

USEMP at MediaEval Placing Task 2014

Adrian Popescu¹, Symeon Papadopoulos², Ioannis Kompatsiaris²

¹CEA, LIST, 91190 Gif-sur-Yvette, France, adrian.popescu@cea.fr

²CERTH-ITI, Thessaloniki, Greece, {papadop,ikom}@iti.gr

ABSTRACT

We describe the participation of the USEMP team in the Placing Task at MediaEval 2014. We submitted four textual runs which are inspired by CEA LIST's 2013 participation. Our entries are based on probabilistic place modeling but also exploit machine tag and/or user modeling. The best results were obtained when all these types of information are combined. The accuracy of automatic at 1km reaches 0.235 when using only training data provided by organizers and 0.441 with the use of external data.

1. INTRODUCTION

The goal of the task is to produce location estimates for a test set of 500,000 images and videos using a set of approximately five million geotagged images and videos and their metadata for training. A full description of the challenge and of the associated dataset is provided in [1]. Our runs were implemented using, for a large part, methods described in CEA LIST's participation at Placing Task 2013 [2]. For this reason, after a short presentation of the methods, runs and obtained results, we focus on failure analysis.

2. METHOD DESCRIPTION

2.1 Probabilistic location models

Language models are successfully introduced in [3] as an alternative to gazetteer-based geolocation and were progressively improved in following years. Test photos can be placed anywhere in the physical world and the training data provided by the organizers is insufficient in order to build robust probabilistic models. To verify the assumption that better results are obtained with the use of more data, we exploited: (1) all geotagged metadata from the YFCC dataset¹, after removing all test items and (2) an additional set of ~90 million geotagged metadata from Flickr.

Similar to last year [2], the surface of the earth was split in (nearly) rectangular cells of size 0.01 of latitude and longitude degree (approximately $1km^2$ size). User counts were used instead of tag counts in order to mitigate the influence of bulk tagging. Both titles and tags were taken into account and are referred to as tags hereafter. Put simply, we

¹<http://webscope.sandbox.yahoo.com/catalog.php?datatype=i&did=67>

computed the probability of a tag in a cell by dividing its user count in that cell by its total user counts in all cells. Given a test item, we simply summed-up contributions of individual tags to find the most probable cell for that item. Finally, the photo was placed at the center of the cell.

2.2 Machine tag modeling

The authors of [4] show that machine tags can improve automatic geotagging quality. In [2] we propose a machine tag processing method which models only machine tags which are strongly associated to locations. (i.e. Foursquare, Lastfm and Upcoming entries) and we exploited it this year.

2.3 User modeling

If images do not have associated tags or if these tags are not geographically discriminant, placing photos with probabilistic models is likely to fail. To overcome this problem, we exploited a simple user modeling technique [2], which computes the most probable cell of a user. Only photos which are at least 24 hours away from any of the user's test set images were exploited to reduce the risk of learning from test data. We downloaded up to 500 geotagged images per user in order to determine her most probable cell.

2.4 Fusion

We propose a late fusion scheme which is empirically derived from tests with a validation dataset. Since they are associated to precise locations or geolocated events, processed machine tags are very reliable and were used in priority. If there were no machine tags, location models were exploited to predict the most probable location of a set of tags. Finally, if there were no tags available or if the prediction score was below a threshold, the photo was placed in the most probable cell of the user who uploaded it. The threshold for replacing location models with user models was empirically determined on the validation dataset. We exploited user models for the 30% of test images which had the lowest placing scores.

3. RUNS

We submitted the following runs: RUN_1 - exploited location models and machine tags from training data provided by the organizers; RUN_3 - combined location models and machine tags from the entire geotagged YFCC dataset, after excluding test items; RUN_4 exploited tags and user models; RUN_5 - exploited YFCC location models, machine tags and user models. We present the performance of the submitted runs in Table 1. The best results were obtained when com-

| Run | P@X km | | | | | |
|-----|--------|-------|-------|-------|-------|-------|
| | 0.01 | 0.1 | 1 | 10 | 100 | 1000 |
| #1 | 0.007 | 0.016 | 0.235 | 0.408 | 0.481 | 0.618 |
| #3 | 0.026 | 0.043 | 0.428 | 0.582 | 0.644 | 0.753 |
| #4 | 0 | 0.012 | 0.418 | 0.597 | 0.679 | 0.779 |
| #5 | 0.026 | 0.043 | 0.441 | 0.613 | 0.691 | 0.787 |

Table 1: P@X precision at X km.

