

DISCOVER POINTS OF INTEREST BASED ON USERS' INTERNET SEARCHES THROUGH AN ONLINE SHOPPING WEBSITE

G. Shahriari Mehr¹, M.R. Delavar^{2,*}, C. Claramunt³, B.N. Araabi⁴, M.R. A. Dehaqani⁴

¹ School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran-g.shahriari@ut.ac.ir

² Center of Excellence in Geomatic Engineering in Disaster Management, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran-mdelavar@ut.ac.ir

³ Naval Academy Research Institute Lanveoc-Poulmic, BP 600, 29240 Brest Naval, France-christophe.claramunt@gmail.com

⁴ School of Electrical and Computer Engineering, School of Engineering, University of Tehran, Tehran, Iran-(araabi,dehaqani)@ut.ac.ir

Commission IV, IV/3

KEY WORDS: Internet behavior, Points of Interest, Marketing, Recommendation, Similarity, Context-awareness

ABSTRACT:

In recent years, the development of the Internet plays a significant role in human's daily activities. One of the most important effects of the Internet is the change in the process of shopping. The advent of online shopping leads to establish a new channel for customers to obtain information about their desired goods and demands. Although many customers collect information from online channel, they also wish to try and search for their required goods at the stores. Besides, discovering this data leads to a new source for spatial analysis to find the users' interests. Therefore, we can consider this data as a contextual information source for spatial analysis or primary source for recommending points of interest (POIs). In this research, our aim is to discover a relation among the users' internet searches and the goods at the stores to recommend the best store to the users. Euclidean distance is used to calculate the similarity between users' searches and the available goods at the stores. The proposed method has been implemented in the city of Tehran, capital of Iran. The results show that the users' internet search behavior plays an essential role in the recommendation system which provides stores to the users based on the similarity among the users' internet searches and the available goods at the stores.

1. INTRODUCTION

Nowadays, the development of the internet affects daily activities such as the shopping process. Online shopping websites have established a new channel to search and find goods. On the other hand, this channel provides appropriate goods for customers to reduce their information load and search costs [1]. This information leads to more efficient shopping or to have an evaluation of the available goods at the stores [2]. Thus, the customers can go to a store with previous knowledge about the goods available at a store. Therefore, the discovery of the relationships between store goods and the preferences of the users provides information to the users.

In order to avoid the customers' confusion in purchasing the goods, the users need to go to stores that have goods tailored to their demands. Therefore, considering users' internet searches can provide better information about the users, this information leads to a better recommendation to the customers. For example, Lee et al. [3] have developed an analysis of clickstream on the online retail store. Also, Cho et al. [4] proposed a personalized recommender system for the internet shopping mall based on pattern recognition on the web usage data. Lopes and Roy [5] proposed a recommender system using web mining and pattern recognition from the historical data which are dynamically changed as per changing users' behavior. Pattern recognition from web data can be applied to some applications such as web personalization, system improvement, business intelligence discovery [6]. We believe that this data can be applied to spatial

* Corresponding Author

analysis. Thus, the objective of this paper is to introduce internet searches as a new source for spatial analysis.

Internet searches can be used as a new source to find the points of interest (POIs) because they consist of users' interests. In recent years, thanks to the development of the location-based systems (LBS), a number of research have used location-based social media to gain the users' interest based on their historical records. For example, Wang et al. [7] suggested recommendation algorithms based on past users' visited places, the location of each venue, the social relationship and the similarity among the users. Cho et al. [8] investigated human mobility pattern on location-based social networks check-in data and cell phone location data. They develop a model of human mobility which is predicted the locations and human's movement in the future. Gonzalez et al. [9] studied the trajectories of 100000 mobile phone users during six months. They show that despite all the diversity of their travel history, people follow simple, predictable patterns. On the other hand, the location based-recommender system apply Global Positioning System (GPS) trajectories, users comments and so on to provide better recommendation [10]. Users do their internet searches not only without any stress but also without time constraint, so this type of data leads to one of the best methods to find users' interests.

The context consists of information that leads to determine a user's status and his or her surrounding environment [11]. There are elements such as physical environment, users' attitude, users' role and the relationship among users and users' interests that can be introduced as a context to the spatial analysis of the location-based system. Therefore, as the users' internet searches can be considered as users' preferences, they can be considered as a context to develop a recommender system based on users' preferences and internet search.

