CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montréal)
June 12 - 15, 2019

VISUALIZATION OF LOCAL MUNICIPAL SATISFACTION BY TWITTER DATA ANALYSIS

Zarei, F. [1] Nik-Bakht, M.[2] Hammad, A.[3]
Concordia University, Canada
[1] f_zaire@encs.concordia.ca
[2] mazdak.nikbakht @concordia.ca
[3] amin.hammad@concordia.ca

Abstract: One of the most important factors in evaluating the performance of boroughs in a city is the satisfaction level of the residents with respect to municipal services. Social networks provide an extensive platform for local residents and communities to express their ideas about the level of service they receive, their problems and their expectations. In this paper, the level of satisfaction of people in different boroughs of Montréal is quantified by analyzing their tweets. In this aim, 3,293 tweets from January 2009 to March 2018 were collected. Then, the implicit viewpoint behind each of them was evaluated via human knowledge approaches. Hence, not only a valid data set was obtained for future automation, but also sufficient information was gained for visualizing and comparing the levels of satisfaction of several services in different boroughs. The results show that Verdun Borough provides more satisfaction to its residents. In spite of the limitations of the heterogeneity of Twitter users and their spread in a city, and the biases that might be associated with the tweets' analytics, our results can add an extra layer to the evaluation of the level of service in cities. Overlaying such results with other performance indicators can lead to significant lessons for governance of smart cities.

1 INTRODUCTION

Nowadays, with the aid of social networks, large amounts of data about the opinions of people regarding different topics are available (Androniceanu 2017). Social media platforms, including Twitter and Facebook, enable people to share their ideas and express their opinions about the issues in which they are involved. Therefore, there is an opportunity for data scientists and researchers to analyze such data and gain beneficial information about the public opinion in different domains. Such information is useful for the governments to recognize and interpret the public demands and expectations (Bello et al. 2017).

Generally, the role of the municipality is providing the services which are necessary and emergent for the communities. These services should be provided for the citizens either by municipality or other organizations. In fact, the municipal services can be defined as complex systems and activities which help people to deal with physical and nonphysical problems (Goodchild, 2007). However, defining the exact definitions of these kinds of services is not always an easy task. There are different services based on the rules that municipal representatives should meet to capture citizens' expectations. Citizen's expectations from the municipalities are increased by the advantage of technology, and this pushes the municipalities to provide better services to meet the new expectations.

In this paper, considering the satisfaction level of municipal services, it is desired to understand if there are significant differences among residents of different boroughs of Montréal. Montréal, the second largest city of Canada and one of the most important cities in Québec, is divided into 19 boroughs (ville.Montréal.qc.ca)

The word "borough" is usually used for the governmental division of a city. Table 1 demonstrates some properties of different boroughs in Montréal including population, area in km^2 and density per km^2 ("Population totale en 2011 et en 2016 - Variation — Densité"). In Montréal, each borough municipality is expected to respond to people's demands such as fire prevention, removal of household wastes and residual materials, funding the communities, social and local economic development agencies, and planning and management of the different aspects of the city (Mitchell et al. 2013).

Table 1 Some information about Montréal boroughs

Borough	Population	Area (in km^2)	Density (per km^2)
Ahuntsic-Cartierville	126,891	24.2	5,252.10
Anjou	41,928	13.7	3,064.90
Côte-des-Neiges–Notre-Dame-de-Grâce	165,031	21.4	7,697.30
Lachine	41,616	17.7	2,348.50
LaSalle	74,276	16.3	4,565.20
Le Plateau-Mont-Royal	100,390	8.1	12,348.10
Le Sud-Ouest	71,546	15.7	4,562.90
L'Île-Bizard–Sainte-Geneviève	18,097	23.6	766.8
Mercier–Hochelaga-Maisonneuve	131,483	25.4	5,174.50
Montréal-Nord	83,868	11.1	7,589.90
Outremont	23,566	3.9	6,121.00
Pierrefonds-Roxboro	68,410	27.1	2,528.10
Rivière-des-Prairies–Pointe-aux-Trembles	106,437	42.3	2,517.40
Rosemont–La Petite-Patrie	134,038	15.9	8,456.70
Saint-Laurent	93,842	42.8	2,194.10
Saint-Léonard	75,707	13.5	5,612.10
Verdun	66,158	9.7	6,809.90
Ville-Marie	84,013	16.5	5,085.50
Villeray–Saint-Michel–Parc-Extension	142,222	16.5	8,624.70

This paper aims to obtain local municipal satisfaction in Montréal. Generally, previous studies in this area used surveys to find out the opinion of the residents. Furthermore, a Matlab script is developed to visualize the ranking of regions through color coding. The result of this research can be used as a training dataset for sentiment analysis tasks within this scope.

