Georeferencing images using tags: application with Flickr

Thomas Wider
Department of Geography
University of Zürich
Winterthurerstrasse 130
8057 Zürich, Switzerland
twider@geo.uzh.ch

Damien Palacio
Department of Geography
University of Zürich
Winterthurerstrasse 130
8057 Zürich, Switzerland
damien.palacio@geo.uzh.ch

Ross S. Purves
Department of Geography
University of Zürich
Winterthurerstrasse 130
8057 Zürich, Switzerland
ross.purves@geo.uzh.ch

Abstract

Flickr has a huge collection of images, but only some 3% have explicit coordinates assigned to them. However, most images have tags assigned by users. Many of these tags can be linked to one or several specific places, and can thus be used to predict the location of images. In this paper we propose an approach based on tag frequency and a geodesic grid to extract tags related to location and thus assign coordinates to non-georeferenced images. The method presented is efficient and robust to tagging behaviours, and produces results comparable to state of the art Naive Bayes based methods.
Keywords: geotagging, flickr, georeferencing.

1 Introduction

Flickr[1], a Yahoo! service, is one of the most popular image hosting websites. It currently has more than six billion images uploaded by users of which around 200 million, some 3%, are georeferenced[2].

Studies on search engine logs have shown that up to 18% of user's queries contain geographic information [2, 12]. For images, it is even more important: in [1] it was shown that around 70% of images included at least one place name tag.

Geotagging the 97% non-georeferenced Flickr images can have many potential applications including:

- identifying the footprints of vernacular place names used regularly as tags [1]
- generating tags for images without tags [10]
- suggesting tags for users submitting images [6]
- improving access of images like organization or browsing of collection [7]
- geographic information retrieval [4]

In this paper, we propose a simple, fast and robust approach to georeference images. It uses tag frequency as well as user characteristics and is based on a machine learning approach. We also present an experiment that evaluates our approach and compares it to a Naive Bayes classification.

This paper is organized as follows. In Section 2, we describe works on defining the location of user generated content, in particular tags. In Section 3, we describe the approach proposed in this paper. In Section 4, we describe the Flickr data used and the results of our experiments. The last section concludes and discusses further work.

2 Previous work

Statistical language models with machine learning (in particular the Naive Bayes approach) are used to estimate the location of user generated contents [8, 13, 14]. In [14] they use Kullback-Leibler divergence on Wikipedia articles; the results are slightly better than the Naive Bayes approach. Wikipedia articles have the particularity of typically focussing on a single subject, often related to a single geographic footprint. User behaviour is considered unimportant in georeferencing such content. [8, 13] propose approaches to georeference Flickr images using tags. Tags associated with images by users can help to specify locations. Such *place semantics* of tags can be automatically extracted by using already georeferenced images [11]. Unlike these previous works that aim to assign image locations to a geodesic grids, [3] use clustering. The drawback of clustering is that it doesn't provide a full spatial coverage.

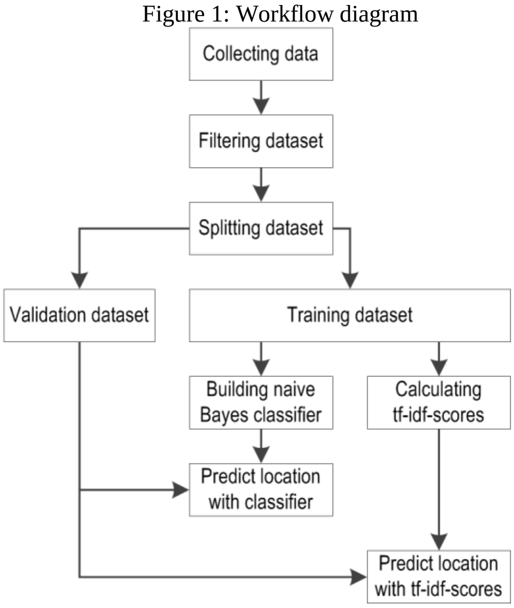
Current implementations of Naive Bayes classification typically don't use parameters such as the importance of a tag among users in a particular cell. Collections like Flickr have considerable user bias as a minority of users upload a large proportion of images (bulk uploads). In [8] 63% of images are bulk uploads. The Naive Bayes approach is intensive in terms of processing, in particular with a grid, where it has to calculate probabilities for each cell for each picture. In [14], it tooked them four months to run the full experiment (around 400 000 training article and 50 000 test articles; six strategies: three baselines, three non-baselines) on a server computer. We now present our approach which aims to predict location for pictures that we will compare to Naïve Bayes approach.

