Where your photo is taken: Geolocation Prediction for Social Images

Bo Liu[†] Quan Yuan[‡] Gao Cong[‡] Dong Xu[‡]
[†] Facebook, U.S.A.
[‡] Nanyang Technological University, Singapore
bol@fb.com, {qyuan1@e.,gaocong@,dongxu@}ntu.edu.sg

April 24, 2014

Abstract

Social image sharing websites, such as Flickr and Zooomr, have attracted a large number of Internet users. These systems allow users to associate geolocation information to their images, which is essential for many interesting applications such as location-aware image search. However, only a small fraction of social images have geolocation information. Thus, an automated tool for suggesting geolocation is essential to help users geotag their images.

In this paper, we investigate how to assist users in geotagging social images, and how to boost the accuracy of geotagging. We use a large dataset consisting of $221M$ Flickr images uploaded by $2.2M$ users. We analyze for the first time user uploading patterns, user geotagging behaviors, and the relationship between the taken-time gap[1] and the geographical distance between two images from the same user. Our analysis shows that the taken-time gaps between the image to be geotagged and historical images are very important for geotagging. Based on the finding, we represent a user profile by historical tags for the user, and build a multinomial model on the user profile for geotagging. We further propose a unified framework to suggest geolocations for images, which combines the information from both image tags and the user profile. Experimental results on the Flickr dataset show that for images uploaded by users who have never done geotagging, our method outperforms the state-of-the-art method by 10.6% to 34.2%, depending on the granularity of the prediction. For images from users who have done geotagging, a simple method is able to achieve very high accuracy.

1 Introduction

Social image sharing services such as Flickr and Zooomr have accumulated a huge number of photos contributed by many users. Photos are taken at specific places and thus are inherently spatial (Crandall, Backstrom, Huttenlocher, & Kleinberg, 2009). Many applications can benefit from the geolocation[2] information of photos. To name just a few, consider 1) With geolocation information, we can organize images by location in a database and enable location-aware queries. For example, users can retrieve photos within a geographical region or photos that are close to the query location (Toyama, Logan, & Roseway, 2003). In addition to the geolocation, a query can also contain a tag component (Rorissa, 2010; Cong, Jensen, & Wu, 2009), for example,

[1]That is, the amount of time between when two photos are taken.
[2]Geolocation refers to geographic coordinates in this paper.

to retrieve photos close to a location and containing the tag "beach". 2) With geolocation information, photos can be browsed using a map-based interface, as currently done in many social image sharing services. For example, photos can be pin-pointed on a map to identify very small regions (Chen, Battestini, Gelfand, & Setlur, 2009). 3) With geolocation information, images can be associated with points of interest in the rapidly developing Location Based Social Network Systems (Ye, Shou, Lee, Yin, & Janowicz, 2011; Ye, Yin, Lee, & Lee, 2011).

The geolocations of photos in current image sharing services come from two sources: 1) With GPS-enabled cameras, geolocations can be automatically associated with images; 2) Users can also manually geotag photos by dragging a photo to a point on a world map interface when uploading photos to a image sharing service. This method, however, is tedious.

Although geolocation information is indispensable to many applications, a large portion of photos uploaded to social image sharing services contain no geolocation information. Our analysis of a large collection of 221M Flickr photos shows that only 7.8% of the photos in the collection are geotagged with *latitude* and *longitude*.

To make photos without geolocation available to location-aware applications, it is natural to consider automatically geotagging the photos. The feasibility analysis of the geotagging task is as follows: 1) Previous work shows that about 29% of Flickr tags[3] are location specific or location relevant (Sigurbjörnsson & Zwol, 2008); furthermore, location tags form the largest part among all tag categories. This indicates that image tags provide good indicators for suggesting geolocations for photos. 2) Although geotagged photos only account for a small portion of the available photos in social image sharing sites, there is still a large number of geotagged photos that have been accumulated which can be utilized to geotag other photos.

Several recent studies (Crandall et al., 2009; Serdyukov, Murdock, & Zwol, 2009) have addressed the problem of geotagging images. In existing works, the world map is divided into grids at different granularities based on *latitude* and *longitude* coordinates, and the tag information of a photo without geolocation is used to predict its cell location. Unfortunately, developing an efficient and effective method that is able to place photos in the correct cells is a challenging problem. As pointed out by (Serdyukov et al., 2009), the huge number of candidate locations makes the use of computationally expensive machine learning approaches unsuitable — when the world map is divided at a $1km$ granularity, we end up with more than $600M$ grid cells. Because of low accuracy and high computational cost, it is also impractical to solve the problem by extracting and utilizing image features (Hays & Efros, 2008). According to the experimental results in (Serdyukov et al., 2009), the accuracy of the state-of-the-art method is low, even though images without tags have been excluded in their experiments.

Moreover, we observe that only 44.9% of Flickr images in our collection have tags, which greatly limits the applicability and accuracy of the existing methods that utilize tags alone for geotagging. Hence, there is a large space for improving on the accuracy of current geotagging approaches.

In this paper, we explore a new approach to the geotagging problem. In addition to the relationship between geolocations and image tags, we propose to utilize the relationship between geolocations and users for geotagging. The intuition is that a user should have preference for some geolocations, which can be described by a user profile extracted from historical images by the user and their tags.

To utilize a user's geolocation preferences as embedded in their profile, we perform an analysis on a large Flickr dataset. We propose to represent a user profile by the set of historical tags the user has used, and then utilize the user profile as supplementary information to image tags

[3]We use the term "tag" to refer to textual tags, and "geotag" for *latitude* and *longitude* information.

for geotagging. We model each user profile and each cell, respectively, by a multinomial distribution over the vocabulary of tags, and then employ KL-divergence to estimate the similarity between the two distributions, which depicts how likely a user is to visit a location. Finally, we combine the information from image tags and the user profile to suggest the geolocation of each image.

The contributions of this paper are fourfold.

1. Based on a large dataset from Flickr containing $221,801,183$ images uploaded by $2,252,758$ users, we analyze for the first time user uploading patterns, user geotagging behaviors, and the relationship between the taken-time gap of two images from the same user and the distance between the images. Based on the analysis results, we find that the taken-time gap between a query image and a historical image is very important for geotagging. We employ this finding in our geotagging algorithm.

2. We represent a user profile using historical tags from the user, and then build a multinomial model based on the user profile information for geotagging.

3. We propose a unified framework to suggest geolocations for images, which combines the information from image tags and the user profile. To the best of our knowledge, no previous work on geotagging social images exploits user profile information.

4. The proposed methods are evaluated on a large Flickr dataset. Experimental results show that, for users who have never done geotagging, our method is able to improve on the accuracy of a state-of-the-art method by 10.6%, 16.8% and 34.2% for $1km$, $10km$ and $100km$ grids, respectively. For users who have done geotagging before, a simple method is able to achieve very high accuracy.

Our proposed method is simple yet effective. It is also practical, as we can efficiently build our multinomial model on a large number of images. Moreover, since the model is based on Naive Bayes and language models, it can be easily updated to accommodate newly geotagged photos generated by users.

