14th International Scientific Conference on Sustainable, Modern and Safe Transport

Car Parking User's Behavior Using News Articles Mining Based Approach

Nabil Arhab[a]*, Md Saroar Jahan[a], Mourad Oussalah[b]

aResearch Assistant, University of Oulu, CMVS, 90570 Oulu, Finland
bResearch Professor, University of Oulu, CMVS, 90570 Oulu, Finland

Abstract

Studying individual's parking choice behavior can considerably contribute towards evidence-based policing in urban area. This study investigates evidence gathered by mining Finland news article API concerning car parking associated topics in order to comprehend user's behavior and identify potential unforeseen circumstances that may impact users' decisions and preferences. The study follows a natural language processing research pipeline, emphasizing word co-occurrence analysis, sentiment score and named-entity monitoring. The results can be exploited by local authorities to develop further evidence based policing in city urban planning.

1. Introduction

With the economic prospects, people's lifestyle conditions have improved, which, in turn, increased the ratio of car per household. On the other hand, the city transportation demands continue to grow due to the rapid increase in private car ownership, bringing the issue of car parking management at crucial state. For instance, a recent research revealed that a driver takes almost 8 minutes to park his vehicle as he invests more time searching for parking lot. This leads to a 30% to 40% increase of traffic congestion (Walvekar et al. 2017). In this context, several solutions have been investigated by urban planners and research community. This includes an expansion of existing parking lot (Walvekar et al. 2017), the use of big data for smart parking (Nguyen et al. 2018), IoT integration with smart parking (Karpenko

---

* Corresponding author. Tel.: +358- 413-682-772.
  E-mail address: Nabil.Arhab@oulu.fi

et al. 2018), smart parking systems (Basavaraju, 2015; Snell, 2015), predictive analytic through mathematical modeling (Zong et al. 2019). However, such studies were very lacking in Finland for at least two reasons. First, the scarcity of populated areas in Finland made such problem less stressful than other European or worldwide populated cities. Second, some urban planners believe that it is still arguable whether the development of enough parking lot or smart parking could ease the parking burden as this may bring extra management related difficulties of other assets of the City as well. Nevertheless, the advances in environment friendly technology as well as the UN sustainable development goals raise the importance of car parking management as an important trigger for achieving such goals. This brings the issue of mining user's behavior regarding car parking at paramount importance as any City development plan cannot be successfully achieved without citizen's active participation. Therefore, comprehending the hidden factors that govern users' preferences and guide their choice has intuitively a substantial impact in achieving such active participation. Indeed, observing parking lot and user's parking behavior can yield optimized-based schemes for efficient utilization of existing parking lots and available solutions (Ying et al. 2020). This motivates the current study and analysis, which aims to identify unforeseen parking issues and pinpoint factors that influence individuals' search decisions and their preferences in Finland area. In the era of social media and internet, which gather many opinionated documentations, we hypothesize that many user's attitudes, opinions and suggestions are already encapsulated in news articles and their associated uses' comments. This explains the choice of Finnish news API as an ideal source for collecting community related behavioral data, which is then analyzed using appropriate natural language processing (NLP) techniques. Indeed, NLP has been used in many kinds of behavior analysis, including patient psychology reading (Kocyigit et al. 2020), online user behavior detection (Guidère 2020), spammer behavior analysis (Kumar 2020), behavior analysis of marketing in transportation systems (Xing et al. 2018), etc. This provides a sound ground for applying NLP based technology for mining car parking users' behavior as well. Especially, in this study, we applied four NLP methods: word count, sentiment analysis, bag-of-word, and word cloud to extract and analyze the collected data. Considerable works have been performed in applying sentiment analysis and word frequency count to finding important parameter for a particular topic. A work using the frequency count method has been applied to find a thematic hierarchy within a group of related texts (Maldonado and Maria 2014). Sentiment analysis has been used in customer reviews analysis (Samuels and Mcgonical 2020), offensive language detection (Jahan and Oussalah, 2020), review analysis (Dey et al., 2016), among others. The paper is structured as follows. Section 2 encompasses dataset collection and preprocessing. Section 3 outlines the employed methodology methodology. In Section 4, the obtained results are detailed and discussed.