1 http://www.flickr.com
2 On the 31th January 2013, 198 570 914 are available from Flickr API

3 Methods

The following section presents our approach to predict the location of photos. Figure 1 shows the workflow.

Figure 1: Workflow diagram



3.1 Filtering and splitting datasets

We obtained two datasets with georeferenced photos for Great Britain and Switzerland using the public API provided by Flickr. The search is only limited to the bounding boxes of Great Britain and Switzerland.

Since our interest lies in the identification of tags people use to describe specific locations, photos without tags are removed from the datasets. Flickr provides an accuracy level for georeferenced photos ranging from 1 (world level) to 16 (street level). Only photos with street level accuracy (16) remain in the datasets which is the case for over 99% of images. Photos outside the 5km buffered border of both countries are discarded. We filter out stop-words and also a manually gathered list of photographic-related words followed by normalization of the tags (e.g., “canon”, “35mm”). Characters with umlauts or accents are altered to their base types and punctuation marks are stripped. To sustain the semantics in tags consisting of multiple words, we remove space characters between words to merge them into a single compound tag (e.g. “hyde park” becomes “hydepark”).

To minimize the influence of single users, bulk uploads are excluded from the datasets. Bulk uploads are multiple photos with an identical set of tags uploaded by a single user. We also remove single users with only one photo and prolific users contributing more than 10’000 photos in Great Britain and 1’000 photos in Switzerland respectively. As shown in [9], prolific users can be a significant source of bias. Since we are interested in a general description of a place, photos of these users were discarded.

Finally, tags only assigned by a single user or used less than ten times are not considered as representative descriptions and thus removed. Subsequently the photos in the filtered datasets are split into a training and a validation dataset.

We use two different splitting approaches:

- Distribute photos randomly
- Distribute all photos of a user either randomly to the training or to the validation dataset

The different splitting approaches allow the analysis of location predictions for photos mainly based on tags from the same user. Size of both sets is adjustable (from 25% to 75%) but the best results are obtained with a training-validation ration of 75-25.

3.2 Extracting relevant tags using term selection

Tags from photos of the training dataset, which are relevant to describe places represented by the discrete cells of a geodesic grid, are identified using a TF-IDF-based term selection method proposed by [11].

According to TF-IDF, the relevance of a tag in a specific grid cell is based on its term frequency (TF) and its inverse document frequency (IDF). The frequency of photos using a specific tag in a cell is represented by TF-component. The distribution of photos using a specific tag among all photos is represented by the IDF-component. The local use of a tag in a cell is considered to be more representative the more users assign it to their photos. This is represented by an additional user frequency (UF) introduced by [11].

The score of a tag (t) in a cell (c) is calculated as the product of all three factors as shown in equation 1 below:

$$score(t,c) = tf(t,c) * idf(t) * uf(t,c) \qquad (1)$$

From all occurring tags among a cell, the 30% highest ranked are selected as relevant placetags to describe the location of the grid cell.

3.3 Predict the location using relevant tags

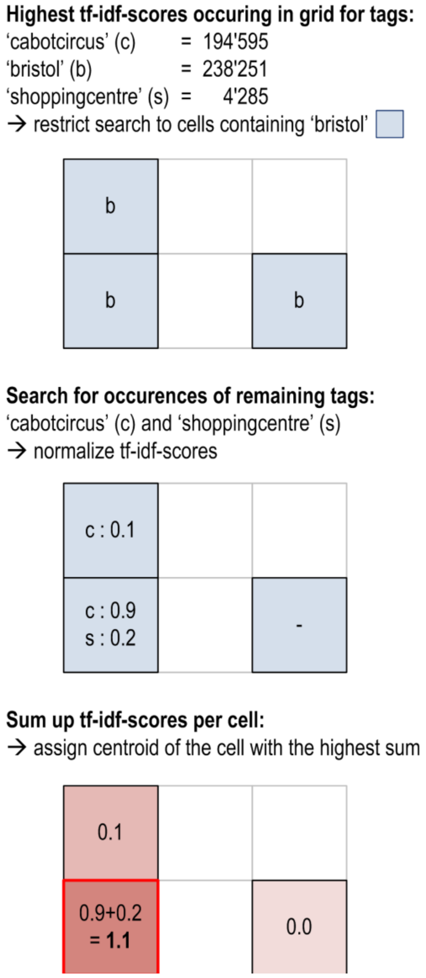
To predict the location for a photo from the validation dataset only using its textual description, the cell with the highest TF-IDF score for the given tags is determined.