For analyzing people's satisfaction about the environmental services, such as water and sewer system, cleaning service, green fields and parks, and traffic management and control in each borough, it is proposed to gather the shared contents of twitter which are about the municipal performance of Montréal. These data should be gathered via Twitter API. The collected data should include some keywords indicating the relevance of the shared content to the municipal issues, as well as being related to Montréal. In the next step, by applying the human knowledge methods, it is investigated if each post indicates a positive or a negative opinion about the municipal performance. Then, based on the location of the posts, they are partitioned to different municipal boroughs. Finally, at each borough, the percentage of positive opinions is evaluated.

After the fulfillment of the data analysis described above, visual representation of the citizens satisfaction of different boroughs are presented using Matlab. Using this software, a map of Montréal with its different boroughs is presented, where each borough is depicted with the color which illustrates the relative success of the local municipality. These results illustrate the public satisfaction, which could be beneficial for house seekers to choose their residential place and also for the municipalities to amend their performance. Moreover, it can be useful to find out the demands in each borough of the city. The remaining of this paper is organized as follows: We begin by discussing about tweet sentiment analysis in Section 2. Preliminaries and a formal problem definition are presented in Section 3. Section 4 presents the obtained results, and finally, section 5 concludes the paper.

2 OPINION MINING IN SOCIAL NETWORK

Opinion mining or sentiment analysis is the scope of analyzing people's opinion, evaluations, attitudes, appraisals, sentiments, and emotions towards entities (Bollen et al. 2011). One of the important methods for analyzing social media data in microblogging, online communities, blogs, and the other online collaborative media is sentiment analysis (Adnan et al. 2013). Moreover, It is an important area in computing research (Piyawongwisal et al. 2011). The main task of sentiment analysis is classifying the texts into two (positive or negative) or sometimes three (positive, negative or neutral) different classes. Although most of the literature is for English texts, there are some multilingual studies dealing with sentiment analysis. Topical and sentiment analysis of social media content is also used for evaluating urban infrastructure systems (among other services) in recent years. The majority of applications in this regard have been on the micro-blogging website twitter. Community engagement for transportation planning (Evans-Cowley & Griffin, 2012, Casasa & Delmelleb, 2017); Evaluation of the public feedback on major infrastructure construction projects (Nik-Bakht & El-Diraby, 2017) and analysis of urban infrastructure sustainability via public tweets (Nik-Bakht et al., 2018) can be mentioned among other major applications.

Generally, sentiment analysis research works could be classified into two principal groups which are: knowledge-based and statistics-based. Such analysis needs to deal with many NLP tasks, such as named entity recognition, concept extraction, sarcasm detection, aspect extraction, and subjectivity detection. Hence, there are several challenges which should be solved in the sentiment analysis.
In this paper, the human knowledge method is used due to the lack of proper training data in the scope of resident satisfaction. The training data play an important role in the accuracy of the sentiment analysis. Moreover, applying sentiment analysis by general training data provides low accuracy in the results.

3 MUNICIPAL SATISFACTION EVALUATION BY TWITTER

This section explains the used methods, which include: gathering the tweets of different boroughs, selecting the tweets which are relevant to the satisfaction of residents, extracting their viewpoints, calculating the total satisfaction of different boroughs by extracted viewpoints, and finally visualizing the obtained results. As mentioned in the previous sections, in this paper, we focused on the satisfaction of residents in different boroughs of Montréal about some topics, which are cleaning services, green fields, parks and traffic management (Casagrande et al. 2018). For this aim, the following three main steps were taken:

- Tweet gathering
- Viewpoint analysis
- Satisfaction Visualizer

The rest of this section discusses these three main steps.

