Investigating summarization techniques for geo-tagged image indexing

Ahmet Aker[1], Xin Fan[2], Mark Sanderson[3] and Robert Gaizauskas[1]

[1] University of Sheffield, UK
[2] Yahoo! Labs Beijing
[3] RMIT University, Australia

Abstract. Images with geo-tagging information are increasingly available on the Web. However, such images need to be annotated with additional textual information if they are to be retrievable, since users do not search by geo-coordinates. We propose to automatically generate such textual information by (1) generating toponyms from the geo-tagging information (2) retrieving Web documents using toponyms as queries (3) summarizing the retrieved documents. The summaries are then used to index the images. In this paper we investigate how various summarization techniques affect image retrieval performance and show significant improvements can be obtained when using the summaries for indexing.

1 Introduction

Many images on the Web are geo-tagged, i.e. assigned with longitude and latitude coordinates. However, this geographical information is of limited usefulness for indexing and retrieval; for this purpose images must be annotated – usually manually – with textual information.

However, even when manually annotated, most images are only annotated with place name which limits retrieval to cases where the search query contains the place name. Sanderson and Kohler [4] analyze geographic search engine queries. They show that more than 40% of the queries contained, in addition to the place name, more general geographic terms which were used to get information about e.g. the "location" or "surrounding" of the place. In [2] it has been also shown that humans appear to have a conceptual model of what is salient regarding geographically situated objects of similar type (e.g. mountains, lakes, etc.) and that they use this model to seek information about such objects. We hypothesize that including the information humans commonly associate with geographic objects of a specific type in a richer caption for indexing purposes could lead to better retrieval.

In this paper we propose to automatically generate richer captions using the geographical information only. The caption generation is realized through summarizing multiple web-documents that contain information related to an image's location. Our summarizer uses "conceptual models" [1] to generate the descriptions. The conceptual models for different object types are learned from collections of texts about specific objects types drawn from Wikipedia. We use the summaries to index geo-tagged images and assess their performance using image retrieval effectiveness. As baseline indexes we use the titles and tags provided with the images and summaries generated by surface-level methods. The results show

that the generated model based summaries in combination with existing textual information lead to significant improvement in the retrieval effectiveness.

2 Experiment

2.1 Data

A set of 6,385 geo-tagged photos from Flickr.com were downloaded together with human-written titles and tags for the evaluations. The image coordinates were within ten National Parks and four big cities in the UK.

2.2 Obtaining Location Information

In [3] automatic ways of deriving location name and scene type information using geo-referencing data are described. For example, from the geo-coordinates <lat: 51,499532> and <long:-0.12913> we derive the location name <Westminster Abbey> and the type <heritage site>. We adopted this method to generate location information for our image set.

2.3 Summary Generation

The image captions are generated from a set of 10 top-ranked documents retrieved from the Web using the location name as query and the Yahoo! search engine. The summarizer is given three inputs: a query (location name), the scene type associated with an image and the retrieved web-documents. We score the sentences in five separate ways using a different feature in each case[1]. For each scored list we start from the top and select sentences into the summary until the limit of 200 words is reached. The features are:

- ***qSim***: Cosine sentence similarity to the query.
- ***cenSim***: Cosine sentence similarity to the 100 most frequently occurring non stop words in the 10 web-documents.
- ***senPos***: Position of the sentence within its document. The first sentence in the document gets the score 1 and the last one gets $\frac{1}{n}$ where n is the number of sentences in the document.
- ***isStarter***: A sentence gets a binary score if it starts with the query or with the object type. We also allow gaps between *the* and the query or object type to capture cases such as *The most magnificent abbey* - where "abbey" is the object type.
- ***conceptModel***: We use n-gram language models to create object type conceptual models. We trained our language models on Wikipedia articles about locations of the same object type.

2.4 Evaluation

A prototype text-based image search engine was implemented using Solr. Each image in the 6,385 image set has the following available fields for indexing and searching: generated summary and Flickr text (including title and tags). We used eight different configurations of indexes to produce image search results:

- **S1**: summary generated using the feature *senPos*
- **S2**: summary generated using the feature *cenSim*
- **S3**: summary generated using the feature *qSim*
- **S4**: summary generated using the feature *isStarter*