Note that the proposed method can be used in an interactive way: when a user uploads an image, we use the map interface to pin-point the cell suggested by the method as a starting point. The user can then choose the right place for the image, which is typically at or near the suggested location.

The remainder of this paper is organized as follows. Related research is reviewed in the next section. In Section 3, we detail the dataset used in this work and the data analysis results. We describe the proposed geolocation suggestion algorithm in Section 4. In Section 5, we present the experimental setup, the evaluation metrics and the experimental results. Finally, we conclude this paper and discuss future directions in Section 6.

2 Related Work

2.1 Image Geolocation Suggestion

Suggesting geographic locations of user generated images has been attracting increasing research interests recently. The studies by Hays et al (Hays & Efros, 2008) and Serdyukov et al (Serdyukov et al., 2009) are the most related to our work.

Hays et al. (Hays & Efros, 2008) infer the geographic location of an image by its nearest neighbors defined by visual features. In their work, they use a special sub-set of Flickr images tagged with at least one name of a country, territory, continent, densely populated city, US states or popular tourist site, and not tagged with some specific tags such as "birthday" or "cameraphone". Their method is able to find the correct locations for 16% of the testing images within $200 * 200km$ area on the restricted dataset.

Instead of using visual features, Serdyukov et al. (Serdyukov et al., 2009) infer geographic locations for Flickr images by user generated textual tags. The authors place a grid over the whole world map, and estimate a language model from tags of images for each grid cell. The estimated language models are then employed to predict which grid cell a testing image resides in. We use this method as the baseline method in our experiments.

GeoFolk (Sizov, 2010) is a framework based on Bayesian latent topic model, which characterizes social images by combining text features with spatial knowledge. Due to the limited space, we do not detail GeoFolk in this paper and interested readers are referred to (Sizov, 2010). Among other applications, GeoFolk can be used for image geotagging by associating image tags with geolocations. However, as we will see in Section 5, the accuracy of GeoFolk for geotagging is low. Furthermore, it does not work for $1km$ granularity due to the high running time complexity.

The recent work (Ostermann, Tomko, & Purves, 2013) presents an evaluation of automatically generated concept keywords and place names for geo-referenced images.

Our work is also related to the work on landmark identification. Crandall et al (Crandall et al., 2009) build a system to place images on a map using a combination of textual and visual feature (the SIFT visual words). Instead of considering the entire world map, they limit their task to deciding which of ten landmarks in a given city is the subject of an image. Their method is a classifier-based approach. Specifically, for each of the ten landmarks of a city, a binary classifier is built, which takes the images taken at that landmark as the positive examples, and other images as the negative examples. It is not clear how to scale this classifier based method to place an image onto a point of the entire world map. There exist other proposals (Y. Li, Crandall, & Huttenlocher, 2009; Zheng et al., 2009; Chen et al., 2009) for landmark classification or recognition. Buscaldi et al. (Buscaldi & Rosso, 2008) employ Geo-WordNet[4] to extract geographical words from textual tags. However their method does not fit the fine-grained case discussed in this work. Our work is related to tag prediction (Hsu & Chen, 2011), which however does not predict the geolocation as we do.

In this work, we propose to find the location of an image by combining the evidence from its tags and the evidence from its user's historical images, whose tags are relevant to the users' activities (Stvilia & Jörgensen, 2010). To the best of our knowledge, none of existing work considers to enhance the accuracy of geotagging by using user profile built from the user's historical images. We use language models to represent user profiles. The user profiling in our work aims to profile the historical tags used by individuals, and the purpose is different from the previous work on profiling users' expertise (e.g., (Liu, Wang, Johri, Zhou, & Fan, 2012)). As we will see in this paper, the proposed methods are able to greatly improve on the prediction accuracy over the method (Serdyukov et al., 2009) which uses image tags alone.

2.2 Geolocating Users or Online Contents Other Than Images

There exist a host of work on studying the geographical scope of users or other online contents.

MediaEval workshops (Ferres & Rodriguez, 2010; Larson & Eskevic, 2010) provide a Flickr video dataset with textual tags, based on which a campaign is hosted to find the most possible

[4]http://wordnet.princeton.edu/wordnet/related-projects/

place for each video. They do not consider user profiles, which will be shown very helpful for the geotagging task in the experimental part of this paper.

Based on user-supplied address information and the social network in Facebook, Backstrom et al. (Backstrom, Sun, & Marlow, 2010) measure the relationship between geography and friendship. Using the measurement, the authors introduce an algorithm that predicts the geolocation of an individual with performance exceeding the IP-based geolocation.

Cheng et al. (Cheng, Caverlee, & Lee, 2010) propose a probabilistic framework for estimating a Twitter user's city-level location based on the content of the user's tweets. They build a classification model for identifying words in tweets with a strong geo-scope, and use a lattice-based neighborhood smoothing model for refining the estimation of a user's geolocation.

The recent work (R. Li, Wang, Deng, Wang, & Chang, 2012) addresses the problem of estimating the location for a twitter user. The work is based on the assumption that a twitter user is likely to follow users living close to her and to tweet nearby locations. The work presents a discriminative influence model to infer the home location of twitter users.

There also exist studies on extracting geographical information from web page (Amitay, Har'El, Sivan, & Soffer, 2004; Ding, Gravano, & Shivakumar, 2000) and search query (Backstrom, Kleinberg, Kumar, & Novak, 2008) (e.g., to exact geographic term "Effel Tower, Paris" from a web page). These approaches are based on the gazetteer taxonomy to identify locations. Hence it is difficult to extract locations when the text has a geographical focus, but does not mention toponyms found in gazetteers explicitly.

3 Geotagging Behaviour for Social Images

3.1 Dataset

We collect a large random subset of the Flickr images to simulate the entire Flickr images for analyzing the image geotagging patterns of users and evaluating the proposed solution. This is in contrast with many existing works (Cha, Mislove, & Gummadi, 2009; Mislove, Koppula, Gummadi, Druschel, & Bhattacharjee, 2008), in which the dataset is collected by starting from an initial set of users and then traversing along these users' social links, aiming to focus on social network structure. We do not adopt this approach, since starting from an initial set of users will be biased towards the set of initially selected users.

To achieve our goal, we choose to collect a dataset of images uploaded by a set of randomly selected users. Since we do not have the list of Flickr users, we make use of the fact that the vast majority of Flickr user identifiers take the form of $[0-9]\{8\}@N00$ (Mislove et al., 2008), i.e., an eight digit number followed by $@N00$. By searching a randomly selected subspace of $30,624,071$ possible Flickr user IDs of the entire space, we get a set of $2,252,758$ valid Flickr user IDs. This indicates that about 6.8% of these generated Flickr IDs are valid, which is consistent with the result reported by (Mislove et al., 2008). Then for each valid user ID, we download the meta information ("taken-time", "uploaded time", "tags", "latitude and longitude pair (if it is geotagged)) of all the publicly accessible images of the user, using the API exported by Flickr. This yields a dataset of $221,801,183$ images uploaded by $973,179$ Flickr users. Note that only $973,179$ out of the $2,252,758$ Flickr users have at least one public image. In other words, about 56.8% Flickr accounts do not have any public photo. The randomly generated dataset in this way is a representative subset of Flickr data, and is able to reflect the Flickr data properties. Note that the Flickr users can specify whether their photos are visible to others when uploading. In this work, we only collect the meta data of these public available images.