In a first step the search for the given tags is limited to cells that contain the most relevant tag (i.e. the one with the highest TF-IDF-score incorporating all possible occurrences in all cells) among these given tags.

We then add up the normalized TF-IDF-scores of occurrences of the remaining tags within all the preselected cells. Therefore we normalize a tag’s TF-IDF-score to the range [0…1].

Finally the centroid of the cell containing the highest sum is chosen as location for the photo.

Figure 2: Example of TF-IDF-scoring approach



3.4 Predict the location using a Naive Bayes classifier (baseline)

As we use Naive Bayes approach as our baseline, we present it succinctly in this section.

Naive Bayes is a supervised machine learning-based text classification method [5]. The classification is based on the assumption of independent occurrence of terms in a document. To predict membership of a document to a class, based on its terms, a set of training documents with predefined class memberships is required. In our case the cells of a grid are the distinct classes, photos are documents and tags can be referred to terms.

The distribution of the georeferenced photos and their associated tags in the training dataset is used to build a statistical learner. The probability of a photo p to be member of cell c is estimated as in equation (2):

$$P(p|c) \propto P(c) \prod_{1 \le t_k \le n} P(t_k|c) \qquad (2)$$

$P(c)$ is the prior probability of a photo p to be in cell c, $P(t_k|c)$ is the conditional probability of tag t_k of photo p in cell c.

Using the photos tags of the validation dataset, the classifier determines the most likely cell by calculating all probabilities $P(p|c)$ and chooses the cell with the highest probability.

We now present the collection on which we use our approach and the results obtained.

4 Results

We collected Flickr photos for two areas: Great Britain (GB) and Switzerland (CH). Table 1 details these two datasets. The GB dataset is seven times bigger than CH. We choose GB and CH to check if our method works on datasets with different characteristics. Table 2 details the dataset after filtering (bulk uploads, photos without tags). As we can see, a large part of the dataset is removed: 77% for GB, 88% for CH.

Table 1: Dataset statistics

Country	GB	CH
Uploaded	1.1.2004 - 31.5.2012	1.1.2004 - 24.11.2012
Total photos	6'663'046	876'182
Photos with tags	81.7%	77.7%
Tags per photo	8.5	8.7
Unique users	92'662	18'395
Photos per user	58.7	37.0
Unique tags	1'484'355	228'957

Table 2: Dataset after filtering

Country	GB	CH
Total photos	1'529'504	106'617
Unique users	48'789	6'472
Photos per user	94'620	12'884

First, we compared the splitting approaches: distribute photos randomly (RP), and, distribute all photos of a user either randomly to the training or to the validation dataset (RU). As we can see in Table 3, RP gives the best results. The difference is even bigger for Naive Bayes approach.

Table 3: Comparison of splitting approaches (distance)

	TF-IDF-scores		Naive Bayes	
	RP	RU	RP	RU
Mean (km)	49.4	78.3	55.4	93.8
Median (km)	**1.0**	2.2	**1.0**	8.1

RP: Random Photos, RU: Random Users

Then we looked at the accuracy of predicting location of photos with our TF-IDF-scores. Table 4 presents the results for different grid sizes. On the finest grid, our approach locates photos in the correct grid cell with an accuracy of 40%. If we compare to the neighbors (correct area but wrong cell), the accuracy is almost 67%. Figure 3 shows how the neighbors of a correct cell are selected.