Therefore, the contribution of this research is to discover a relation among the users' internet searches and physical stores. It also considers users' internet searches as a new source for spatial analysis, especially location-based recommender system. This information is suitable for spatial analysis because web interaction does not have any sparse data like social networks data[12].

2. METHODOLOGY

The objective of this paper is to consider the users' internet searches as a source to improve the location based recommendation system based on online shopping websites. The proposed methodology is explained in two steps. First, the required data is introduced and data preparation is elaborated in section 2.1. At the next step, the data analysis to discover the relationships among the users' internet searches and the goods available at stores are explained in section 2.2.

2.1 Data collection

Two independent data including (1) users' internet searches on the online shopping websites for discovering users' preferences and (2) the available goods at stores and positions are explained below.

2.1.1 Users' internet history searches

As the scope of this research is marketing, so tracking users' internet searches on an online shopping website is considered. The internet shopping search data have been collected from a company, Fan Gostaran Pishro Bazar, that consists of some of their users' searches which have been done over a period of time on the www.torob.com website. A Comma - separated values (CSV) format file which included the users of the website and their clickable URL acquired from the company. These data are shown in Figure 1. Although these data include some detailed information about the users and their searches, these data cannot be stored in the PostgreSQL database. Besides, there is no direct access to the company database, so a web scraping process should be applied to extract the required information from the website.

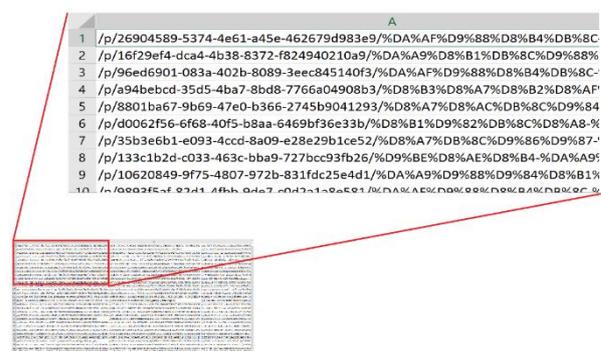


Figure 1- A piece of the raw data acquired from the company

Web scraping is a technique to extract structured data set from the websites and provide human-readable content. A web scraper can extract important elements of the web page which is saved in a structured data set [13]. Therefore, the user's search information such as category, subcategory and products price are obtained with this technique and stored in the PostgreSQL database. The web-scraping procedure is shown in Figure 2.

2.1.2 Stores characteristics

For our analysis, we need to know the available goods at the stores and stores positions. Although some of these types of data can be accessible from volunteered geospatial/geographic information (VGI) source. These data are collected by applying field observation in the study area. The stores positions that recorded in a Google map application are extracted using Google Takeout service.

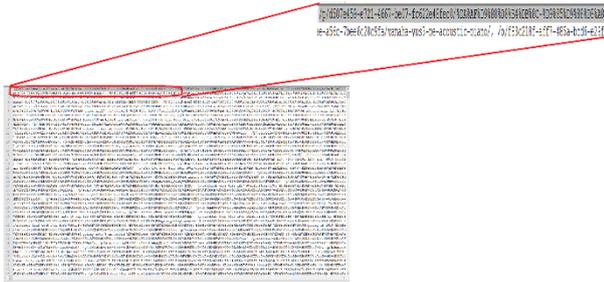


Figure 2(a) An overview of the webpage link

user_id [PK] integer	user_search_id [PK] integer	numeric_first_category integer	numeric_second_category integer	numeric_third_category integer	product_code integer	product_name character varying (50)	
1	1	1	12	1	458	Bluboo S8 Plus	
3	1	2	12	4	166	Samsung Galaxy A5 (2017) Dual SIM Mobile Phone	
3	1	8	12	1	9	31	37
4	1	4	12	2	7	128	تابلو شارژر کفد لایبرو TSCD
5	1	5	12	1	9	70	Phone SE 512GB
6	1	7	12	1	9	68	Phone SE 64GB
7	1	8	12	5	11	2	LG Tone Infinix HDS-610 Wireless Stereo Headset
8	1	9	12	5	1	30	[null]
9	1	10	12	5	394	1	EarPods MD625
10	1	11	12	1	9	48	Phone 6s 64GB
11	1	12	12	1	14	2	18.5GB
12	1	47	12	2	4	1	[null]
13	1	48	12	2	4	646	Exoam M4 / M4 Ac TPU Armor Back Cover
14	1	49	12	2	4	1	[null]

Figure 2(c) The User Searches table in the database

Google Takeout service, or Download Your Data is a new service for Google users to download the data, and storing it in a variety of standard formats. Thus, the field observation data are downloaded in a JavaScript Object Notation (JSON) format, and stored the extracted data in the PostgreSQL database.

