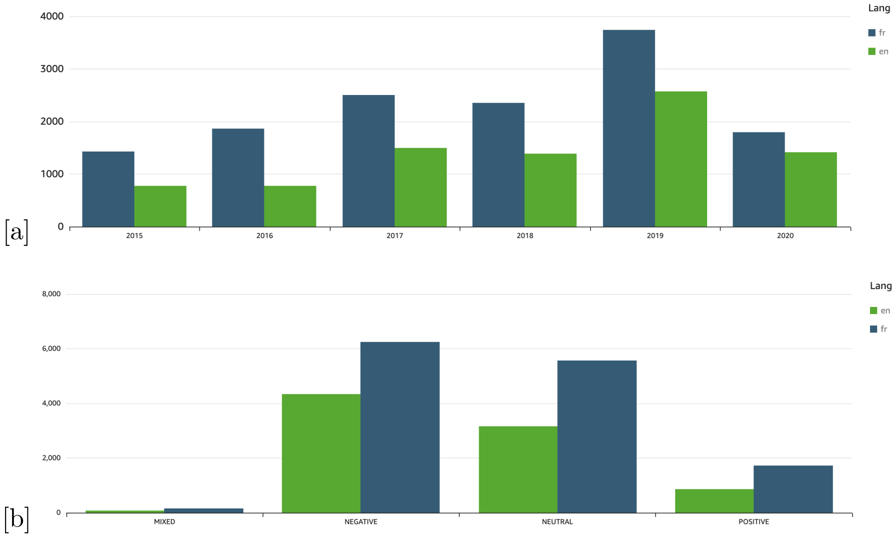

Fig. 8 (a) Distribution of tweets in both languages over years. (b) Distribution of tweets in four sentiment categories per language.

The project was carried out in a bilingual urban community with a French-speaking majority population and as Figure 8 shows majority of tweets are in French language. Also tweets in all categories are mostly in French as shown in Figure 8. In addition, we see more negative tweets in general compared to other sentiment categories. This finding is aligned with the assumption that users usually tweet complaints when they interact with public transit official accounts [13].
Figure 9 (a) and (b) illustrates the incidents on the STM lines and the tweets posted related each line of metro along with user reactions. Although the number of incidents that occurred in the Orange and Green lines were similar, we see much more tweets related to the Orange line than to the Green line. That might suggest that because the Orange line is the longest metro line amongst STM network and it serves more passengers than Green line, incidents on Orange line might provoke more user interaction than incidents on other metro lines. Hence, to have a more definitive understanding, we can normalize the number of tweets by the length of the line, or by traffic over a given period.

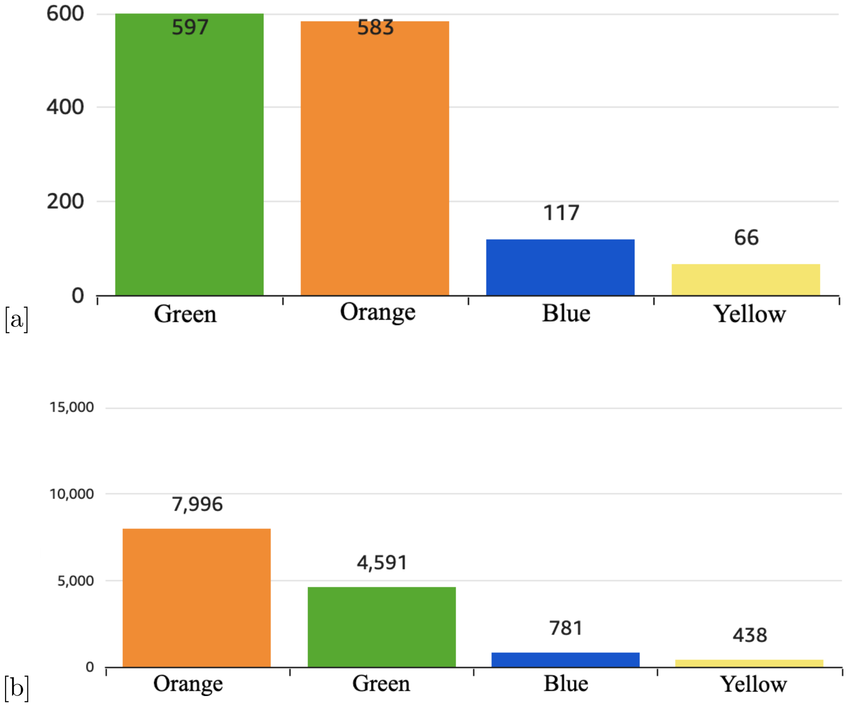

Fig. 9 (a) Incidents per lines of metro. (b) Tweets per lines of metro.

