TALP-UPC at MediaEval 2014 Placing Task: Combining Geographical Knowledge Bases and Language Models for Large-Scale Textual Georeferencing

Daniel Ferrés, Horacio Rodríguez
TALP Research Center
Computer Science Department
Universitat Politècnica de Catalunya
{*dferres,horacio*}*@cs.upc.edu*

ABSTRACT

This paper describes our Georeferencing approaches, experiments, and results at the MediaEval 2014 Placing Task evaluation. The task consists of predicting the most probable geographical coordinates of Flickr images and videos using its visual, audio and metadata associated features. Our approaches used only Flickr users textual metadata annotations and tagsets. We used four approaches for this task: 1) an approach based on Geographical Knowledge Bases (GeoKB), 2) the Hiemstra Language Model (HLM) approach with Re-Ranking, 3) a combination of the GeoKB and the HLM (GeoFusion). 4) a combination of the GeoFusion with a HLM model derived from the English Wikipedia georeferenced pages. The HLM approach with Re-Ranking showed the best performance within 10m to 1km distances. The GeoFusion approaches achieved the best results within the margin of errors from 10km to 5000km.

1. INTRODUCTION

The MediaEval 2014 Placing task requires that participants use systems that automatically assign geographical coordinates (latitude and longitude) to Flickr photos and videos using one or more of the following data: Flickr metadata, visual content, audio content, and social information (see [1] for more details about this evaluation). The Placing Task training data consists of 5,000,000 geotagged photos and 25,000 geotagged videos, and the test data consists of 500,000 photos and 10,000 videos. Evaluation of results is done by calculating the distance from the actual point (assigned by a Flickr user) to the predicted point (assigned by a participant). Runs are evaluated finding how many videos were placed at least within some threshold distances.

2. SYSTEM DESCRIPTION

We used four approaches for the MediaEval 2014 Placing Task (see more details about the approaches in [3]):

1) Geographical Knowledge Bases (GeoKB). We used this approach in MediaEval 2010 and 2011 Placing Tasks [2] [4] after some improvements (see [3]).The GeoKB approach uses the Geonames[1] Gazetteer for detecting the place names, stopwords lists, and an English Dictionary. The system uses the following rules from Toponym Disambiguation techniques [3] to get the geographical focus of the photo/video: 1) select the most populated place that is not a state, country or continent and has its state apearing in the text, 2) otherwise select the most populated place that is not a state, country or continent and has its country apearing in the text, 3) otherwise select the most populated state that has its country apearing in the text 4) otherwise apply population heuristics.

2) Hiemstra Language Model (HLM) with Re-Ranking. This approach uses the Terrier[2] Information Retrieval (IR) software (version 3.0) with the HLM weighting model [5]. The HLM default lambda (λ) parameter value in Terrier (0.15) was used. See in [3] more details about the Terrier implementation of the HLM weighting model (version 1 [5]). The indexing of the metadata subsets were done with the coordinates as a document number and some metadata fields (Title, Description and User Tags) as the document text. For each unique coordinate a document was created with all the textual metadata fields content of all the photos/videos that pertain to this coordinate.

The indexing process uses a multilingual stopwords list to filter the tokens that are indexed. The following metadata fields (lowercased) from the photos/videos were used for the query: User tags, Title and Description. A Re-Ranking process is applied after the IR process. For each topic their first 1000 retrieved coordinates pairs from the IR software are used. From them we selected the subset of coordinates pairs with a score equal or greater than the two-thirds (66.66%) of the score of the coordinates pair ranked in first position. Then for each geographical coordinates pair of the subset we sum its associated score (provided by the IR software) and the score of their neighbours at a threshold distance (e.g. 100km). Then we select the one with the maximum weighted sum.

3) GeoFusion: Hiemstra Language Model with Re-Ranking and GeoKB. This approach is applied by combining the results of the GeoKB approach and the IR approach with Re-Ranking. From the GeoKB system are selected the predicted coordinates that come only from the heuristics 1, 2 and 3 (avoiding predictions from the population heuristics rules). When the GeoKB rules (applied in priority order: 1, 2, and 3) do not match then the predictions are selected from the HLM approach with Re-Ranking

[1] **Geonames (downloaded in 2011).** `http://www.geonames.org`

[2] **Terrier.** `http://terrier.org`

MediaEval 2014 Workshop, October 16-17, 2014, Barcelona, Spain

4) GeoFusion+GeoWiki: GeoFusion combined with a HLM model of Georeferenced Wikipedia pages. This is the only improvement with respect to the system used at MediaEval 2011. This approach uses a set of 857,574 Wikipedia georeferenced pages[3] that were indexed with Terrier. The coordinates of the top ranked georeferenced Wikipedia page are used as a prediction. The predictions from the georeferenced Wikipedia based HLM model are used only in case that the HLM model with Re-Ranking based on the Training data gives an score lower than 7.0. This threshold was found empirically training with the MediaEval 2011 test set. The system uses the coordinates of one of the most photographied places in the world as a prediction when the approaches cannot give a prediction.

