

A Text-Based Approach to the ImageCLEF 2010 Photo Annotation Task

Wei Li, Jinming Min and Gareth J. F. Jones

Centre for Next Generation Localisation
School of Computing, Dublin City University
Dublin 9, Ireland
{wli, jmin, gjones}@computing.dcu.ie

Abstract. The challenges of searching the increasingly large collections of digital images which are appearing in many places mean that automated annotation of images is becoming an important technology. We describe our participation in the ImageCLEF 2010 Visual Concept Detection and Annotation Task. Our approach used only the textual features (Flickr user tags and EXIF information) provided with the images to perform automatic annotation. Our method explores the use of a combination of techniques to address the annotation problem. Our results indicate that the techniques works reasonably given the limitations inherent in using only textual data for this task. We identify the drawbacks of our approach and how these might be addressed and optimized in further work.

Keywords: Photo annotation, Document expansion, Feature extraction

1 Introduction

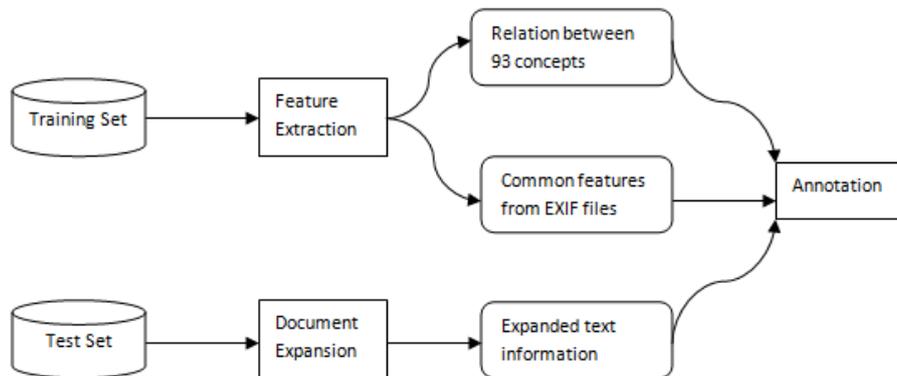
The exponential increase in the number of images available on the World Wide Web has led to a great interest in the topic of Automatic Image Annotation (AIA) to support applications such as effective search of online image collections. This paper describes details of our participation in the ImageCLEF 2010 Photo Annotation task which aims to explore methods for automatic annotation of large photo collections. The task involves assigning 93 concepts to images from the MIR Flickr 25,000 image dataset. The training and test sets consist of 8,000 and 10,000 images respectively. The Flickr images in each collection include user assigned tags and EXIF data for the photos where they are present. Automatic image annotation can broadly be classified into three approaches: visual, textual and hybrid models. In our work for ImageCLEF 2010 we concentrated only on use of text metadata for this task.

We submitted one run for the annotation task. The focus of our work was to attempt to exploit different methods to derive more text information from available resources to do the automatic annotation. In this our participation in the task we extracted features from the training set; used document expansion to enrich the existing text information resources; and extracted identified additional features. This paper is organized as follows: Section 2 describes our indexing and retrieval methods, Section 3 gives our experimental results and finally Section 4 concludes the paper.

2 Metadata Processing and Retrieval Strategies

Attempting to annotate images based on the available text information poses a significant challenge. Images are provided with tags of varying quality and scope manually assigned by users and with standard EXIF information. Investigation revealed of the provided Flickr dataset revealed that some images do not in fact have any user tags at all. In our experiments, we investigated approaches to making use of the limited information which was available to capture more features from both the training set and test set to assist with the annotation. These methods included document expansion and feature extraction which are introduced in the following subsections. The stages of processing and annotation are summarised in Figure 1.