We collected the available goods at the stores by using a questionnaire. This questionnaire filled by the sellers. Some seller provided us a file form their store's databases which included the available goods at the stores

2.2 Data Analysis

The analysis includes recognizing the similarity between the user's preferences and the available goods at stores. The similarity analysis is done at three steps including creation of feature class, normalizing the feature class and calculating Euclidean distance between the user feature class and the stores feature classes. The first and second steps are done independently on the users and the stores, while the third step combined the results of the second step to calculate the similarity.

Step one: In this step, feature classes are created based on the website category. The website categories are shown in Table 1.

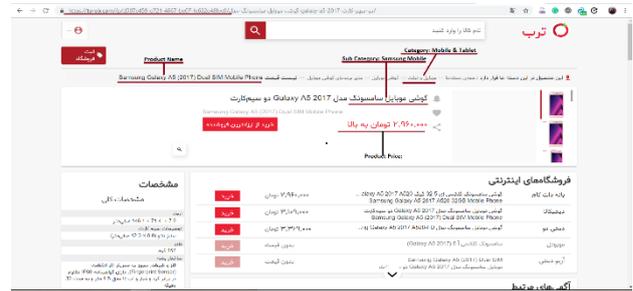


Figure 2(b) An overview of the website for the specific webpage link

user_id [PK] integer	user_search_id [PK] integer	numeric_first_category integer	numeric_second_category integer	numeric_third_category integer	product_code integer	product_name character varying (50)	
1	1	1	12	1	458	Bluboo S8 Plus	
3	1	2	12	4	166	Samsung Galaxy A5 (2017) Dual SIM Mobile Phone	
3	1	8	12	1	9	31	37
4	1	4	12	2	7	128	تابلو شارژر کفد لایبرو TSCD
5	1	5	12	1	9	70	Phone SE 512GB
6	1	7	12	1	9	68	Phone SE 64GB
7	1	8	12	5	11	2	LG Tone Infinix HDS-610 Wireless Stereo Headset
8	1	9	12	5	1	30	[null]
9	1	10	12	5	394	1	EarPods MD625
10	1	11	12	1	9	48	Phone 6s 64GB
11	1	12	12	1	14	2	18.5GB
12	1	47	12	2	4	1	[null]
13	1	48	12	2	4	646	Exoam M4 / M4 Ac TPU Armor Back Cover
14	1	49	12	2	4	1	[null]

Figure 2(d) The numeric user searches table in the database

Table 1-Category types of the website (www. torob.com)

Category Name	Category Code
Tools and Home appliances	1
Clothes, Shoes, Accessories	2
Automotive	3
Beauty and Personal Care	4
Health	5
Electronics	6
Books	7
Baby, Toys and Games	8
Computers and Office Devices	9
Home and Kitchen	10
Mobile and Tablet	11
Hypermarkets	12

According to our assumption, each store belongs to only one category. The user and the store of selected category are in the same feature space so that we can consider the similarity between the user and their associated stores as a measurement for allocating the store to the user. In the N-dimensional feature space, the similarity is computed using Equations 1, 2, 3 and 4.

$$\vec{X} = (x_1, x_2, x_3, \dots, x_n) \quad (1)$$

$$\vec{X}^u = (x_1^u, x_2^u, x_3^u, \dots, x_n^u)$$

$$\vec{X}^s = (x_1^s, x_2^s, x_3^s, \dots, x_n^s)$$

where X^s define the store feature vector and X^u is the user feature vector.

$$\|\vec{X}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (2)$$

$$\vec{Z} = \frac{X}{\|X\|} \quad (3)$$

where \vec{Z} is a normalized vector of \vec{X}

The similarity between the user's interests and the available goods at stores are calculated by the Euclidean distance using Equation 4.

$$d(u, s) = \|Z^u - Z^s\| = \sqrt{(Z_1^u - Z_1^s)^2 + \dots + (Z_n^u - Z_n^s)^2} \quad (4)$$

3. IMPLEMENTATION

In this paper, Six hundred stores are considered which are located in Tehran District 6. This area is located in a crowded area with commercial land uses. The stores locations are shown in Figure 3. Besides, the information of one thousand users have been acquired from Fan Gostaran Pishro Bazar company. There is no problem with data sparsity because the users whose shopping information have been acquired from the company have undertaken 100 to 200 internet searches.

