Journal of Engineering and Applied Sciences 13 (18): 7716-7720, 2018
ISSN: 1816-949X

Semantic Based Geotagged Photos Similarities for Location's Ranking Purposes

Ghaidaa A. Al-Sultany
Department of Information Network, Babylon University, Babel, Iraq

Abstract: Nowadays, large amount of user-generated data can be obtained from social media (e.g., Instagram and Flicker). People sharing their travel experiences with Geotagged photo through such this kind of media. In addition, the taken photo can provide significant information such as title, location and tags from which a new perspective to comprehend the contexts of users can be reflected. In this study, the ranking of touristic attractions based on semantic similarity among retrieved Geotagged photos from social networking media has been investigated. We focused on tourism service by collecting and analyzing Geotagged photo from the social media to identify the most popular tourist places and rank them based on user location.We used HITS algorithm that rank locations based on the relation between the locations and tags that revealedvia weighting scheme. Thus, intelligent location-based tourist services can be provided to people.

Key words: Semantic analysis, Geotagged photos, social network data, ranking algorithm, networking media, tourist services

INTRODUCTION

In recent decades, many of smart applications have been diversely utilized in mobile devices in which the mobile and web based services have increased abilities human exploration (Nguyen *et al*., 2017; Memon *et al*., 2015). The use ofsocial media is a growing phenomenon in contemporary society and its platform used as a means of communications for sharing information (Townsend and Wallace, 2016). The huge amounts of shared data on social network (e.g., Flicker, Twitter, Instagram) that being Geotagged has reflected activities and interests of people's. It has known thatassociating geographic information with the social media helps significantly to understand what geographic areas people are interest in it (Sadilek *et al*., 2012). As the data of social media associated with geographic coordinates, it will be called Geotagged resources (Xia *et al*., 2014; Flatow *et al*., 2015).

People equipped with smart tourism services can share their travel experience such as Geotagged photos on social network and interact with tourist objects (Nguyen *et al*., 2017). Photo sharing web service and social networking have been regarded as one of significant source of tourist resources as it contains billion images that taken in different places of the world and various kinds of information can be obtained from (Memon *et al*., 2015). Consequently, the behaviors of tourists can be observed through social Geotagged data can effect positively on enhancing many applications such as travel recommendations and tourism resource (Peng and Huang, 2017).

In this study, the ranking of touristic attractions based on semantic similarity among retrieved Geotagged photos from social networking media has been proposed through collecting and analyzing Geotagged photo to identify the most popular tourist places and rank them based on user location.

Literature review: Recently, there are many studies that benefit from the social network service resources such as Geotagged photos which aim to provide a specific tourist service to users and help in the development of tourism. Peng and Huang (2017) was analysing set of Geotagged photos for retrieving the most popular tourist attractions in Beijing through applying text mining approaches on spatial clustering. The social network services such as Flickr have been exploited and have the attention of researchers for information retrieval purposes (Yin *et al*., 2011; Mousselly-Sergieh *et al*., 2014). Memon *et al*. (2015) the researchers had utilizedthe time and preference of tourists for locations travel. They recommended a new city to user in terms of his past preferences time by examining a Geotagged photos collectiondataset from Flickr that comes with spatial and temporal context (Memon *et al*., 2015). Other researcher showed that, it's possible to use metadata that attached with resources to provide another type of service (Tri and Jung, 2015) presented a method for finding the famous places related to food through exploiting attached tags annotated with geographical photos. Therefore, a list of ordered restaurants/places are retrieved with respect to food key words. They proposed a novel method (called Geo-HITS) as extending to HITS algorithm forannotated graphs analysis.

The unlabeled resources on social media also taken into consideration by some researcherslike Jason Jung whose proposed a resources classification method for predicting the location of unlabeled resources on social network using support vector machine and Naive Bayes and tags frequency (Adhianto *et al*., 2010). Other researchers by Lee and Lee (2014) focused on practical and structural aspects of Flickr mining. Through concentrating on Point of Interest (PoI) identification and mining the relationships among them (Tran *et al*., 2015).

