OCCUPANTS' COMFORT AT URBAN SCALE: ANALYZING CITIZENS' OPINING USING CONVOLUTIONAL NEURAL NETWORKS

Farzaneh Zarei, and Mazdak Nik-Bakht
Department of Building, Environment and Civil Engineering, Concordia University, Canada
E-mail: f_zaire@encs.concordia.ca, mazdak.nikbakht@concordia.ca

ABSTRACT

Moving more and more toward urbanization, citizens' comfort is widely considered as a potential source for demand flexibility. In this regard, getting feedbacks and engaging occupants is recognized as a key component to ensure that urban spaces are capable of providing acceptable functionality. In this paper, we mainly focus on the "parks" in the cities under the three main categories of amenities, visual and acoustics. In this way, the comfort level of the citizens is evaluated by using the citizens' feedbacks. Accordingly, more than 10,000 reviews about different parks in Canada are gathered from the google map. On the next step, by applying the latent Dirichlet allocation (LDA), the google map reviews are classified into various topics. On the other hand, each review has a rate between 1 to 5 in google map which indicates the amounts of the importance of the issue. By applying convolutional neural networks (CNN), the reviews classified based on their polarities (positive, negative and neutral). Upon improving its accuarcy, this classifier can be used for labeling other texts and reviews generated by citizens that do not have explicit rating score (such as tweets).

Moreover, by analyzing reviews assigned to each topic and through the polarity of the comments, the most significant issues related to Canadian parks in terms of citizens' comfort were detected and discussed. The result of this paper can provide a break down of comfort conditions assesment (from prespective of space end-users), to be used as an input by urban space designers.

INTRODUCTION

Public space always played a significant role in cities, since citizens, regardless of their gender, age, religion, income, can meet each other. Hence, one of the main goals of urban planning and designing is shaping outdoor spaces according to the citizens' needs and making these spaces attractive to people. Many different factors are important when defining which place is named attractive such as climate condition (temperature, humidity and wind speed), vegetation, lighting, noise, bike paths, street furniture, location of sidewalks, availability and the condition of infrastructure (Geht 2011).

Defining comfort is subjective and different people have different preferences in terms of outdoor environmental conditions, accordingly one important issue about public space is defining proper comfort level (Chen et al. 2012). In other words, outdoor comfort is a result of interaction between peoples' preferences.

Using the citizens' feedback is one efficient way for measuring comfort level. Citizens can report their observations and experiences either intentionally by sharing their feedback on social media, or unintentionally through the information sharing devices they carry, such as cellphones.

The main goal of this paper is to identify citizens' comfort levels in terms of amenities, visual, and noise level conditions and identifying their issues related to "parks", using citizens' feedback. For this aim, we use (i) google-map reviews to fetch the comfort level reported in parks; (ii) Latent Dirichlet Allocation (LDA) algorithm to extract the topics; and (iii) convolutional neural network (CNN) as a tool to detect the sentiment of the reviews. While google-map reviews normally assign a score to the space (in our case, parks) such a scoring system is to holistic and cannot be necessarily reflective of end-users' comfort. Using LDA helps to revile different aspects of the environment discussed by comments, and CNN can identify the perspective reflected by the comments (positive, negative or neutral). Hence, both aspect extraction (LDA) and sentiment analysis (CNN) are needed to identify the positive and negative aspects of each places.

This paper organizes as follows: Related works summarize different categories of comfort levels such as climate conditions, lighting, noise, views, and amenities. Next, the proposed method will be introduced which consists of different steps including data-collection, pre-processing, topic modeling and classification step. Then, the experiments and their results will be discussed in the next section. Finally, discussion, result analysis, and conclusion will be presented.

RELATED WORKS

Parks as an important component of each city which bring-about clean air, improve our mental health, provide space for physical and social activities among lots of other advantages. In terms of climate change, parks are playing an important role in moderating the city temperatures and sequester carbon trapped by the urban heat island that may otherwise alter the local and global atmospheric composition.