2. Dataset description

The dataset used in this study are collected using Webhose API. Since Finland news and blog articles are mostly written in Finish, Swedish, and English, we have used all the three languages as part of input parameters where the query focused on car parking in the Finland area. Multiple lists of words related to parking were used to fetch the articles and related attributes from the API. A combination of English and Finnish wording has been applied to obtain the wide-spread and diverse car-parking occurrences in news articles. This list includes parking, parkade, park, carport, carParking, carpark, car park, car stall, park of vehicle, pysakointi, parkkipaikka, pysakointitila, paikoitus, and ajoneuvojen puisto.

Multiple attributes have been collected (i.e., titles, texts, dates, site types, URLs, authors, site names, and language) that are saved into a single CSV file. Table 1 provides some description and exemplification of the collected data.

Table 1. Dataset summary of 100 News Websites.

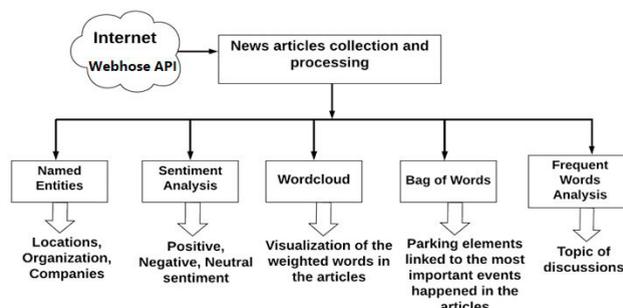| Attribute	Value
Start, date	22 April 2020.
End date	22May 2020.
Site names	ls24.fi, ts.fi, forssanlehti.fi, oulunylioppilaslehti. fi, yle.fi, sss.fi, muusikoiden.net, mtvuutiset.fi, lianapress.com, rakennusmaailma.fi, kaleva.fi, jyvaskyla.fi, pori.fi, pietarsaarensanomat.fi, iltalehti.fi, turku.fi, parma.fi, janakkala.fi, reska.fi, goodnewsfinland.com

3. Dataset processing

Fig. 1: Overview of the implementation process.

Before analyzing the dataset, it was essential to preprocess and make it suitable for further NLP uses. Besides, Google translation API was used to translate non-English text (Finnish, Swedish) into English to optimize the use of existing advanced parsers developed mainly for English text. Preprocessing includes removing unidentified characters, symbols, tab-token, lower-casing, tokenization, and fixing spelling. Some articles related to advertisements and amusement parks are also removed since they are related to this study. Besides, stop-words (i. e., as, a, on, with, in, etc.), which are generally the most common words in a language and often do not harm other subsequent NLP reasoning, were removed. The last step is related to part-of-speech (PoS) tagging that recognizes categories such as verbs, nouns, adjectives, adverbs. After the cleaning was performed, the number of articles in the dataset was reduced to 25, holding exclusively articles related to car-parking, which are then preserved for further analysis.

4. Proposed method and methodology

Four distinct NLP-based approaches were performed to identify parking issues and factors influencing individuals' search decisions and their preferences; namely, (i) word-count, (ii) sentiment analysis, (iii) bag-of-word, and (iv) word cloud. For this purpose, we first determine the word-count in both main text of article and its title. Second, named-entity tagging is performed to distinguish organizations, companies, and locations. This expects to grasp key trends that might be raised in the articles. The named entity tagging also intends to distinguish claims linked to economical, governmental and individual / personal aspects. While the sentiment analysis together with emotional polarity (Li and Wu 2010) computation brings user's attitude towards the car-parking raised in the news article. Specifically, three different emotional states were computed: positive, negative and neutral sentiment score. To visualize the result of text queries, the word-cloud technique was adopted. This is a well-known NLP visualization technique widely explored in research community. This extract relevant information in a categorized way with unique fonts and sizes, indicating word occurrences in the dataset. This results in a clean diagram displaying the most interesting words by their presence. The final technique employed in this analysis is the Bag of Words approach, a usually used method in text analysis or image analysis. This technique was adapted to this dataset to deeply analyze the texts and determine the word's collections with the parking keywords. Figure 1 summarizes the different approaches and techniques and dataset applied in this work.