Table 1: Basic Statistics of the Dataset

	Total	Textual Tagged	Geotagged	Both Tagged
#Images	$221,801,183$	$99,649,530(44.9\%)$	$17,355,876(7.8\%)$	$13,268,992(5.9\%)$
#Users	$2,252,758$	$468,555(20.7\%)$	$106,289(4.7\%)$	$97,061(4.3\%)$

The numbers of images (resp. users) having textual tags, geotags, or both are summarized in Table 1. Table 1 shows that, though Flickr has provided geotagging service for years, as of Aug 2011 (the last date for collecting our dataset), only 7.8% Flickr images were geotagged, and only 4.7% Flickr users ever geotagged their images. As discussed in Introduction, many applications benefit from geolocation information of social image. Hence, an effective automated image geolocation suggesting tool is of great value. In contrast, much more Flickr images (44.9%) are annotated with textual tags by 20.7% of Flickr users.

3.2 Taken-time Gap vs Distance

We present an important analysis on the relationship between the distance and taken-time gap of two images uploaded by the same user. The analysis aims to identify guidelines to design geotagging methods.

Intuitively, *if the taken-times of two images by a user are close, they are likely to be spatially close*. To investigate whether this is true, we proceed to perform the following analysis. We sort the geotagged images of each user in ascending order of their uploaded times. Then for each geotagged image I_i, we find the image I_j with the closest taken-time among images that are taken by the same user and uploaded before I_i, and calculate the Euclidean distance ΔD_{ij} based on their coordinates.

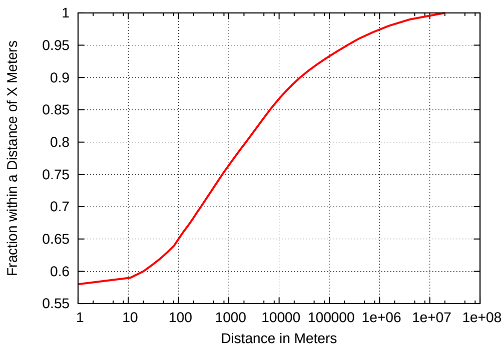


Figure 1: CDF of Distance between two Images with Close Taken-time by the same User

Figure 1 shows the Cumulative Distribution Function (CDF) of all possible ΔD_{ij}, from which Observation 1 can be made.

Observation 1 *More than half of images are within one meter from the historical images with the closest taken-time to them, and the percentage value reaches* 93% *when the distance scales up to* $100km$.

Observation 1 indicates a simple way to suggest geolocations for images, i.e., *a new image can be geotagged by the geolocation of the historical image with the closest taken-time by the same user*. Indeed, as we will see in Section 5, this method gives an impressive performance.

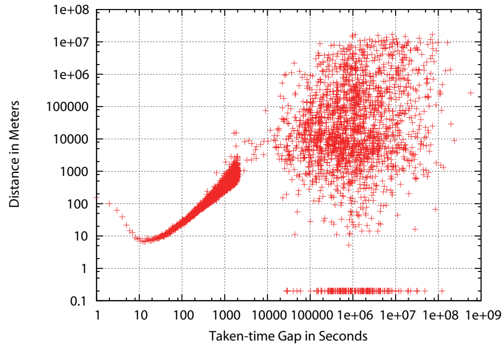


Figure 2: Mean Distance vs Different Taken-time Gaps

However, as shown in Table 1, only 4.7% of Flickr users did geotagging before. Thus a large portion (95.3%) of users cannot benefit from this simple method. For the remaining 95.3% of users, we cannot use the geolocation of historical image to tag a new image taken by the same user simply because his historical images have no geolocation. However, the textual tags of the historical images are still of great value for suggesting geolocation for a new image, since it is observed that most images are spatially close to their preceding images taken by the same users, and close images share similar tags (Serdyukov et al., 2009). Thus historical tags are related to geolocations of new images to some extent. Hence, we generate the profile of a user using the tags of his historical images. The inherent assumption is that a new photo of a user is likely to reside in the geolocations represented by the user's profile.

An open problem is whether we should treat a user's historical tags equally when suggesting geolocation for the user's new image? Intuitively, the images with shorter taken-time gaps are more likely to be close to the new image to be geotagged. Thus, their tags are perhaps more indicative for geotagging the new image. To verify the intuition, we analyze the dataset further by calculating taken-time gaps of all pairs of geotagged images for each user. Then for each taken-time gap, we compute the average distance corresponding to the taken-time gap, i.e., the expected spatial distance for a pair of images from the same user and with the given taken-time gap. The relationship between taken-time gap and the expected distance is shown in Figure 2. As we cannot plot zero values on a log scale axis, all distance values are increased by 0.2 meters.

It can be observed from Figure 2 that image pair with a taken-time gap of 10 seconds has the smallest expected distance, which is about 10 meters on average. As taken-time gap increasing to $2,000$ seconds, the distance increases accordingly to $1,000$ meters. This observation indicates that, *within a range between* 10 *to* $2,000$ *seconds, shorter taken-time gap between two images uploaded by a user comes with shorter geographic distance.* Figure 2 also shows that when taken-time gap is larger than $100,000$ seconds, distances fluctuate a lot, indicating that *a taken-time gap larger than* $100,000$ *seconds (more than* 27 *hours) provides little information about the geographic distance between two images.* This might come from the fact that people might either move far away or still stay at the same place between the long taken-time gap. We also note from Figure 2 that, when taken-time gaps are less than 10 seconds, mean distances are larger than that of 10 seconds. This might be caused by noise in the dataset (e.g., taken-time is not correct). However, the distances are still within 100 meters, which is still consistent with the observation that the expected distance for a short taken-time gap is short. From above discussions, Observation 2 can be made.

Observation 2 *Images contributed by the same user with close taken-time (within* $2,000$ *sec-*

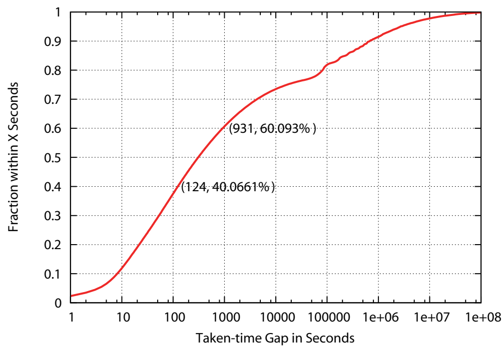


Figure 3: CDF of Taken-time Gap

onds) to the image to be geotagged are more likely to be spatially close to it, and thus their tags are more relevant to the geolocation of the image to be geotagged.

