Location Familiarity Based Flickr Photographer Classification for POI Mining

Chenyi Zhuang, Qiang Ma, Masatoshi Yoshikawa
Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan
zhuang@db.soc.i.kyoto-u.ac.jp, {qiang, yoshikawa}@i.kyoto-u.ac.jp

ABSTRACT

In this paper, we propose and compare three ways of modeling photographers' location familiarity: a social network driven model, a time driven model and a location driven model. Then, the integration of the three models is further discussed. Experimental evaluations and analysis on a real data set consisting of 14,112 images collected from three cities well demonstrate the performance of the proposed classification methods. Many applications could benefit from information about the location familiarity, such as personalized geo-social recommendation, epidemic dispersion, urban computing, and so on.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models; H.2.8 [**Database Applications**]: Data mining

Keywords

User profiling, location familiarity, geo-tagged image, probabilistic model, social network.

1. INTRODUCTION

A number of algorithms have been proposed to detect users' different attributes. For instances, the location inference related works have been done from a wide variety of social media, such as Tweets [4], geo-referenced pages on Wikipedia [8], image tags on Flickr [10], and social network structures on Facebook [2]. Besides the location information, a variety of other user attributes have also been investigated by using NLP related technologies, including gender, age, and even political orientation [9].

However, to the best of our knowledge, how to estimate a user's location familiarity by analyzing geo-tagged images has not yet been well studied. In [15], we are trying to develop a sightseeing recommendation system by dividing sightseeing spots into four quadrants on the basis of their "popularity" and "sightseeing quality". We have verified that

the location familiarity is essential for discovering the obscure spots. Located in the quadrant with high sightseeing quality but low popularity, an obscure sightseeing spot is a considerable choice for in-depth travel to not only enjoy the beautiful scenery but also experience local culture.

By analyzing the image taken time information, Fischer [1] presents the mobility behaviors of Flickr photographers to a city scale. Compared with tourists, locals always have very different mobility behaviors when taking photographs. By further analysis, we find that the reason for this phenomenon occurred is the information asymmetry between different groups. Intuitively, the familiar ones (e.g., locals) are always having more choices for visit. On the basis of this observation, we assume that the obscure spots can be discovered by comparing the visiting frequencies of different photographer groups. The key problem is how to identify these groups.

In this paper, we focus on the photographer classification task. By analysis of geo-tagged images from Flickr, we propose three models to estimate photographers' location familiarity, based on which we can detect the spots that are obscure to travelers who are unfamiliar with the city. First, we introduce a social network driven model, where a graph based method is used. Second, we introduce a time driven model, which estimates someone's location familiarity based on their visiting frequency. Third, considering the fact that different photographers have different mobility behaviours, a location driven model is devised. In addition to our goal, many other applications could benefit from information about the location familiarity, such as personalized geo-social recommendation [13], epidemic dispersion [7], urban computing [14], and so on.

2. MODELS

Our problem is defined as: given a Flickr photographer, his/her location familiarity will be estimated. Three totally different models will be introduced in this section.

2.1 Social Network Driven Model

If a photographer is a resident, we assume s/he is familiar with the target city. However, few users specify their residency information on Flickr. Due to this, we propose a social network based method to calculate the familiarity of a photographer with a given city c. We assume that a photographer with many friends who are familiar with city c will also be familiar with the city c.

$$F_s(c) = \alpha_B \cdot N \cdot F_s(c) + (1 - \alpha_B)d_c. \qquad (1)$$

Represented by a transition matrix, N is a directed graph $N(V, E)$, where V is the photographer set of v_i and E is the friendship set of $e(v_i, v_j)$ from v_i to v_j. For the normalized bias distribution d_c, compared with the TruskRank [6], our system can automatically select good seeds due to whether a v_i is a resident or not as detected from his/her profile. While the others are assigned to 0, the entries of the vector d_c that correspond to good seeds sum up to 1. α_B is the decay factor. Given a city c, $F_s(c)$ is initialized to d_c before the iteration of the above computation. The computation would be finished when either it reached the maximum number of iterations (i.e., M_B) or $F_s(c)$ was converged.