MATERIALS AND METHODS

In this research, identifying ranked locations for user in a given area with respect to the public tags retrieved from social network. The research has focused on the touristic locations in which the tags that have been annotated with the user location are compared with set of general keywords (popular touristic terms). The similarity among the given tags are computed in term of the syntax similarity (cosine similarity), semantic similarity (synonym and polysemy) and co-occurrence based similarity (pointwise mutual). The output touristic places in user location are passed to the ranked process with respect to the tourism resources obtained from Geotagged social network media. Ranking based undirected graph HIT algorithm are implemented to rank locations in which the ranked locations have been depended on relationships between tags and locations (Fig. 1).

Data representation: In this research, a data of Geotagged photos on a world-wide scale was provided by Mousselly-Sergieh *et al*. (2014). It contains a sample of 5000 thousand Geotagged photos posted by the users around the world and crawled rom Flickr with the corresponding metadata such as title, tags, latitude, longitude, date taken, date uploaded and other attached properties belong to that photos.

In this research, the dataset has been divided into two partitions (training and test set) based on 70% for training process and 30% for testing. The collected tags have been pre-processed via removing noise data such as (stop words) and applying the stemming process on the chosen tags.

Tags relatedness

Syntax based tags similarity: The tags annotated with the user location are compared with set of general keywords (popular touristic terms). The similarity among the given tags are computed in term of the syntax similarity (cosine similarity) as in Eq. 1 (Garcia and Co, 2015):

$$\cos(\theta) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sqrt{\sum_{i=1}^{n} B_i^2}}} \quad (1)$$

where, A and B are the tags to be compared. After calculating the similarity among the touristic keywords and the tags of social network photos, we calculated the number of locations that contain the tourism tags to get a set of common tags that occurred in many locations.

Semantic based tags relatedness: Leveraging the tags with their semantics can enhance the process of ranking asthe tags are mainly used to determine a list of ranked locations using HITS algorithm. Therefore, to overcome the weakness in the result of ranking, the tags with their to form new subsets of tag's vocabularies has been proposed. The extracted features have adopted the

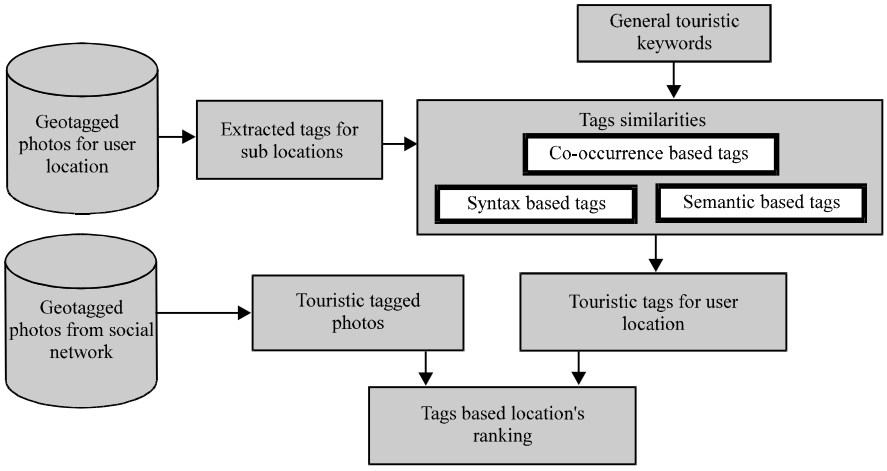

Fig. 1: Tags based location's ranking

shortest path measure to calculate the closeness of the tags words in the WordNet taxonomy as shown in Eq. 2:

$$sim_{path(t_i,t_j)} = 2*deepMax*len(t_i,t_j) \quad (2)$$

where, deepMax is the maximum path length between two concepts in the WordNet taxonomy and len is the minimum number of links between tags t_i and t_j.