Table 2 Topics and most probable terms shown for each topic detected by Neural Topic Model

Topic	High probability terms
Normal service	service, normal, métro,stminfo, metro
Planned service interruption	stop, prévue, arrêt, delay, problème
Service slowdown	ralentissement, slowdown, minute, breakdown
Gradual resume	resume, métro, reprise, ligne, graduelle
Service back to normal	normal, état, station, covid19, state

Table 2 shows five inferred topics that have been detected by Neural Topic Model algorithm along with the five terms per topic with highest probability. We linked topics with subjects in subway transportation based on the word distribution. The results of topic modeling shows that tweets are not only related to the incidents and interruptions but are also related to normal operation of transit services as well.
The number of users' reaction for each topic is normalized by dividing by the number of official STM tweets per topic. Figure 10 (a) depicts more tweets

related to service slowdown which is aligned with STM incident data that reports more incidents occurred in 2019. As expected and also confirmed by figure 10 (b) the less positive and also one of the most negative user reaction belongs to the "service slowdown" topic. In addition, the most positive user reaction falls under "planned service interruption" which suggests that passengers feel less negative towards service interruption when the interruptions are communicated with them beforehand. Furthermore, we have more "gradual resume" followed by "service slowdown" on Orange line and Green line which are respectively the two longest metro lines in STM network.