We then, compared our approach to the Naive Bayes classification. Because of the heavy resources needed for Naive Bayes approach, we only have preliminary results on a sample of the dataset (1000 photos). As we can see on Table 5, Naive Bayes approach performs slightly better but the results are quite similar (47% vs. 44%). Comparing accuracy with adjacent neighbors both methods produce similar results (68% vs. 67%). This experiment need to be explored further to confirm results of Naive Bayes approach on the full dataset. If we look at previous work, [8] obtained 17% accuracy for 1 km and 40% for 3 km, while [13] obtained 7% for 1 km 14% for 5 km. Nevertheless, their datasets covered the world, and not individual countries like us, but the size of their dataset for the world is quite small; comparable to our GB dataset for [8] (10 millions) and to our CH dataset for [13] (400K). They also use the RU splitting approach which gives worse results as we presented in Table 3. If we use the RU splitting we obtain 28% of accuracy for TF-IDF-scores. Importantly, previous works make no comments on the geographic variability in the efficacy of the methods applied.

Figure 4 and Figure 5 show, respectively, median error par cell for TF-IDF-scores in Great Britain (GB) and Switzerland (CH). The more accurate places are dense areas like cities (e.g. London, Zurich) or locations with specific properties (e.g coast).

Figure 3: Neighbouring of a correct cell

n3	n3	n3	n3	n3	n3	n3
n3	n2	n2	n2	n2	n2	n3
n3	n2	n1	n1	n1	n2	n3
n3	n2	n1	n0	n1	n2	n3
n3	n2	n1	n1	n1	n2	n3
n3	n2	n2	n2	n2	n2	n3
n3	n3	n3	n3	n3	n3	n3

Table 4: Prediction accuracy (GB+CH) on full dataset for TF-IDF-scores

	1 km	5 km	10 km
Correct cell (n0) (%)	**40.75**	**54.75**	**58.9**
1st neighbor (n1) (%)	17.5	16.3	16.4
2nd neighbor (n2) (%)	5.5	3.85	3.65
3rd neighbor (n3) (%)	3.05	2.45	1.95
n0, n1, n2, n3 (total)	**66.85**	**77.3**	**80.9**

Table 5: Prediction accuracy on a sample (1000 photos)

	TF-IDF-scores	Naive Bayes
Correct cell (n0) (%)	**44.55**	**47.35**
1st neighbor (n1) (%)	15.3	14.35
2nd neighbor (n2) (%)	5.65	4.7
3rd neighbor (n3) (%)	2.3	2.55
n0, n1, n2, n3 (total)	**67.8**	**68.95**

Figure 4: Median error per cell in Great Britain

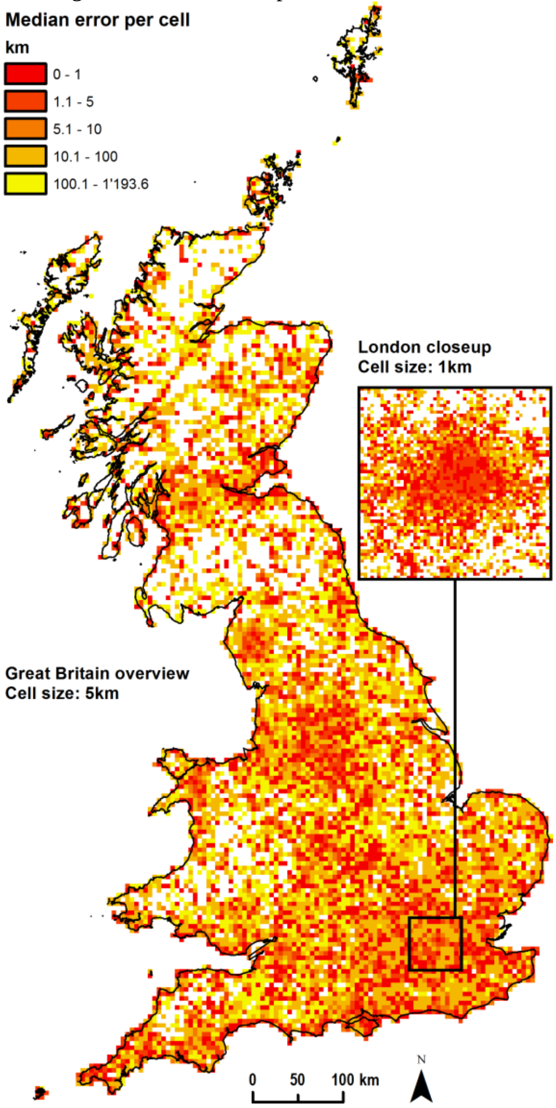
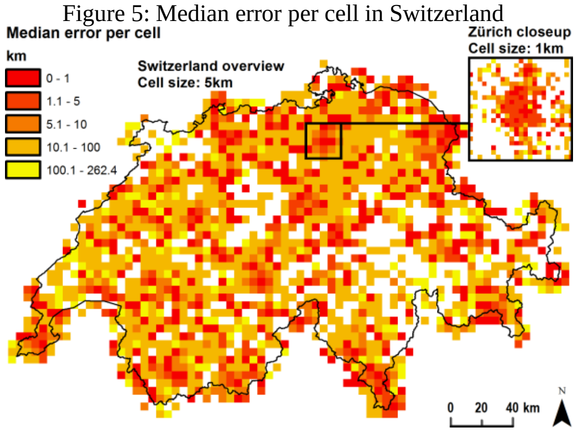