Figure 1. Flow diagram of image annotation approach



2.1 Document Expansion

The limitations of the text descriptions provided with images can lead to significant problems for reliable processing of the images in applications such as search tools and classifiers. Particular problems can arise due to mismatch between the manually assigned tags when comparing images and when attempting to identify images relevant to user queries in retrieval application, and due to the general inadequacy of the tags assigned by users. In our approach to this task we sought to enrich text information about images by using a process of document expansion [6]. In document expansion the existing text metadata for an image is used as a query to a text information resource. Items retrieved in response to the query are then processed to identify terms strongly associated with the image's metadata. These words can then be added to the metadata, in the same manner as queries are expanded in traditional query expansion methods. For our work we use DBpedia as an external information resource for expansion of the image metadata "documents".

We used document expansion to expand the image metadata, but also the concepts which are to be used to annotate the images. Each concept usually consists

of only 1 or 2 words. Thus it is hard to reliably match concepts to image metadata. Thus it is interesting to try to expand concepts to include words related to the concept or which describe the concept. We thus hoped that after this expansion, concepts could be more reliably matched to image metadata. To perform concept expansion each of the concepts was treated as a query and again applied to external DBpedia information resource. Selected expansion terms were then added to the concept.

Our document expansion method uses the Okapi feedback method [7]. For expansion of the concepts, we assumed that the top 100 retrieved ranked DBpedia documents were relevant to the concept, we then added 10 top scoring words from the retrieved documents to the concept. For user tags a slightly more complex procedure was used. We still added 10 words to the metadata data of each image. However, since some user tags are sentences, they may contain stop words or other words which are not central to the focus of the tag. If we use the simple document expansion method which treats every word with the same weight, some stop words or other words not related to the topic of the tag may be added to user tags. To help avoid this problem, we used the document expansion method introduced in [1]. In this procedure user tags are first reduced by removing stop words and other words not likely to be significant to the document. The document expansion stage is then performed to add additional words to the image metadata. To perform the concept assignment, words in the expanded concepts and metadata documents were first stemmed, the similarity between each expanded image tag and concept was then computed to perform the annotation.

While this approach has the potential to assign good concept annotations for images which have manual tags to seed the expansion process, it does not work well for images which do not have manual tags as a starting point for expansion. In order to be able to annotate these images another method is required.

2.2 Feature Extraction

The annotation scheme has been setup in such a way to make it easy to extend it with new keywords without having to go through all images again [2]. In this part, we present a further method we used to refine the annotation process. The ImageCLEF 2010 task provides 93 annotation concepts. The relation between these concepts is another useful way for us to perform the annotation.

2.2.1 Affiliation Between Concepts

From the training set, some general concepts can be found. They cover a number of proper subtopics, see Table 1. We used a simple greedy algorithm method to assign an affiliation relation selection. The algorithm operates as follows:

Greedy Algorithm 1 (affiliation):

1. for each concept c_i ($0 \leq i \leq 92$), count how many times it appears in image collection, Nc_i
2. when concept c_i appears, count how many times another concept c_j ($0 \leq j \leq 92, j \neq i$) appears, Nc_j
3. compute $P_{ij} = Nc_j / Nc_i$
4. if $P_{ij} \geq 0.97$, then we assume, concept c_j is the subtopic of concept c_i

The value 0.97 was selected empirically for this collection. According to this relationship, if any subtopic is annotated in one photo, then its corresponding general topic will be annotated in the same photo.

Table 1. Some examples of the affiliation in 93 concepts

General Concept	Sub Concept
Sky	clouds, shadow
Water	lake, rive, sea
City_life	car, vehicle, bicycle, ship, train, airplane
Animals	dog, cat, bird, horse, fish
winter	snow
architecture	building-sight, church, bridge

2.2.2 Opposite Relation Between Concepts

In addition to the affiliation, an opposite relationship was also identified, see examples in Table 2. Similar to the affiliation relation method, a greedy algorithm was used to identify these relations. The algorithm operates as follows:

Greedy Algorithm 2 (opposite relationship):