5. Results

5.1. Word Frequencies

Both the title and the (body) text of the articles are taken into consideration when counting the most common words. The approach adopted is straightforward and follows simple word frequency approach, creating a dictionary where keyword and word-count are reported in Table 2 and Table 3 for the title and core text of articles, respectively.

Table 2. Most frequent words occurrence in Titles.

Word	Occurrence	Word	Occurrence
car	3	Helsinki	3
position	2	Parking	2
park	2	gun	2

Table 3. Most frequent words occurrence in Text.

Word	Occurrence	Word	Occurrence	Word	Occurrence
park	66	project	22	player	13
park	52	center	21	development	12
Area	52	kilometers	19	tour	12
parking	41	turku	18	association	12
city	40	road	18	example	11
car	40	sports	18	company	11
time	31	space	16	manager	11
lot	29	golf	15	energy	11
route	26	police	14	technology	11
people	25	museum	14	trees	11
traffic	23	trip	13	building	11
nature	22	services	13		

Looking at the frequent terms in Table 2 and 3, it can be inferred that four different sectors and activities are correlated with the parking as follows: *traffic*, *construction*, *travel*, and *residential*. For instance, the words trees, trail, traffic, route, and intersection can be assigned to the category "Traffic". Similarly, Construction of new parking projects can be interpreted by words like economy, project, energy, and development. The travel aspect is defined by a list of words such as trip, destination, excursion, and travel. Words such as studio, apartment, service, market and building can be cast to Residential area. These sectors and interests reflect several new car parking and bike projects in various locations, particularly in Oulu, Helsinki, and Turku. The travel and tour are also important topics since most travellers select destinations outfitted with park and parking amenities, and sometimes they utilize these parking facilities for camping. In some posts, parking has also been mentioned as tourist villages, in which parking is thought to be one of the traveling company's primary providers and the village. Additionally, the parking is recognized in certain posts as an additional feature for somebody who would like to rent his apartment. This is because parking was found liked to renting flats and studios. In summary, this enabled us to generate a categorization of the parking concept from news articles. This can provide insights to urban planners in terms of appropriate management standards and practices taking into account region constraints and users' views in a way to ease traffic burden and promote tourism according to travellers' preferences.

5.2. Named entities.

The Named entities of the locations and organizations are retrieved using Stanford Named entity tagger. Table 4 and Table 5 indicate the frequency of occurrences of Location and Organization entities, respectively, in the articles.

It can be seen that most of the organizations are either Finnish municipalities, companies (Fabulos, Dong Feng Motor, Mayra, Pietasaari) or smart city projects such as the Turku project Wise. This may identify the strong presence of Helsinki, Oulu, and Turku regions and certain tourist regions and villages such as Leivo, Siikaneva, Halikonlahti with regard to locations. These findings support the results obtained when evaluating the word frequencies in tables 2, 3 by showing that parking-related travel, traffic, and growth aspects are listed as key topics of the news articles' discussions. Furthermore, this technique permitted us to shed light on the municipalities' work and different companies

specializing in parking in different city zones. Another aspect highlighted in this study is related to tourist villages and how parking facilities influence travelers' decisions and comfort. One solution that can ease and increase the quality of services in these areas by is developing a multi-function parking as a design demand for the urban planners to increase the parking spaces' efficiency and enable the broader sharing of parking spaces and thus adapt more with the concept of smart mobility.

Table 4. Location (cities, countries) entities and occurrence in the text. Excluding Finland, as it is a common location.

