

What Can I Do There? Extracting Place Functionality Based on Analysis of User-Generated Textual Contents

Mina Karimi*, Mohammad Saedi Mesgari*, Omid Reza Abbasi*

* GIS Department, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran

Abstract.

Recently, due to the rapid increase of social networks, user-generated textual contents have grown significantly. Sense of place emerges when a place is functionally different from its surroundings and is therefore recognizable. This paper aims to extract place functionality by analyzing user-generated textual contents. Therefore, using different natural language processing and machine learning methods, we show how much place functionality can be represented through the whole text with lemmatized words and only action verbs in describing place. The evaluation results demonstrate that using the whole text achieves better performance and among different classifiers, SVM leads to better accuracy and f1-score.

Keywords. Place-Based GIS, Place Functionality, User-Generated Textual contents (UGTC), Natural Language Processing (NLP).

1. Introduction

In GIScience, spatial information is mostly presented in the form of space and coordinates while human reasoning, behavior and perception is based on place, not space (Goodchild 2015). Place is related to the human experience of the world (Couclelis 1992) and is usually ambiguous and context-dependent. In recent decades, the speed of urbanization has increased, enabling cities to engage in a wide range of human functions and activities. Sense of place emerges when a place is functionally different from its surroundings and is therefore recognizable (Purves et al. 2019). In this case, GIS should link functionality and space using data and knowledge about human activity to answer questions such as "What can I do there?" or



Published in "Proceedings of the 16th International Conference on Location Based Services (LBS 2021)", edited by Anahid Basiri, Georg Gartner and Haosheng Huang, LBS 2021, 24-25 November 2021, Glasgow, UK/online.

<https://doi.org/10.34726/1789> | © Authors 2021. CC BY 4.0 License.

"How can I find places that provide a particular functionality?" In particular, GIS should be able to extract place functionality (Papadakis et al. 2019) that do not necessarily exist explicitly in the stored data.

Today, due to the rapid increase in using social networks, textual contents have grown significantly. Among the different types of user-generated contents, textual contents are often the most unstructured form and highly dependent on context and application. In the case of unstructured or semi-structured texts, first, they have to be structured and then different methods of natural language processing (NLP) can be applied to extract information and knowledge. Various studies have used user-generated textual contents such as place descriptions and user reviews to extract place information (Alazzawi 2012, Ballatore 2015, Chen et al 2018, Gao2021, Richter et al 2013).

This paper aims to extract place functionality based on analysis of user-generated textual contents such as users' reviews on travel blogs. Khoury et al. (2006) believe that the activity described by a sentence shows most of semantic information. Also, we are going to find how much place functionality can be represented through verbs that people use to describe or comment on a place. Therefore, the performance of action verbs on extracting place functionality is compared with the case when the whole sentence with lemmatized words is used.

2. Methodology

The general framework of the proposed method is summarized in *Figure 1*.