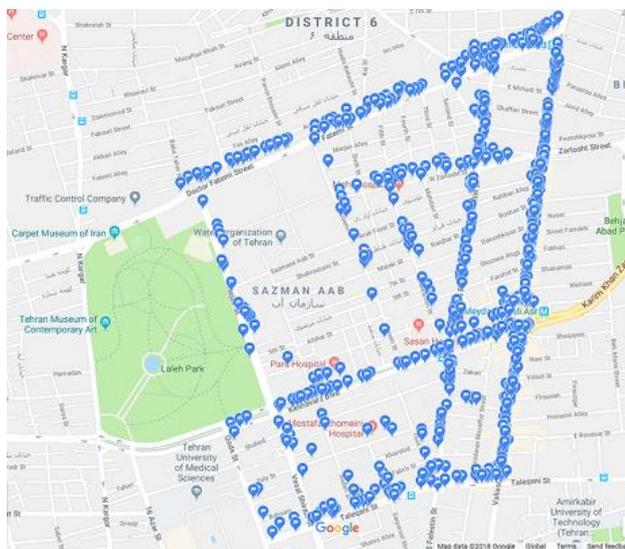


Figure 3- The registered stores in Google map by field observation

For example, the user with ID 73 searches are shown in Table 2. According to the user's internet searches, ten feature classes should be created. For example, If the user wants to buy a cloth, we have to extract a feature class in category two not only from the user but also from the stores. The user's feature vector is $X^u = (21, 0, 3, 17, 1, 0)$. The user's feature vector and some stores feature vectors are illustrated in Figure 4.

Table 2- User 73 searches

Category	Total	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	31	0	1	0	30					
2	21	0	3	17	1	0				
3	6	3	0	1	2					
4	3	1	0	0	0	0	0	0	2	0
6	5	0	5	0	0	0	0			
7	1	0	0	0	0	0	0	0	1	0
9	2	1	0	0	1	0	0	0		
10	26	1	3	1	0	0	0	1	20	0
12	36	16	13	2	5	0				
13	1	1	0	0	0					
all	132 Searches									

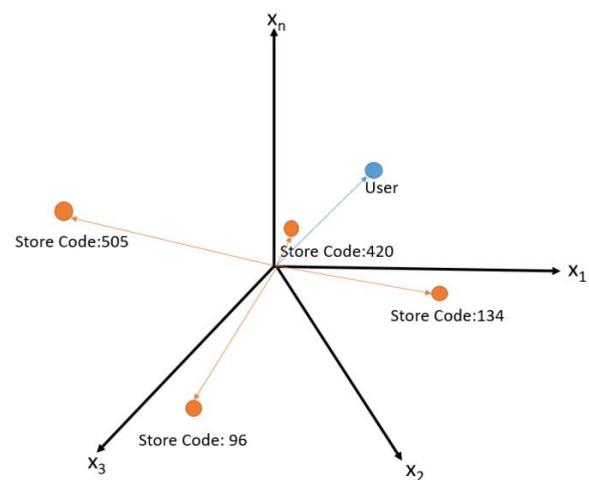


Figure 4- The user's feature vector in clothes category and four clothes stores

In the last step, the similarity among the available goods at the stores and the user's desired goods have been calculated based on Euclidean distance. Now, the user has a right perspective of the available goods at the neighbor stores, so he or she can get a better experience by going to the stores. The final ranked list, which is included a ranked list of the stores based on the similarity, is shown in Figure 5.