Observation 2 indicates that tags of images with taken-time gaps less than $2,000$ seconds are useful for geotagging. To investigate it, for an image to be geotagged, there exist a historical image of the same user and taken $2,000$ seconds close to it, we plot in Figure 3 the Cumulative Distribution Function (CDF) of taken-time gaps for image pairs used in Figure 1. Most taken-time gaps fall into a range from 10 to 100 seconds. As labeled in Figure 3, for more than $40\%(resp.60\%)$ images there exists a historical image of the same user, which was taken within $124(resp.931)$ seconds of their taken-times. We arrive at Observation 3.

Observation 3 *For more than* 60% *images we can find a historical image of the same user, which was taken within* $2,000$ *seconds of their taken-times.*

3.3 Summary and Discussion

In summary, the three observations offer important guidelines for us to develop geotagging methods.

- In particular, Observation 1 suggests a very simple yet effective method for geotagging a photo uploaded by the user who has done geotagging before, i.e., using the geolocation of the historical geotagged image with the closest taken-time to it, which is uploaded by the same user.
- For 95.3% of Flickr users, who have not done geotagging before, Observation 2 and 3 suggest that 1) tags of historical images with taken-time gap less than $2,000$ seconds are relevant to the geolocation of the image to be geotagged; 2) for most of images (more than 60%) we can find such tags.

4 Proposed Methods for Geolocation Suggestion

We present the geolocation suggestion methods using image tags in Section 4.1, the methods of utilizing user profile in Section 4.2, and the unified framework for combining both image tags and user profile for geolocation suggestion in Section 4.3.

4.1 Using Image Tags

4.1.1 Language Model Based Method

The Language Model (LM) method exploits image's tags, and is used by Serdyukov et al. (Serdyukov et al., 2009) to suggest an image's geolocation. Suppose that the world map is divided into grids of equal size, and each grid cell represents a location l. The LM method (Serdyukov et al., 2009) models each cell by a multinomial probability distribution over the vocabulary of tags. Specifically, the tag distribution of a location l is estimated by the tags of images located at this location as in Equation 1:

$$\begin{aligned} P(t|l) &= \frac{|l|}{|l|+\lambda}P(t|l)_{ML} + \frac{\lambda}{|l|+\lambda}P(t|G)_{ML}, \text{ where} \\ P(t|l)_{ML} &= \frac{tf_{t,l}}{\sum_{t'\in l} tf_{t',l}} \\ P(t|G)_{ML} &= \frac{tf_{t,G}}{\sum_{t'\in G} tf_{t',G}} \end{aligned} \tag{1}$$

Here $P(t|l)_{ML}$ and $P(t|G)_{ML}$ are maximum likelihood estimates of tag generation probabilities given location l and the global language models, respectively; $tf_{t,l}$ is the frequency of tag t in the image tags at location l, $|l|$ is the total number of tags of images at location l, and λ is the parameter of Dirichlet smoothing. Similarly, $tf_{t,G}$ is the frequency of tag t in the collection G.

Given an image's tag set T, the most likely location L in which the image was taken is estimated as Equation 2.

$$L = \arg\max_{l} P(l|T), \text{ where} \tag{2}$$

$$P(l|T) = \frac{P(T|l)P(l)}{P(T)} \propto P(T|l) = \prod_{t\in T} P(t|l) \tag{3}$$

4.1.2 Naive Bayes Method

The LM method implicitly assumes that images are evenly distributed across all locations, i.e., $P(l)$ is identical for any l. However, images are not evenly distributed on the map, as more photos would be taken at popular places. The strong assumption could degrade the accuracy of the LM method. We remove this assumption by using the Naive Bayes (NB) method given in Equation 4.

$$P(l|T) = \frac{P(T|l)P(l)}{P(T)} \propto P(T|l)P(l) = (\prod_{t\in T} P(t|l))P(l), \tag{4}$$

where $P(l)$ is estimated by the fraction of images located in l out of the entire image set and $P(t|l)$ is estimated as Equation 1.

In the NB method, the location for a given image with tag set T is predicted using Equation 2, where $P(l|T)$ is calculated using Equation 4.

4.1.3 Image Tag Expanding

As shown in Table 1, over half of images do not have any tag. We also observe from our dataset that a significant portion of images have very few tags. Obviously, it is very hard to predict geolocations of these images by the methods described in Sections 4.1.1 and 4.1.2.

For a query image with few than K tags, we propose to expand its tag set with tags relevant to the test image until its tag set contains K tags. If the test image already has K tags or more, we do nothing. The reason we set a limit K is that the historical tags may contain both useful information and noise for geotagging an image since the historical tags may correspond to images taken at multiple places.

We consider two methods to select tags for expansion. The first method is to randomly select tags from the historical tag set of the user for the test image. Based on Observation 2, we propose to select the tags with closer taken-time as the expanding tags and ties are broken arbitrarily. This method performs better than the first method in our preliminary study.

4.2 Using User Profile

4.2.1 Modeling User Profile by Language Model

Recall that we represent a user profile by the set of historical tags used by the user. We refer to a user profile as p. We build a language model $\mathcal{U}$ to model a profile p, which describes the location preferences of the user.

Given the set of historical tags of a user, i.e., the user's profile, denoted by p, we consider two methods to build language model from p. The first one treats equally each historical tag in user profile p. This can be modeled by replacing l with p in Equation 1.

The second method is motivated by Observation 2 that tags with small taken-time gaps to the image to be geotagged are more useful for geolocation suggestion. Hence, in the second method we give historical tags different weights based on the taken-time of their corresponding images—the closer an image is taken, the higher weight its tags have. In the model, each historical tag is given a weight of $(\text{gap}+1)^{-\alpha}$, where gap is the taken-time gap between the test image and the historical tag, and $\alpha \geq 0$ is a tuning parameter to be set using a development set. If $\alpha = 0$, the second method reduces to the first method.

4.2.2 KL Divergence Between User profile and Location

To incorporate the user profile information into the proposed framework (to be presented in Section 4.3) to suggest geolocation for an image, we need a way to measure the similarity between the user profile and each location. We propose to use KL-divergence for the purpose.

To measure the similarity between language models of a user's profile $\mathcal{U}$ and a place $\mathcal{L}$, the KL-divergence is computed by Equation 5.

$$D_{KL}(\mathcal{U}||\mathcal{L}) = \sum_i \mathcal{U}(i) \log \frac{\mathcal{U}(i)}{\mathcal{L}(i)}, \tag{5}$$

where $\mathcal{U}(i)$ and $\mathcal{L}(i)$ are distribution probabilities of tag i in language model $\mathcal{U}$ and $\mathcal{L}$, respectively. If $\mathcal{U}(i) = 0$, we assume $\mathcal{U}(i) \log \frac{\mathcal{U}(i)}{\mathcal{L}(i)} = 0$. Since $\mathcal{L}$ is smoothed by global language model, $\mathcal{L}(i) \neq 0$ for all possible i. Thus Equation 5 is well defined.

