Computer Science and Information Systems 12(1):185–201 DOI: 10.2298/CSIS141015091T

Exploiting Geotagged Resources to Spatial Ranking by Extending HITS Algorithm

Tuong Tri Nguyen[1] and Jason J. Jung[2*]

[1] Department of Computer Engineering
Yeungnam University
Gyeongsan, Korea 712-749
[2] Department of Computer Engineering
Chung-Ang University
Seoul, Korea 156-756
{tuongtringuyen, j2jung}@gmail.com

Abstract. With a large amount of geotagged resources from smart devices, it is important to provide users with intelligent location-based services. Particularly, in this work, we focus on spatial ranking service, which can retrieve a set of relevant resources with a certain tag. This paper designs ranking algorithm in order to find out a list of locations which are collected from geotagged resources on SNSs. As extending HITS algorithm [13], we propose a novel method (called GeoHITS) that can analyze an undirected 2-mode graph composed with a set of tags and a set of locations. Thereby, meaningful relationships between the locations and a set of tags are discovered by integrating several weighting schemes and HITS algorithm. To evaluate the proposed spatial ranking approach, we have shows the experimental results from the recommendation applications.

Keywords: Spatial ranking, Geotagged resource, GeoHITS, Information ranking, Recommendation service.

1. Introduction

Social networking service (SNS) is known as an application system in which made up of individuals connected by one or more specific types of relationship, reflects the real-life among people through online platforms such as a blog, website, bring many ways for users to share ideas, events, activities, and interests via internet [9,10,1,11].

Today, the use of the digital camera during traveling is becoming increasingly popular. Besides, sharing photos on the SNS is also an inevitable trend, and it became a huge source of data. Exploiting the data source tagging geographical location to introduce tourist attractions, to recommend wonderful places to travellers, to promote the tourism services, etc are more and more interesting and necessary up today. There are many studies mentioned about this topic [5,7,12,15,2,3,14,16].

Using location resources focus on three folds [2], which are *geo-tagged-media-based, point-location-driven* and *trajectory-centric*. Geo-tagged-media-based method focuses on using data on SNS and enable users to add a location label to media content such as text,

* Corresponding author

photos, and videos generated in the physical world (e.g., data enriched with location information [17], giving personalized travel recommendations for users with a clear preference for a specific type of landmark [5]). Point-location-driven method refers to encourage people to share their current locations, e.g., restaurants and museums. By this way, users can recommend other users many things to do, see, or eat at the location [18,11,15,14]. Trajectory-centric scheme is focusing on basic spatial information of a certain location (e.g., distance, duration) and velocity about a particular trajectory. It can show user experiences represented by tags, tips, and photos for the trajectory.

The data more and more increasing that implies the information of the provided locations more complete and detail. Fig. 1 shows the chart of geotagged resources which are collected in 9 years from 2005 to 2013 on Flickr[3], demonstrated that.

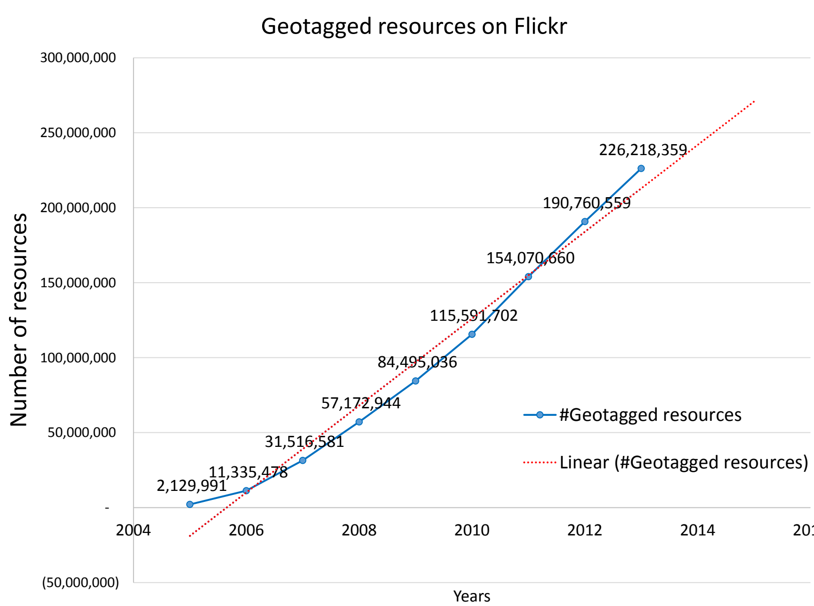


Fig. 1. Geotagged resources on Flickr from 2005 to 2013.

