Ghent University at the 2010 Placing Task

Olivier Van Laere
Department of Information Technology, IBBT
Ghent University, Belgium
olivier.vanlaere@ugent.be

Steven Schockaert*
Dept. of Applied Mathematics and Computer Science
Ghent University, Belgium
steven.schockaert@ugent.be

Bart Dhoedt
Department of Information Technology, IBBT
Ghent University, Belgium
bart.dhoedt@ugent.be

ABSTRACT

We present a two-step approach to georeferencing tagged resources. First, language models are used to find an area which is likely to contain the location of the resource. In the subsequent second step, the location of the most similar resources in that area are determined.

Categories and Subject Descriptors

I.2 [**ARTIFICIAL INTELLIGENCE**]: Miscellaneous; H.3.7 [**INFORMATION STORAGE AND RETRIEVAL**]: Digital libraries

General Terms

Experimentation

1. INTRODUCTION

Web 2.0 systems such as Flickr describe resources using both structured and unstructured forms of meta-data. In this context, unstructured meta-data mainly takes the form of tags, i.e. short textual descriptions. For resources such as photos or videos, geographic location forms an important type of structured meta-data, which is unfortunately not available for the majority of photos or videos. Recently, there has been an increasing interest in techniques that could automatically estimate the geographic location of photos and videos, by looking only at the tags that users have provided for them [1, 2, 3]. In this paper, we tackle this problem by combining two strategies. The first strategy is to transform this task into a classification problem by clustering the locations of the photos in the training set, and then use a standard language modeling approach to find the cluster that is most likely to contain the actual location of a previously unseen (tagged) resource. The second strategy is based on identifying the photos, from the training set, that are most similar to an unseen resource and using their location as an estimation of its location.

2. GEOREFERENCING RESOURCES

*Postdoctoral Fellow of the Research Foundation – Flanders (FWO).

MediaEval 2010 Workshop, October 24, 2010, Pisa, Italy

Data acquisition and representation.

As training data, we used a collection of 8 685 711 photos, containing the 3 185 343 photos that were provided to participants of the Placing Task, together with an additional crawl of 5 500 368 georeferenced Flickr photos. In addition to the coordinates themselves, Flickr provides information about the accuracy of coordinates as a number between 1 (world-level) and 16 (street level). The locations of these photos were then clustered in a set of disjoint areas $\mathcal{A}$ using the k-medoids algorithm with geodesic distance (using a varying number of clusters k; see below). Subsequently, a vocabulary V consisting of 'interesting' tags is compiled, which are tags that are likely to be indicative of geographic location. We used χ^2 feature selection to determine for each area in $\mathcal{A}$ the m most important tags.

Language models.

For each area $a \in \mathcal{A}$, we write X_a to denote the set of images from our training set that were taken in area a. Given a previously unseen resource x, we try to determine in which area x was most likely taken by comparing its tags with those of the images in the training set. Specifically, using a language modeling approach, the probability of area a, given the tags that are available for resource x is given by

$$P(a|x) \propto P(a) \cdot \prod_{t \in x} P(t|a) \quad (1)$$

The prior probability $P(a)$ of area a is estimated using maximum likelihood, i.e. $P(a) = \frac{|X_a|}{\sum_{b \in \mathcal{A}} |X_b|}$. To obtain a reliable estimate of $P(t|a)$, some form of smoothing is needed. We have experimented with Laplace, Jelinek-Mercer smoothing, and Bayesian smoothing with Dirichlet priors, the latter yielding the best results in general (with Jelink-Mercer producing similar results); see [4] for more details on smoothing.

Similarity search.

Once the area a maximizing the right-hand side of (1) is found, we still need to determine an appropriate location within that area. A basic method would be to use the medoid of that area as the estimation of the location of resource x. However, our results were substantially improved by instead returning the location of the most similar resource within that cluster, where similarity between resources x and y is quantified using the Jaccard measure: $s_{jacc}(x, y) = \frac{|x \cap y|}{|x \cup y|}$ (identifying a resource with its set of tags, before feature selection). Interestingly, using a weighted centre-of-gravity of the k most similar resources did not yield

any improvements for any $k > 1$.