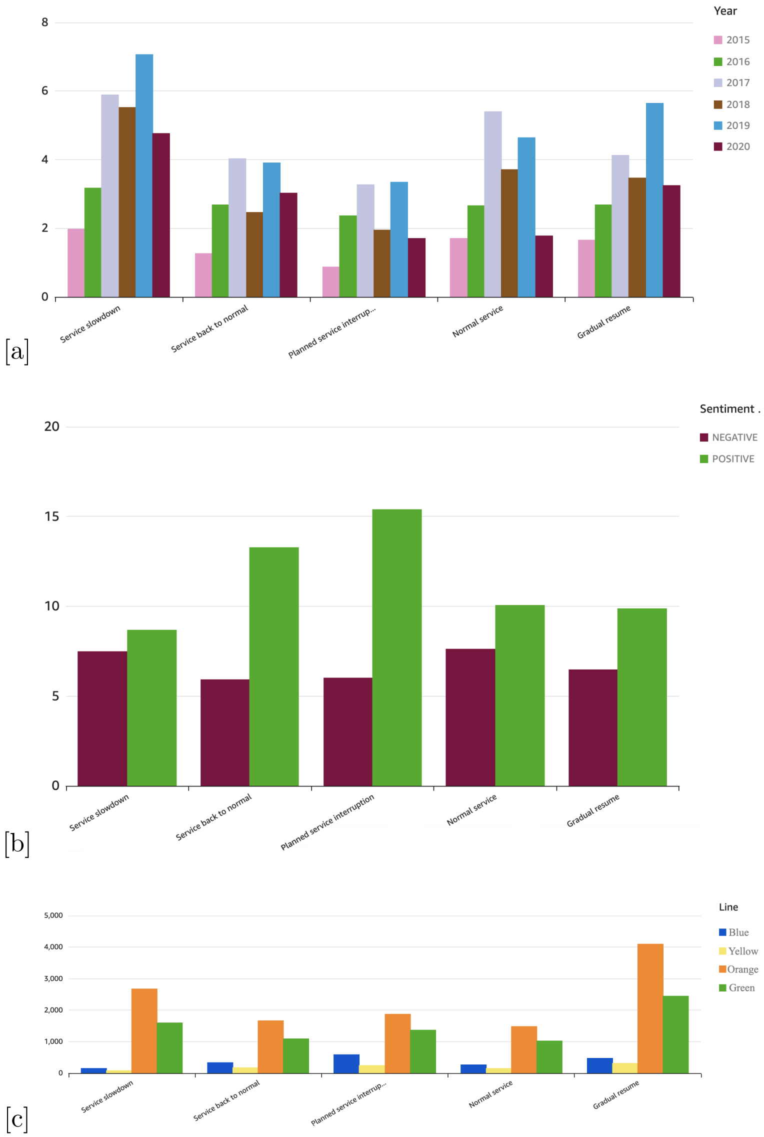

Fig. 10 (a) Distribution of topics in users' reaction over years. (b) Negative and Positive users' sentiments per topic. (c) Distribution of tweets over topics and metro line .

Regarding the associated circumstances (day of the week, time, etc) figure 11 (a) shows the distribution of STM incidents per hour and figure 11 (b) depicts the distribution of STM tweets and users' reaction. In both incidents and tweets, we see a peak around 8:00 am and 5:00 pm which the sentiment of users' reaction on those periods of the day is mostly negative, as figure 11 (c) shows. In addition, Saturday and Sunday are less disturbed days in both STM incident and tweets. However, the weekdays with the most reported STM incidents are not exactly the same as the weekdays that most communication from STM and most reactions from users occurred. This suggests that not all the incidents in STM database are considered as eligible to be communicated with users. Hence, not much reaction from users as well.

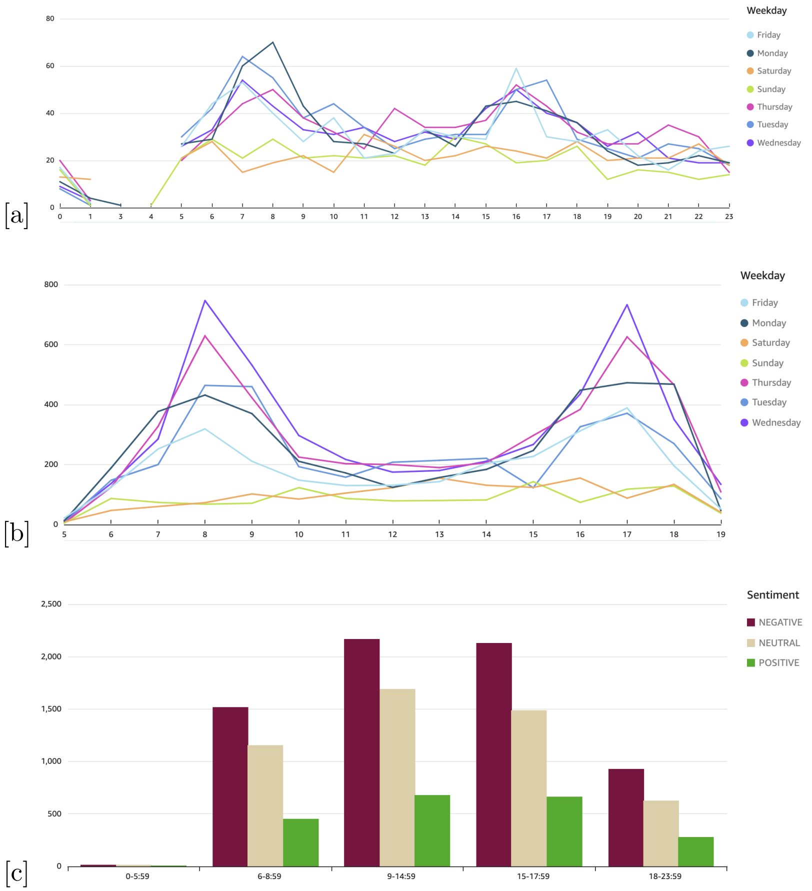

Fig. 11 (a) Incidents per hour. (b) Tweets per hour. (c) Sentiment category of users' reaction per period of the day.

5 Conclusion

In this study we examined the possibility of social media data usage in addition to traditional data sources in transit studies to model public transit scenarios.