We note that the more similar $\mathcal{U}$ and $\mathcal{L}$ are, the smaller is the value of Equation 5. Equation 5 gets a value of zero when $\mathcal{U}$ and $\mathcal{L}$ are identical. Thus we suggest the location with the smallest KL-divergence as the geolocation.

We did consider other possibilities of using user profile: 1) other similarity measures, including cosine similarity and Jaccard similarity, and 2) using the methods in Section 4.1 on user profile, i.e., by replacing T with user profile p. However, the accuracies of these alternative methods are worse than the KL-divergence according to our preliminary study.

4.3 Unified framework for Combining Image Tags and User Profile

We proceed to present the proposed unified framework for combining the evidence of image tags and the evidence of the profile of the user who took the image.

Let S_{tag} be the similarity score between query image tags and a candidate location, which can be computed using the methods in Section 4.1. Let S_{user} be the similarity score between user profile and a candidate location computed using the method in Section 4.2.2. We propose a unified framework, as described in Equation 6, to combine the aforementioned two scores for geolocation suggestion.

$$S = (1 - b^{-n})S_{tag} + b^{-n}S_{user} \tag{6}$$

Here, b is a tuning parameter larger than 1, and n is the tag number of the query image to be geotagged. The rationale that we introduce n and b is that for an image with fewer tags, its tags might contain less location information, and thus we increase the weight of its user profile. When $b = 1$, Equation 6 reduces to the method that uses user profile alone. When $b \to \infty$, it reduces to the method that uses image tags alone.

We normalize S_{tag} and S_{user} to an identical value range. In our experiments, we use Equation 5 to measure the similarity between location and user profile, i.e., S_{user}, and use Equation 4 to measure the similarity between location and image tags, i.e., S_{tag}. Before using them in Equation 6, we normalize them into range $[0, 1]$ by Equation 7 and Equation 8, respectively.

$$S_{D_{KL}(\mathcal{U}||\mathcal{L})} = 1 - \frac{D_{KL}(\mathcal{U}||\mathcal{L}) - \min D_{KL}}{\max D_{KL} - \min D_{KL}} \tag{7}$$

$$S_{P(l|T)} = \frac{P(l|T) - \min P}{\max P - \min P}, \tag{8}$$

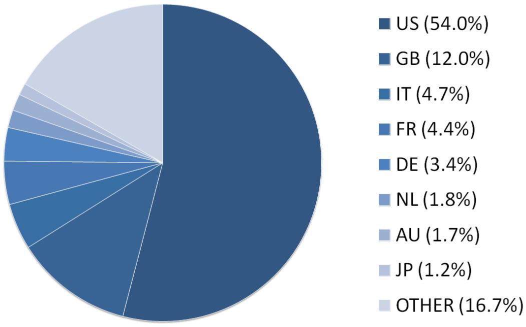
where max (resp. min) is the maximum (resp. minimum) value among all candidate locations.

5 Experiments

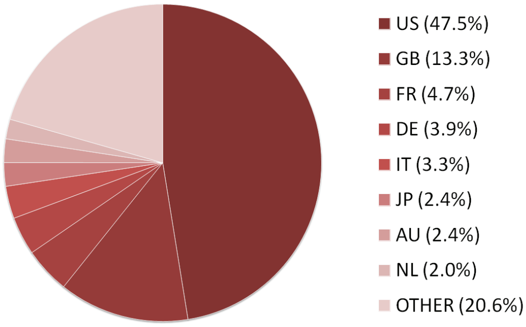
5.1 Data Filtering

To suggest geolocations for images, we need to discover the relationships between geolocations of images and their tags and/or user profiles. Thus only geotagged images can provide us information to analyze these relationships and build geolocation suggestion method. In the rest of this paper, we focus on the subset of geotagged images only. Note that focusing on geotagged dataset is only for model construction and performance evaluation (Geotagged images provide the ground truth for evaluation). Obviously, our model is applicable to geotagging new query image which has no geolocation information.

Each Flickr image is associated with an integer value to describe the location accuracy ranging from 0 to 16, among which 0 means there is no location information of the image, and 16 indicates the most accurate location information, i.e., street level location information. To make our models more accurate, we filter out images with location accuracy less than



Training Data



Testing Data

Figure 4: Image Distribution Over Counties

14. As tags used by few users or with small total occurrences are unlikely to carry location information, following the preprocessing method used in (Rattenbury & Naaman, 2009), we only keep tags occurring more than 25 times and used by at least two users in the experiments. Flickr allows users to apply the same set of tags to images uploaded at the same time, which has a negative effect on the building of model. Thus we follow the work (Serdyukov et al., 2009) and apply a filter to remove this bulk uploading effect. After filtering, only one image from each bulk uploading is kept. We separate the remaining images into two subsets. The first subset comprises images uploaded before March $1st$ 2011, and is used for model building. From the other subset comprising images uploaded after March $1st$ 2011, we randomly select $10,000$ images for parameter tuning, and $10,000$ images for performance testing. As to be explained in Section 5.2, our methods are designed for users who never geotagged before, and thus we further remove from the first subset (training set) the images that are uploaded by users whose images are also contained in the $20,000$ selected images (test set and parameter tuning set).

Finally, we obtain a dataset comprising $3,491,429$ images with $38,376$ unique tags for building our model, which is still much larger than the datasets used in previous work (Serdyukov et al., 2009). Figure 4 shows the distributions for training images and test images across different countries. We observe that about half of the images reside in US, followed by GB, and the distributions of training images and test images are similar. Figure 5 shows the number of images in the training dataset per square kilometers for each country, indicating that, although US has the most number of images, its image density is diluted by its large area of territory and is much smaller than those of the relatively small countries.

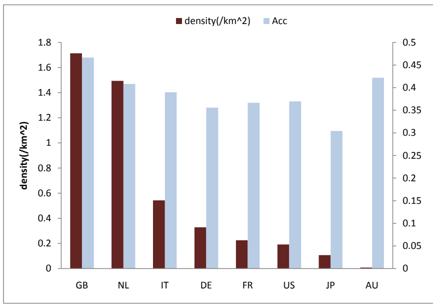


Figure 5: Image Density For Each Country

5.2 Suggesting Geolocations for Images Uploaded by Users Who Did Geo-tagging

As revealed by Observation 1, most images are spatially close to the historical images with the closest taken-time to them, which leads to a simple but effective method to suggest geolocation for a test image, i.e., among the geotagged historical images contributed by the same user, we select the one with the closest taken-time to the test image and use its geolocation as the prediction result. Note that to avoid the effect of bulk loading of users (since it is not reasonable to assume that the geolocation of images in the same bulk loading is known), we exclude the images with the same uploading time when we find the image with the closest taken-time.

By experimenting on the dataset described in Section 5.1, we find that this simple method is able to achieve an accuracy of 63.5%, 78.2% and 89.1% for $1km$, $10km$ and $100km$ grids, respectively. Compared to the results reported in previous works (Hays & Efros, 2008; Serdyukov et al., 2009; Crandall et al., 2009), the performance of this simple method is very impressive.