1. for each concept c_i ($0 \leq i \leq 92$), count how many times c_i appears in the image collection, Nc_{ia}
2. when c_i occurs, count how many times another concept c_j ($0 \leq j \leq 92, j \neq i$) does **not** occur, Nc_{jn}
3. compute $P_{ij} = Nc_{jn} / Nc_{ia}$
4. for each pair of concepts c_i and c_j , compute $P_{ji} = Nc_{in} / Nc_{ja}$
5. if $P_{ij} \geq 0.7$ & $P_{ji} \geq 0.7$, we assume concept c_i and c_j are an opposite pair of concepts

Nc_{ia} = number of times concept c_i appears in the image collection

Nc_{jn} = the number of times concept c_j does not appear in the image collection when c_i appears

This relationship means that if one concept occurred in a photo, its opposite concept is unlikely to have occurred in the same photo. The value 0.7 was again chosen empirically. In this experiment, only two of these opposite pairs were found (the pair with ‘*’ mark in Table 2). How to find more opposite pairs is another challenge for our future work in this kind of task.

Table 2. Some examples of the opposite relation in 93 concepts

Concept	Opponent Concept
Indoor	Outdoor
*day	*night
No_visual_Time	day, night
*no_person	*single_person, female, male, baby, child, teenager, adult, old_person

2.2.3 Extract features from EXIF file

For concept classification, assignment of each concept was treated as an individual classification task. Thus for each concept, we consider that there is an annotated image collection. We find the common features of all images in this collection from their EXIF information file. Then this common feature is used to annotate this concept on test set.

EXIF metadata represents a number of properties and settings of the digital camera at the time of taking picture [2]. This includes the information:

- Camera itself: brand...
- Camera settings: exposure, aperture, focal length, ISO speed...
- Image settings: orientation, resolution, compression...
- Time and Date

Because not all of these fields are present in every EXIF file and the time restrictions to perform this task. We did not use EXIF collection effectively. We only extracted the Date and Time properties from EXIF metadata. Pictures which were taken at times between 08.00 and 17.00 were annotated as the day time concept and other times are assumed to be associated with a night concept. Further features could be extracted if more time were available to analyze this EXIF metadata. Further study of EXIF metadata is also planned in future work for this task.

2.3 Feature Combination

To calculate the concept assignments the features need to be combined to produce a final result. Following the application of document expansion, we get a binary result matrix

A. All other feature functions are then applied on this matrix. The final combination result is calculated using the following equation:

$$\text{Final result} = (A + Rb + Rc) \otimes D$$

where:

A represents the document expansion binary result matrix;
 $+$ represents application of the following method on the previous matrix;
 Rb is the affiliation relation method;
 Rc is the opposite relation method;
 \otimes is exclusive or symbol;
 D is the binary result matrix achieved by the EXIF metadata method.

3 Task Submission and Evaluation

We made only one submission for this task. This used all the methods introduced above in combination to annotate the test dataset. The official result of this run is reported in Table 3.

For this task, 64 runs were submitted in total, only two runs chose to use the text-based approach (our submission and another from the MLKD group). Based on the reported MAP measure, these two runs got very close results, and were ranked at positions 42 (MLKD group) and 45 (our run) out of 64 submitted runs, respectively. The best run used a hybrid approach.

Table 3. Result of Runs evaluated by MAP, EER and AUC

Submission run	MAP	Avg. EER	Avg. AUC
Text-Based Run (DCU__1277149866992_ _test_annotation.txt)	0.2284	0.4508	0.1944

For each concept, the EER (Equal Error Rate) and AUC (Area Under Curve) were calculated. The results of each concept are shown in Figure 2 (the x axis indicates the 93 concepts; the y axis indicates the Accuracy Rate). From the figure we can see that the results of our experiment are variable, it can be noted that some concepts are not detected at all. One of the main reasons underlying poor results is that the text resource available for some concepts is not sufficient for this task. In particular, some images do not have tags and EXIF files at all. This is obviously a big problem when using a text only based approach to doing the annotation task. Another issue is that both the EER and AUC evaluation methods require confidence scores of each annotated concept. However, our method cannot provide this score information.