In this framework, Twitter data is notable for its capacity to provide useful information about public transit users perceptions and sentiments over different topics. In this study we worked with Montreal transportation (STM) official tweets as well as users' reply and retweets. We mapped the distribution of metro users with incidents reported in the STM network, whereas prior publications are mostly focused only on sentiment analysis and not concerned with cross-checking the findings with spatial dimension in terms of reported incidents in the transit networks. In this investigation, we applied a sentiment analysis algorithm to our data and customized the probability scores of the algorithm to our unique application based on subject matter experts' feedback at STM. In addition to the sentiment analysis on users' tweets, this study has investigated the latent topics that users discuss about and react to the most, via Neural Topic Model.

Furthermore, we searched to extract information out of users' reaction such as circumstances like the day of the week, time, etc. Findings confirmed that users interact less with STM during weekends than during weekdays. Two peak periods during the weekdays in terms of users' reaction in general and negative tweets in particular coincide with the time when passengers travel from home to work and vice versa which is consistent with the incidents reported on STM official data as well. These insights might be utilized by STM to enhance services in the mentioned periods of time. Using social media data and combining its insight with traditional transit data sources can provide a new perspective on public transit issues and enable fast and focused answers.

Highly accurate sentiment analysis algorithms on users' tweet may help transit agencies to better understand the emotions, expectations and requirements of Metro passengers. The discovery of topics on users' tweet may enable better understanding of subjects that make metro passengers react the most and the least due to changes in the volume of users' tweet per topic regarding transit services. This information is necessary for accurate evaluation of temporal patterns and consumer feedback.

Although in this research we revealed the potential capacity of Twitter data for transit studies and the ability of advanced natural language processing techniques to extract useful information out of this textual data, only non geotagged twitter data was used. However, in order to perform a more granular spatial analysis, geolocating the tweets in the study area is required. Using geotagged tweets can help to find the location and name of stations on the metro line where the incident occurred or where an important amount of negative scored tweets are published. Therefore future works include the usage of geotagged data and/or extracting geo-locations from the text of user tweets to also investigate spatiotemporal tweeting patterns of passengers and their feedback. Additionally, the proposed approach could also be applied to other transit services (bus, car-sharing, bike-sharing, etc.), which could contribute to the improvement of these services based on the feedback of their users. Therefore, applying the same frame work to other transit means and comparative analysis of the public metro network with other public transit systems is part of the future investigations.

References

[1] A. M. Kaplan and M. Haenlein, Users of the world, unite! the challenges and oppor- tunities of social media, Business horizons, vol. 53, no. 1, pp. 59–68, 2010.

[2] M. F. Goodchild, Citizens as sensors: the world of volunteered geography, GeoJour- nal, vol. 69, no. 4, pp. 211–221, 2007.

[3] S. Goldsmith, La's testing ground for transportation efficiency, mar. 2016, 2017.

[4] R. Wei et al., Evaluating public transit services for operational efficiency and access equity, Journal of transport geography, vol. 65, pp. 70–79, 2017.

[5] S. K. Fayyaz, X. C. Liu, and R. J. Porter, Dynamic transit accessibility and transit gap causality analysis, Journal of Transport Geography, vol. 59, pp. 27–39, 2017.

[6] K. Group et al., Transit capacity and quality of service manual, 2013.

[7] N. Wanichayapong et al., Social-based traffic information extraction and classifica- tion, in 2011 11th International Conference on ITS Telecommunications. IEEE, 2011, pp. 107–112.

[8] R. Kosala and E. Adi, Harvesting real time traffic information from twitter, Procedia Engineering, vol. 50, pp. 1–11, 2012.

[9] H. Gao, J. Tang, and H. Liu, Exploring social-historical ties on location-based social networks, in Sixth International AAAI conference on weblogs and social media, 2012.

[10] R. Steur, Twitter as a spatio-temporal source for incident management, Master's thesis, 2015.

[11] Y. Tian et al., Quality assessment of social media traffic reports-a field study in austin, texas, Tech. Rep., 2016.

[12] S. A. Golder and M. W. Macy, Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures, Science, vol. 333, no. 6051, pp. 1878–1881, 2011.

[13] D. Tasse and J. I. Hong, Using social media data to understand cities, 2014.