3.1 Tweet gathering

To collect the relevant tweets, in the first step, tweets which contained the name of the boroughs in Montréal were collected. By crawling Twitter using Twitter API, we collected our dataset. This data set consists of the tweets, their authors, the date and time of the tweets, hashtags and some other properties of the tweets provided by the API. The resulted data set contains tweets from January 2009 until March 2018 whose hashtags include the name of the boroughs of Montréal (e.g. #Lasalle or #Lachine).

3.2 Viewpoint analyzer

In this section, we filter-out those tweets which are not relevant to the municipal issues. Given the scope of our study, topics such as cleaning service, green fields and parks and the traffic management are selected in this step. Then, we attempt to classify the tweets based on their sentiment in three classes, which are positive, negative and neutral. The satisfaction measure used for polarity detection is as follows:

$$\text{Satisfaction Measure} = \frac{\#\ \text{Positive tweets} - \#\ \text{negative tweets}}{\#\ \text{relevant tweets}}$$

3.3 Satisfaction visualizer

The satisfaction visualizer provides a platform to visualize the satisfaction of different boroughs on a map via color coding. In the provided map, boroughs with more saturated green color are those with higher levels of satisfaction. On the contrary, boroughs with more saturated red color are the boroughs with higher levels of dissatisfactory. The visualizer can be adapted by other cities to assess the level of satisfaction of the residents in different boroughs.

4 MODEL AND RESULTS

In this step, it is explained how the data are gathered, how they are processed and what are the obtained numerical results. Finally, there is a result analysis part at the end of this section.

4.1 Data Collection

Tweets from people of different boroughs of Montréal are gathered by the aid of the popular Python API, *tweepy*. This API is able to search tweets based on their geo tag (latitude and longitude) of the tweets. In the first step, we searched tweets by the exact latitude and longitude of Montréal with an accuracy and granularity specified in meters. The results, however, were not satisfactory. One main reason was that the majority of people turn off their device location when they share a post in twitter. Twitter uses a combination of device/GPS coordinates, user-provided profile location, and network/IP address location. However, if people turn off their GPS or do not allow Twitter to find their location, the geo information of the place from where that tweet was posted is not accessible. To resolve this problem, hashtags associated with different boroughs of Montréal were considered as the inputs for this API.

At the end of this step, about 10,000 tweets were collected. However, not all of them were directly relevant to our scope (i.e. the satisfaction of residents about their boroughs). Gathered tweets were in different languages, but the majority was in English and French. So, we filtered the tweets and selected those which were in either of these two languages. As the result, about 2,163 English and 1,129 French tweets were passed to the next step.

4.2 Analyzing the tweets

To analyze the gathered tweets, we applied expert knowledge-based approach. Tweets were categorized as: irrelevant tweets; relevant to citizen’s satisfaction with a neutral view point; relevant to citizen’s satisfaction with a positive view point; and relevant to citizen’s satisfaction with a negative view point. In spite of the high accuracy of this approach, it is a time-consuming process since the speed of human is limited. After such a classification, only the relevant data to municipal services were considered. By assigning a score of +1 for positive view point; -1 for negative view point; and 0 for natural viewpoints; the relative satisfaction of each borough is calculated.

As shown in Table 1, “Côte-des-Neiges–Notre-Dame-de-Grâce” is the most populated borough in Montréal while “Saint-Laurent” and “Le Plateau-Mont-Royal” are the most expanded and dense boroughs in Montréal, respectively.

Table 2 provides some examples of positive, negative or neutral tweets. Table 3 illustrates the numeric details of sentiment analysis for different boroughs including the number of related, positive, negative and neutral tweets for each borough. We can determine that “Saint-Laurent” has the highest number of related tweets while “Villeray–Saint-Michel–Parc-Extension” has the lowest rank in this term. We can compare the difference between the number of positive, negative and neutral tweets for each borough to determine the citizen’s satisfaction. Figure 1 demonstrates this difference for each borough.