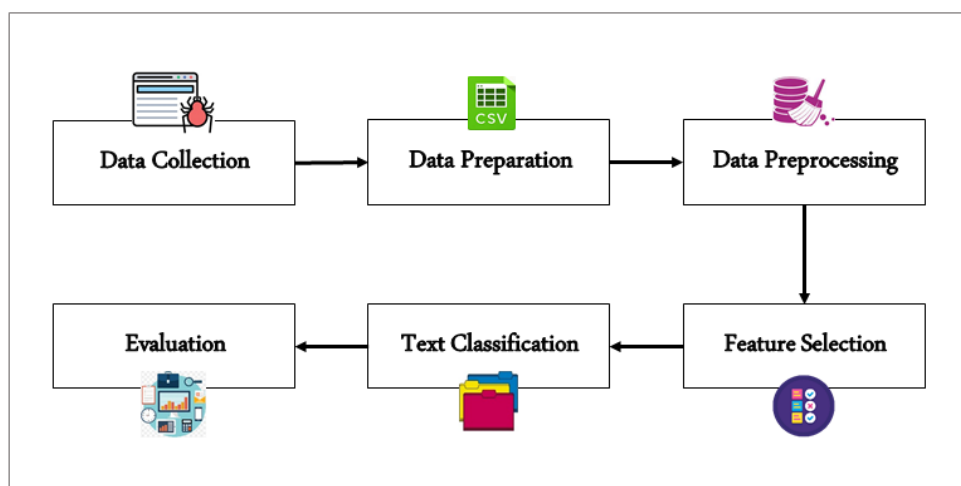


Figure 1. The general workflow of the proposed method

First of all, places and users' reviews on TripAdvisor were collected through web crawling by python. New York City (NYC) which is the most populous and also the most densely populated major city in the United States is chosen as our study area. These data were collected in October, 2020, and only English reviews are considered. For each place, place ID, name, type, sub-type, and geographical location of all realted places are extracted. Then, for each place, place ID, review ID, review title, review text and review rating of a maximum of 1000 top reviews are extracted. Finally two files are created, one for places and one for users' reviews. These two files can be joined using place ID.

It is clear that these data were unstructured and unprepared. In order to prepare data, places without geographic coordinates, those that were not located at the study area, NYC, duplicates, and places whose types are unknown were removed. There are five place types in TripAdvisor blog. *Table 1* shows the number of places and user's reviews in each category after preparing data.

Place Type	Attraction	Food Place	Hotel	Shopping	Vacation Rental
No of Places	1203	13282	842	1049	1080
No of Reviews	95661	325774	157625	31404	3334

Table 1. Number of places and user's reviews in each category

Data preprocessing is the most time consuming and important part in any text mining and machine learning task (Han & Kamber 2001). Therefore, first of all, the users' reviews should be preprocessed. NLTK library is used for this purpose. First, each review is converted to lower case and tokenized, then punctuations and stop words were removed. Afterward, all tokens are stemmed and lemmatized. In the first approach, these review texts with lemmatized words are considered. In the second approach, for each review verbs are extracted using NLTK and WordNet part-of-speech (POS) tagging. Then state verbs and auxiliary verbs are then removed and only action verbs in the review are selected for prediction process.

In the next step, the features should be generated and selected. As candidate words for feature selection, we only consider sufficiently frequent words. Bag of Words (BoW), Word2Vec, and Doc2Vec are different methods for feature selection that are used in our method.

Finally, place functionality of each user's review can be predicted using text classification methods in NLP. Crucially, our method is fully supervised, requiring a bag of words/word2vec/doc2vec representation of the objects

and place types as input. Logistic regression (LogReg), support vector machine (SVM), stochastic gradient decant (SGD), multinomial naïve bayes (MNB), k nearest neighbors (KNN), random forest (RF), gradient boosting (GB), decision tree (DT), and multi-layer perception (MLP) classifiers are different machine learning classifiers that are tested on both review text with lemmatized words and just action verbs of the review using three different methods of feature selection (BoW, Word2Vec, Doc2Vec). The performance of the text classifiers are evaluated using common machine learning evaluation measures such as accuracy, precision, recall, and f1-score by computing confusion matrix.

3. Results

Evaluation results demonstrate that in all the three feature selection methods and in all the different classification algorithms, using whole sentence with lemmatized words achieves better performance rather than considering just action verbs. This shows that words with differnt part-of-speech and the sementic relations between words are important for extracting place functionality. Among different classifiers, although SVM leads to better accuracy and f1-score, the execution time of the algorithm is about 317 times and 758 times that of LogReg for lemmatized words and verbs with just 1% less accuracy. Also, MNB is the fastest algorithm. The best prediction results are achieved for Food Places, Hotels, Attractions, Shops, and Vacation Rentals, respectively. Figure 2 shows the confusion matrix of SVM for lemmatized words and verbs using BoW feature selection method.