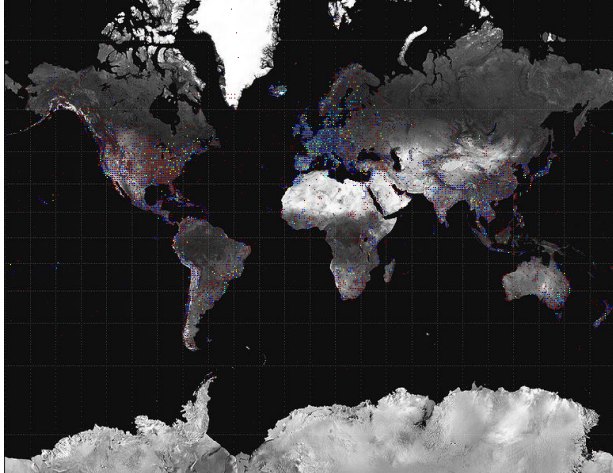


Figure 1: Average error plot for RUN_1 . Red/blue dots correspond to large errors/precise geotagging respectively.

binning all types of available information. As expected, the largest contribution was due to location models. The large gap between RUN_1 and the others confirms that the use of supplementary training data is very beneficial. The difference of precision at close range (P@0.1) between RUN_3 and RUN_4 confirms that machine tags are very useful for precise geolocation. Inversely, if larger errors are admitted, user models become more useful than machine tags. The combination of these types of cues in RUN_5 gives the best performance for all precision ranges. The results obtained this year are in the same range as those we reported in 2013 [2], confirming thus that our geolocation pipeline has consistent behavior over different datasets.

4. FAILURE ANALYSIS

In addition to the submitted runs, we tested other configurations which gave lower results and briefly describe them here. We notably tried a combination of location models and gazetteer information in order to give a privileged role to toponyms such as administrative division names (i.e. countries, regions, cities). The addition of the gazetteer gave lower results compared to the sole use of location models. This negative result could be explained by the strong ambiguity which characterizes the geographic domain. As we mentioned, we also tried to add a dataset of ~ 90 million geotagged metadata to the YFCC full training data. Contrarily to existing literature [4, 2], the use of this supplementary dataset actually degraded the overall quality of results. This negative result might indicate that probabilistic models reach saturation when too much metadata are available.

In Figures 1 and 2, we present a visualization of geotagging performance for RUN_1 and RUN_5 and the performance

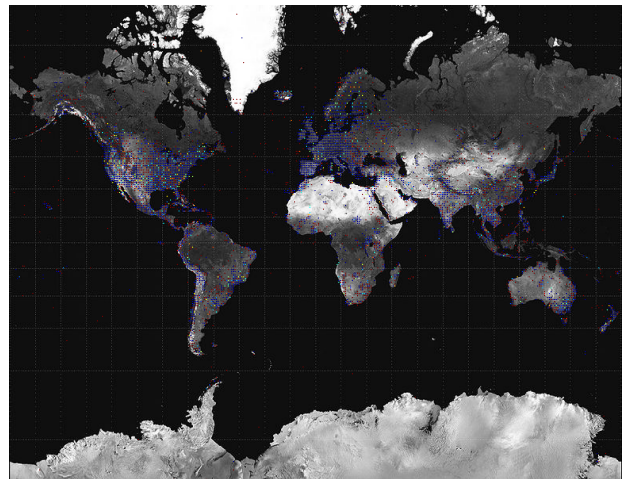


Figure 2: Average error plot for RUN_5 . Red/blue dots correspond to large errors/precise geotagging respectively.

difference between the two runs is clearly reflected. Geotagging is precise in most European regions and worse for the other regions. Low performances can be easily explained by sparse data for Africa, Asia or South America. However, the imprecision is also high for the United States, the region of the world which concentrates the largest number of geotagged images. In this case, poor geotagging could be due to a very high ambiguity of place names. For instance, there are dozens of places called *London* or *Paris* in the US. If there is not enough disambiguation information associated to them in annotations, photos tagged with these toponyms will be placed in Europe.

5. FUTURE WORK

Due to lack of time, we did not submit a visual run this year. While visual geotagging still lags well behind textual geotagging, it would be interesting to explore if it is possible to predict coordinates accurately at least for visually distinctive objects such as Points of Interest. Regarding text models, we would like to investigate in more depth why adding more data from outside YFCC degrades performance. Equally interesting, it would be interesting to investigate ways to select reliable annotations before computing location models.

6. ACKNOWLEDGMENT

This work is supported by the USEMP FP7 project, partly funded by the EC under contract number 611596.

7. REFERENCES

- [1] J. Choi and al. The placing task: A large-scale geo-estimation challenge for social-media videos and images. In *Proc. of GeoMM'14*.
- [2] A. Popescu. Cea list's participation at mediaeval 2013 placing task. In *MediaEval*, 2013.
- [3] P. Serdyukov and al. Placing flickr photos on a map. In *Proc. of SIGIR 2009*.
- [4] M. Trevisiol and al. Retrieving geo-location of videos with a divide & conquer hierarchical multimodal approach. In *ICMR*, pages 1–8, 2013.