Table 2 Examples of collected tweets

Sentiment	Examples
Positive	“#Anjou Borough mayor argues his local officials and workers are faster/more efficient at snow removal than central city of #Montréal #polmtl pic.twitter.com/Kuoyv9Wnjc
Positive	“The lowest price for gas in #LaSalle #Montréal for a VERY LONG TIME pic.twitter.com/WttraC78X5”
Negative	“help! Snow removal in Lasalle is a disgrace! 8th Avenue had one snow removal since the 1st snow storm 8 days ago. No where to park. What is going on? What a mess! #Snowmageddon #Lasalle #Montréal #Cyrenne”, “Highway 40 East closed after #TractorTrailer crash in #Anjou http:// globalnews.ca/news/2516164/highway-40-east-closed-after-tractor-trailer-crash-in-anjou/”
Negative	“Construction on #ruedollard in #LaSalle dragging on #traffic guys at every corner #cityscape #urbanphoto #Montréal infrastructure pic.twitter.com/a9YmmhueuF”
Neutral	“The Last February Storm. #snow #neige #winter #hiver #canada #Québec #Montréal #anjou #dannyvb #trees http:// flic.kr/p/dYLFsm”
Neutral	“Earthquake in #Anjou #Montréal”

Table 3 Results of sentiment analysis for each boroughs

Borough	Related Tweets	Positive	Negative	Neutral
Ahuntsic-Cartierville	196	62	104	30
Anjou	57	12	37	8
Côte-des-Neiges–Notre-Dame-de-Grâce	262	130	113	19
Lachine	239	77	145	17
LaSalle	351	98	229	24
Le Plateau-Mont-Royal	315	150	134	31
Le Sud-Ouest	61	26	35	0
L'Île-Bizard–Sainte-Geneviève	67	18	40	9
Mercier–Hochelaga-Maisonneuve	168	73	77	18
Montréal-Nord	110	45	44	21
Outremont	159	63	81	15
Pierrefonds-Roxboro	126	61	38	27
Rivière-des-Prairies–Pointe-aux-Trembles	63	19	28	16
Rosemont–La Petite-Patrie	18	7	9	2
Saint-Laurent	389	165	201	23
Saint-Léonard	233	85	121	27
Verdun	297	180	71	46
Ville-Marie	167	56	98	13
Villeray–Saint-Michel–Parc-Extension	14	5	9	0
Total	3292	1332	1614	346