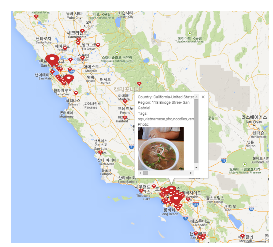
In this study, we show an idea in order to exploit the users' tags which attached on geographical photo to help finding the most appropriate famous places by using keyword. Our approach is similar to the top-down method. At the beginning, users input a keyword

[3] www.flickr.com

(e.g., 'pho'). The system collects data from SNS and implement to show results on the map same as Fig. 2. In this, we show three map for finding 'pho'- a famous dishes of Vietnam. On the world map (Fig. 2a), we easy to find out where is the best country for enjoy this food (e.g., Vietnam, United States, Canada relate to the big icons). And by selecting United States, the map (Fig. 2b) indicates California is one of the best places for 'pho'. By doing this way with California on the map (Fig. 2c), there are many photos (including 'pho') which are related to restaurants, parties, meals, etc, that we can see on the map including address, photos, tags. So, by giving a keyword and limited locations, for example 'pho' in state of 'California', we can find out a list of ordered restaurants/places to enjoy 'pho' (a traditional food of Vietnam).

(a) 'Pho' on the world map

(b) 'Pho' on the United State map



(c) 'Pho' on California map

Fig. 2. Spatial distribution of the geotagged resources with 'pho' on the map

For finding the best thing, ranking is one of the easy ways to do this issue. Our work is introduced through a workflow as in Fig. 3. The components are described as follows:

– *Collecting data*: Dataset is collected by using Open API of SNS;
– *Tag analysis*: The list of tags is refined by removing the stop words and classified based on the geotagged photos (all tags will be determined the number of locations it's belongs to) and determined the set of common tags

– *Extending HITS algorithm*: According to [4,13], we apply the HITS algorithm by using a set of nodes which are tags or locations for an undirected graph, called $LocHITS$; For comparing with this algorithm, we use similarity measurement between tags of each location and a set of common tags (called $LocHITS_S$ algorithm). Besides, we use tag frequency (TF) in order to add value for tag nodes (called $LocHITS_{TF}$ algorithm) as a case study for evaluating results.

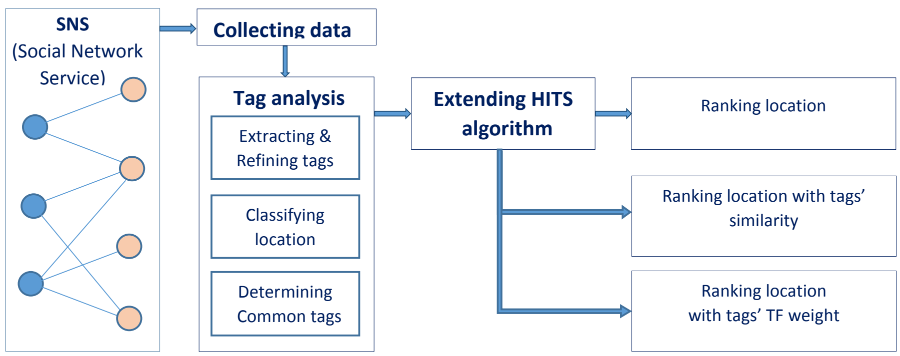


Fig. 3. The workflow of the ranking location system

Herein, we present the experimental results with the set of tags from the geotagged photos. After collecting data, we collect all locations for computing the relations between tags and locations. For all tags, we determine the value of term frequency of tag in each location.

The paper is organized as follows. Sect. 2 introduces some issues related to background for ranking location based on geotagged resources. Sect. 3 shows some use case of HITS algorithm combined with tags similarity and term frequency of tags occurs in each location. Based on the weight of tags in order to ranking location. Thereby, we could determine the rank of the locations which is found by the given keyword through SNS. In Sect. 4, experimentation was conducted to evaluate the results. Then, in Sect. 5 focus on some related studies and discuss about our works. Finally, Sec.t 6 draws a conclusion and future work of this study.

2. Problem Description

In this section, we focus the fundamental issues and problems related to ranking locations based on relationships between tags as well as locations. Additionally, we introduce the way to apply HITS algorithm to rank locations by using undirected graph instead of directed graph as the original HITS algorithm [13]. Thereby, we can determine the method to solve the problem posed.

2.1. Notation and definition

Ranking location problem focus on using tags on SNS in order to determine the features of locations. Thereby, the system can update information for finding out a suitable ordered list of locations related to the keyword that people used for searching.

Table 1. Notation

Symbol	Notation
t_i	a tag
l_i	a location
Θ	a set of tags
Λ	a set of locations
Θ^{l_j}	a set of tags of location l_j
Λ^{t_i}	a set of locations contains tag t_i

Definition 1 (LocTag). *In order to determine tag t_i is contained in location l_j or not, a function ϕ given as follow:*

$$\phi(t_i, l_j) = \begin{cases} 1, & \textit{if tag } t_i \textit{ is contained in location } l_j \\ 0, & \textit{otherwise}. \end{cases} \tag{1}$$

Definition 2 (LocRank). *LocRank is a location (l_j) for ranking, is name of a country, region, city, or place where is identified by a set of tags (Θ^{l_j}) over a social network service.*

$$\Theta^{l_j} = \{t_i | t_i \in \Theta, \left|\Lambda^{t_i}\right| \geq 2, \phi(t_i, l_j) = 1\} \tag{2}$$

In this issue, we use a set of geotagged photos in order to find out the best location by ranking method. These tags which occur on many locations should be selected for using in this work. In addition, if a tag only occurs in a location that is worthless for ranking locations. Because, they have no meaning with other location that did not contain them. So, in order to rank locations, we need to use a set of common tags, they are defined as follows.