By increasing the trend of urbanization, more issues affected by parks as one important place in cities. However, smart cities offer some solutions for resolving these problems. Using sensors to monitor the conditions of city furniture and other components of the built environment can help us in this way. For instance, there is a product from "Street Furniture Australia" which tells if the trash bin must be emptied (Tietz et al. 2019). In terms of street furniture, another product can be detect defected or broken benches by monitoring the frequency of its use. Furthermore, it can realize whether a bench is being used by children, to identify the locations that are not children-friendly (Tietz et al. 2019). The other example in this regard is related to finding a parking space. There are applications such as "Parking Patch" which help the citizens to find elusive vacant spots using mobile apps and wireless sensors (Graham-Rowe 2012). The practical version of this idea was applied in some parking spaces and shopping centers. However, the broad spatial spread of the infrastructure network does not allow placing sensors on every single piece of infrastructure physical asset.

On the other hand, apart from attaching sensors to every street furniture, however, the analysis of citizens' feedback can also be helpful. In other words, we can use citizens as sensors. Citizens are distributed enough with the same pattern as the infrastructure network and continuously interact with the infrastructure service. Hence, developing a "Human sensor network" can provide an effective tool for constantly monitoring the system. As an example of this approach can be mentioned the new service of the Google Map about public transportation vehicles. Recently, Google started to collect the number of available seats on buses at different hours. After gathering this data for few months, by applying machine learning algorithms, Google is now able to tell us the approximate number of free seats to be expected in a bus we search on Google Maps (Graham-Rowe 2012).

The same idea can be applied to evaluating citizens' comfort levels in parks. Based on the literature in terms of comfort levels, there are five physical factors that affect human satisfaction and productivity such as climate condition, lighting, noise and acoustics, views and amenities. In the rest of this section, each factor will be described.

Climate condition (temperature, humidity and wind speed)

This factor is dependent on several dynamic parameters including clothes, activities, and moods (Han et al. 2007). Achieving overall climate comfort is not a simple task, as this comfort is a result of different parameters. Furthermore, it depends on both human characteristics (such as age, gender, and metabolism) and geographical features (Zhai et al. 2015). However, most of the standards related to amenities comfort are appropriate for uniform amenities conditions and do not consider personal characteristics such as age, sex and regional behavioral actions and expectations (Han et al. 2007) (Zhai et al. 2015).

Lighting, Daylighting, Solar radiation

Previous works related to the lighting can be divided into two main groups. First, those works that aim to protect people from solar radiation and providing suitable shading in terms of amenities comfort (Middel et al. 2016). This is due to the fact that both solar radiation and mean radiant temperature affected how citizens perceive outdoor amenities comfort. The second group is referring to those papers which their goals were analyzing the quality and availability of the artificial lights in urban areas at night. Since some places in the cities are designed for using in the evening and night hours and it is essential that artificial light is adapted to the needs of citizens. In this way, street lights are important not only in terms of providing light but also for the quality of the light they provide (Xi et al. 2012).

Noise and Acoustics

Noise is one of the most intrusive pollutants in urban areas. Some papers in the literature addressed this issue through at-source interventions such as traffic management (Boer et al. 2007) or through building protective installations like roadside noise barriers (Dzhambov et al. 2015). After studying the previous works on outdoor noise comfort levels, we divided the literature into five main categories. First, papers on sound propagation in urban environments (Berglund et al. 2006) and corresponding methods for reducing noise like using vegetation and understanding the interaction between vegetation and sound waves, such as absorption, scattering, reflection and ground effect (Onaga et al. 2007). The next group is relating to those works on sound propagation through tree belts to determine the combining effects of individual trees, mainly against road traffic noise (Fang et al. 2003). Finally, papers on noise control by using natural

materials such as vegetation in urban situations (Wong et al. 2010).

Views

The look and feel of the citizens' environment can have a great effect on citizens' sense of well-being. Hence, the demand for shaping the cities more attractive and comfortable is becoming increasing. In this regard, there are many studies which aim to identify citizens' perspective in terms of the current condition of the city elements, organizing and understand their ideas for creating new elements such as new forest trails, and parks. Based on the literature, some important solutions to achieve the desired city landscape are improving the urban environment using the principles of color and light (Zhang et al. 2004), balancing the street view, using of graffiti and graphics on the street body (Talebi et al. 2016).