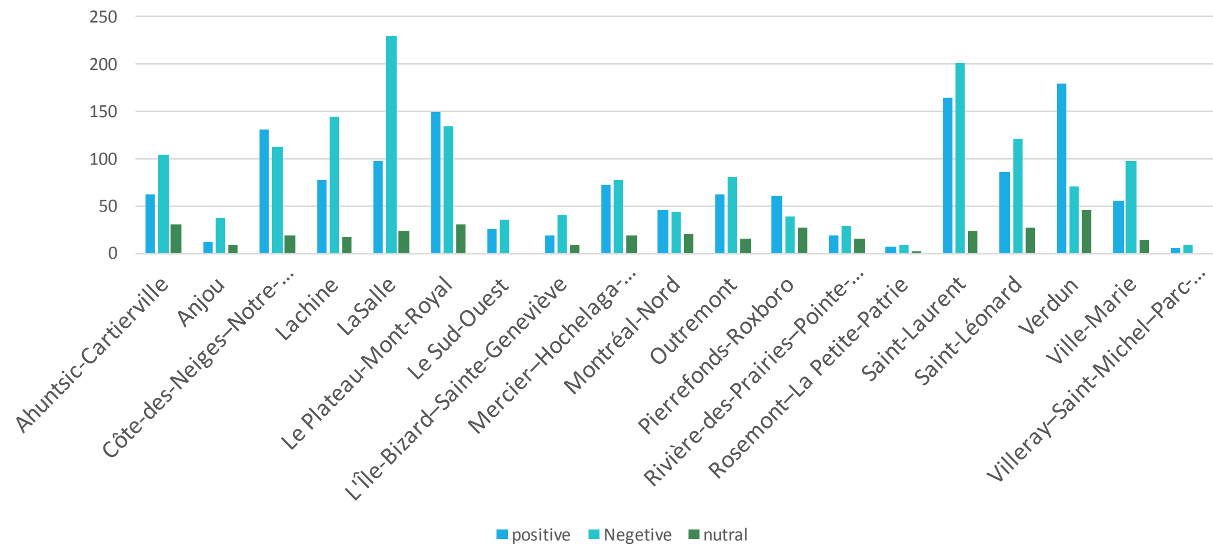

Figure 1 Montréal boroughs sentiment analysis results

4.3 Visualization of the results

After calculating the relative satisfaction of each borough, the results are encoded as colors and are visualized by the aid of Matlab. Figure 2 presents visual presentation of relative satisfaction for different boroughs, obtained by the expert knowledge-based approach. The red color represents the low level of satisfaction of citizens in different boroughs while the green color illustrates the high level of satisfaction during 9 years, from January 2009 to March 2018.

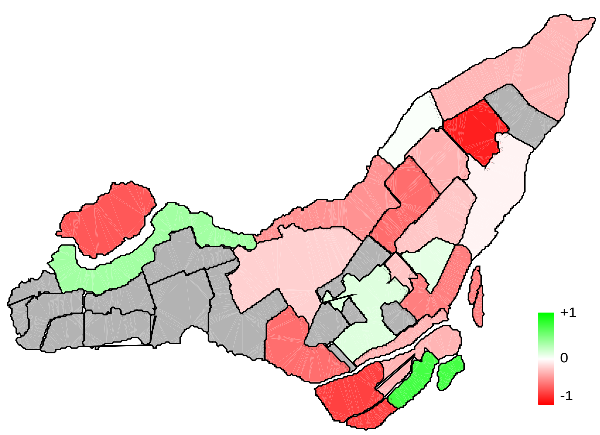

Figure 2 Visual presentation of relative satisfaction of different boroughs obtained by expert knowledge-based approach

4.4 Word cloud analysis

To analyze the results, we tried to determine frequent words in tweets, and as a result, we could find out the main concerns of citizens in different boroughs. This helps the municipalities to better understand their citizen's demands and also evaluate their own performance. In this way, all the hashtags and words which were used for extracting tweets were omitted, and then tokenization was done for the rest of the words using NLTK (Bird, 2006) to find out the frequency of these words. We considered those words with the frequency of 100 and above as frequent words and then we demonstrated them as a word cloud. In this way, the terms are colored based on their frequency; thus, we can easily determine the high-frequency words. The result is shown in Figure 3. As seen, “metro”, “bus”, “fire”, “neige”, “clear” are the most frequent words that citizens in Montréal talked about in the collected tweets which are mostly related to transportation.

Figure 3 word cloud analysis for extracted tweets

5 CONCLUSION AND FUTURE WORK

Analyzing social opinion about the municipal services is one of main influencing criteria in their management and planning, as well as providing new services. However, it is a challenging process to gather valid samples of social opinion to fulfill the analysis in this aim. Nowadays, with the emergence of social networks, a powerful framework for gathering a social opinion is available.

In this paper, we gathered the tweets with hashtags of different boroughs in Montréal, and analyzed them to determine whether they indicate citizen’s satisfaction; and if they did, whether they express a positive or negative view-point. We realized such sentiment analysis by the expert knowledge approach. As the result of the procedure, it was determined that the residents of Verdun Borough have the highest level of satisfaction from the municipal services; while the residents of Anjou have the less satisfaction level regarding these services. The main contributions of this paper can be summarized as: (i) providing a novel approach, which can replace commonly used surveys, to obtain municipal satisfactions through the residents; (ii) the results of this research can be used as a training dataset for sentiment analysis tasks in this scope; and (iii) the Matlab script which was developed can visualize the ranking of regions through color-coding.

Following this research, it is proposed to develop a framework which gathers tweets in a real-time procedure and demonstrate the satisfaction of people in different boroughs. Accordingly, by the aid of such framework, it would be possible to amend the performance of weaker boroughs in a more responsive way. Moreover, it is suggested to apply the generated training data to make an automatic sentiment analyzer for evaluating the satisfaction of the residents of different boroughs. Another extension is to consider the timestamp of tweets to track the variation of satisfaction of people in different boroughs.