However, as shown in Table 1, only 4.7% Flickr users did geotagging before. Thus a large portion (95.3%) of users cannot benefit from this method, simply because they do not have geotagged historical images. The rest experiment will focus on methods for users who never geotagged before.

5.3 Experimental Setup

We evaluate the following methods over the dataset described in Section 5.1.

- GeoFolk (GeoFolk) (Sizov, 2010).
- Language Model based method (LM), which is described in Section 4.1.1 and used in (Serdyukov et al., 2009).
- Naive Bayes Model (NB), which is described in Section 4.1.2.
- Randomly (resp. taken-time aware) expanding tag set before applying Naive Bayes model ($\mathsf{NB}+\mathsf{RT}$ (resp. $\mathsf{NB}+\mathsf{CT}$)), which is described in Section 4.1.3.
- Using user profile (resp. weighted profile) alone (P (resp. WP)), which is described in Section 4.2.2.
- Combining NB and user profile (reps. weighted user profile) by the proposed unified framework ($\mathsf{NB}+\mathsf{P}$ (resp. $\mathsf{NB}+\mathsf{WP}$)), which is described in Section 4.3.

In our experiments, three different cell granularities, namely, $1km$, $10km$ and $100km$, are used for a fair comparison, since these three granularities are also used in previous work (Serdyukov et al., 2009). In fact, users would be interested in different granularities for different images: for the images of landmarks, e.g., the Golden Gate Bridge, the $1km$ is appropriate; however, for the images of nature senary, e.g., the Montes Alps, $100km$ would be better. Thus, it is reasonable to use these three granularities by following previous work.

After training models described in Section 4, the $10,000$ held-out tuning images are used to tune parameters in these models. All parameters (λ, α and b) for the evaluated methods are optimised on the held-out data by maximizing accuracy. We first fix λ as the optimum value for LM and NB, and then tune the other parameters independently.

The metric used for training models and tuning parameters is accuracy (**Acc**), which is the percentage of correctly predicted images out of all testing (training) images, and is the most important metric for our problem. Apart from **Acc**, we use another two types of metrics to evaluate the proposed methods: 1) fraction of images predicted correctly within k-cell distance (**Acc@k**); 2) fraction of images predicted correctly among top-k locations (**Top-k**). The rationale to introduce **Acc@k** and **Top-k** is that our task is actually a likelihood estimation problem, and thus apart from the suggested location, other top locations are important as well. An example application is to present multiple locations with the highest probabilities to users, a user can select one location out of them to geotag his image.

5.4 Experimental Results

Table 2 shows the results of all methods described in Section 5.3 except GeoFolk for $1km$, $10km$ and $100km$ sized grids.

With the most fine granularity ($1km$), $\mathsf{NB} + \mathsf{WP}$ is able to correctly predict locations for 6.37% of testing images, while for $100km$, this value increases to 32.77%. Not surprisingly, suggesting accuracies of all methods increase as cell size increases. Moreover, $\mathsf{NB} + \mathsf{WP}$ improves on **Acc** by 10.6%, 16.8% and 34.2% over the method (LM) for $1km$, $10km$ and $100km$, respectively. We note that the results of LM are better than those reported in (Serdyukov et al., 2009), which could be attributed to the larger training data we use. Note that even 10.6% improvement (for $1km$ grid) is very significant considering the huge number of social images without geolocations(e.g., as of Aug 2011 Flickr held 6 billion photos[5]).

Though $\mathsf{NB} + \mathsf{WP}$ achieves the best accuracy (**Acc**) among all methods, it is not an all-round winner for all metrics. It can be noticed that $\mathsf{NB} + \mathsf{CT}$ achieves better **Top-2** and **Top-3** than **NB+WP** for $1km$ and $10km$ grids, which means $\mathsf{NB} + \mathsf{CT}$ could be a good choice for application accepting multiple suggested locations with size $1km$ or $10km$.

The usefulness of tag expanding By comparing the results of $\mathsf{NB} + \mathsf{CT}$ and NB in Table 2, we find that tag expanding improves on the performance of NB. The improvement increases as the grid size increases. The reason would be the fact that expanded historical tags are more likely be relevant to a larger cell in which the test image resides.

The usefulness of user profile By comparing the performance of $\mathsf{NB} + \mathsf{WP}$ and NB, we find that utilizing user profile is able to greatly improve the performance for location suggestion.

$\mathsf{NB} + \mathsf{WP}$ outperforms NB by 6.5%, 14.5% and 33.0% for $1km$, $10km$ and $100km$ granularities, which increases as the granularities of cell become coarser. The reason for this is that a user takes photos within a big region, which is better captured by cells in $10km$ and $100km$.

[5]http://thenextweb.com/socialmedia/2011/08/05/flickr-hits-6-billion-total-photos-but-facebook-does-that-every-2-months/

Table 2: Performance for All Methods. NB + WP improves on LM by 10.6%, 16.8% and 34.2% for $1km$, $10km$ and $100km$, respectively, in terms of **Acc**.

	Acc	Acc@1	Acc@2	Acc@3	Top-2	Top-3	Top-4
			$1km$				
LM	0.0576	0.0958	0.1173	0.1297	0.0815	0.0935	0.1029
NB	0.0598	0.0989	0.1217	0.1354	0.0811	0.0947	01033
NB + RT	0.0572	0.0946	0.1197	0.1351	0.0771	0.0906	0.0995
NB + CT	0.0634	0.1065	**0.1352**	0.1521	**0.0866**	**0.1011**	0.1123
P	0.0121	0.0270	0.0438	0.0599	0.0187	0.0236	0.0286
WP	0.0389	0.0737	0.0960	0.1178	0.565	0.696	0.775
NB + P	0.0577	0.0968	0.1230	0.1430	0.0800	0.0934	0.1046
NB + WP	**0.0637 (+10.6%)**	**0.1067**	0.1334	**0.1546**	0.0858	0.1007	**0.1138**
			$10km$				
LM	0.1550	0.2014	0.2187	0.2263	0.1996	0.2135	0.2228
NB	0.1581	0.2058	0.2245	0.2312	0.2010	0.2218	0.2368
NB + RT	0.1602	0.1941	0.1985	0.2005	0.2084	0.2311	0.2447
NB + CT	0.1794	0.2145	0.2193	0.2212	**0.2323**	**0.2575**	**0.2722**
P	0.0710	0.1214	0.1557	0.1697	0.1004	0.1164	0.1280
WP	0.1309	0.1982	0.2328	0.2493	0.1734	0.1996	0.2132
NB + P	0.1685	0.2301	0.2589	0.2717	0.2138	0.2365	0.2503
NB + WP	**0.1811 (+16.8%)**	**0.2442**	**0.2721**	**0.2851**	0.2285	0.2515	0.2669
			$100km$				
LM	0.2442	0.2766	0.2976	0.3081	0.2758	0.2898	0.3006
NB	0.2463	0.2818	0.3056	0.3231	0.2977	0.3300	0.3490
NB + RT	0.2839	0.3349	0.3644	0.3849	0.3399	0.3685	0.3854
NB + CT	0.3056	0.3585	0.3912	0.4151	0.3596	0.3870	0.4068
P	0.2196	0.2828	0.3209	0.3456	0.2729	0.3044	0.3311
WP	0.2978	0.3643	0.4020	0.4283	0.3560	0.3893	0.4185
NB + P	0.3023	0.3633	0.3973	0.4189	0.3640	0.3967	0.4214
NB + WP	**0.3277 (+34.2%)**	**0.3899**	**0.4254**	**0.4491**	**0.3856**	**0.4193**	**0.4439**

