QUT at the NTCIR Lifelog Semantic Access Task

Harrisen Scells, Guido Zuccon, Kirsty Kitto
Faculty of Science & Technology, Queensland University of Technology, Brisbane, Australia
harrisen.scells@connect.qut.edu.au, g.zuccon@qut.edu.au, kirsty.kitto@qut.edu.au

ABSTRACT

This notebook paper describes the submissions to the 2016 NTCIR Lifelog Semantic Access Task made by the Queensland University of Technology (QUT).

1. INTRODUCTION

Lifelogging, i.e., the practice of digitally record and document life and daily activities, is becoming increasingly popular [1]. This activity often involves wearing a camera attached to a shirt or a lanyard to take several photos a minute. This allows the user to passively monitor their biometric activities, communication activities, and record their life. The problem that arises when the recording of lifelog data is searching through the vast number of images that are taken each day. This is the basis for the NTCIR Lifelogging pilot task. In this paper we describe our participation to this task.

Our approach is mainly based on the idea of annotating images using long, descriptive paragraphs rather than tag based approaches, and attaching some measure of importance to each concept for an image when performing retrieval. Due to the fact that each image will possibly have more concepts associated with it when compared to tag based systems, it is important to be able to give each individual concept a level of importance. Not all concepts in an image will have the same relevance to an image. A long description of the image allows the retrieval system the ability to treat an image as a document rather than a set of separate tags.

2. METHODS

The approach we used in the pilot task involves three stages. Firstly, the images are annotated, followed by propagating descriptions to unannotated images, and finally performing information retrieval on the images. Many of the keywords provided with the concept annotations by the task were found be be irrelevant; thus we generated an alternative set of keywords. To do so, textual descriptions are used in place of tags and keywords are extracted by tokenising the descriptions.

2.1 Annotation

In order to prepare the images for annotation, a reasonable subset of images which are distinct from every other image in the subset are selected from the initial set. Not every image is going to have a unique annotation associated with it as the dataset is too large to annotate manually. One image is chosen at random from each cluster for annotation. An assumption is made that descriptions are associated to M images, and that $M < N$, where N is the total number of images. The process involved in creating this subset is a temporal clustering algorithm designed for this task. The algorithm groups visually similar images together into a relatively small group of clusters. The algorithm makes two passes over the collection, detailed as follows:

First Pass.

1. Sort images in order of time taken and then on who took the photo, start at the beginning
2. Look at the current image
3. Use a similarity measure to compare the histogram vectors for the current image and the next image
4. If the images are similar, group them together, if not create a new cluster
5. Continue points 2-4 until no more images can be looked at

Second Pass.

1. Pick a group of un-merged images from the set of clusters
2. Take the average histogram vector for all images in that cluster, and compare this with all other clusters which have not been merged using a similarity measure
3. If the two clusters pass over a threshold, merge them and mark them as merged
4. Continue 1-3 until there are no more clusters which have not or cannot been merged

2.2 Propagation

Only a small subset of the initial set of selected images are annotated. Clustering will result in M images to annotate. If k is the number of annotated images, and $k < M$, then propagation is needed. This is outlined as follows:

1. Spread the annotations intra-cluster := that is, for each cluster with an annotated image, give all other images in that cluster the same annotation
2. Spread the annotations inter-cluster := that is, for each cluster where no images have been annotated, get the visually similar nearby clusters which have annotations and propagate these intra-cluster, concatenating if already annotated
3. Recursively repeat step 2 until either there are no more unannotated clusters, or there are no more annotatable clusters (that is, clusters which have not been annotated but there exists a set of annotations for which it can be given)

2.3 Retrieval

The information retrieval system ranks results using concepts in the query. A concept is the combination of a keyword and a weight. The weight of a keyword represents the importance of that keyword to the information need. A query in the system is an unordered set of these concepts. Since concepts are formed from the descriptions of images, it may not be possible to answer some queries due to the fact that no relevant concepts exist in the collection.