Location	Occurrence	Location	Occurrence	Location	Occurrence
kristiina	3	malmberg	1	us	1
herralahti	2	kaarina	1	sweden	1
rutajarvi	2	marshland	1	vaskio	1
czech republic	1	tureng	1	jatkasaari	1
germany	1	lasturjarvi	1		
helsinki	1	north paijanne region	1		

Table 5. Organization (companies, agencies, and institutions) entities and occurrence in the texts.

Organization	Occurrence	Organization	Occurrence	Organization	Occurrence
fabulous	2	residential area	1	yard bring coziness	1
kirveslampi	2	cm7	1	club	1
Salo excursion destinations	1	eu	1	dubai world chalenge	1
city center rekijoki	1	Helsinki fabulous project	1	golf association evergreen	1
Marynummi village associato	1	tieto julkaisi jasta	1		

5.3. Sentiment analysis

For the different article's text, the sentiment analysis is performed using VaderSentiment library. The technique adopted to extract the sentiment is based on considering each article's text as a sum of each word sentiment. Ultimately, by assigning an integer value for each word, the number of positive, negative, and neutral words is counted. These values reflect the polarity's power or type conveyed (Li and Wu 2010). From Table 6, one notices that positive sentiment accompanies parking facilities, and services in Finland. Twenty-two (88%) of the articles were positive, two articles were negative, and one was neutral. The positive sentiment was attributed to the launch of various projects related to new parking spaces for cars and bicycles in various areas, particularly in Turku and the Wise Smart Cities project. After the corona pandemic, optimistic emotions was also identified to be linked to the launch of golf clubs. Similarly, this was also linked to tourism, nature trails, despite limited parking lots. Regarding the negative sentiment, there were two articles, one related to closing parking spaces in a nature trail in April due to corona. The other one was linked to police arrest in the vicinity area. The neutral sentiment occurred in one article discussing parking buildings' closure in the Torroncuo area for maintenance. This closure caused the frustration of people as there are no other parking alternatives. Furthermore, the sentiment revealed good parking management, and these results contribute by highlighting the various parking solutions accommodated by the parking provider and the municipalities' actions. The results also highlight some parking factors or topics that positively influence the driver's decisions, such as improving parking places in sports fields and nature trails, facilities adaptation for self-driving cars, and satisfaction with different smart city projects. On the other hand, the negative sentiment revealed some negative parking factors such as the parking closure and the temporary parking maintenance without providing a subordinate solution.

Table 6. Sentiment analysis of the Text's articles.

Title of article	Sentiment	Title of article	Sentiment
salon excursion destinations afford choose	1	nyrola nature trail closed notice 15.4	-1
golfer's scrutiny greens open easter	1	mr. lahti sports field parking traffic arrangements improve	1
04/13/2020 rent studio mode sound inn herttoniemi Helsinki	1	pietarsaari square park largest shareholder bankruptcy parking operations continue unchanged cave convince shareholders	1
summer comes ready start jogging avoid unnecessary injury	1	video police chased gun raplaajaa porvoo catch unearthed gun	-1
residential area model whole world finnoossa passenger car secondary position geothermal energy solution first local heating network'	1	laitila intersection cycle path uusikaupunki revive economy offered long list construction projects	1
sensible 4 launches fleet self-driving vehicles busy area helsinki	1	new car bike parking services development starts turku	1
expert advises make yard second home living room	1	west harbor parking house frame delivery consolis parmalta	1
aurajoen moorings piling progress know piles blue color dates back	1	technical committee wood lake beach campfire	1
linnanmaa located new central fire station given planning permission	1	time stands still siikaneva boundless landscapes eye seeing trip ends earth	1
reader concert hall position consider	1	transfer sara hild´en art museum s¨ark¨anniemi tampere city center assured	1
kiljamo parking area the closing torronsuo cause temporary trouble parking spaces	0	council wants decide market parked without delay	1
faarinpuisto ayhonjarven north side official car park	`1	leivonmaki national park hiker central	1