Thus, user profile built on the larger cells contains more accurate information for a user. This fact is also supported by suggesting accuracies (**Acc**) of WP and NB for grids of different sizes. As can be found in Table 2, NB outperforms WP by 53.7% with regard to **Acc** for $1km$ grids, while in contrast the **Acc** of WP beats NB by 20.9% for $100km$ grids. This gives a surprising finding that, when suggesting $100km$ sized locations, user's history profile is more accurate than the tags of test image for location prediction. This is due to that images uploaded by the same user are very likely to be within $100km$, and some tags of an image might not be location related, which introduce noise.

The usefulness of time factor Also, it can be observed that NB + WP (resp. NB + CT) outperforms NB + P (resp. NB + RT) greatly, which verifies the effectiveness of utilizing image taken-time in location suggestion. This result is consistent with Observation 2 that historical images with shorter taken-time gaps to the test image are more likely to be spatially close it.

Comparing with GeoFolk For GeoFolk, the **Acc**s for $10km$ and $100km$ grids are 0.0758 and 0.1606, respectively, which are much worse than that of the baseline method LM. Moreover, the time complexity of GeoFolk is too high, which makes GeoFolk infeasible for $1km$ sized grids. The reason why GeoFolk does not work well for geotagging problem might be due to the huge number of candidate cells. As usually topic model uses tens or hundreds of latent topics (and it becomes too expensive to use with larger number of topics), latent topic model is not able to well discriminate such a large number of classes. The number of topics used by GeoFolk in this paper is set to 100. We tried to use 300 topics to achieve better results. However, we cannot obtain the results after 5 days running due to the high time complexity of GeoFolk.

Looking into country To better understand the performance for images from different countries, we select eight countries with the largest numbers of images, and compare the Acc of each using NB + WP on $100km$ grids. The results are shown in Figure 5. We can find that Great Britain (GB) achieves the highest **Acc** of 46.65%, which also has the largest image density. Netherlands (NL) has the second largest image density, which achieves an **Acc** of 40.82%. Other countries in Figure 5 have much sparser image distributions than do GB and NL. **Acc**s of these countries are lower than that of GB. Though it is not a strict rule, we can find a trend that denser image distribution leads to higher suggestion accuracy. An outstanding counter example is Australia (AU), which has a very sparse image distribution while having a relatively high **Acc** of 42.19%. The reason could be that the majority of territory of Australia is not human occupied, and thus images from Australia are limited to some small areas. As reported in Table 2, the **Acc** for all countries obtained by NB + WP with $100km$ is 32.77%. Seven out of the eight countries having the largest numbers of images in our dataset (except for Japan) have higher accuracies than this overall value, which indicates that we are able to obtain better performance if focusing on countries with a large number of Flickr images.

Looking into user moving region Some Flickr users move around a lot and upload images with diversified geolocations, while some users move in a relatively small region and upload images with close geolocations. To understand the effect of the sizes of user moving regions on the performance gain for NB + WP over NB, we divide users into different groups based on the number of unique geolocations (cells) where their images reside in, and calculate the relative **Acc** improvements of NB + WP over NB for different groups of users.

For $1km$, group 1 (resp. 2, 3 and 4) contains users with 1 to 9 (resp. 10 to 39, 40 to 106 and more than 106) unique geolocations. For $10km$, group 1 (resp. 2, 3 and 4) contains users with 1 to 5 (resp. 6 to 19, 20 to 49, and more than 49) unique geolocations. For $100km$, group 1 (resp. 2, 3 and 4) contains users with 1 to 3 (resp. 4 to 9, 10 to 23 and more than 23) unique geolocations. Users are divided such that users in each group have similar number of images in the testing set. Figure 6 depicts the relative **Acc** improvements of NB + WP over NB for different groups of users with $1km$, $10km$ and $100km$ granularities. We find that the method (NB + WP) combining evidences from both user profile and image tags improves on the **Acc** of the method (NB) employing image tags alone for all user groups (all relative improvements are positive), which shows the effectiveness and robustness of incorporating user profile for image geotagging. For users with larger number of geolocations, NB + WP achieves relatively less improvement over NB(except Group 2 and 3 for $10km$). The reason would be the fact that the profiles of users with smaller number of geolocations contain more accurate geolocation information. This is consistent with the result, indicated in Table 2, that NB + WP achieves

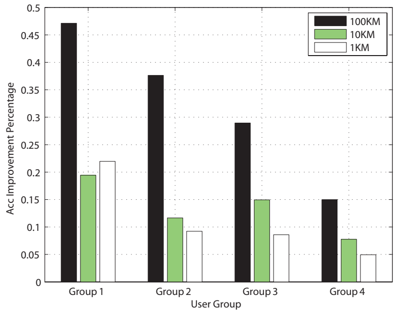


Figure 6: **Acc** Improvement of $\mathsf{NB} + \mathsf{WP}$ over NB

larger improvement over NB for grids of larger size.

5.5 Experimental Summary

Section 5.2 shows that a simple method can achieve impressive accuracy for images from users who did geotagging. However, as shown in Table 1, only 4.7% Flickr users did geotagging before. Thus majority of users cannot benefit from this method.

In summary, experimental results for images from users who never geotagged before show the following.

- Combining image tags and user profile together is able to achieve much better accuracy than utilizing either of them alone for geolocation suggestion.
- The taken-time stamps of historical images play an important role in utilizing user profile.
- While $\mathsf{NB} + \mathsf{WP}$ achieves the best performance for most scenarios, $\mathsf{NB} + \mathsf{CT}$ is a better choice for applications accepting multiple suggested geolocations at the size of $1km$ or $10km$ granularity.
- For large granularity ($100km$), it achieves better accuracy to use a user profile alone than using tags of the query image alone.
- The proposed methods are able to achieve better accuracy if we focuse on countries with the largest numbers of Flickr images.
- The method ($\mathsf{NB} + \mathsf{WP}$) combining evidences from both user profile and image tags improves on the **Acc** of the method (NB) employing image tags alone for all the users with different moving regions. However, the improvement for images from users whose movements cover fewer cells is greater.