Figure 5: Median error per cell in Switzerland



5 Discussion

In this paper we presented an approach to georeference images based on a TF-IDF method. The main goal was to increase the number of images with known location for applications like information retrieval or collection visualization and browsing.

The approach is robust as changing the training set has no effect on the results. It's also an efficient approach as it can be calculated on a desktop computer. Indeed, the TF-IDF approach uses only cells containing the most relevant tag instead of using the whole grid. Furthermore almost 41% of images are correctly located in the correct one kilometer by one kilometer cell, and, 58% in a neighboring cell. Our preliminary experiment with Naive Bayes classification, our baseline, shows that our approach gives similar results, but also that it is more efficient and robust than this baseline. We are continuing this experiment with more data to confirm these results.

In [8, 14], a smoothing approach based on cell neighbors was applied and significantly improved the results. In future work we will explore whether this is also the case for our method.

References

[1] Hollenstein, L. and Purves, R. 2010. Exploring place through user-generated content: Using Flickr to describe city cores. *Journal of Spatial Information Science*. 1, 1 (Jul. 2010), 21–48.

[2] Jones, R. et al. 2008. Geographic intention and modification in web search. *International Journal of Geographical Information Science*. 22, 3 (2008), 229–246.

[3] Van Laere, O. et al. 2011. Finding locations of flickr resources using language models and similarity search. *ICMR-11: Proceedings of the 1st ACM International Conference on Multimedia Retrieval* (Trento, Italy, 2011), 1–8.

[4] Larson, R.R. 1996. Geographic information retrieval and spatial browsing. *Geographic Information Systems and Libraries: Patronsand Mapsand and Spatial Information* (1996), 81–124.

[5] Manning, C.D. et al. 2008. *Introduction to Information Retrieval*. Cambridge University Press.

[6] Moxley, Emily and Kleban, Jim and Manjunath, B.S. 2008. SpiritTagger : A Geo-Aware Tag Suggestion Tool Mined from Flickr. *MIR '08: Proceedings of the 1st ACM international conference on Multimedia information retrieval* (Vancouver, British Columbia, Canada, 2008), 24–30.

[7] Naaman, M. et al. 2004. Automatic organization for digital photographs with geographic coordinates. *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries* (New York, NY, USA, 2004), 53–62.

[8] O'Hare, N. and Murdock, V. 2012. Modeling locations with social media. *Information Retrieval*. (Apr. 2012), 1–33.

[9] Purves, R. et al. 2011. Describing place through user generated content. *First Monday*. 16, 9 (2011).

[10] Purves, R.S. et al. 2010. Automatically generating keywords for georeferenced images. *Proceedings of the GIS Research UK 18th Annual Conference GISRUK 2010* (London, UK, 2010), 203–207.

[11] Rattenbury, T. and Naaman, M. 2009. Methods for extracting place semantics from Flickr tags. *ACM Transactions on the Web*. 3, 1 (Jan. 2009), 1–30.

[12] Sanderson, M. and Kohler, J. 2004. Analyzing Geographic Queries. *Proceedings of the Workshop on Geographic Information Retrieval, SIGIR* (2004).

[13] Serdyukov, P. et al. 2009. Placing flickr photos on a map. *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Boston, MA, USA, 2009), 484–491.

[14] Wing, B.P. and Baldridge, J. 2011. Simple supervised document geolocation with geodesic grids. *ACL'11: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (Portland, Oregon, USA, 2011), 955–964.