REFRENCES

A. ANDRONICEANU. 2017. "Improving Citizens' Satisfaction Concerning the Social Welfare Services at Urban Level" *Research Center in Public Administration and Public Services*, **vol. 12**, pp. 67-82.

Bello MU, Martin D, Kasim R, 2017, " A Review of Effects of Quality Municipal Services on Citizen Satisfaction in Malaysian Municipal Council: A Content Analysis ", *Glob J Res Rev*, **vol. 4**, no. 3.

Bollen, J., Pepe, A., & Mao, H. 2011. "Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena ", in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (ICWSM , 2011), Barcelona.

Casasa, I., Delmelle E. C. 2017 "Tweeting about public transit — Gleaning public perceptions from a social media microblog" *Case Studies on Transport Policy*, **vol. 5 (4)**, PP 634-642

D. Montalvo, "Citizen Satisfaction with Municipal Services," *LAPOP- AMERICASBAROMETER*, **vol. 1**, pp. 75-80, 2009.

Evans-Cowley, J. S., & Griffin, G. 2012. "Microparticipation with social media for community engagement in trasportation planning". 2307, 90-98

G. Evans, 2009, "Creative Cities, Creative Spaces and Urban Policy," *Urban Studies*, **vol. 46**, no. 5-6, pp. 1003 - 1040.

Hana, A., Akram, S., & Abd El-Aziz, A. 2016. "A literature review on twitter data analysis”, *International Journal of Computer and Electrical Engineering*, **vol. 8**, no. 3, pp. 241-249.

http://ville.Montréal.qc.ca/portal/page?_pageid=5977,86679930&_dad=portal&_schema=PORTAL

M. Goodchild. 2007, "Citizens as Sensors the World of Volunteered Geography". *GeoJournal*, **vol. 69**, p. 211–221.

Macke, J., Casagrande, R.M., Sarate, J.A.R., Silva, K.A. 2018. "Smart city and quality of life: Citizens’ perception in a Brazilian case study ", *Journal of Cleaner Production*, **vol. 182**, pp. 717-726.

Madgin, R., Webb, D., Ruiz, P., Snelson, T. 2018. "Resisting relocation and reconceptualising authenticity: the experiential and emotional values of the Southbank Undercroft*," International Journal of Heritage Studies*, **vol. 24**, no. 6, pp. 585-598.

Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. 2013. "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place," *PLoS ONE*, **vol. 8**, no. 5, p. e64417.

Muhammad Adnan, Paul A. Longley, Shariq M.Khan. 2013. "Analysis of Twitter Usage in London, Paris, and New York City," in AGILE: Leuven, Belgium.

Nik-Bakht, M., and T. El-Diraby. 2017. “Project collective mind: Unlocking project discussion networks” Autom. Constr. 84: 50–69.

Nik-Bakht, M., T. El-Diraby.and Hossaini, M. 2018. “Game-based crowdsourcing to support collaborative customization of the definition of sustainability” *Advanced Engineering Information*, **vol. 38**, pp. 501-513.

"Population totale en 2011 et en 2016 - Variation — Densité" Canada 2016 Census (in French). Ville de Montréal. 2016. Retrieved 1 December 2018.

S. Bird, "Nltk: the natural language toolkit". 2006. In Proceedings of the *COLING/ACL on Interactive presentation sessions*, *COLING-ACL*, Stroudsburg, PA, USA.

Y. Liu, P. Piyawongwisal, S. Handa, L. Yu, Y. Xu, A.Samuel. 2011. "Going Beyond Citizen Data Collection with Mapster: A Mobile+Cloud Real-Time Citizen Science Experiment," *in IEEE Seventh Int. Conf. E-Science Work*.

Zeynep Filiz, Veysel Yilmaz, Ceren YagÄ±zer, 2010, " The Measurement Of Service Quality In Municipalities By Servqual Analysis: A Case Study In Eskiåžehir Municipalities ", *Anadolu University Journal of Social Sciences, Anadolu University*, **vol. 10**, no. 3, pp. 59-76.