6 Conclusion and Future Direction

In this paper, we propose a solution to the problem of geolocation suggestion for social images. On a large Flickr dataset, for the first time we analyze the user uploading patterns, user geotagging behaviors, and the relationship between the taken-time gap of two images from the same user and their spatial distance. Based on the analysis, we represent a user profile by the historical tags used by the user, with the taken-time of historical tags being considered. Then we build a multinomial model based on the user profile for geotagging. A unified framework is proposed to combine information from user profile and image tags to suggest geolocations for social images. This is the first work that utilizes the user profile to improve the geolocation suggestion for social images. The proposed methods are evaluated on the Flickr dataset. Experimental results show that, for users who have never done geotagging, our method is able to improve on the performance of baselines significantly for grids at different granularities. For users who have done geotagging, a simple method is able to achieve very high accuracy.

This work suggests a number of promising directions for future work. First, in addition to the user profile, some other possible evidences can be incorporated into this work, such as users' social links, month or season of images. Second, it appears promising to develop an adaptive algorithm to automatically decide the size of suggested location. If we have sufficient evidences to decide the geolocation of an image, we can suggest a more specific geolocation. Otherwise, we suggest a larger geolocation. Finally, it would be interesting to study the user satisfaction for the geotagging task, in addition to the evaluation based on the groundtruth.

Acknowledgement
This work is supported in part by a grant awarded by a Singapore MOE AcRF Tier 2 Grant (ARC30/12), a Singapore MOE AcRF Tier 1 Grant (RG66/12), and a grant awarded by Microsoft Research Asia. This work was done when Bo Liu worked at Nanyang Technological University. Quan Yuan would like to acknowledge the Ph.D. grant from the Institute for Media Innovation, Nanyang Technological University, Singapore.

References

Amitay, E., Har'El, N., Sivan, R., & Soffer, A. (2004). Web-a-where: geotagging web content. In *Proceedings of the 27th annual international acm sigir conference on research and development in information retrieval* (pp. 273–280). New York, NY, USA: ACM.

Backstrom, L., Kleinberg, J., Kumar, R., & Novak, J. (2008). Spatial variation in search engine queries. In *Proceeding of the 17th international conference on world wide web* (pp. 357–366). New York, NY, USA: ACM.

Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on world wide web* (pp. 61–70). New York, NY, USA: ACM.

Buscaldi, D., & Rosso, P. (2008). A conceptual density-based approach for the disambiguation of toponyms. *International Journal of Geographical Information Science*, *22*(3), 301–313.

Cha, M., Mislove, A., & Gummadi, K. P. (2009). A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on world wide web* (pp. 721–730). New York, NY, USA: ACM.

Chen, W. C., Battestini, A., Gelfand, N., & Setlur, V. (2009). Visual summaries of popular landmarks from community photo collections. In *Proceedings of the 17th acm international conference on multimedia* (pp. 789–792). New York, NY, USA: ACM.

Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th acm international conference on information and knowledge management* (pp. 759–768). New York, NY, USA: ACM.

Cong, G., Jensen, C. S., & Wu, D. (2009, August). Efficient retrieval of the top-k most relevant spatial web objects. *Proc. VLDB Endow.*, *2*, 337–348.

Crandall, D. J., Backstrom, L., Huttenlocher, D., & Kleinberg, J. (2009). Mapping the world's photos. In *Proceedings of the 18th international conference on world wide web* (pp. 761–770). New York, NY, USA: ACM.

Ding, J., Gravano, L., & Shivakumar, N. (2000). Computing Geographical Scopes of Web Resources. In *26th international conference on very large databases.* Cairo, Egypt.

Ferres, D., & Rodriguez, H. (2010). TALP at MediaEval 2010 Placing Task: Geographical Focus Detection of Flickr Textual Annotations. In *Working notes proceedings of the mediaeval 2010 workshop.*

Hays, J., & Efros, A. A. (2008). IM2GPS: estimating geographic information from a single image. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, *0*, 1–8.

Hsu, M.-H., & Chen, H.-H. (2011). Efficient and effective prediction of social tags to enhance web search. *JASIST*, *62*(8), 1473-1487.

Larson, M., & Eskevic, M. (2010). Overview of MediaEval 2011 Rich Speech Retrieval Task and Genre Tagging Task. In *Working notes proceedings of the mediaeval 2011 workshop.*

Li, R., Wang, S., Deng, H., Wang, R., & Chang, K. C.-C. (2012). Towards social user profiling: unified and discriminative influence model for inferring home locations. In *Kdd* (p. 1023-1031).

Li, Y., Crandall, D. J., & Huttenlocher, D. P. (2009, September). Landmark Classification in Large-scale Image Collections. In *Proceedings of international conference on computer vision* (pp. 1957–1964). Kyoto, Japan.

Liu, X., Wang, G., Johri, A., Zhou, M., & Fan, W. (2012). Harnessing global expertise: A comparative study of expertise profiling methods for online communities. *Information Systems Frontiers*, 1-13.

Mislove, A., Koppula, H. S., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2008). Growth of the flickr social network. In *Proceedings of the first workshop on online social networks* (pp. 25–30). New York, NY, USA: ACM.

Ostermann, F. O., Tomko, M., & Purves, R. (2013). User evaluation of automatically generated keywords and toponyms for geo-referenced images. *JASIST*, 480-499.

Rattenbury, T., & Naaman, M. (2009). Methods for extracting place semantics from Flickr tags. *ACM Trans. Web*, *3*(1), 1–30.

Rorissa, A. (2010). A comparative study of flickr tags and index terms in a general image collection. *JASIST*, *61*(11), 2230-2242.

Serdyukov, P., Murdock, V., & Zwol, R. van. (2009). Placing flickr photos on a map. In *Proceedings of the 32nd international acm sigir conference on research and development in information retrieval* (pp. 484–491). New York, NY, USA: ACM.

Sigurbjörnsson, B., & Zwol, R. van. (2008). Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on world wide web* (pp. 327–336). New York, NY, USA: ACM.

Sizov, S. (2010). GeoFolk: latent spatial semantics in web 2.0 social media. In *Proceedings of the third acm international conference on web search and data mining* (pp. 281–290). New York, NY, USA: ACM.

Stvilia, B., & Jörgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in flickr. *JASIST*, *61*(12), 2477-2489.

Toyama, K., Logan, R., & Roseway, A. (2003). Geographic location tags on digital images. In *Multimedia '03: Proceedings of the eleventh acm international conference on multimedia* (pp. 156–166). New York, NY, USA: ACM.

Ye, M., Shou, D., Lee, W. C., Yin, P., & Janowicz, K. (2011). On the semantic annotation of places in location-based social networks. In *Proceedings of the 17th acm sigkdd interna-*

tional conference on knowledge discovery and data mining (pp. 520–528). New York, NY, USA: ACM.

Ye, M., Yin, P., Lee, W. C., & Lee, D. L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international acm sigir conference on research and development in information* (pp. 325–334). New York, NY, USA: ACM.

Zheng, Y.-T., Zhao, M., Song, Y., Adam, H., Buddemeier, U., Bissacco, A., et al. (2009). Tour the world: Building a web-scale landmark recognition engine. In *Proceedings of international conference on computer vision and pattern recognition* (pp. 1085–1092).