Definition 3 (Common tag). *Common tags (denoted Θ_C) are a set of tags whose occurrence is greater than a threshold α.*

$$\Theta_C = \{t_i | t_i \in \Theta, \left|\Lambda^{t_i}\right| \geq \alpha, \alpha \geq 2\} \tag{3}$$

In our work, the number of tags grown in large scale. So we need to employ a selection process of the most popular tags. These tags not only form to a set of features in order to represent for locations but also have a relationship closed to the keyword. We use a set of frequent tags in locations to build a set of common tags. Thereby, we can compute the weight of tags, locations for our issue about ranking locations. For example, someone who would like to know where is the best place to enjoy 'Kimchi' can find out a good answer. Using the keyword *'Kimchi'* for collecting data from SNS, we collected 1416 geotagged photos. (as shown in Tab. 2). In which, there was 10 locations ($|\Lambda| = 10$) and 23395 tags ($|\Theta| = 23,395$). With $\alpha = 5$, we have $\Theta_C = \{$*kimchi, spicy, soup, food, lunch, korean, noodles, tofu, rice, salad, restaurant, hot, egg, travel, dinner, pork, market, kimchee*$\}$ ($|\Theta_C| = 18$), and the list of tags of locations is showed in Tab. 5.

2.2. Relationship between tags and locations

The HITS algorithm (Hypertext Induced Topics Search)[13] based on the relationship between a set of relevant authoritative pages and the set of hub pages that join them together in the link structure of webpages. In [13], webpages are described by a directed graph $G = \langle V, E\rangle$: the nodes correspond to the pages, and a directed edge $(p, q) \in E$ indicates the presence of a link from p to q. Each webpage has both a hub score y and an authority score x (as shown in Equ. 4).

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

In a directed graph, there are two types of degree of vertex p:

- In-degree (called $I_{d(p)}$): the number of links points to vertex p;
- Out-degree (called $O_{d(p)}$): the number of links from p point to others.

On the contrary, an undirected graph $G = \langle V, E\rangle$, $I_{d(p)} \equiv O_{d(p)}, \forall p \in V$. In this study, for ranking locations based on tags, the data is described by an undirected graph $G = \langle V, E\rangle$: a node corresponds to a tag t_i or a location l_j, and an edge $(t_i, l_j) \in E$ indicates tag t_i occurs in location l_j or in other words, location l_j contains tag t_i.

2.3. Computing the weight of tags in location

Using term frequency to compute the term weight is popular such as classifying document [22]. They can determine the valued class of a document by using term frequency from a set of words in that document.

In our work, the weight of tag is computed for each location. Thus, we calculate the occurrence of tag t_i in location l_j (denoted w_{ij}). The value of w_{ij} is computed as follows:

$$w_{ij} = \frac{\mathcal{P}(t_i, l_j)}{max\{\mathcal{P}(t_k, l_j)|t_k \in \Theta^{l_j}\}} \tag{5}$$

with $\mathcal{P}(t_i, l_j)$ 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.

3. Extending HITS algorithm for ranking locations

This section shows some use case of HITS algorithm combined with tags similarity and term frequency of tags occurs in each location. Based on the weight of tags in order to ranking location. Thereby, we could determine an order for a list of location which is found by keyword through SNS.

3.1. Formalization of ranking locations

According to [13], the HITS algorithm is used for webpages ranking. The hyperlinks from these webpages form a directed web graph $G = \langle V, E\rangle$, where V is the set of nodes representing webpages, and E is the set of hyperlinks. The hyperlink topology of the web graph is contained in the asymmetric adjacency matrix $L = \{l_{ij}\}$, where $l_{ij} = 1$ if $page_i \rightarrow page_j$ and $l_{ij} = 0$ otherwise. And each webpage p_i has both a hub score p_i^{Hub} and an authority score p_i^{Aut}.

In the paper, we use the relationship between tags and locations same as hubs and authorities in [13] and [6]. However, we use an undirected graph $G = \langle V, E\rangle$, where V is the set of nodes representing tags (Θ) or locations (Λ), and E is the set of edges. $V = \Theta \cup \Lambda$, and $E = \{(t_i, l_j) : \phi(t_i, l_j) = 1\}$, where $\phi(t_i, l_j)$ is defined as in Def. 1.

At the beginning, $\forall v \in V$ have value equal to 1 ($\psi_v = 1$). For each iterations, they are computed by the formulas as follows

$$\psi_{l_j} = \sum_{i=1}^{m} \frac{1}{|\Lambda^{t_i}|}\psi_{t_i}; \qquad \psi_{t_i} = \sum_{j=1}^{n} \frac{1}{|\Theta^{l_j}|}\psi_{l_j}; \tag{6}$$

The formula to compute the value of nodes as follows:

$$\theta = A\lambda \quad and \quad \lambda = A^T\theta \tag{7}$$

where A is an adjacency matrix ($m \times n$) with a_{ij} is determined by Equ. 1,
and $\theta = \{\psi_{t_1}, \psi_{t_2}, \ldots, \psi_{t_m}\}^T$, $\lambda = \{\psi_{l_1}, \psi_{l_2}, \ldots, \psi_{l_n}\}^T$.