5.4. Bag of words

A bag of word techniques is employed to intensify the analysis of the dataset further. This technique permits linking the parking with the most important aspects or events in the text, which is reflected by producing elements of bag of words like sentences without respecting the words' grammatical order. For the implementations of such technique, the wordVectorizer feature extractor was used. The input to this extractor is a corpus of news articles, and the output is elements of the bag of words, which are a group of 5 tokens for each BoW element. After fitting the corpus to the wordVectorizer extractor, and generates the corresponding dictionary, a proper cleaning phase was performed, by filtering out only the BoW elements that hold specific words linked directly to the parking, i.e., any element which contains at least one word from this list (parking, car park, carpark) is filtered as the wanted output from the entire BoW technique. The technique of the BoW has produced some exciting results for the study, see Table 7. The extractors have learned the data well, and it combined the parking words with the actual parking events that occurred with the articles, such as parking preferences and parking issues. Example of some of this events and issues are the parking closure, parking maintenance, new car-bicycle, river parking, reduction of parking space, nature trail parking, parking projects, new parking pricing, street parking issues, the necessity to provide additional parking spaces. In addition, the BoW has confirmed some conclusions drawn from the previous techniques. Moreover, the results permitted revealing the key sentences that represent the most discussed subject within each article. And here it allowed us to expose the impact of parking management to traffic, nearby events and tourism.

Table 7. Example results for Bag of words.

Bag of words examples	Bag of words examples
April car park used occasionally	car park closed 26 april
new car bicycle parking services	path parking lot four kilomete
people gathering parking lots front	people pietarsaari use parking prices
renovation work kiljamo car park	renovations car park closed 26
prate car park lower expected	pietarsaari lot free parking spaces

5.5. Word cloud

Word cloud is a suitable technique that helps in dealing with large scale reviews and facilitates the reading and interpretations of a massive number of articles (Wang et al. 2014). The python library wordcloud was used to create the word cloud for the available dataset to evaluate the text of the different articles. Each tag corresponds to a word in the visual representation, and the font-size and color indicate the weight, while each word's value represents the meaning. The word with a larger size exemplifies a more important value and meaning, while words with a small size imply a less impact and effect on the context. Word cloud in Figure 2 has revealed the important words or terms and their significance values in the articles. Most of the big size words such as, area, city, traffic, project, new, time, route, solution, trip, development, good. These terms reflect and confirm the development and construction of new parking spaces as the main topic discussed in the articles. Moreover, the approach has validated some analysis that found the

Fig. 2: Word cloud representation of the dataset text.

named entities already identified, such as locations, names of projects, or parks. Further, this representation provides a resource that helps the parking providers to determine the parking's various critical use cases. Also, it facilitates the categorization of the concept into different urban subcategories. For instance, one can consider here the construction, traffic, mobility, and smart city topics as the main subcategories formed and influenced by some use cases that are linked to parking such as travel, events, and tourism.

6. Conclusion

In this study, a parking investigation in the Finland regions is done by analyzing the news articles related to parking. Various NLP techniques were considered to extract and retrieve the essential parking events linked to either a parking issue, new parking projects, or individuals' parking preferences. Furthermore, the articles have discussed the parking from various modalities. There were four different general categories linked to parking, new parking systems, tourism, travel, and residential aspects. Most articles have found to bear a positive sentiment, which reflects the global satisfaction trend due to a good parking planning in the region, and the launching of multiple new parking projects either for cars or bicycles. Regarding the parking issues, the writers have reported problems such as: closure or

