VIT@MediaEval 2013 Placing Task : Location Specific Tag Weighting for Language Model Based Placing of Images

Sandeep Subramanian, Vivek Vidyasagaran
School of Computer Science and Engineering
VIT University
Vellore, India
{sandeep.subramanian,vivek.v.sagar}@gmail.com

Krishna Chandramouli
Division of Enterprise and Cloud Computing
School of Information Technology and Engineering
VIT University
Vellore, India
krishna.c@vit.ac.in

ABSTRACT

This paper describes our participation in the Placing Task at MediaEval 2013. The goal of the task is to predict the geographical coordinates of a set of images using already geo-tagged user annotated ones. Our approach to solve this problem relies purely on textual metadata present in the geo-tagged images to place images. We used a frequency based filter followed by a tag spread measure as our feature selection technique. The approach produced 26% accuracy within a 500km radius.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Placing, Geo-Spread, Language Model

1. INTRODUCTION

Pervasive growth in the amount of multimedia content on the internet especially in the form of pictures and videos has opened new avenues for research. Social media analytics has gained recent interest amongst researchers. Social media websites such as Flickr[1], Instagram[2], Twitter[3] and Facebook[4] are great resources for obtaining user annotated multimedia content. This information can be put to use in several domains and one such domain is the placing task [2] which is concerned with automatically assigning geographical coordinates to an image using textual metadata in the form of image tags and visual features associated with it.

Our model is an attempted improvement to the solution proposed by [1] at the 2011 Placing Task. We sought to improve the term filtering method adopted by them. Specifically, we incorporated new parameters into the calculation of their tag geopsread measure. We calculated geospread measures for tags that are specific to an area on the globe.

[1]www.flickr.com
[2]www.instagram.com
[3]www.twitter.com
[4]www.facebook.com

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

2. MODEL OVERVIEW

8.5 million training images obtained from Flickr were provided to us for this task. The images contained user annotated textual metadata along with visual features. The ground truth location for every image was also supplied. To start, we looked to divide the world into grids [4] to able to spread the images across the globe based on their ground truth coordinates. We placed them in grids of varying sizes, recursively subdividing grids starting from a base grid that spanned the entire globe to accommodate a maximum of 10,000 images per grid. The threshold value for the number of images per grid was determined after experimentation to ease computation without compromising on placing accuracy. A grid, on exceeding its maximum capacity was split into four equally sized sub-grids. This resulted in 3118 unique grids across which the images were distributed. Once every image was assigned a gird, a bag of words language model was generated for each grid which would serve as an accurate identifier for a grid. Identifying if an image belongs to a particular grid could then be done by searching for the best possible match score grid_{max} between the image tags and the language models of the generated grids. Inside a particular grid however, to pin point a particular image that could be the closest match, we perform a tag wise comparison of images with every image in the gird. The image with the best match score img_{max} is deemed the closest match. The image is assigned the same geographical coordinates as that of matched image.

In the unlikely event that none of the tags in the image are present in any of the language models constructed, or formally when $\text{grid}_{max} = 0$, the image is placed in the location of a random image from a randomly selected grid.

3. FEATURE SELECTION

Feature selection techniques effectively reduce the magnitude of data being dealt with for efficient computation and at the same time improve accuracy by filtering out noisy information. It is therefore imperative in a problem such as this to have a robust feature selection technique. Our approach to feature selection solves the tradeoff between the computation required to place images versus the time it takes to train the model itself by investing a large chunk of computation in training the model. We used a two tier feature selection technique, initially taking only the 20 most frequently occurring tags from each grid. A list of the most frequent tags in a grid located in the close proximity of London is shown in Table 1. This was an attempt at filtering out noisy and irrelevant tags. The threshold of 20 tags was chosen solely for simplicity of computation and often 20 tags is insufficient to represent a grid entirely. We then computed a tag spread measure for the selected tags specific to the grid. The motivation behind a grid specific tag spread measure is to incorporate a degree of association between a tag and a grid. Our estimate borrows from conventional feature selection techniques such as mutual information and χ^2 [3] but produces more meaningful measures of geographical spread. With N_{11} representing the number of occurrences of a tag in a grid, N_{10} representing the number of tag in other grids, N_{01} representing the number of other tags in the grid and N_{00} representing the number of other tags in other grids, our baseline spread measure can be defined as follows:

$$(N_{11} + N_{01} + N_{10})/N_{11} \tag{1}$$

The N_{00} term is excluded from calculations. Additionally, we penalized occurrences of a particular tag outside a grid by multiplying N_{10} by a factor $D_{g1,g2}$ directly proportional to the distance between the two grids. This factor can be computed as α * $D_{g1,g2}$ where α is a constant assigned manually and $D_{g1,g2}$ is the distance between the two grids. Since all ground truth estimates have been provided using Mercator Projections, it is possible to calculate the distance between the centers of two grids by using the Euclidean distance measure. The improved tag spread measure is defined as:

$$(N_{11} + N_{01} + \sum(N_{10(g2)} * \alpha * D_{g1,g2}))/N_{11} \tag{2}$$

The summation is not computed on grids where the particular tag is not present. Spread measures for various tags have been shown in Table 1.

S.No	Tag	Spread
1	london	14.201
2	england	60.265
3	crouchend	68.037
4	permaculture	71.633
5	northlondon	75.836
6	garden	119.134
7	meadoworchardproject	78.466
8	community	95.175
9	highgate	91.592
10	live	161.816
11	mop	132.123
12	music	288.666
13	snow	249.955
14	gig	210.775
15	dirtywaterclub	198.154
16	rock	288.407
17	uk	217.601
18	gardening	218.916
19	hampsteadheath	211.820
20	hampstead	218.379

Table 1: Tags and their geographic spread scores in decreasing order of frequency of occurrence in the grid

Tag	Spread(Grid 1)	Spread(Grid 2)	Spread(Grid 3)
Wedding	372.141	96.916	629.671
Film	287.741	259.562	390.104
Building	426.145	286.761	1162.522
Colorado	10.755	17.472	11.2436
Barcelona	45.632	192.216	14.145
Chicago	8.222	7.828	10.808

Table 2: Sample geographic spread scores for geographically significant and geographically insignificant tags in 3 arbitrarily selected unique grids

4. RESULTS

Results of our run on test data set #1 containing 5300 images have been detailed in Table 2. for a single baseline approach. Accuracies within 1, 10, 50,100,500 and 1000km of the true geo-coordinates have been specified. The reported median error in placing is 6168.307 Kilometers.

1km	10km	100km	500km	1000km
0.74%	3.9%	15.24%	26.3%	30.14%

Table 3: Placing accuracy at different distances

5. CONCLUSION

The primary motivation behind our approach of incorporating a degree of association between a grid and tag showed good consistency with geographically significant tags receiving low spread measures and noisy tags a higher spread value. As expected, the spread measures showed significant variations in some cases across different grids as indicated clearly by the tag 'building' in Table 1. However, selecting the 20 most frequently occurring tags as a primary feature selection technique could be modified to supplying a hard cut-off value for the frequency of tags to avoid noise during selection. The computation of spread values was very computationally intensive and therefore, we managed to use only roughly 50% of training dataset. It took us roughly 5 days of non-stop computation to compute the spread values for half the training dataset. We hope to run the model using the entire training dataset and study its performance in detail. We also hope to use visual features coupled with textual metadata to aid in pin pointing images within a grid. 4

6. REFERENCES

[1] C Hauff and G Houben. WISTUD at MediaEval 2011: Placing Task. In *MediaEval 2011* Workshop, 2011.

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

[3] O. Van Laere, S. Schockaert, and B. Dhoedt. Finding locations of flickr resources using language models and similarity search. In *ICMR '11* , pages 48:1-48:8, 2011.

[4] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In SIGIR '09, pages 484-491, 2009.