Amenities

The location, accessibility, and quality of each city element such as public transportation, intersections, pedestrians, and green spaces can directly affect the comfort level of citizens. As an instance, number of potholes and safety are some important factors for pedestrians (Tolstrup 2019). Furthermore, in terms of comfort in public transportation, criteria such as traveling and transferring time, reliability, cost need to be considered (İmrea et al. 2017).

After a brief review of the outdoor comfort levels, the next section will present the proposed methodology of this research.

METHODOLOGY

In this paper the proposed method consists of five main steps such as data collection, pre-processing, topic modeling, keyword extraction, relevance detection and classification. The workflow of these steps illustrated in Figure 1. In the rest of this section, the detailed explanations about each step will be discussed.

Figure 1 Workflow of the proposed method

Data collection

For finding the relevant comments to park and green spaces, google-map reviews are gathered. Google-map provides a platform for people to share their ideas about different public and private places such as restaurants, bars, hotels, hospitals, malls, parks, etc. In this way, the google-map application programming interface (API) is applied to gather the google-map reviews. In this regard, around 50 parks in Canada are selected and about 10,000 reviews are gathered from these selected parks. Some examples of these reviews are presented in Table1.

Table 1 Some samples of the collected reviews

Reviews	Rates
Great place to walk.	5
Nice place for outing with family and friends.	5
Beautiful, clean great for kids area lots of picnic tables...	4
Great place for walking and cycling, also for an evening picnic.	4
Getting parking is the only problem	3
Beautiful day in the summer to come out here. Very busy. Lots of people laying out to get tans. Predominantly one group of patrons,. not multi-ethnic.	3
No escape from the noise from the highway behind.	2
No public water? Fountain? No bbq? Little to no parking. Toronto could learn a lot.	2
Probably one of the worst parks in the city. It's constantly full of people disrespecting the space. The smoke, the litter, it's like a singles bar for people who say they're not into singles bars.	1
It used to be well setup but now (this year, feels VERY DIRTY and as the water level has risen, all the slides and swings have been removed	1

Data pre-processing

In this step, in order to improve the quality of the gathered data, some pre-processing algorithms are applied to remove the noise from the raw data. In this way, stop-words, emojis, and URLs are removed. Finally, the lemmatization algorithm, which is implemented by NLTK[1], is applied to transfer the words into a dictionary format using.

Topic modeling

In this step, the latent Dirichlet allocation (LDA) algorithm is used for extracting topics. Topic modeling is referring to the process of retrieving the topic of a text based on the distance, automatically (Blei et al. 2003). Generally, LDA is a probabilistic point model in which every record is indicated as a random mixture of idle subjects. In this algorithm, every single latent topic is named as a distribution over a fixed arrangement of words and is used to identify the underlying latent subject structure based on watched information. The process of generating words consists of two main steps. Firstly, a circulation over topics is haphazardly chosen for each word of the document. Next, the back circulation of the hidden variable should

[1] Natural Language Toolkit

be estimated. In this way, algorithms such as Gibbs sampling (Porteous et al. 2008), Laplace (Yangqiu et al. 2009), and Markov chain (Bolelli et al. 2009) can be applied.

Keyword extraction & Relevance detection

After applying the LDA to the dataset, some words are suggested as the extracted topics. Next, skimming the suggested keywords, it is seen that there are not many relevant words in terms of thermal comfort and lightening. Hence, based on the available data, we mainly focus on 3 different comfort factors such as amenities, noise, visual comfort. In this regard, 20 relevant topics related to each group are selected manually and the sentences which consist of these topics are assigned to that cluster. Furthermore, for those reviews which contain more than one selected word, we manually decide about the cluster that the review belongs to. Finally, at the end of this step, the collected data is clustered into 4 different clusters.

On the next step, a convolutional neural networks algorithm is used for identifying the polarity of the reviews. The next section will describe the details of this algorithm.

Classification

In this step, a convolutional neural networks algorithm, which is a multi-layer feed-forward network is used for text classification. The architecture of the algorithm consists of different layers including input layer, which embeds words into lower dimension, the convolutional layer, which applies convolutions over the embedded word vectors and uses multiple filter sizes, the pooling layer which converts the output of the convolutional layer into a long feature vector, the fully connected layer with dropout regularization and finally, the softmax layer which provides the classification labels. The details of the implementation of this algorithm will be discussed in the next section.