From Equ. 7, they can be computed by recursive as follows

$$\theta = AA^T\theta \quad and \quad \lambda = A^TA\lambda \tag{8}$$

3.2. $LocHITS$ algorithm

From studying results of [4,13], we have proposed the $LocHITS$ algorithm (as shown in Alg. 1). In which each node is represented by a tag or a location.

In practice, using the Equ. 6 to compute values for nodes, we found that if one tag only appears in a location for a few times then it has worthless for ranking with another location. Besides a tag appears in many locations that has high value for ranking (it will be close to the keyword). It means that this tag is called co-occurrence with keyword. For that reason, we build a set of common tags in order to rank locations. The process is implemented based on a set of tags on the dataset and filtered by condition as in Def. 3.

Data: $\theta, \lambda, \varepsilon$
Result: R
Initialization;
Determine a set of common tags;
Determine matrix $A(m \times n)$;
Let $\theta = \{\psi_{t_1}, \ldots, \psi_{t_m}\}$ denote the vector {1,. . . ,1};
Let $\lambda = \{\psi_{l_1}, \ldots, \psi_{l_n}\}$ denote the vector {1,. . . ,1};
Iterations=0;
while *max of* $|\lambda_k - \lambda_{k-1}| \geq \varepsilon$ *or Iterations=0* **do**
    Computing $\psi_{t_i} = \sum_{j=1}^{n} (a_{ij}.\psi_{l_j})$ for $\forall i \in [1:m]$;
    Computing $\psi_{l_j} = \sum_{i=1}^{m} (a_{ij}.\psi_{t_i})$ for $\forall j \in [1:n]$;
    Normalize(θ);
    Normalize(λ);
    Iteration+=1;
end
$R \leftarrow arsort(\lambda)$;
Return(R)

Algorithm 1: $LocHITS$ Algorithm

3.3. Using similarity and term frequency

The use of tags to determine a list of ranking locations has been done by LocHITS. However, if the number of tags increases significantly, the process will find out certain difficulties: Slow processing speed; Information noise, because of many tags are not related to the keyword but they are still used to rank location. This leads to the ranking will no longer accurate.

To overcome the weaknesses, we need to know how to:

– Reduce the number of irrelevant tags in ranking location;
– Determine the weight of each tag in each specific location;
– Accelerate using tags which are close to the search keyword;
– Consider these tags which often occur in many locations in order to build a set of common tags.

In this study, we solve this issue by focusing on two aspects: *Relationship between tags of locations (tags similarity)* and *Tags occur in each location (tags frequency)*.

Tags similarity. We try to impact to change the weight of vertices (tags or locations) for applying $LocHITS$ algorithm. Using statistics methods [8,21] with the dataset, we select a set of common tags following the Def. 3. Besides, to calculate the ranking coefficient for each location based on its tags, we have used the Jaccard similarity method[19] for determining the similarity between the set of common tags with a set of tags of each location. By this way, $LocHITS$ is extended with tags similarity coefficient (called $LocHITS_S$ Algorithm).

Tags frequency. Both $LocHITS$ and $LocHITS_S$, the weight of each tag has not been considered. In actually, the value of each tag in each location is distinguished. Consequently, each location will be affected by his tags being completely right. Thus, we need to compute the TF weight [20] for each tag follow by Equ. 5 and propose the $LocHITS_{TF}$ algorithm. The weight of one tag reveals relationship between tag and keyword, the values more high than the tag more closed to the keyword. We implement three algorithms ($LocHITS$, $LocHITS_S$, $LocHITS_{TF}$) with the dataset that is described in Tab. 2 on the next section.

4. Experimental results

This section introduces the dataset and shows some implementation steps to rank locations based on tags of them. From dataset which is collected from SNS in order to answer the question 'Where is the most famous place in the world for to do something (extracted from keyword)?'

4.1. Dataset

We collect data and perform basic processes to get the data in Tab. 2 as the basis dataset to experiment.

Table 2. The dataset

Keyword	#Collected photos	#Geotagged photos*	#Tags of (*)	#Locations
kimchi	15,230	1,416	23,395	10
pho	34,247	5,141	72,612	46
pizza	9,457	1,544	23,805	40
poutine	7,658	1,027	10,352	24
sushi	23,140	2,615	28,063	57

In Tab. 2, there are 5 keywords (*kimchi, pho, pizza, poutine and sushi*) of 5 traditional dishes which are from 5 different countries: *Sounth Korea, Vietnam, Italy, Canada and Japan*. With each keyword, we calculate the number of locations for ranking based on geotagged photos. We split the dataset for each keyword based on geotagged photos. For example, the value of Tab. 3 shows a list of locations (as definited in Def. 2) for finding out the best location by ranking. It means that from dataset (as shown in Tab. 3) we can find out the answer for the question Which is the best country for enjoy 'kimchi' ('pho', 'pizza', 'poutine' or 'sushi')? Similar to this issue, we can work with regions, city of a country as an option to implement (as shown in Tab. 4).