3. EXPERIMENTS

3.1 Annotation

The clustering algorithm has two parameters, τ_1 and τ_2. For the first pass, τ_1 is used as the minimum similarity threshold to start adding images to a new cluster, and τ_2 is used as the minimum similarity threshold for merging clusters. τ_1 is set to 0.86 and τ_2 is set to 0.95. These numbers are found, through experimentation, to give well distributed sets of image clusters for the dataset at hand. When calculating inter-cluster similarity, the same 0.86 was used for the τ threshold. This was due to the fact that a wide range of visually similar clusters was desirable, and most clusters are not annotated, thus a very small proportion of clusters will have anything to propagate with.

When calculating the similarity between images, the cosine similarity measure is used. This is defined as:

$$\cos(a,b) = \frac{\sum_{i=1}^{m} a_i b_i}{\sqrt{\sum_{i=1}^{m} a_i^2 \cdot \sum_{i=1}^{m} b_i^2}}$$

Where a and b are vectors of the histograms for two images respectively.

3.2 Collection and Mapping

A custom built web interface is used to annotate images with descriptions. In this interface, ten images at a time are shown and the user is asked to describe what they see in each. When all ten images are annotated, another set is shown and the cycle repeats. Before the annotations are propagated, the set of concepts are formed by tokenising the descriptions. In this process, concepts are extracted by removing all syntax, lowercasing, tokenising by splitting each word by spaces, and finally removing any irrelevant concepts using a stop word list.

Concepts are mapped to images by taking the keywords already associated with an image and calculating the term frequency for each concept. This allows labels that appear more than once in the description for an image to have a higher importance. All weightings for the concepts attached to an image sum to 1.

At this stage in the task, 444 images from the 16196 clusters had user-created annotations associated with them. Intuitively, this small number of annotations considerably limits the performance of the system. For example, when manually mapping image concepts to queries, some queries did not have any concepts which could be considered relevant.

3.3 Retrieval

The retrieval process produces six runs. Five of these runs were submitted to the task for evaluation. The first three use a concept list provided with the task whereas the last three use concepts derived from the descriptive annotations. For each set of three runs, the first run does not take weighted concepts into account, the second run takes does take the weight of each concept into account, and the final run not only takes into account the weight of each concept, but also the inverse document frequency (IDF) of the concept. The additive scoring system does not take into account weights and will simply rank images based on the number of times a concept appears on an image. Weighted scoring multiplies the score of an image by the weight of a concept. Weighted scoring with IDF first multiplies the weight of the concept by that concepts' IDF and then an images' score is multiplied by this new value. IDF for a concept in an image is calculated by using:

$$\log(\frac{N}{n_t + 1})$$

Where N is the total number of images and n_t is the number of images the concept appears in.

4. LIMITATIONS

Our initial work on this pilot task has a number of limitations. Each image in the dataset has a location, date and time associated with it. This is useful metadata which is intended to be used in future work. A small number of queries suffer from the system not having any notion of temporal or location attributes of images. When ranking images, this time and location data should boost certain images higher than images with inconsistent meta data. An example of one such query is “times when I was on the bus to work”. The current system is able to retrieve images of times when somebody is on a bus, but has no additional information to associate the time and location of images with the specified time and location in the query.

Other limitations include:

- Descriptions are currently treated as a bag of words, and concepts as keywords.
- Querying images currently requires manual mapping of query to concepts. This can introduce human error and subjectivity for how important a concept is to a query.

Our future work will be directed towards addressing these limitations. We Further plan to (1) collect more annotations and testing whether descriptive annotations are more effective than tag-based annotations; (2) integrate a query language into our search tool that would allow to express powerful queries posing constrain on concepts, times and locations matching.

5. REFERENCES

[1] C. Gurrin, A. F. Smeaton, and A. R. Doherty. Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014.