Co-occurrence based tags similarities: Typically, the associated tags that occur mostly in same context reflect their relatedness and appearance in similar locations. Hence, the co-occurred words can be extracted via. dealing with them as one token in the text. The Point wise Mutual Information (PMI) has been applied to compute the association strength among the co-occurred words in the context. It calculates the frequency of any two tags appeared together proportional to their frequencies separately as shown in Eq. 3 (Bilal and Shalan, 2016):

$$PMI(t_i,t_j) = \log\left(\frac{P(t_i,t_j)}{P(t_i),P,(t_j)}\right) \quad (3)$$

Ranked based HITS algorithm: Basically, the framework of HITS algorithm (Aggarwal, 2016; Ricci *et al*., 2011) depends on the relationship among a set of nodes that are treated as authoritative and hub pages. The joined nodes are linked using directed graph link structure of webpages in which directed edge (p, q)∈E indicates the presence of a link from (p-q) with a score for hub webpage and score for authority webpage as shown in Eq. 4 and 5:

$$x^p = \sum_{q:(p,q)\in E} y^q \quad (4)$$

$$y^p = \sum_{q:(p,q)\in E} y^q \quad (5)$$

Based on the research by Nguyen *et al*. (2017), the relationship between tags and location are represented as an undirected graph G = (V, E) where the given tag t_i or a location l_j (is used such as hubs and authorities) can be described as vertexes and the edge $(t_i, l_j)\in E$ indicates the given location l_j may contain tag t_i. The tag's weights (w_{ij}) in terms of each location are calculated through Eq. 6 which reflect the occurrence of tag t_i in location l_j:

$$w_{ij} = \left(n(t_i,l_j)\right)/\left(\max\left\{n(t_k,l_j) \mid t_k \in T^{lj}\right\}\right) \quad (6)$$

where $n(t_i, l_j)$ is the number of occurrence of tag t_i in location l_j and t_k is a tag in the set of tags of location l_j. Typically, the initial rank value for each in node in the hub and authority is equal to one. In our research, depending on the normal distribution of initial rank used in PageRank algorithm (Aggarwal, 2016), the initial rank value has reflected the ratio of given tag's location -($1/T^l$) and the ratio of a particular locations that contain tags t-($1/L^t$). At each iteration, they are calculated based on Eq. 7 and 8:

$$R_{lj} = \sum_{i=1}^{m} \frac{1}{|L^{ti}|} R_{ti} \quad (7)$$

$$R_{ti} = \sum_{j=1}^{n} \frac{1}{|T^{lj}|} R_{lj} \quad (8)$$

The problem of ranking location focus on data that can obtained from SNS by using tags that attached with. The tags (t) determine the features of locations as an attached label to location or photo in order to identify or describe the photo's contexts. A location (l_j) refer to any region or place that is intended to be ranked and it may be identified through a set of tags obtained from social network media.

RESULTS AND DISCUSSION

Experimental results: In our research, a Geotagged photos dataset (Mousselly-Sergieh *et al*., 2014) has been utilized for examining the results of the research. It was collected from the user's postsin Flicker website on over the world. It has been divided into two parts, the firstonewas for the photos of a location of a given user or area in terms of their latitude and longitude. We assumed that a user is in New Orleans City which is on of Louisiana State in United States of America. Retrieving the positions of the around area of the selected location was collected using Google map API via. The latitude and longitude metadata. The second part was used the tags of the photos related with tourism words. The tags have been extracted from Geotagged photos that posted by users when they want to. We experienced the proposed work to test the results with 1300000 Geotagged photos which has different number of touristic locations and non-touristic locations. The first dataset contains a set of locations that near user locations, we assume that a given user is located in New Orleans City as shown in Table 1 while the second dataset contains tourist tags that used by public users whose published Geotagged photos social network as described in Table 2.