Table 3. The list of locations for ranking (by country) with 5 keywords

Keyword	# (location): List of locations
kimchi	(10): Canada, China, France, Japan, North Korea, South Korea, Taiwan, Thailand, United Kingdom, United States
pho	(46): Argentina, Australia, Austria, Belgium, Bhutan, Brazil, Cambodia, Canada, China, Colombia, Croatia, Cyprus, Czech Republic, Denmark, France, Germany, Hong Kong, Hungary, India, Indonesia, Italy, Japan, Laos, Mexico, Myanmar, Netherlands, New Zealand, Philippines, Poland, Reunion, Russia, Saint Pierre and Miquelon, Singapore, Slovakia, Slovenia, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, United Arab Emirates, United Kingdom, United States, Venezuela, Vietnam
pizza	(40): Algeria, Argentina, Australia, Austria, Belgium, Brazil, Cambodia, Canada, China, Cuba, Czech Republic, Denmark, Estonia, Finland, France, Germany, Guatemala, India, Ireland, Italy, Japan, Kenya, Malaysia, Myanmar, Netherlands, New Zealand, Peru, Philippines, Poland, Romania, Slovenia, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, United Kingdom, United States, Vatican City
poutine	(24): Australia, Belgium, Canada, China, Dominican Republic, Estonia, France, Germany, Guatemala, Hong Kong, Ireland, Japan, Macau, Mexico, Netherlands, Netherlands Antilles, Romania, Russia, Serbia, South Korea, Taiwan, Thailand, United Kingdom, United States
sushi	(57): Argentina, Australia, Austria, Barbados, Belgium, Brazil, Canada, Chile, China, Costa Rica, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Guatemala, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Japan, Latvia, Macau, Malaysia, Malta, Mexico, Moldova, Mongolia, Morocco, Netherlands, New Zealand, Northern Mariana Islands, Norway, Philippines, Poland, Portugal, Russia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Thailand, Trinidad and Tobago, Uganda, United Kingdom, United States, Venezuela, Vietnam

Table 4. The list of location for ranking (by regions in South Korea) with *'kimchi'*

# Location	The list of locations
27	Busan, Chungcheongbuk-do, Chungcheongnam-do, Daegu, Gangwon-do, Gwangju, Gyeonggi-do, Gyeongsangbuk-do, Gyeongsangnam-do, Incheon, Jeollanam-do, Seoul,Busan,Chungcheongbuk-Do, Chungcheongbuk-do, Chungcheongnam-Do, Daegu, Daejon, Gwangju, Gyeongsangbuk-Do, Gyeongsangnam-Do, Incheon, Jeollabuk-Do, Jeollanam-Do, Kangwon-Do, Kyeongki-Do, Seoul

4.2. Results on ranking locations

To evaluate the proposed algorithm, we have to implement with five keywords as described on the dataset. We used the $LocHITS$ algorithm for ranking locations based on tags. Giving a set of tags for ranking locations, these tags and locations are described as Def. 2 and Def. 3.

We conduct the experimentation with two methods for ranking which are called $online$ and $offline$. With offline ranking, we use the dataset in order to calculate all variables before using iterations of $LocHITS$. Using the algorithm 1 with $\varepsilon = 10^{-8}$, we found that the convergence of iterations is very rapidly. The results is showed in Fig. 4 with keyword 'kimchi' (with the number of iteration $k = 9$).

With online ranking, we use the dataset as the same collecting time. We add more tags at each iteration step as well using ε value, and the convergence value is obtained at $k = 395$ (with 'kimchi') as shown in Fig. 5.

For the purpose of comparing and solving the feasibility of this approach for ranking locations based on tags, we use tags frequency and determine tags similarity to expand the $LocHITS$ algorithm. We has tried with 'kimchi' and showed results as shown in Tab. 6. In order to make more clearly for comparing, the values in Tab. 6 are normalized at $[0 \ldots 1]$.

For comparing the results, we focus on Tab. 5 and Tab. 6. Based on the results of Tab. 6, we can comment that the $LocHITS$ got early convergence so that is the best algorithm. However, in this work we are considering using tags for ranking locations. Thus, we consider the analyzed data in Tab. 5 and conclude that it isn't an exact answer. Indeed, the ranking with $LocHITS_{TF}$ converge slower than $LocHITS$ and $LocHITS_S$ but the result is more suitable than two algorithms above.

The experimental results which are presented in Tab. 7 show that there are four values matching with the expecting results (*'kimchi'-South Korea; 'pho'-Vietnam; 'poutine'-Canada and 'sushi'-Japan*).

Additionally, we implemented the dataset and got five top positions that are constant. With these results, we believe that this approach is useful to rank locations based on geotagged resources from SNS.

5. Related work and discussion

Nowadays, SNSs (e.g, such as Facebook[1], Photobucket[2], Flickr[3] and so on) contains billions of images and videos which have been annotated and shared among friends. In fact, users annotate in a form of tags, ratings, preferences etc. These annotations are updated daily, become a huge data source that reflects events (society or individual) and contains many useful information to explore, discover and forecast what is going to happen in the future.