EXPERIMENT

In order to detect relevant reviews to the comfort level from the collected dataset, the topic modeling is fulfilled. In the LDA algorithm, the number of topics parameter is set as 40. Next, after applying LDA algorithm, 20 words related to each cluster are selected manually. Table 2 presents some of the extracted topics in each group.

Table 2 Some samples of the extracted topics in each cluster

AMENITIES COMFORT	NOISE COMFORT	VISUAL COMFORT
BBQ	train	trees
washroom	quiet	blossoms
dirty	people	grass
parking	crowded	walking
small	calm	colors
garbage	lake	green
playground	road	beautiful
accessible	relaxing	attractive
friendly	river	path
nature	events	amazing
boardwalk	noisy	flower
pool	peaceful	trails
tables	activities	pictures
small	cozy	nice
clean	areas	awesome
picnic	weekends	environment
rides	loud	views
pond	fresh	enjoyable
family	reads	garden

Based on the selected keywords, the collected data are divided into 3 groups. In this way, those reviews which contain these selected keywords are labeled based on Table 2 as relevant to amenities comfort, noise comfort and visual comfort. As a result, 6001 reviews are labeled as relevant to the comfort level. Table 3 illustrates the number of reviews in each category.

Table 3 Number of extracted reviews in each comfort category

	AMENITIES COMFORT	NOISE COMFORT	VISUAL COMFORT
# reviews	1703	350	3948

The collected reviews are accompanied by a rate ranging from 1 to 5 stars. Here, the reviews are considered as three main classes which are positive (4 or 5 stars), negative (1 and 2 stars), and natural (3 stars). Based on this class definition, the data are prepared for the classification step.

On this step, the entire dataset is divided randomly into training and test sets. The training set consists of 90% of the sentences while the test set includes 10% sentences. The CNN algorithm is implemented in Pytorch, which is a Python library. For word embedding step, glove algorithm is used which is an unsupervised algorithm for transferring words into vectors. Furthermore, the Adam optimization algorithm is selected to optimize the moment of features' values and the loss function was cross-entropy.

RESULT ANALYSIS

In order to evaluate the performance of the suggested classifier, the diagnostic test is done. In this way, three

main metrics such as precision, recall, and F- measure are considered using Eq. (1), (2), and (3).

$precision =$ ——

(1)

——

(2)

——————

(3)

Here, *tp* is true positive, *tn* is true negative, *fp* is false positive, and *fn* is false negative. Table 4 demonstrates the results of the proposed classifier.

Table 4 Results of the trained classifier

	AMENITIES COMFORT	NOISE COMFORT	VISUAL COMFORT
Precision	85.11	85.67	87.33
Recall	86.49	86.09	89.55
F-measure	88.84	89.69	84.67

As a result, for a new review about the park, the main topic can be extracted in terms of different comfort measures such as amenities, noise and visual. Moreover, the polarity of that comment also can be extracted.

In terms of the sentiment, Figure 2 summarizes the obtained results.

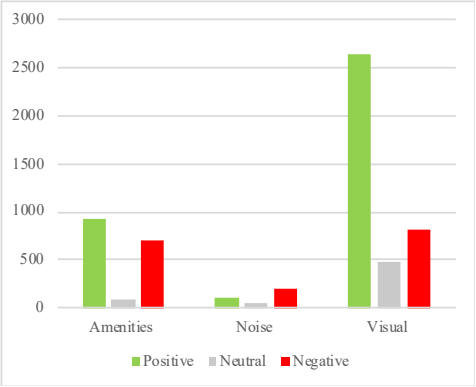

Figure 2 Sentiment of the obtained results

As seen in Figure 2, most of the discomfort reviews are related to the amenities. This is due to the fact that the reviews mostly discuss the facilities of the parks such as clean washrooms, enough benches and tables, and etc. Furthermore, the beauty of the parks is the subject of some reviews, especially in some specific seasons. However, few comments are discussed the noise levels in the parks, just some positive comments which believe the park is a quiet place and some negative ones which state that the citizens are annoyed about the number of people there and believe that the place is too crowded. Moreover, some parks are close to the train station and the noise is annoying.