2.2 Time Driven Model

In the time driven model, we devise a method by considering photographers' visiting frequencies. Intuitively, if a photographer has visited the target city c frequently and recently, s/he may be familiar with c. Considering the time sensitivity of location familiarity, we calculate the score of each location before a year λ.

$$F_t(v_i, c, \lambda) = diag(r_i^T \cdot (w_i \cdot M_i))[c] \qquad (2)$$

A matrix M_i is a mapping between a photographer v_i and all the cities C_i that s/he has been to. For example, $M_i[y = 2014, c = kyoto] = 10, c \in C_i$ means that among v_i's images taken in Kyoto in 2014, there are 10 different dates detected from the images' taken dates. Before λ, the reproducibility of cities is denoted as a matrix r_i. Assuming that we collect 10 years' worth of image data of v_i, for each year x, $r_i[x, c] = \frac{5}{10}$ if we detect that v_i went to c in five of the years; We also introduce a diagonal matrix, $\omega_i(x, y) \, \forall x, y \in \{1, 2, ..., n\}$, to characterize the staleness of information. n equals the number of detected years before λ.

Obviously, the accuracy of F_t highly depends on the sparsity of the familiarity matrix M_i. For a photographer who has uploaded few images, this method would not be effective.

2.3 Location Driven Model

The geographical information about where a user has been to in a city c is also essential for the estimation of their location familiarity. By introducing Bayes' theorem, we model user mobility behavior based on the following observations.

For each user v_i, when given a city c, there exists an inherent property which describes his/her familiarity with the target city. For instance, there are three different groups labeled by red, blue, and yellow in [1]. Accordingly, in our location driven model, we introduce a latent variable *group*, denoted as g, to represent this property. Furthermore, for the users from different groups, they prefer visiting different regions of a city. Namely, some regions appealed to more users from the *red group*, while some others attracted many users from the *blue group*.

Table 1 shows the notations used in this model. To generate such a map illustrated in [1], Algorithm 1 presents the generative process. The whole generative process above can be explained by the following joint distribution.

$$p(t, g \mid \vec{\alpha}, \vec{\beta}, \Lambda) = p(l_t \mid r, \Lambda)p(v_t \mid g, \vec{\alpha})p(g \mid \vec{\beta}). \qquad (3)$$

Given the priors $\vec{\alpha}, \vec{\beta}$, and Λ, by introducing a latent variable *group* g, the model is aiming to generate each tuple $t = \{v_t, l_t\}$ (i.e., t of an image in our dataset) in the set T. Λ denote all the priors of the Gaussian Mixture Model (GMM). More specifically, there are a Dirichlet distribution

Table 1: Notations used in the model

	Size	Description
V	$[1, V]$	The user set, v is a particular user in V
R	$[1, R]$	The region set, r is a region in R
G	$[1, G]$	The *group* set, g is a group in G
T	$[1, T]$	$t \in T$ is a user-coordinate tuple: $\{v_t, l_t\}$
μ_r	$\mathbb{R}^2$	Mean location of a latent region r
Σ_r	$\mathbb{R}^{2 \times 2}$	Covariance matrix of a latent region r
$\vec{\vartheta_r}$	$1 \times \|G\|$	A region-dependent group distribution
$\vec{\varphi_g}$	$1 \times \|V\|$	A group-dependent user distribution
$\vec{\alpha}, \vec{\beta}$		Hyperparameters of Dirichlet priors
Λ		Denote all the priors of the GMM

Algorithm 1 A probabilistic generative process

1: $R \sim \sum_{r=1}^{R} \pi_r \mathcal{N}(x \mid \mu_r, \Sigma_r)$ ▷ initialize R using GMM
2: **for** all groups $g \in [1, G]$ **do**   ▷ Observation 3
3:     sample mixture components $\vec{\varphi_g} \sim Dir(\vec{\alpha})$
4: **end for**
5: **for** all regions $r \in [1, R]$ **do**   ▷ Observation 2
6:     sample mixture proportion $\vec{\vartheta_r} \sim Dir(\vec{\beta})$
7:     **for** all tuples $t \in [1, N_r]$ in region r **do**
8:         sample a coordinate $l_t \sim \mathcal{N}(\mu_r, \Sigma_r)$
9:         sample a group $g_{rt} \sim Mult(\vec{\vartheta_r})$
10:         sample a user $v_{rt} \sim Mult(\vec{\varphi_{g_{rt}}})$
11:     **end for**
12: **end for**

---

over the mixing coefficients π_r, and a Gaussian-Wishart distribution governing the mean μ_r and precision Σ_r^{-1} of each Gaussian component.