3. EXPERIMENTS AND RESULTS

We designed a set of four experiments for the MediaEval 2014 Placing Task (Main Task) test set of 510,000 Flickr photos and videos (see results in Figure 1 and Table 1):

1. The experiment *run1* used the HLM approach with Re-Ranking up to 100 km and the MediaEval 2014 training set metadata as a training data. From a set of 5,050,000 photos and videos of the MediaEval 2014 training set, a set of 3,057,718 coordinates pairs with related metadata info were created as textual documents and then indexed with Terrier.

2. The experiment *run3* used the GeoKB approach.

3. The experiment *run4* used the GeoFusion approach with the MediaEval training corpora.

4. The experiment *run5* used the GeoFusion approach with the MediaEval training corpora in combination with the English Wikipedia georeferenced pages HLM model.

Table 1: Percentage of correctly georeferenced photos/videos within certain amount of kilometers and median error for each run.

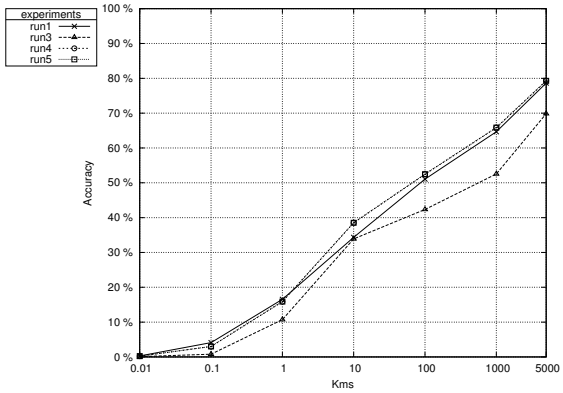
Margin	run1	run3	run4	run5
10m	**0.29**	0.08	0.23	0.23
100m	**4.12**	0.80	3.00	3.00
1km	**16.54**	10.71	15.90	15.90
10km	34.34	33.89	38.52	**38.53**
100km	51.06	42.35	**52.47**	**52.47**
1000km	64.67	52.54	**65.87**	65.86
5000km	78.63	69.84	**79.29**	79.28
Median Error (kms)	83.98	602.21	64.36	64.41

4. CONCLUSIONS

We used four approaches at MediaEval 2014 Placing Task. The GeoFusion approaches achieved the best results in the experiments clearly outperforming the other approaches. These approaches achieve the best results because combine high precision rules based on Toponym Disambiguation heuristics and predictions that come from an HLM models. The GeoKB rules used in the GeoFusion approach achieved 81.17% of accuracy (131,207 of 161,628 photos/videos) predicting up to 100km. The most difficult cases for prediction with our textual based approach are the ones with few textual information and tags. In this evaluation we tried an approach that uses sometimes the English Wikipedia Georefenced pages to handle these cases. The GeoFusion+GeoWiki approach (that uses an HLM model of English Wikipedia georeferenced pages) does not generally offers better performance than the original GeoFusion approach. This approach only improved very slightly the results for estimations at 10km. The HLM approach with Re-Ranking obtained the best results in the 10m to 1km range because the model takes some benefits of relating non-geographical descriptive keywords and place names appearing in the geographical coordinates' associated metadata.

[3] http://de.wikipedia.org/wiki/Wikipedia:WikiProjekt_Georeferenzierung/Hauptseite/Wikipedia-World/en

Figure 1: Accuracy against margin of error in kms



Acknowledgments

This work has been supported by the Spanish Research Department (SKATER Project: TIN2012-38584-C06-01). TALP Research Center is recognized as a Quality Research Group (2014 SGR 1338) by AGAUR, the Research Department of the Catalan Government.

5. REFERENCES

[1] J. Choi, B. Thomee, G. Friedland, L. Cao, K. Ni, D. Borth, B. Elizalde, L. Gottlieb, C. Carrano, R. Pearce, and D. Poland. The Placing Task: A Large-Scale Geo-Estimation Challenge for Social-Media Videos and Images. In *Proceedings of the 3rd ACM International Workshop on Geotagging and Its Applications in Multimedia*, 2014.

[2] Daniel Ferrés and Horacio Rodríguez. TALP at MediaEval 2010 Placing Task: Geographical Focus Detection of Flickr Textual Annotations. In *Working Notes of the MediaEval 2010 Workshop*, Pisa, Italy, October 2010.

[3] Daniel Ferrés and Horacio Rodríguez. Georeferencing Textual Annotations and Tagsets with Geographical Knowledge and Language Models. In *Actas de la SEPLN 2011*, Huelva, Spain, September 2011.

[4] D. Ferrés and H. Rodríguez. TALP at MediaEval 2011 Placing Task: Georeferencing Flickr Videos with Geographical Knowledge and Information Retrieval. In *Working Notes Proceedings of the MediaEval 2011 Workshop, Santa Croce in Fossabanda, Pisa, Italy, September 1-2, 2011*, 2011.

[5] D. Hiemstra. *Using Language Models for Information Retrieval*. PhD thesis, Enschede, Netherlands, January 2001.