Using both topic modeling and sentiment classification of reviews can help to identify the weaknesses and strengths of different environments (e.g parks, in the present case study). As an example, based on the collected reviews on "High Park", a popular park in downtown Toronto, Canada, we can understand although the overal google maps ranking for this park is high (4.7/5), it is mostly contribiuted to visual comfort (including trails and green spaces). However, in terms of amenities comfort such as washrooms, parking, trashbeens, and noise comfort, the collective sentiment of reviews is not positive. Some examples of these reviews are as below:

- "Good park but the washroom smells like literal death"
- "Because of bad city management, the park is not designed to cope with holiday traffic."
- "This place used to be nice but man oh man, finding a parking spot is a nightmare! It's overcrowded nowadays. I've seen people getting into fist fights over a mere parking spot."

DISCUSSION

In this paper, we present a two-level classification method for analyzing comfort level in parks. In this regard, at first, we need to identify the main topics in each comfort aspect and then evaluate the sentiment. The reason that why we have comfort aspect detection at a higher level is that it might affect the sentiment that is expressing the dissatisfaction in each of these classes is different. Applying such a two-level classification is expected to improve the accuracy of the classification.

Skimming through the reviews, it was observed that while most gathered reviews are directly addressing one specific issue, some of them belong to more than one category. To resolve this issue, a combination of the proposed method with fuzzy classification techniques can be helpful. However, that would require a considerable amount of data. In this study, due to the lack of enough data, we failed to perform such a classification.

Even though the main focus of this study was on parks, we can readily extend this method to other environments such as indoor spaces. Given the availability of comments, looking at correlation between comments and measured sensory data can help for better understanding and analyzing occupants' comfort.

CONCLUSION

This paper aims to evaluate the level of citizen comfort in parks based on google-map reviews. In this way, some of the reviews related to the parks were collected. After some preprocessing, the LDA algorithm was carried out to extract the main topic of each review. Next, in order to detect that each review was related to which cluster of the defined comfort level clusters (amenities, noise and visual), we manually select top most relevant extracted topics to each cluster and if a review contains those selected words, it labeled as relevant to that defined cluster.Finally, CNN was used to classify the reviews based on their polarities (positive, negative, neutral).

As a result of this research, the proposed method enables us to find the relevant reviews to the comfort levels in parks and also detect the polarity of them. This enables us to identify the weaknesses and strengths of different parks (as components of urban built environment) using citizens as distributed sensors. This research can also help the decision-makers to make more effective decisions rather than just using surveys. Moreover, it can be useful to evaluate the position of different city elements from the citizens' viewpoint. Furthermore, the methodology is universal and can be applied to other components of the built environment rather than parks.

Following this work, by gathering more reviews and implementing a two-level classification, we can improve the performance of the classification. Furthermore, this approach can be applied to other places of the cities such as bus stations and parking. Also we can detect the correlation between results and the level of citizens' satisfaction in different indoor and outdoor spaces of cities.

REFERENCES

Angel M. Dzhambov, Donka D. Dimitrova, "Green spaces and environmental noise perception," Urban Forestry & Urban Greening, no. 14, pp. 1000-1008, 2015.

Ariane Middel, Nancy Selover, Björn Hagen, Nalini Chhetri, "Impact of shade on outdoor thermal comfort—a seasonal field study in Tempe, Arizona," International Journal of Biometeorology, vol. 60, no. 12, p. 1849–1861, 18 May 2016.

Berglund, Birgitta; Nilsson, Mats E., "On a Tool for Measuring Soundscape Quality in Urban Residential Areas," Acta Acustica united with Acustica, vol. 92, no. 6, pp. 938-944, November/December 2006.

Chih-Fang Fang, Der-Lin Ling, "Investigation of the noise reduction provided by tree belts," Landscape and Urban Planning, vol. 63, no. 4, pp. 187-195, 2003.