In accordance with our goal of user classification, the target of model inference is the distribution $p(g \mid t)$, where the hyperparameters are omitted. By seeing the Eq. 3, when a region r is given in the initialization stage, g is independent ($\perp\!\!\!\perp$) of l_t. Intuitively, for all tuples t in a specific region r, we can regard them as *bag of instances of different users* by ignoring the location information. Note that, in spite of $g \perp\!\!\!\perp l_t \mid v_t, r$, our model is flexible enough to describe the situation that different instances v_t of the same user v could be assigned to different g. Accordingly, we divide the inference procedure into two steps as follows.

Step 1. $l_t - generation$: Since a city c is described by a GMM, and each region r is a component of the GMM, we utilize a variational Bayesian machinery for the inference of GMM. As mentioned above, because it is not necessary to identify the exact location information, given r a random coordinate is assigned to each l_t in the r.

Therefore, we simply utilized the implementation introduced in the section 10.2 of [3]. It has the advantage that the number of mixture components (i.e., $|R|$) can be automatically identified with a relatively large initial value.

Step 2. $v_t - generation$: After eliminating the l_t, the joint distribution in Eq. 3 has been simplified as follows:

$$p(t, g \mid \vec{\alpha}, \vec{\beta}) = p(v_t \mid g, \vec{\alpha})p(g \mid \vec{\beta}). \qquad (4)$$

There have been three strategies, EM with variational inference, EM with expectation propagation, and Gibbs sampling, which can be applied in our inferential problem. We use the Gibbs sampling method, whose performance is comparable with the other two but is tolerant better to local optimization.

Table 2: Datasets of the three cities

	Beijing	Kyoto	SF
Geo-tagged images	4,002	6,121	3,989
Target photographers	230	300	536
% of images from top 20%	85.7%	84.4%	78.6%

Table 3: Context data used in Eq. 1 and 2

Eq. 1:	A friendship based directed social network.
	Nodes: 837,633; Edges: 3,082,408.
Eq. 2:	1066 familiarity matrices.
	Year range: $2005 \sim 2015$; Cities: 17,460.

2.4 Model Integration

By model integration, one important goal is to identify the latent groups generated in the location driven model. Because only photographer ids and geo-coordinates are used, the system cannot automatically know the meanings (i.e., close to the familiar or unfamiliar groups) of the groups when using this model alone. Another goal is to make the classification much more stable and robust, by considering the different available information for each photographer.

In this integration method, we simply use the scores obtained from the three models as features to train a classification model. For instance, if the number of groups $|G| = 2$ in the location driven model, there would be four features for the SVM model training. In the process of implementation, we utilize SVM with the RBF kernel. Then, a grid search method [11] is applied on a 3-fold cross validation to find the best RBF kernel's parameter γ and the penalty parameter C in SVMs. At last, by comparing the average classification accuracy, the most powerful model is chosen.

To make a summary, since the three original models are established using completely different information, predictably the integration method would obtain a better performance. However, the only drawback is that a manually labeled training data set is needed. Although how to establish a comprehensive training data set could be further discussed, currently we do a random sampling and manually label them.

3. EXPERIMENT

We first show an outline of the retrieved data sets and parameter selections in our experiments. Then, the performance of all the three photographer classification methods (i.e., Eq. 1, 2, 3) will be presented.