maintenance of parking in various areas, rise of parking fares in some residential areas, self-driving car projects, individual parking skills, among others. Other parking issues are related to the unsatisfactory parking providers' inadequate occupancy rate with underground parking, where people prefer on-street parking over underground ones even in cold winter. The study outcome can help to improve parking systems and suggest potential solutions for the better quality of services in a way to accommodate drivers' needs and population expectation. Moreover, the finding highlights the need for the municipality to investigate the use of a multifunctioning parking solution to accommodate people's needs in some circumstances. For example, this could be addressed by designing the multifunction parking solutions adjusted to users' demands in event-places and tourist villages. Another important parking aspect that the municipality needs to address concerns the parking policies used to enhance smart mobility and smart vehicles' integration by managing a convenient solution for such cars and systems to be used safely by everyone. Finally, this study provides prospects for applying big data analytics, including sentiment analysis, work co-occurrences and named-entity tagger to address car parking issues from open news articles. This can easily be accommodated to address other burden issues in urban planning, transportation and smart city management.

Acknowledgment

References

Walvekar, M. O., Kulkarni, M. A., Waghmode, M. A., & Mane, M. R. (2017). Automatic Car Parking.International Research Journal of Enginerring and Technology, 4(3), 1070-1073

Basavaraju S, R. (2015). Automatic smart parking system using Internet of Things (IOT). International Journal of Scientific and Research Publications, 5(12), 629-632.

Snell, R. (2015). Implications for planning smart parking schemes in Espoo.

Zong, F., Yu, P., Tang, J., & Sun, X. (2019). Understanding parking decisions with structural equation modeling. Physica A: Statistical Mechanics and its Applications, 523, 408-417.

Ying, M., & Sun, Y. (2020, April). Discussion on Parking Management System Based on Parking Behavior. In 2020 International Conference on Urban Engineering and Management Science (ICUEMS) (pp. 513-516). IEEE.

Kocyigit, D., Milinovich, A., Lee, C. M., Hanna, M., Gabrovesk, A., Jin, J., ... & Cho, L. (2020). THE ADDITIONAL VALUE OF NATURAL LANGUAGE PROCESSING TO IDENTIFY DISEASE PREVALENCE BASED ON ELECTRONIC MEDICAL RECORD IMAGING REPORTS. Journal of the American College of Cardiology, 75(11 Supplement 1), 3642.

Guidère, M. (2020, March). NLP Applied to Online Suicide Intention Detection. In HealTAC 2020, Paris, London, INSERM-02521389.

Kumar, A., Chatterjee, J. M., & Díaz, V. G. (2020). A novel hybrid approach of SVM combined with NLP and probabilistic neural network for email phishing. International Journal of Electrical and Computer Engineering, 10(1), 486-493.

Maldonado García, M. I. (2014). Frequency Counts in South Asian Poetry; Sultan Bahu's Concept of Mystical Love. South Asian Studies, 29(1), 333-344.

Samuels, A., & Mcgonical, J. (2020). Sentiment Analysis on Customer Responses. arXiv preprint arXiv:2007.02237.

Jahan, M. S., Oussalah M. (2020). Team Oulu at SemEval-2020 Task12: Multilangual Identification of Offensive Language, Type and Target of Twitter Post Using Translated dataset. Proc. of 14th International Workshop on Semantic Evaluation, 1628-1637.

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. arXiv preprint arXiv:1610.09982.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision support systems, 48(2), 354-368.

Wang, J., Zhao, J., Guo, S., North, C., & Ramakrishnan, N. (2014). ReCloud: semantics-based word cloud visualization of user reviews. In Proceedings of Graphics Interface 2014 (pp. 151-158).

Nguyen, S., Salcic, Z., & Zhang, X. (2018). Big Data Processing in Fog-Smart Parking Case Study. In 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom) (pp. 127-134). IEEE.

Karpenko, A., Kinnunen, T., Madhikermi, M., Robert, J., Främling, K., Dave, B., & Nurminen, A. (2018). Data exchange interoperability in IoT ecosystem for smart parking and EV charging. Sensors, 18(12), 4404. doi:10.3390/s18124404.

Xing F.Z., Cambria E., Welsh E. (2018). Natural Language Based Financial Forecasting, Artificial Intelligence Review 50(1), 49-73