Studying of dependence of tag characteristics on spatial scale of aggregation is very important to discover process and mine data on SNS[7,16]. The authors [7] analysed tag frequency for a set of geotagged photos across multiple scales using Flickr data obtained

[1] www.facebook.com
[2] www.photobucket.com
[3] www.flickr.com

Table 5. The dataset for ranking locations with 'kimchi'

Location	#Photo	#Tag	Top-10 popular tags
Canada	26	244	kimchi, food, korean, toronto, spicy, banchan, market, scallions, finch, foodie
China	26	432	kimchi, china, beijing, duck, daejeon, travel, noodles, airport, friends, cold
France	10	337	food, kimchi, soup, noodles, pickled, rice, cuisine, ginger, balls, pepper
Japan	32	435	kimchi, japan, food, korean, tokyo, dinner restaurant, pork, hot, lunch, shrimp
North Korea	212	2208	kimchi, pyongyang, korea, dprk, juche, arirang, koryo, travel, cold, fun
South Korea	504	10983	kimchi, korea, food, korean, seoul, restaurant, culture, red, spicy, dinner
Taiwan	16	225	kimchi, taiwan, taipei, geotagged, food, ł, egg, tomato, soup, pot, watermelon, hot, tainan, shrimp
Thailand	13	93	kimchi, teachingsagittarian, kimchee, korean, thailand, bangkok, delicacies, chilli, food, vegetables
United Kingdom	82	607	kimchi, food, korean, kimchee, london, gimchi, season, extenstion, fermentation, restaurant
United States	495	7831	kimchi, korean, food, bulgogi, dinner, ssam, banchan, shrimp, pork, restaurant

Table 6. The results on locations ranking with 'kimchi' (offline)

$LocHITS$	$(k = 9)$	$LocHITS_S$	$(k = 11)$	$LocHITS_{TF}$	$(k = 30)$
SouthKorea	0.16494845	SouthKorea	0.16439629	SouthKorea	0.16573591
UnitedStates	0.15979381	UnitedStates	0.15937351	UnitedStates	0.16394372
Japan	0.10824742	Japan	0.10835741	Japan	0.11062289
UK	0.10309278	UK	0.10056532	UK	0.10655475
France	0.09793814	Taiwan	0.09865156	France	0.10188920
Taiwan	0.08247423	China	0.08690820	NorthKorea	0.07972374
Canada	0.07731959	Thailand	0.08109632	China	0.07706780
NorthKorea	0.07216495	France	0.07255086	Taiwan	0.06713576
China	0.06701031	Canada	0.06695163	Canada	0.06549773
Thailand	0.06701031	NorthKorea	0.06114889	Thailand	0.06182849

Table 7. Top-10 locations ranking with 5 keywords ($LocHITS_{TF}$-$online$)

Pos.	kimchi (KR)	pho(VN)	pizza(IT)	poutine(CA)	sushi(JP)
1	**SouthKorea**	**Vietnam**	USA	**Canada**	**Japan**
2	USA	USA	**Italy**	USA	USA
3	Japan	Canada	UK	SouthKorea	Canada
4	UK	Thailand	Canada	France	UK
5	France	Australia	Australia	UK	Australia
6	NorthKorea	UK	Germany	Ireland	China
7	China	SouthKorea	Spain	HongKong	Taiwan
8	Taiwan	Philippines	China	China	Spain
9	Canada	HongKong	France	Netherlands	Germany
10	Thailand	China	Netherlands	Taiwan	Singapore

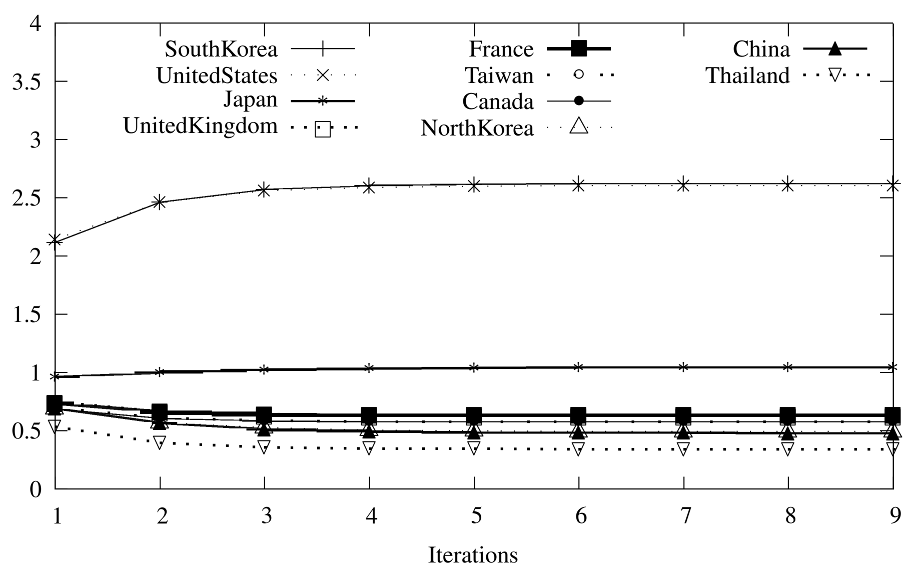


Fig. 4. Ranking location with 'kimchi' ($LocHITS$-$offline$)