Table 1: The user location's places in New Orleans City with their tags

Location name	Location tags
2900-2998 Loyola Ave	Cemetery-grave-louisiana-neworleans-vault-crypt-fuchs
102 City Park Ave	Door-cemetery-rust-Louisiana-neworleans-chain doorknob-padlock-crypt-forlom-grave-warning-tomb-gravesite
5100 Canal Blvd	Flowers-cemetery-grave-gold-golden-louisiana-paint-neworleans-vault-bouquet-crypt-cemetery-grave headstone-tomb-brotherhood-boilermakers-shipbuilders
425 Basin St	Flowers-blue-roses-strange-grave-Louisiana-purple neworleans-eerie-vault-crypt-cemetery-cult vault-priestess-voodoo-marieleveau-ochre-soldier-memorial-revolution-revolutionarywar-black-dark-magic-occult-sinister grave death crypt new orleans louisian aeeriemacabre-death-macabre
1725 St Roch Ave	Strange-grave-death-louisiana-god-neworleans-cemetary-religion-jesus-eerie-graves-macabre-hdr-gravesite-bigeasy-stroch-camposanto
1914 Tea Room Dr	Bird-nature-animals-zoo-louisiana-wildlife-neworleans-parakeet-audubonzoo-orangutan-monkey-primate-feline-jaguar-growl-cat-aligator-bayou-albino-gator- swamp-tropical-red-colorful-toucan

Table 2: Touristic tags with number of occurrences

Touristic tags	No. of locations that contain it	Tourist tags	No. of locations that contain it
Louisiana	12	Slfashion	12
Germany	16	Porta	21
Deutsch land	14	Secondlife	11
Square	52	Ilfordxpii	13
Square format	52	Kodakportranc160	39
Instagram App	52	Ikon	10
Vienna	91	Uploaded: by = instagram	52
Europe	10	Nature	11
Container	14	Nonplace	11
Stpauli	14	Milano	11
December	13	Nonluoghi	11
Kran	14	Speicherstadt	14
Segelboot	14	Rotlicht	14
Miniatur Wunderland	14	Feuerschiff	14
Reihenhaus	14	Davidstrasse	14
Panoptikum	14	Raddampfer	14
Herbertstrasse	14	Modeleisenbahn	14
Canon	17	Weimar	13
Tauchmaske	14		

Table 3: The locations with its tourist tags

Location names	Tourist tags
2900-2998 Loyola Ave	Louisiana
102 City Park Ave	Louisiana
5100 Canal Blvd	Louisiana
425 Basin St	Louisiana
1725 St Roch Ave	Louisiana
1914 Tea Room Dr	Louisiana-nature

Table 4: The rank of tourist places

Rank of locations	Rank value for the locations
1914 Tea Room Dr	0.18452332055959883
2900-2998 Loyola Ave	0.1630953358880802
102 City Park Ave	0.1630953358880802
5100 Canal Blvd	0.1630953358880802
425 Basin St	0.1630953358880802
1725 St Roch Ave	0.1630953358880802

We have taken the data that resulted from dataset 1 and 2 in order to rank the locations that belong to the tourism. The tags which are closed to tourism termsare exploited to filter the information ona given location, so, from the two dataset we noticed that each location have a set of touristic tags as shown in Table 3.

By applying HITS algorithm depending on the relations between tags and locations, the most attractive touristic places and tourist tags in New Orleans City are retrieved. HITS algorithm has removed all the locations that not contain any tourist tag and remove the tags that don't found in any location, hence, more accurate outcomes resulted as shown in Table 4 and 5.

Table 5: The rank of tourist tags

Rank of tags	Rank value for the tags
Louisiana	0.5535699616787966
Nature	0.44643003832120337

CONCLUSION

In this research, ranked locations for user in a given area with respect to the public tags retrieved from social network has been proposed. The tags have been analysed and compared against set of most popular tourism terms in order to support the ranking process. The similarity among the given tags are computed in term of the syntax similarity (cosine similarity), semantic similarity (synonym and polysemy) and co-occurrence based similarity (pointwise mutual). In addition, we proposed HITS algorithmand modified the values of nodes with tags and relevant locations of which finding useful information for tourists, travel agents and enhancing the tourism will be developed.

REFERENCES

Adhianto, L., S. Banerjee, M. Fagan, M. Krentel and G. Marin *et al.*, 2010. HPCToolkit: Tools for performance analysis of optimized parallel programs. Concurrency Comput. Pract. Experience, 22: 685-701.