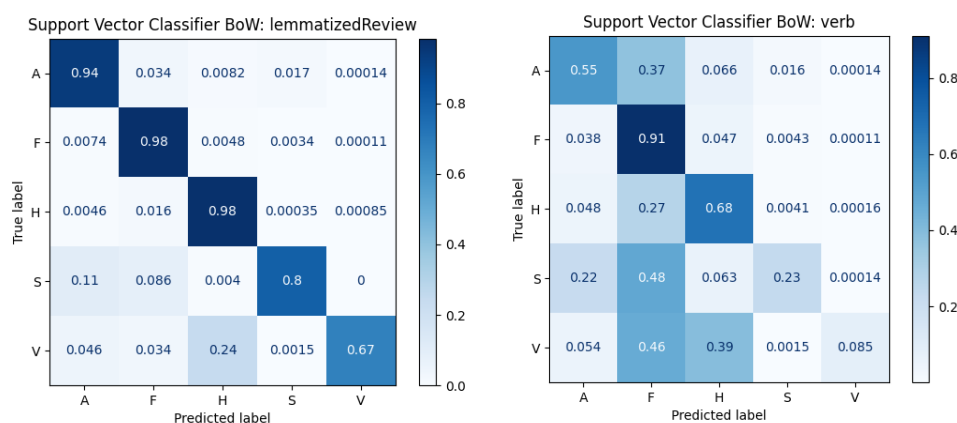


Figure 2. The confusion matrix of SVM for lemmatized words and verbs using BoW feature selection method.

4. Conclusion

Applying human cognition in defining and extracting place information is necessary. Today, this would be possible due to the growth of technologies related to artificial intelligence such as natural language processing and machine learning algorithms. Place functionality is considered important information in defining a place by humans. The propose of this paper was extracting place functionality through analyzing user-generated textual contents. Therefore, the activity of people in the place was extracted using action verbs, and then by using machine learning methods, the place functionality was predicted from the whole text and only action verbs. The results show that using whole sentence with lemmatized words achieve better performance rahter than considering just action verbs. Also, among different classifiers, SVM leads to better accuracy and f1-score.

References

- Alazzawi, A. N., Abdelmoty, A. I., & Jones, C. B. (2012). What can I do there? Towards the automatic discovery of place-related services and activities. *International Journal of Geographical Information Science*, 26(2), 345-364
- Ballatore, A., & Adams, B. (2015). *Extracting place emotions from travel blogs*. Paper presented at the Proceedings of AGILE.
- Chen, H., Vasardani, M., Winter, S., & Tomko, M. (2018). A graph database model for knowledge extracted from place descriptions. *ISPRS International Journal of Geo-Information*, 7(6), 221.
- Couclelis, H. (1992). Location, place, region, and space. *Geography's inner worlds*, 2, 15-233.
- Gao, S., Liu, Y., Kang, Y., & Zhang, F. (2021). User-Generated Content: A Promising Data Source for Urban Informatics.
- Goodchild, M. F. (2015). Space, place and health. *Annals of GIS*, 21(2), 97-100
- Khoury, R., Karray, F., & Kamel, M. (2006). *Extracting and representing actions in text using possibility theory*. Paper presented at the Proceedings of the 3rd annual e-learning conference on Intelligent Interactive Learning Object Repositories (i2LOR 2006).
- Papadakis, E., Baryannis, G., Petutschnig, A., & Blaschke, T. (2019). Function-based search of place using theoretical, empirical and probabilistic patterns. *ISPRS International Journal of Geo-Information*, 8(2), 92
- Purves, R. S., Winter, S., & Kuhn, W. (2019). Places in information science. *Journal of the Association for Information Science and Technology*, 70(11), 1173-1182.
- Richter, D., Winter, S., Richter, K.-F., & Stirling, L. (2013). Granularity of locations referred to by place descriptions. *Computers, environment and urban systems*, 41, 88-99.

Han, J., & Kamber, M. (2001). Data mining: Concepts and techniques. San Francisco, CA: Morgan Kaufman.