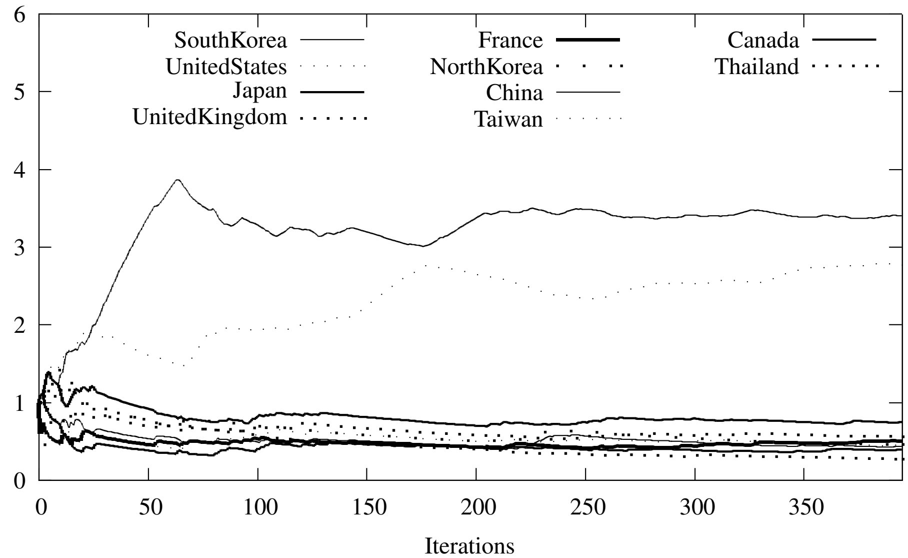


Fig. 5. Ranking location with 'kimchi' ($LocHITS$-$online$)

for the city of Vancouver, Canada. In the context of this data, they determined the tag space is dominated by a few frequent tags that describe large geographies, whereas more place-specific tags emerged at local scales. With this results, they determined the spatial distribution of geotagged photos is clustered in the main urban tourist and entertainment oriented core of the city.

Nowadays, searching information which are relevant to geographic locations is becoming increasingly important[5,15,14,16]. The authors [15] study the issue of ranking web pages using geographically-sensitive link analysis algorithms. They proposed several geographically-sensitive link analysis algorithms which exploit the geographic linkage between pages.

The HITS algorithm has known well as a famous method for ranking webpages [13,6]. Recently, the HITS algorithm is accelerated and development by many researchers with various fields [15,18,4,6]. In this study, we have employed this algorithm for ranking locations based on tags from SNS. We emphasize the role of occurrence of tag in many defining locations 2. Therefore, if one tag is contained in many locations, then it gets high value in ranking.

To discuss about this study, we want to mention several important issues that we have realized from this work. Firstly, tag appears in many locations having more valuable than tag only appears in some locations. Because, in this work, we want to rank locations by keyword searches, so if one tag appears in multiple locations, it demonstrates that tag is very close to the keyword search, so it would be better. Furthermore, if for each tag is considered a criterion for ranking location, then the location that achieves a lot of criterions would be more appreciated. Secondly, the calculation of the value of tag for ranking is necessary. Obviously if you do not count the value for each tag, it mean, all node (for HITS) are of equal value and equal 1. The case happened with the same original HITS algorithm. This is not fair to the tag that appears many times. For example, the keyword searches ‘pho’, we find two tag, tag ’noodle’ appears in 3 location (e.g., A, B, C), with the number of occurence is 3, 2, 20, tag ‘food’ also appears in 3 locations as above with the number of occurence is 4, 3, 2. If using these tags to rank for 3 locations above, the ranking values should be based primarily on the tag ‘noodle’, that is more appropriate (it means C is the first position). However, if not considering the frequency of two tags, both A, B and C have the same value when ranking - this is not true.

6. Concluding remarks and future work

For the purpose of ranking locations based on geotagged resources, we propose using $LocHITS$ algorithm and modifying value of nodes with tags and relevant locations. Besides, we extend $LocHITS$ by using similarity between the set of common tags and a set of tags of each location, called $LocHITS_S$ algorithm. Moreover, we use term frequency of tags in each location in order to compute the weight of tags and apply them into the $LocHITS_{TF}$ algorithm.

In our experiments, more importantly, we empirically showed that the GeoHITS algorithm (offline case) converge quickly. The obtained results with three algorithms are quite interesting and suitable. Although we could not determine the precision of these results, but based on the reality with keywords belong to countries, our results have obtained the high ranking values.

In spite of the imbalance of dataset with many locations for ranking (a large part of dataset belongs to United States due to users of Flickr), our method found out locations which hold traditional dishes for each keyword as shown in collected dataset. We appreciate using term frequency of tags in each location (called $LocHITS_{TF}$ algorithm).

As future work, we plan $i)$ to propose a location recommendation system for traveler based on tags from SNS; $ii)$ to detect events based on geotagged photos from SNS as a our new approach.

Acknowledgments. This research was supported by the Chung-Ang University Research Grants in 2014. Also, this work was supported by the ICT R&D program of MSIP/IITP [10035348, Development of a Cognitive Planning and Learning Model for Mobile Platforms]. Moreover, this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2014R1A2A2A05007154).