[14] Z. Yin et al., A scalable framework to detect personal health mentions on twitter, Journal of medical Internet research, vol. 17, no. 6, p. e4305, 2015.

[15] E. Cho, S. A. Myers, and J. Leskovec, Friendship and mobility: user movement in location-based social networks, in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, pp. 1082–1090.

[16] Z. Cheng et al., Exploring millions of footprints in location sharing services, in Pro- ceedings of the International AAAI Conference on Web and Social Media, vol. 5, no. 1, 2011.

[17] S. Hasan, X. Zhan, and S. V. Ukkusuri, Understanding urban human activity and mobility patterns using large-scale location-based data from online social media, in Proceedings of the 2nd ACM SIGKDD international workshop on urban computing, 2013, pp. 1–8.

[18] T. Sakaki, M. Okazaki, and Y. Matsuo, Earthquake shakes twitter users: real-time event detection by social sensors, in Proceedings of the 19th international conference on World wide web, 2010, pp. 851–860.

[19] B. R. Lindsay, Social media and disasters: Current uses, future options, and policy considerations, 2011.

[20] E. Mai and R. Hranac, Twitter interactions as a data source for transportation inci- dents, Tech. Rep., 2013.

[21] S. V. Ukkusuri et al., Use of social media data to explore crisis informatics: Study of 2013 oklahoma tornado, Transportation Research Record, vol. 2459, no. 1, pp. 110–118, 2014.

[22] K. Fu, R. Nune, and J. X. Tao, Social media data analysis for traffic incident detection and management, Tech. Rep., 2015.

[23] C. Collins, S. Hasan, and S. V. Ukkusuri, A novel transit rider satisfaction metric: Rider sentiments measured from online social media data, Journal of Public Trans- portation, vol. 16, no. 2, p. 2, 2013.

[24] L. Schweitzer, Planning and social media: a case study of public transit and stigma on twitter, Journal of the American Planning Association, vol. 80, no. 3, pp. 218–238, 2014.

[25] E. Steiger, T. Ellersiek, and A. Zipf, Explorative public transport flow analysis from uncertain social media data, in Proceedings of the 3rd ACM SIGSPATIAL interna- tional workshop on crowdsourced and volunteered geographic information, 2014, pp. 1–7.

[26] T. T. Luong and D. Houston, Public opinions of light rail service in los angeles, an analysis using twitter data, IConference 2015 Proceedings, 2015.

[27] A. Gal-Tzur et al., The potential of social media in delivering transport policy goals, Transport Policy, vol. 32, pp. 115–123, 2014.

[28] T. El-Diraby, A. Shalaby, and M. Hosseini, Linking social, semantic and sentiment analyses to support modeling transit customers' satisfaction: towards formal study of opinion dynamics, Sustainable Cities and Society, vol. 49, p. 101578, 2019.

[29] F. Ali et al., Transportation sentiment analysis using word embedding and ontology- based topic modeling, Knowledge-Based Systems, vol. 174, pp. 27–42, 2019.

[30] N. Korfiatis et al., Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews, Expert Systems with Applications, vol. 116, pp. 472–486, 2019.

[31] B. Qi, A. Costin, and M. Jia, A framework with efficient extraction and analysis of twitter data for evaluating public opinions on transportation services, Travel behaviour and society, vol. 21, pp. 10–23, 2020.

[32] H. M. Wallach, Topic modeling: beyond bag-of-words, in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 977–984.

[33] D. M. Blei, Probabilistic topic models, Communications of the ACM, vol. 55, no. 4, pp. 77–84, 2012.

[34] L. Hong and B. D. Davison, Empirical study of topic modeling in twitter, in Proceedings of the first workshop on social media analytics, 2010, pp. 80–88.

[35] D. P. Kingma and M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114, 2013.

[36] D. J. Rezende, S. Mohamed, and D. Wierstra, Stochastic backpropagation and ap- proximate inference in deep generative models, in International conference on machine learning. PMLR, 2014, pp. 1278–1286.

[37] S. Kullback and R. A. Leibler, On information and sufficiency, The annals of math- ematical statistics, vol. 22, no. 1, pp. 79–86, 1951.