Aggarwal, C.C., 2016. Recommender Systems: The Textbook. Springer, Berlin, Germany, ISBN:978-3-31929657-9, Pages: 298.

Bilal, G. and R. Shalan, 2016. Semantic analysis based semantic analysis based textual features extraction and clustering textual features extraction and clustering. Res. J. Appl. Sci., 11: 1115-1121.

Flatow, D., M. Naaman, K.E. Xie, Y. Volkovich and Y. Kanza, 2015. On the accuracy of hyper-local geotagging of social media content. Proceedings of the 8th ACM International Conference on Web Search and Data Mining, February 02-06, 2015, ACM, Shanghai, China, ISBN:978-1-4503-3317-7, pp: 127-136.

Garcia, E. and A. Co, 2015. Cosine similarity tutorial. Inf. Retr. Intell., 2015: 4-10.

Lee, I., G. Cai and K. Lee, 2014. Exploration of geo-tagged photos through data mining approaches. Expert Syst. Appl., 41: 397-405.

Memon, I., L. Chen, A. Majid, M. Lv and I. Hussain *et al.*, 2015. Travel recommendation using geo-tagged photos in social media for tourist. Wirel. Pers. Commun., 80: 1347-1362.

Mousselly-Sergieh, H., D. Watzinger, B. Huber, M. Doller and E. Egyed-Zsigmond *et al.*, 2014. World-wide scale geotagged image dataset for automatic image annotation and reverse geotagging. Proceedings of the 5th ACM Conference on Multimedia Systems, ACM, Singapore, March 19, 2014, ACM, Singapore, ISBN:978-1-4503-2705-3, pp: 47-52.

Nguyen, T.T., D. Camacho and J.E. Jung, 2017. Identifying and ranking cultural heritage resources on geotagged social media for smart cultural tourism services. Pers. Ubiquitous Comput., 21: 267-279.

Peng, X. and Z. Huang, 2017. A novel popular tourist attraction discovering approach based on geo-tagged social media big data. ISPRS. Intl. J. Geo Inf., Vol. 6, 10.3390/ijgi6070216

Ricci, F., L. Rokach and B. Shapira, 2011. Introduction to Recommender Systems Handbook. In: Recommender Systems Handbook, Ricci, F., L. Rokach, L. Shapira and P. Kantor (Eds.). Springer, Berlin, Germany, ISBN:978-0-387-85819-7, pp: 1-35.

Sadilek, A., H. Kautz and V. Silenzio, 2012. Modeling spread of disease from social interactions. Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM'12), June 4-8, 2012, AAAI Press, Menlo Park, California, USA., pp: 322-329.

Townsend, L. and C. Wallace, 2016. Social media research: A guide to ethics. Master Thesis, University of Aberdeen, Aberdeen, Scotland.

Tran, V.C., D. Hwang and J.J. Jung, 2015. Semi-supervised approach based on co-occurrence coefficient for named entity recognition on Twitter. Proceedings of the 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS'15), September 16-18, 2015, IEEE, Ho Chi Minh, Vietnam, ISBN:978-1-4673-6639-7, pp: 141-146.

Tri, N.T. and J.J. Jung, 2015. Exploiting geotagged resources to spatial ranking by extending hits algorithm. Comput. Sci. Inf. Syst., 12: 185-201.

Xia, C., R. Schwartz, K. Xie, A. Krebs and A. Langdon *et al.*, 2014. CityBeat: Real-time social media visualization of hyper-local city data. Proceedings of the 23rd International Conference on World Wide Web, April 07-11, 2014, ACM, Seoul, Korea, ISBN:978-1-4503-2745-9, pp: 167-170.

Yin, Z., L. Cao, J. Han, J. Luo and T. Huang, 2011. Diversified trajectory pattern ranking in geo-tagged social media. Proceedings of the 2011 SIAM International Conference on Data Mining (ICDM'11), April 28-30, 2011, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA., ISBN:978-0-89871-992-5, pp: 980-991.