Geo_ML @ MediaEval Placing Task 2015

Nghia Duong-Trung, Martin Wistuba, Lucas Rego Drumond, Lars Schmidt-Thieme
Information Systems and Machine Learning Lab
University of Hildesheim, Germany
{duongn,wistuba,ldrumond,schmidt-thieme}@ismll.uni-hildesheim.de

ABSTRACT

We participated in the MediaEval Benchmarking whose goal is to concentrate on the multimodal geo-location prediction on the Yahoo! Flickr Creative Commons 100M dataset - the placing task. It challenges participants to develop models and/or techniques to estimate the geographic locations of the Flickr resources based on textual metadata, e.g. titles, descriptions and tags. We aim to find a procedure that is conceptual to understand, simple to implement and flexible to integrate different techniques. In this paper, we present a three-step approach to tackle the locale-based sub-task.

1. INTRODUCTION

The placing task is the challenge offered by the MediaEval Multimedia Benchmarking [1] Initiative that proposes motivations for working with geotagged applications and solutions [2]. The task focuses on the development of models to predict the geo-location, i.e. the latitude and longitude, of multimedia items based on their metadata and/or visual features. Estimating the geo-location accurately will enable us to provide a wide range of applications such as geo-aware recommendations and targeted advertisements.

The Yahoo Flickr Creative Commons 100 Million Dataset[1] (YFCC100M) which is the largest public multimedia collection contains a total of 100 million photos and videos captured over 10 years [8]. Under the umbrella of geo-location prediction, we focus on the locale-based placing task which aims to estimate the geographic coordinates of a given photo/video. This year's task dataset is based on a subset of the YFCC100M. The training data consists of 4,695,149 items, while the test set contains 949,889 items. The challenge baseline is described in [6].

In this paper we exploit the availability and plurality of textual metadata, especially the titles, users tags, machine tags and descriptions to develop our three-step approach: (i) K-means clustering of multimedia items by their latitude and longitude coordinates; (ii) learning a linear support vector machine on textual contents to predict cluster membership; and (iii) exploiting a K-nearest neighbor regression to find the closest item in the same predicted cluster and return its geo-location as prediction. The theoretical purposes why we split our system into 3 steps are discussed in section 3.2. Moreover, we discuss what has been learned in comparison with the baseline in section 4.

[1]http://bit.ly/yfcc100md

MediaEval 2015 Workshop, Sept. 14-15, 2015, Wurzen, Germany

2. TASK DESCRIPTION

We have m, v geotagged multimedia items in the training and test data respectively, and n features describing each item. These features are drawn from textual metadata. Each item is annotated with a geo-location $\mathbf{y} \in \mathbb{R}^2$, $\mathbf{y} = (y^{lat}, y^{lon})$ where $y^{lat} \in \mathbb{R}$ is the latitude and $y^{lon} \in \mathbb{R}$ is the longitude. Given some training data $X^{train} \in \mathbb{R}^{m \times n}$, and the respective labels $Y^{train} \in \mathbb{R}^{m \times 2}$, we aim to find a model $f : \mathbb{R}^n \rightarrow \mathbb{R}^2$ such that for some test data $X^{test} \in \mathbb{R}^{v \times n}$, the error $\sum_{i=1}^{v} d(f(X_i^{test}), Y_i^{test})$ is minimal, where $Y^{test} \in \mathbb{R}^{v \times 2}$ is the true geo-location matrix and d is the Karney distance [5].

3. PROPOSED APPROACH AND RESULTS

In this section, we discuss the data preprocessing techniques we employed. Then, we present our proposed three-step approach.

3.1 Data preprocessing

Before feeding the dataset to our three-step approach, we pre-processed the data as follows. All given metadata description was converted into a bag of words representation, consisting of all words/unigrams that mutually appear in both training and test set. Then, term frequency - inverse document frequency features were computed to reflect how important a word is to a description in a collection. The features with low-variance were discarded. The number of features after data preprocessing is 20,000.

3.2 Proposed approach

The following part is the paper's main contribution. We simultaneously explain the theoretical purposes and describe how our three-step approach works for the aim of finding the f model mentioned in section 2. We devised a three-step procedure.

1. **K-means clustering**. The target geo-location $\mathbf{y}$ consists of two labels y^{lat} and y^{lon}. The basic idea in the first step is to transform a multi-target prediction task into a multi-class classification task. The idea of an equally squared grid is not applicable since geographic coordinates of items are spread all over the world. In order to find regions of interest we cluster the items on the training set using K-means [3]. At the end of this step, we have a cluster assignment vector $\mathbf{c} \in \mathcal{C}^m$,

where the i-th element c_i contains the cluster assigned to the i-th instance based on its geo-location $\mathbf{y}_i$.

