CEA LIST's Participation at MediaEval 2013 Placing Task

Adrian Popescu
CEA, LIST, Vision & Content Engineering Laboratory, 91190 Gif-sur-Yvette, France.
adrian.popescu@cea.fr

1. ABSTRACT

At MediaEval Placing Task 2013 [2], we focused on improving our last year's participation in four directions: (1) exploit a larger geotagged dataset in order to improve the quality of a standard geolocation language model, (2) model machine tags, (3) estimate the geographicity of tags associated to geolocated photos and (4) exploit user cues in order complement language models whenever these last are likely to fail. Obtained results show that all modifications proposed this year have a positive effect. A "standard" based only on the training data (cues (1)+(2)) has the poorest performance, with P@1km = 0.268, while P@1km = 0.434 when all cues are used.

2. INTRODUCTION

Language models were successfully introduced in [4] as an alternative to gazetteer based geolocation and refined progressively in different editions of MediaEval Placing Task. The best performing state of the art systems combine language models and user modeling [5], [3]. The search space in Placing Task is very wide (the physical word) and it is only partially covered by the training data provided by the organizers [2]. We complemented this training set with external Flickr data, from which we removed all test set items, in order to study the effect of dataset size. If properly modeled, machine tags give very precise information about a photo's location [5] and here we propose a method to exploit them in priority. Geographicity (i.e. the geographic intent) of textual annotations was poorly studied and we the geographicity of individual tags using spatial statistical technique. Finally, user modeling introduces a supplementary constraint since we need to have user data available but this condition is fulfilled for most social networks.

3. LANGUAGE MODELS

Similarly to last year [3], the surface on the Earth is split in (nearly) rectangular in rectangles of size 0.01 of latitude and longitude degree (approximately 1km size). Since the proposed tag modeling is independent of the presence/absence of other tags, we kept only the tags that appear in the test set #5 (262,000 test images), totaling over 155000 tags in order to speed up computation. In order to estimate the effect of rare tags and contrarily to last year, we considered all tags regardless of their user frequency. To mitigate the effects of bulk tagging, cell tag probability is computed as the number of different users that used the tag in the cell divided by the overall tag's user count. We computed models based only on internal training data provided by organizers (around 8.5 million Flickr metadata pieces) [2] but also by adding an external training data set of 90 million items. The first model includes 78897 unique tags and the second contains 128488 unique tags from the test set. The difference with the total number of tags from the test set indicates that an important number of Flickr tags are used by only one user and have little social relevance. Given a test item, we simply sum up contributions of individual tags to find the most probable cell of that item. For each cell, we determine the most probable location by averaging the latitudes and longitudes of photos in that cell and use these coordinates after the detection of the most probable cell.

4. MACHINE TAGS PROCESSING

Machine tags are associated to Flickr data either manually or automatically and some of them give very precise geolocation information. Geotags (latitude and longitude triples) are an obvious type of machine tags that can be exploited. Since they are provided in a modified format (no information about the sign of the coordinate and no decimals), we learned their correlation with real coordinates from the internal and the external training sets. P@1km varies from 0.99 for *foursquare* to 0.97 for *upcoming*. We obtained the following coverage with internal and respectively external models: *foursquare* - 1604, respectively 6031 test items; *geotags* - 10954, respectively 13783 items; *lastfm* - 90, respectively 1347 items; *upcoming* - 292, respectively 955 items. Whenever a photo has associated machine tags, we exploit them instead of standard language models and they are cascaded in descending P@1km scores obtained on a part of the training set.

5. TAG GEOGRAPHICITY

Geographicity is a property that was studied [1] but is still a hot topic, especially for ambiguous and rare tags, which are targeted with our method. The objective here is to find a criterion that separates tags that are well localized from other tags and, consequently, to be able to estimate if a photo can be geolocalized precisely or not. For instance, *Cat* is not spatially discriminant, *Cambridge* is discriminant but is highly ambiguous while *Torre Agbar* is spatially discriminant and appears in a single place. These differences should be reflected by the geographicity score which is calculated by computing the probability of a tag to appear around its most probable cells from the language models. At most 10% of the top cells (i.e. cells with most photos in them) but no more than five are retained as seeds, with a minimum distance of 50 km between two seeds. Then we compute the probability of a tag to appear in a radius of 15 km around all seeds. Several cells are retained in order to deal with ambiguous tags, such as *Cambridge*. The radius is chosen

MediaEval 2013 Workshop October 18-19, 2013, Barcelona, Spain
.

Table 1: Geographicity score vs. geolocation precision in 0.2 intervals.

Decile	0.2	0.4	0.6	0.8	1
Items	28163	23732	22463	16880	122971
P@1km	0.005	0.02	0.051	0.169	0.334
P@10km	0.024	0.071	0.128	0.439	0.716

in order to cover tags whose geographical span goes from very localized to city scale, which are exploitable in order to localized items with city scale precision.