Figure 2(a). EER and AUC of concept 0 to 46

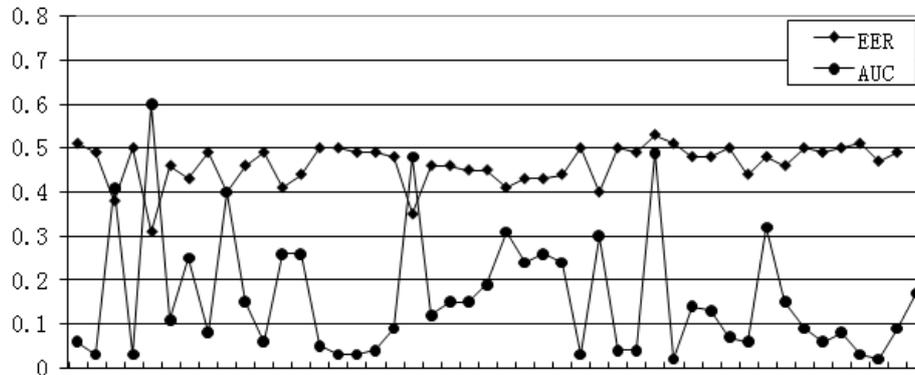
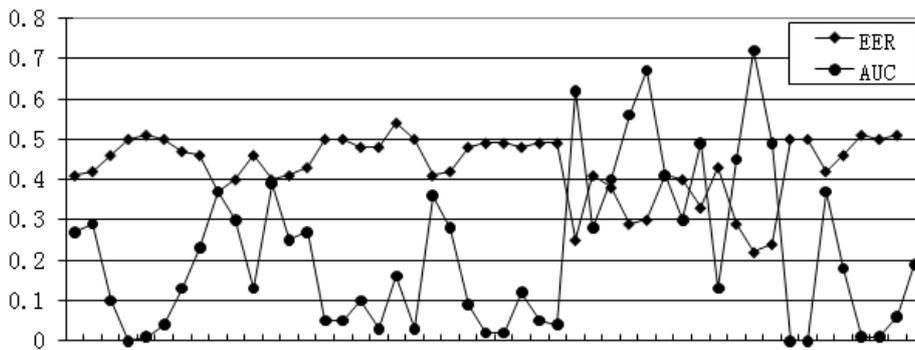


Figure 2(b). EER and AUC of concept 47 to 93



4 Conclusion

We have presented and analysed our submission to the ImageCLEF 2010 Photo Annotation Task and compared our results to those of other participants. Although the text-based approach only achieves moderate and inconsistent results, it has potential to be improved further. In this experiment we used document expansion to enhance image text metadata. In future work we plan to explore used of other external information resources for this task. Some images do not have tags and EXIF information and thus cannot be annotated at all. How to identify more features and information from this limited resource is a big challenge for text-based approach. All of these problems define our future work.

5 Acknowledgements

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (CNGL) at Dublin City University.

References

1. J. Min, J. Leveling, G. J. F. Jones: Document Expansion for Image Retrieval. Proceedings of RIAO 2010, Paris, France (2010)
2. MIRFLICKR Image Collection Website. <http://press.liacs.nl/mirflickr/>
3. T. Tsirikia and J. Kludas: Overview of the WikipediaMM Task at ImageCLEF 2009, In Working Notes of CLEF 2009, Corfu, Greece (2009)
4. J. Ngiam and H. Goh: I2R ImageCLEF Photo Annotation 2009, In Working Notes of CLEF 2009, Corfu, Greece (2009)
5. S. Sarin and W. Kameyama: Joint Equal Contribution of Global and Local Features for Image Annotation. In Working Notes of CLEF 2009, Corfu, Greece (2009)
6. J. Min, P. Wilkins, J. Leveling, and G. J. F. Jones: DCU at WikipediaMM 2009: Document expansion from wikipedia abstracts. In Working Notes of CLEF 2009, Corfu, Greece (2009)
7. S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In Proceedings of the Third Text REtrieval Conference (TREC-3), Gaithersburg, USA, page 109-126 (1995)