2. **Linear support vector machine**. Now that we have identified clusters, we need to learn a model on X^{train} and $\mathbf{c}$ in order to map the test instances to those clusters. For that reason, we use a classifier which has $\mathbf{c}$ as the target and $\mathcal{X}^{train}$ as the training domain. From now on, the task of geo-location prediction can be treated as a multi-class classification problem. The dataset associated with corresponding clusters $\mathbf{c}$ is trained by the linear SVM $g : \mathbb{R}^n \rightarrow c$ with L2 regularization [4].

3. **K-nearest neighbor regression**. Once we have estimated to which cluster c_i a test instance X_i^{test} should belong to, its predicted geo-location $\hat{\mathbf{y}}_i^{test}$ is that of the nearest neighbor in the same cluster $g(\mathcal{X}_i^{test})$. The coordinates of X_i^{test} are predicted using 1-NN regression [7] on all the training instances belonging to $g(X_i^{test})$.

The evaluation metric is the median Karney distance $d : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}_+$ between the actual $\mathbf{y}_i$ and predicted locations $\hat{\mathbf{y}}_i$. We apply grid search to find the best value combination of all hyperparameters that minimize the distance error $\sum_{i=1}^{v} d(f(X_i^{test}), Y_i^{test})$. At the end of the evaluation, we have the number of clusters $k = 1000$, and the cost $s = 0.01$ for the linear SVM. Those aforementioned steps yield the pseudocode below.

Algorithm 1 Three-step approach

INPUT: X^{train}, X^{test}, Y^{train}, cost s, number of clusters k

```
# Step 1: k-means clustering
c ← Kmeans(Y^train, k)
# Step 2: Linear SVM
g ← LinearSVM(X^train, s, c)
# Step 3: k-nearest neighbor
for i = 1 ... v do
    c_i ← g(X_i^test)
    X, Y ← rows of X^train, Y^train belonging to cluster c_i
    ŷ_i ← 1NNRegression(X, Y)
end for
return ŷ_i
```

4. EXPERIMENTAL RESULTS

Our implementation achieves a median error of 352.47 km to the test set. The baseline median error is 71.45 km. In table 1, we present our evaluation results in more details. To compare what has been done with the baseline, we only apply K-means on Y^{train} without any textual knowledge or language models. We also do not apply feature ranking. Those issues will lead to further improvement and we would like to discuss it in section 5.

5. CONCLUSION AND OUTLOOK

We have presented our three-step approach to the geo-location prediction problem based on only textual metadata without exploiting any language models and topic discovery to investigate how reliable and robust this approach actually is. We have split the geo-location prediction into a sequence of conceptual steps. This architecture enables improvement in future experiments. We can easily replace and integrate new techniques in the workflow without redesigning the complete system. For example, we can replace K-means clustering by K-medoids clustering or mean-shift clustering. In addition, we can also apply feature selection or dimension reduction on X^{train} before feeding it into Step 2.

distance	# items	percentage
0.001 km	504	0.05 %
0.01 km	1051	0.11 %
1 km	11849	1.25 %
10 km	287807	30.03 %
100 km	418831	44.09 %
1000 km	566791	59.67 %
10000 km	911364	95.94 %
40000 km	949889	100.00 %

Table 1: Details of our challenge submission. With the median error of 352.47 km, we are at the 4th position over all participants in the leaderboard.

6. ACKNOWLEDGMENTS

We would like to thank the MediaEval organizers for their baseline code and instructions[2]. Nghia Duong-Trung is sponsored by a grant from Ministry of Education and Training (MOET) of Vietnam under the national project no. 911.

7. REFERENCES

[1] Jaeyoung Choi, Claudia Hauff, Olivier Van Laere, and Bart Thomee, *The placing task at mediaeval 2015*, (2015).

[2] Jaeyoung Choi, Bart Thomee, Gerald Friedland, Liangliang Cao, Karl Ni, Damian Borth, Benjamin Elizalde, Luke Gottlieb, Carmen Carrano, Roger Pearce, et al., *The placing task: A large-scale geo-estimation challenge for social-media videos and images*, Proceedings of the 3rd ACM Multimedia Workshop on Geotagging and Its Applications in Multimedia, ACM, 2014, pp. 27–31.

[3] John A Hartigan and Manchek A Wong, *Algorithm as 136: A k-means clustering algorithm*, Applied statistics (1979), 100–108.

[4] Thorsten Joachims, *Text categorization with support vector machines: Learning with many relevant features*, Springer, 1998.

[5] Charles FF Karney, *Algorithms for geodesics*, Journal of Geodesy **87** (2013), no. 1, 43–55.

[6] Olivier Van Laere, Steven Schockaert, and Bart Dhoedt, *Georeferencing flickr resources based on textual meta-data*, Information Sciences **238** (2013), 52 – 74.

[7] Leif E Peterson, *K-nearest neighbor*, Scholarpedia **4** (2009), no. 2, 1883.

[8] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li, *The new data and new challenges in multimedia research*, arXiv preprint arXiv:1503.01817 (2015).

[2]https://github.com/ovlaere/placing-text/tree/mediaeval2015