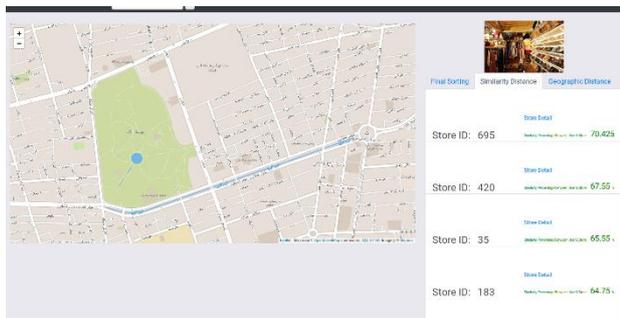


Figure 5- A view of the website that illustrates the similarity

4. EVALUATION

According to the aim of this paper, the system performance to discover the pattern among the neighbor stores and the users' internet searches is evaluated based on the similarity percentage among a user and stores. In order to evaluate the system performance, the role of the users has been taken by 10 users among the company's users. The evaluation results are shown in Table 3.

User Num	Cat 1	Cat 2	Cat 4	Cat 6
3	60	57	74	36
73	76	58	75	76
23	34	56	84	62
21	49	71	65	87
578	58	53	64	58
954	48	58	49	56
654	58	43	65	54
639	58	47	73	61
83	60	54	84	74

The results show that the performance of the system is relatively satisfactory. Max similarity among the stores and the users is 87. Moreover, the results indicate a relation among the internet searches and points of interests (POI)s.

5. CONCLUSION

This research recognized a pattern among the users' internet searches and the available goods at the stores. This research has been undertaken at the three steps including collecting users' internet searches and the associated stores' characteristics, recognizing the similarity between the stores and the user and sorting the stores based on their similarity. The objective of this

research is that it considers internet searches as a new contextual information source for the improvement of the spatial analysis. The results reveal that considering the user's internet search behavior can provide better information about users and their shopping requirements. On the other hand, with the development of the internet, internet search behaviors provide some efficient information to the users that can be employed for their shopping and location based spatial analysis. The approach is based on a better consideration of the similarity among the users' internet searches on online shopping websites and physical stores.

Future research should combine this data with geospatial data such as distances from the users' position to the stores' position for a better recommendation. Also, Future works should consider a collaborating filter in order to provide a recommendation based on similar users.

6. REFERENCES

- [1] D.-N. Chen, P. J.-H. Hu, Y.-R. Kuo, and T.-P. Liang. (2010). A Web-based personalized recommendation system for mobile phone selection: Design, implementation, and evaluation. *Expert Systems with Applications*, 37(12), 8201-8210.
- [2] A. Romadhony, S. Al Faraby, and B. Pudjoatmodjo. (2013). *Online shopping recommender system using hybrid method*. Paper presented at the 2013 International Conference of Information and Communication Technology (ICoICT).
- [3] J. Lee, M. Podlaseck, E. Schonberg, and R. Hoch. (2001). Visualization and analysis of clickstream data of online stores for understanding web merchandising. *Data Mining and Knowledge Discovery*, 5(1-2), 59-84.
- [4] Y. H. Cho, J. K. Kim, and S. H. Kim. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications*, 23(3), 329-342.
- [5] P. Lopes and B. Roy. (2015). Dynamic recommendation system using web usage mining for e-commerce users. *Procedia Computer Science*, 45, 60-69.
- [6] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *Acm Sigkdd Explorations Newsletter*, 1(2), 12-23.
- [7] H. Wang, M. Terrovitis, and N. Mamoulis. (2013). *Location recommendation in location-based social networks using user check-in data*. Paper presented at the Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.
- [8] E. Cho, S. A. Myers, and J. Leskovec. (2011). *Friendship and mobility: user movement in location-*

based social networks. Paper presented at the Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.

- [9] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. (2008). Understanding individual human mobility patterns. *nature*, 453(7196), 779.
- [10] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. (2010). *Collaborative location and activity recommendations with GPS history data*. Paper presented at the Proceedings of the 19th international conference on World wide web.
- [11] I. Afyouni, C. Ray, and C. Claramunt. (2012). Spatial models for context-aware indoor navigation systems: A survey.
- [12] M. Habayeb, B. Soltanifar, B. Caglayan, and A. Bener. (2016). *A novel point of interest (poi) location based recommender system utilizing user location and web interactions*. Paper presented at the 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService).
- [13] G. Boeing and P. Waddell. (2017). New insights into rental housing markets across the united states: web scraping and analyzing craigslist rental listings. *Journal of Planning Education and Research*, 37(4), 457-476.