3.1 Data Preparation

Table 2 summarizes three target datasets of Beijing, Kyoto, and San Francisco, respectively. For all the datasets, the approximate top 20% photographers are vital, while the photographers at the long tail remain very inactive. In Table 3, it shows the context data that would be used in the social network driven model and the time driven model. For instance, we totally collected 837,633 photographers to construct the social network N, in which all the 230+300+536 target photographers are included. Also, we retrieved all the images uploaded by the target photographers to create the familiarity matrices M. As a result, these photographers have visited about 17,460 cities during the last ten years.

3.2 Parameter Selection

Table 4: Labelling results of the 300 photographers

Cities	familiar No.	unfamiliar No.
Beijing	≈ 36	≈ 64
Kyoto	≈ 31	≈ 69
San Francisco	≈ 76	≈ 24

Table 5: Accuracy on integration method: SVM

(γ, C)	All models	Time driven model excluded
$(2^{-7}, 2^3)$	0.877	0.798
$(2^1, 2^{-1})$	0.880	0.820

Social Network Driven Model: We applied our biased PageRank by setting the decay factor $\alpha_B = 0.85$, and the number of iterations $M_B = 1000$.

Time Driven Model: Because the w_i in Eq. 2 are used to character the staleness of information, we set them to the practical exponential decay: $0.8^{(\lambda - x)}$.

Location Driven Model: We set the number of groups $|G| = 2$. Based on the experimental results below, these two latent groups well demonstrate our intentions. In accordance with the conclusion in [5], as the relationship between hyperparameters and group number is a mutual one, we set the symmetric Dirichlet priors as $\alpha = 0.01$ and $\beta = 50/|G|$.

3.3 Photographer Classification Evaluation

ROC based Evaluation: The receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier model as its discrimination threshold is varied. In our experiments, we use the two criteria, "sensitivity (true positive rate)" and "1-specificity (false positive rate)" shown in Figure 1, to do the evaluation. By using these two criteria, the area under the ROC curve (AUC) is recognized as the measurement of a test's discriminatory power. Intuitively, the maximum value for AUC is 1.0 indicating 100% sensitivity and 100% specificity, while an AUC value of 0.5 means no discriminative value.

Ground Truth: We invited 9 subjects to manually align photographers' familiarity with a specific city using all their online information. For each city (i.e., Beijing, Kyoto, and San Francisco), 3 different subjects are assigned to label 100 photographers that have been picked out. Table 4 shows the labeling results, from which about 36, 31 and 76 familiar ones are returned by majority voting.

To make sure that the labeling is credible, all the subjects are familiar with the corresponding cities and recorded reasons for each of their labels. Furthermore, when selecting the target 300 photographers, we chose those who have uploaded a lot of images on Flickr to make subjects' tasks much easier. Even so, there are some differences between labeling results from different subjects. The Jaccard index similarities between any two of the subjects vary between 0.75 and 0.87.

Experiment Results: Figure 1(a) shows the performance of the social network driven model. By proposing a similar assumption, the state-of-art location prediction method in [12] obtained an accuracy of 68.2% for predictions on the city-scale. Considerably, our model achieved a better improvement with $AUC = 0.81$ on average. Moreover, we analyze the situations of different cities to gain a deep insight into this model. By investigating the lower AUC for Kyoto, we find that, since Kyoto is a very traditional city, few online

(a) Social network driven model   (b) Time driven model   (c) Location driven model
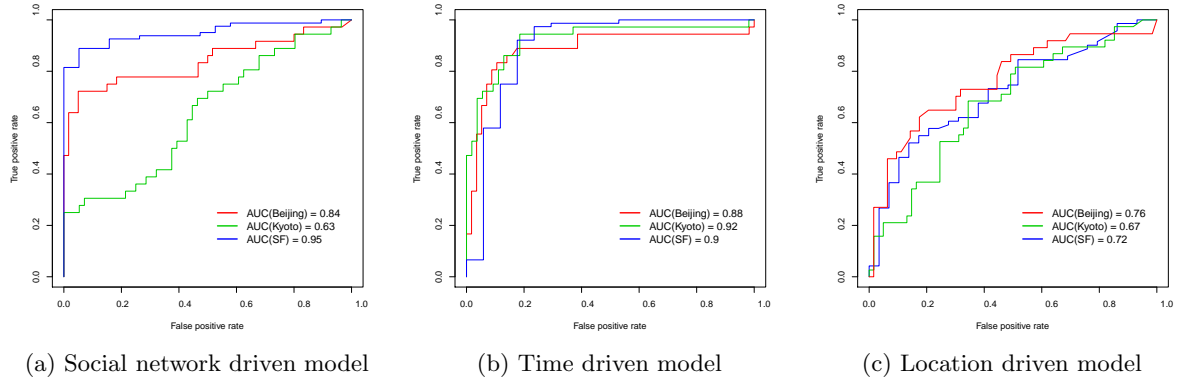
Figure 1: Performance of our models evaluated by ROC and AUC.