An important question is which resources y to consider. In principle, all resources from cluster a could be considered. Surprisingly, however, our results were substantially improved by only considering those resources whose accuracy level is sufficiently high. In the basic version, the location that is determined for resource x is the location of the most similar resource, among all resources in a with accuracy level 16.

Fallback mechanism.

A problem with the presented approach is to find the right number of clusters. Generally, as long as sufficient training data is available for each cluster, a higher number of clusters should result in more accurate results. However, given a bounded amount of memory, increasing the number of clusters means that less features per area could be retained, which increases the chances that none of the tags of a given resource occurs in the resulting vocabulary. Our solution is to use a fall-back mechanism. In particular, we have used different clusterings, partitioning the training data in respectively 2000, 500, and 50 clusters. After applying feature selection on the three partitionings, we arrive at three vocabularies V_{2000}, V_{500} and V_{50} respectively. As V_{2000} is obtained by retaining a small number of features from a large number of clusters, it contains more specific terms, which are indicative for a narrow location, while V_{50} contains a large number of features from a small number of clusters, which leads to more generic terms that correspond to wider areas, but are more likely to be present in most resources.

Our overall strategy to georeference a given resource x is then as follows. If x contains at least one tag from V_{2000} we use the finest clustering in 2000 areas, identify the most likely area a using (1) and then return the location of the most similar photo in a, whose accuracy level is at least 16; we return the medoid of a if there is no such photo. If $x \cap V_{2000} = \emptyset$ we use the clustering in 500 areas instead, or the clustering in 50 areas if also $x \cap V_{500} = \emptyset$. If $x \cap V_{2000} = x \cap V_{500} = x \cap V_{50} = \emptyset$, we simply determine the most similar resource (with accuracy level 16) in the entire training set.

3. RESULTS AND DISCUSSION

We have experimented with 4 different variants of our approach:

run1 uses Bayesian smoothing with Dirichlet priors and $\mu = 1750$. The number of features that were retained was 175 per area for the 2000-areas clustering, 3200 per area at level 500, and 32000 per area at level 50.

run2 uses Jelinek-Mercer smoothing with $\lambda = 0.8$ and is otherwise identical to run1.

run3 is identical to run1, except that less features are retained: 100 at level 2000, 400 at level 500 and 4000 at level 50.

run4 is identical to run1, except that the location is determined by finding the most similar photo among those with accuracy level at least 14 instead of 16.

The results of the four runs are provided in Table 1. In particular, the table shows how many of the 5091 videos in the test collection were localized within 1km, 5km, 10km, 50km and 100km of the correct location.

	1km	5km	10km	50km	100km
run1	2204	2761	2980	3310	3422
run2	2202	2761	2979	3302	3411
run3	2140	2715	2943	3259	3357
run4	2203	2762	2993	3314	3423

Table 1: Overview of the results on the test collection of 5000 videos.

The results show that our overall approach is rather robust to changes in the parameters involved (number of features, type of smoothing, etc.). In general Dirichlet smoothing performed slightly better than Jelinek-Mercer, but both techniques provided nearly-optimal results for a wide range of parameter values (μ and λ respectively). Regarding the number of features, it seems that the higher this number, the better the results; note that run1 uses the maximum number of features that could be handled on a machine with 8GB of internal memory. Finally, it is not clear whether allowing photos with accuracy levels 14 and 15 when determining the most suitable location within a given cluster is beneficial.

To find plausible locations of videos, our approach only looks at the tags that have been provided. We have experimented with several gazetteers (Geonames, DBpedia, and the US and world sets of USGS/NGA), but have not been able to improve our results. It thus remains unclear whether (or how) such resources could be useful for this task. In addition to gazetteers, other types of information could be taken into account, which we have not examined, including visual features and information about the profile and social network of the corresponding user. Another important avenue for future work is to automatically determine for each photo at which resolution it is best localized; always attempting to assign a precise location, even if the tags that are available are not informative at all, is not likely to be useful in practical applications.

4. REFERENCES

[1] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *Proceedings of WWW*, pages 761–770, 2009.

[2] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In *Proceedings of ACM SIGIR*, pages 484–491, 2009.

[3] O. Van Laere, S. Schockaert, and B. Dhoedt. Towards automated georeferencing of flickr photos. In *Proceedings of the 6th Workshop on Geographic Information Retrieval*, 2010.

[4] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*, 22(2):179–214, 2004.