References

1. Atzori, L., Iera, A., Morabito, G., Nitti, M.: The social internet of things (siot)– when social networks meet the internet of things: Concept, architecture and network characterization. Computer Networks 56(16), 3594–3608 (2012)
2. Bao, J., Zheng, Y., Mokbel, M.F.: Location-based and preference-aware recommendation using sparse geo-social networking data. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. pp. 199–208. ACM (2012)
3. Begelman, G., Keller, P., Smadja, F., et al.: Automated tag clustering: Improving search and exploration in the tag space. In: Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland. pp. 15–33. WWW2006 (2006)

4. Benzi, M., Estrada, E., Klymko, C.: Ranking hubs and authorities using matrix functions. Linear Algebra and its Applications 438(5), 2447–2474 (2013)
5. Clements, et al.: Using flickr geotags to predict user travel behaviour. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 851–852. ACM (2010)
6. Ding, C., He, X., Husbands, P., Zha, H., Simon, H.D.: Pagerank, hits and a unified framework for link analysis. In: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 353–354. ACM (2002)
7. Pham, H.X., Jung, J.J.: Recommendation System Based on Multilingual Entity Matching on Linked Open Data. Journal of Intelligent & Fuzzy Systems 27(2), 589–599 (2014)
8. Giannakidou, E., Koutsonikola, V., Vakali, A., Kompatsiaris, I.: Co-clustering tags and social data sources. In: Proceedings of The 9th International Conference on Web-Age Information Management (WAIM'08), pp. 317–324 (2008)
9. Jung, J.J.: Exploiting semantic annotation to supporting user browsing on the web. Knowledge-Based Systems 20(4), 373 – 381 (2007)
10. Jung, J.J.: Knowledge distribution via shared context between blog-based knowledge management systems: A case study of collaborative tagging. Expert Systems with Applications 36(7), 10627 – 10633 (2009)
11. Jung, J.J.: Ubiquitous conference management system for mobile recommendation services based on mobilizing social networks: A case study of u-conference. Expert Systems with Applications 38(10), 12786 – 12790 (2011)
12. Jung, J.J.: Cross-lingual query expansion in multilingual folksonomies: A case study on flickr. Knowledge-Based Systems 42, 60–67 (2013)
13. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM) 46(5), 604–632 (1999)
14. Kurashima, T., Iwata, T., Irie, G., Fujimura, K.: Travel route recommendation using geotags in photo sharing sites. In: Proceedings of the 19th ACM international conference on Information and knowledge management. pp. 579–588. ACM (2010)
15. Lee, H.C., Liu, H., Miller, R.J.: Geographically-sensitive link analysis. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence. pp. 628–634. IEEE Computer Society (2007)
16. Lee, I., Cai, G., Lee, K.: Exploration of geo-tagged photos through data mining approaches. Expert Systems with Applications 41(2), 397–405 (2014)
17. Nguyen, T.T., Hwang, D., Jung, J.J.: Social tagging analytics for processing unlabeled resources: A case study on non-geotagged photos. In: Proceedings of the 8th International Symposium on Intelligent Distributed Computing (IDC 2014), Madrid, Spain, September 3-5, 2014. pp. 357–367 (2014)
18. Pham, X.H., Nguyen, T.T., Jung, J.J., Hwang, D.: Extending HITS algorithm for ranking locations by using geotagged resources. In: Proceedings of the 6th International Conference on Computational Collective Intelligence (ICCCI 2014), Seoul, Korea, September 24-26, 2014, pp. 332–341 (2014)
19. Rokach, L., Maimon, O.: Clustering methods. In: Data mining and knowledge discovery handbook, pp. 321–352. Springer (2005)
20. Sebastiani, F.: Machine learning in automated text categorization. ACM computing surveys (CSUR) 34(1), 1–47 (2002)
21. Zhang, H., Korayem, M., You, E., Crandall, D.J.: Beyond co-occurrence: discovering and visualizing tag relationships from geo-spatial and temporal similarities. In: Proceedings of the fifth ACM international conference on Web search and data mining. pp. 33–42. ACM (2012)
22. Zhang, W., Yoshida, T., Tang, X.: Tfidf, lsi and multi-word in information retrieval and text categorization. In: Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on. pp. 108–113. IEEE (2008)

Tuong Tri Nguyen is currently a Ph.D. student at the Graduate School of Computer Engineering, Yeungnam University, Korea. His main research interest include social network services, big data analytics, data mining, and machine learning.

Jason J. Jung is a corresponding author of this paper. He is an associate professor in Chung-Ang University, Korea, since September 2014. He was a postdoctoral researcher in INRIA Rhône-Alpes, France in 2006, and a visiting scientist in Fraunhofer Institute (FIRST) in Berlin, Germany in 2004. He received the B.Eng. in Computer Science and Mechanical Engineering from Inha University in 1999. He received M.S. and Ph.D. degrees in Computer and Information Engineering from Inha University in 2002 and 2005, respectively. His research topics are knowledge engineering on social networks by using machine learning, semantic Web mining, and ambient intelligence.

Received: October 15, 2014; Accepted: December 18, 2014.