social relationships are established compared with an open society, like San Francisco. In some senses, the performance of this model depends on different kinds of social culture.

Figure 1(b) shows the performance of the time driven model. Because all the 300 target photographers uploaded a lot of images, this model obtained very good results that close to the ground truth. However, for the photographers who have shared few images, obtaining the very sparse matrices M_i, apparently its performance would decrease sharply.

Figure 1(c) presents the location driven model's performance. Although this model didn't achieve an extremely good result as the time driven model did, it has proved that the location information is also an important feature for getting the stable and robust results. Less influenced by the number of images, this method tends to be good at classifying the photographers who have strong preferences.

At last, Table 5 shows the accuracy of our integration method. Since the time driven model would always obtain a good result, we also present the accuracy when only the scores from the other two models are used. By seeing the average accuracy of the 3-fold cross validations, this method performed as well as expected.

4. CONCLUSION

In this paper, to solve the problem of discovering obscure sightseeing locations, we propose three methods to estimate photographers' location familiarity, including a social network driven model, a time driven model, and a novel location driven model that establishes a relationship between people and space. In the experiments, we employ human effort to obtain ground truth and compare the performance of each proposed method. By analyzing the results, we also discuss the utility of different kinds of photographers' available information for the classification task. Our work is intended to inspire more interest in information asymmetry analysis and diverse sightseeing resources discovering.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] https://www.flickr.com/photos/walkingsf/sets/72157624209158632/.

[2] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *WWW*, pages 61–70, 2010.

[3] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.

[4] H.-w. Chang, D. Lee, M. Eltaher, and J. Lee. @ phillies tweeting from philly? predicting twitter user locations with spatial word usage. In *ASONAM*, pages 111–118, 2012.

[5] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235, 2004.

[6] Z. Gyöngyi, H. Garcia-Molina, and J. O. Pedersen. Combating web spam with trustrank. In *VLDB*, pages 576–587, 2004.

[7] V. Lampos, T. De Bie, and N. Cristianini. Flu detector-tracking epidemics on twitter. In *ECML PKDD*, pages 599–602, 2010.

[8] M. D. Lieberman and J. Lin. You are where you edit: Locating wikipedia contributors through edit histories. In *ICWSM*, pages 106–113, 2009.

[9] M. Pennacchiotti and A.-M. Popescu. Democrats, republicans and starbucks aficionados: user classification in twitter. In *SIGKDD*, pages 430–438, 2011.

[10] A. Popescu, G. Grefenstette, et al. Mining user home location and gender from flickr tags. In *ICWSM*, pages 307–310, 2010.

[11] C. wei Hsu, C. chung Chang, and C. jen Lin. A practical guide to support vector classification, 2010.

[12] D. Xu, P. Cui, W. Zhu, and S. Yang. Find you from your friends: Graph-based residence location prediction for users in social media. In *ICME*, pages 1–6, 2014.

[13] J. Zhang and C. Chow. igslr: personalized geo-social location recommendation: a kernel density estimation approach. In *SIGSPATIAL*, pages 334–343, 2013.

[14] Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: Concepts, methodologies, and applications. *ACM TIST*, 5(3):38:1–38:55, 2014.

[15] C. Zhuang, Q. Ma, X. Liang, and M. Yoshikawa. Anaba: An obscure sightseeing spots discovering system. In *ICME*, pages 1–6, 2014.