Christian Tietz., Christine Steinmetz., Homa Rahmat; Kate Bishop., Linda Corkery; Miles Park., Nancy Marshall, "Sensors in public spaces can help create cities that are both smart and sociable," 09 04 2018. [Online]. Available: http:// theconversation.com/ sensors-in-public-spaces-can-help-create-cities-that -are-both-smart-and-sociable-93473 . [Accessed 07 10 2019].

D. Graham-Rowe, " Parking sensors to take pain out of finding a space," 02 2012. [Online]. Available: https://www.newscientist.com/article/mg21328506-100-parking-sensors-to-take-pain-out-of-finding-a-space/. [Accessed 07 October 2019].

David M. Blei, Andrew Y. Ng, Michael I. Jordan, "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, pp. 993-1022, 2003.

den Boer, L.C., Schroten, A., "Traffic Noise Reduction in Europe," 2007.

Hiroshi Onaga, Jens Holger Rindel, "Acoustic characteristics of urban streets in relation to scattering caused by building facades," Applied Acoustics, vol. 68, no. 3, pp. 310-325, 2007.

J. Geht., "Life between buildings. Using public space," in Washington- Covelo - LondonIsland Press 2011, 2011.

J. Han, G. Zhang, Q. Zhang, J. Zhang, J. Liu, L. Tian, C. Zheng, J. Hao, J. Lin, Y. Liu, "Field study on occupants' thermal comfort and residential thermal environment in a hot-humid climate of China," Build. Environ., vol. 42, pp. 4043-4050, 2007.

Levent Bolelli, Şeyda Ertekin, C. Lee Giles, "Topic and Trend Detection in Text Collections Using Latent Dirichlet Allocation", in Advances in Information Retrieval, Springer Berlin Heidelberg, pp. 776-780, 2009.

Liang Chen, Edward Ng, "Outdoor thermal comfort and outdoor activities: A review of research in the past decade," Cities, pp. 118-125, 2012.

M. M. Tolstrup, "Pedestrian wind comfort in urban areas," [Online]. Available: https:// forcetechnology .com/en/articles/pedestrian-wind-comfort-in-urban-areas. [Accessed 13 11 2019].

Negar Talebi, Maryam Farokhi, "The Role of Light and Color in Urban Beautification and its Impact on the Perception of Citizens," Bulletin of Environment, Pharmacology and Life Sciences, pp. 36-43, 2016.

Nyuk Hien Wong, Alex Yong Kwang Tan, Puay Yok Tan, Kelly Chiang, Ngian Chung Wong, "Acoustics evaluation of vertical greenery systems for building walls," Building and Environment, vol. 45, no. 2, pp. 411-420, 2010.

Porteous, Ian and Newman, David and Ihler, Alexander and Asuncion, Arthur and Smyth, Padhraic and Welling, Max, "Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation", Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, pp. 569-577, 2008.

Shu Zhang, Yijie Liu, Hong Nie, "Geographical Feature Based Research on Urban Color Environment – Taking Wuhan as an Example," in 2014 International Conference on Environment Systems Science and Engineering , 2014.

Song, Yangqiu and Pan, Shimei and Liu, Shixia and Zhou, Michelle X. and Qian, Weihong, "Topic and Keyword Re-ranking for LDA-based Topic Modeling", Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09, pp. 1757- 1760, 2009.

Şükrü İmrea, Dilay Çelebi, "Measuring Comfort in Public Transport: A case study for İstanbul," Transportation Research Procedia, pp. 2441–2449, 2017.

Tianyu Xi, Qiong Li, Akashi Mochida, Qinglin Meng, "Study on the outdoor thermal environment and thermal comfort around campus clusters in subtropical urban areas," Building and Environment, vol. 52, pp. 162-170, 2012.

Y. Zhai, Y. Zhang, H. Zhang, W. Pasut, E. Arens, Q. Meng, "Human comfort and perceived air quality in warm and humid environments with ceiling fans," vol. 90, pp. 178-185, 2015.

Yongchao Zhai, Yufeng Zhang, Hui Zhang, Wilmer Pasut, Edward Arens, Qinglin Meng, "Human comfort and perceived air quality in warm and humid environments with ceiling fans," Building and Environment, vol. 90, pp. 178-185, 2015.