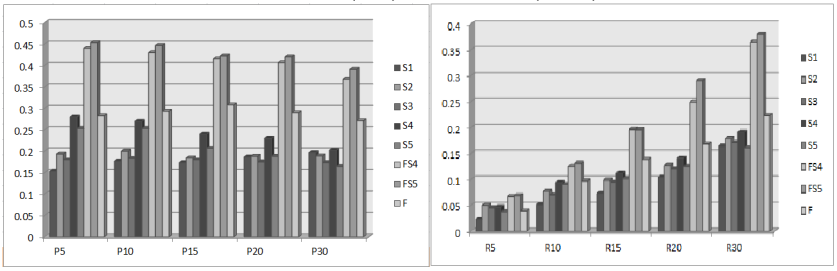
- **S5**: summary generated using the feature *conceptModel*
- **F**: using image title and tags from Flickr
- **FS4**: using summary (S4) plus image title and tags from Flickr (F)
- **FS5**: using summary (S5) plus image title and tags from Flickr (F)

Three participants formulated a total of 30 queries by selecting 10 images from the image set and writing an appropriate query to retrieve the image. We employed the "pooling" approach used in TREC[4] and ImageCLEF[5] to build an image pool for each query. A maximum of 150 top-ranked images were retrieved for each query by each of the retrieval systems with the eight configurations *S1*, *S2*, *S3*, *S4*, *S5*, *F*, *FS4* and *FS5*. The retrieved images were combined and duplicates were removed to create an image pool for the query. The three participants judged the results (relevant vs non-relevant).

2.5 Results

The comparison of precision (P) and recall (R) by the eight configurations is illustrated in Figure 1. From R10 to R30 we see moderate improvement for summaries of type S4 compared to the other four (S1, S2, S3 and S5) summary types. A similar picture can be observed in P scores; however, this time S4 summaries are better than the other summaries at all levels. The conceptual model summaries (*S5*) lead to better results compared to *S1, S2* and *S3* ones (for both P and R at levels 5 to 20). However, when each image is indexed with existing Flickr title and tags (F) then better results are obtained compared to all summary-only index types.

Fig. 1. Precision (left) and Recall (right) results.



We also combined the two best performing summaries (*S4* and *S5*) with the Flickr texts (F) leading to *FS4* and *FS5*. The results show a substantial improvements in both P and R at all levels which demonstrates the usefulness of the summaries for indexing. Furthermore, we can see from both P and R figures that the *FS5* indexes perform best overall. The results are even better than when the best performing summary (*S4*) is combined with *F*. For this *FS4* and *FS5* pair, we ran a two-tail paired T-test whose results indicate that the *FS5* indexes are significantly better than the *FS4* indexes at all P and R levels ($p < 0.05$).

[4] http://trec.nist.gov
[5] http://www.imageclef.org

2.6 Discussion

The results show that the Flickr title + tag indexes (F) lead to better scores as compared to all summary type indexes. However, these titles and tags are manually added to the images, not automatically generated. The results also show that in absence of such manual data summaries can be usefully used to index the images. Furthermore, if the textual information is available then it can be combined with automatic summaries to achieve even better results. Combining *S5* as opposed to *S4* summaries with the Flickr texts to index the images leads to better results. We speculate that this is because the conceptual model biased summaries favour sentences which contain information commonly associated by users with specific geographical object types, such as where the object is located, what it is surrounded by, etc. In addition to the place name such information was also noted in geographical search queries by Sanderson and Kohler [4]. The Flickr texts are likely to contain the name of the place. The place name combined with *S5* type summaries are likely to be matched by queries of the sort analyzed by Sanderson and Kohler. Of course more analysis is needed to confirm this.

3 Conclusion

In this paper we presented automatic rich image caption generation using summarization techniques starting with only geo-tagging information. We evaluated summaries generated by different summarization techniques in the image retrieval effectiveness task. We also used existing Flickr textual information as a baseline. We showed that combining the Flickr texts with conceptual model biased summaries performs significantly better compared to all other index types. In the future we plan to extend our work by using the conceptual modeling summarization method for images of other object types such as persons.

References

1. Aker, A., Gaizauskas, R.: Generating image descriptions using dependency relational patterns. Proc. of the ACL 2010, Upsala, Sweden (2010)
2. Aker, A., Gaizauskas, R.: Understanding the types of information humans associate with geographic objects. In: Proceedings of the 20th ACM international conference on Information and knowledge management. pp. 1929–1932. ACM (2011)
3. Fan, X., Aker, A., Tomko, M., Smart, P., Sanderson, M., Gaizauskas, R.: Automatic Image Captioning From the Web For GPS Photographs. In: Proc. of the 11th ACM SIGMM International Conference on Multimedia Information Retrieval (2010)
4. Sanderson, M., Kohler, J.: Analyzing geographic queries. In: SIGIR Workshop on Geographic Information Retrieval (2004)