The geographicity score of a tag (*geo*) is defined between 0 (non-discriminant) and 1 (perfectly discriminant). For instance, a photo tagged with *cat* is a priori harder to pin on the map than a photo tagged with *Cambridge*, which is in its turn harder to localize than a photo tagged with *Torre Agbar*. We select 214214 tagged photos from the training set and use the rest of it to create location models. The results presented in table 1 indicate that there is a correlation between geographicity scores and localization precision. Photos whose max geographicity score (*geo* <= 0.2) is small are very hard to localize and, in this case, user cues could be used instead of location models.

Although obtained scores often make sense, we noticed two pitfalls that are probably due the incompleteness and noisy character of Flickr annotations. First, there are some very rare tags whose geographicity score is 1 while they are not geographically discriminant. Second, the proposed approach is not fitted to large entities such as regions or countries since their surface is much larger than the radius chosen to model geographicity. One interesting finding is that around 35% of the test set only contains tags with *geo* <= 0.6. In such cases, accurate geolocation would be difficult regardless of the location models used since there is no precise spatial information associated to the images.

6. USER MODELING

Last year [3], we proposed a simple user modeling that extracted the user's top cell (i.e. the cell including the highest number of user photos). Since test and training sets provided by the organizers don't share users, the modeling was realized with external resources. We have downloaded metadata for each user and, to avoid overfitting, we removed all items whose time stamp is less than 24 hours from any of the test item. Unlike last year, when nearly 25% of data could was placed in the top cell, this year only 4% of user annotations are in the top cell. However, this percentage is much higher than that of photos with *geo* <= 0.3 and, in such cases, user models replace language models. In addition, geographicity is also used in conjunction with temporal metadata. If two images shared by the same user have timestamps within a 24 hours interval and their geographicity score difference is at least 0.2, we transfer coordinates from the item with the larger score to the other.

7. RESULTS AND DISCUSSION

We have submitted the following runs, using a cascade of techniques in the order presented below: **RUN1** - machine tag detection and location models based exclusively on internal training data; **RUN2** - machine tag detection and location models based all training data; **RUN3** - machine tag detection, location models based on all training data, geographicity and user modeling; **RUN4** - RUN3 and the use of temporal cues.

Table 2: P@X precision at X km. err@1 - median error at 1 km.

Run	P@0.1	P@1	P@10	P@100	P@1000	err@1
#1	0.074	0.26	0.43	0.5	0.63	98.8
#2	0.067	0.38	0.58	0.67	0.79	3.45
#3	0.133	0.43	0.62	0.71	0.81	2.07
#4	0.132	0.43	0.63	0.72	0.83	2.08

We present the performances of the different runs in Table 2. The exploitation of decoded geotags, introduced in [5], is debatable since one could claim they can be assimilated to training information. They make for around 4% of the dataset for internal models and 5% for external location models. Without their use, geolocation scores would be reduced by less than 4% and 5% and performances remain interesting with respect to scores reported in past campaigns.

The comparison of RUN1 with the other runs indicates that adding external data to language models has a positive effect on performances. In particular, RUN2 is similar to RUN1 but it exploits a much larger training set. The use of supplementary data results in more robust language models and we hypothesize that adding even more supplementary training data would further improve results. The superior performances of RUN3 indicate that adding user modeling is beneficial since precision is improved at all scales. RUN3 and RUN4 have nearly equal performances up to 10 km precision and the introduction of temporal cues is only useful at larger scales. This result is probably explained by the fact that it is usually improbable for users to move in regions of size greater than tens of kilometers.

We didn't have time to submit visual runs but we plan to implement a two stage approach in which a global feature is used to retrieve a number of similar images and then a geometric check is performed to find images that depict the same object. In function of the performances of visual processing, we will decide about its integration in the geolocation cascade. Currently, rare tags all have high geographicity scores while only a part of them are actually useful. We will study ways to separate useful rare tags from the others in order to improve geolocation precision. Finally, we will build language models that don't include any contributions from test users to evaluate the effect of removing any prior knowledge about the test set.

8. REFERENCES

[1] Z. Cheng and al. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proc. of CIKM 2010*, 2010.

[2] C. Hauff, B. Thomee, and M. Trevisiol. Working Notes for the Placing Task at MediaEval 2013, 2013.

[3] A. Popescu and N. Ballas. Cea list's participation at mediaeval 2012 placing task. In *MediaEval*, 2012.

[4] P. Serdyukov and al. Placing flickr photos on a map. In *Proc. of SIGIR 2009*.

[5] M. Trevisiol and al. Retrieving geo-location of videos with a divide & conquer hierarchical multimodal approach. In *ICMR*, pages 1–8, 2013.