Targeting Diversity in Photographic Retrieval Task with Commonsense Knowledge

Supheakmungkol Sarin and Wataru Kameyama

Graduate School of Global Information and Telecommunication Studies, Waseda University
1011 Okuboyama, Nishi-Tomida, Honjo-shi, Saitama-ken 367-0035, Japan
{mungkol@fuji.waseda.jp, wataru@waseda.jp}

Abstract

Image search engines have a very limited usefulness since it is still difficult to provide different users with what they are searching for. This is because most research efforts to date have only been concentrating on relevancy rather than diversity which is also a quite important factor, given that the search engine knows nothing about the user's context. In this paper, we describe our approach for ImageCLEF 2008 photographic retrieval task. The novelty of our technique is the use of AnalogySpace [3], the reasoning technique over commonsense knowledge for document and query expansion, which aims to increase the diversity of the results. Our proposed technique combines *AnalogySpace* mapping with other two mappings namely, *location* and *full-text*. We then re-rank the resulting images from the mapping by trying to eliminate duplicate and near duplicate results in the top 20. We present our preliminary experiments and the results conducted using the IAPR TC-12 photographic collection with 20,000 natural still photographs. The results show that our integrated method with AnalogySpace yields slightly better performance in terms of *cluster recall* and the *number of relevant photographs retrieved.* We finally identify the weakness in our approach and ways on how the system could be optimized and improved.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [Database Management]: Languages – Query Languages

General Terms

Algorithms, Experimentation, Performance

Keywords

Commonsense knowledge, Image Retrieval, Diversity, AnalogySpace, Query/Document Expansion, Re-rank

1 Introduction

The affordability of digital camera and the ease of use of content publishing tool have pushed for the rapid growth of everyday photographs on the web with a large percentage coming from the amateur photographers. These published amateur photographs usually come with either a short description or a few keywords. This shows potentials for image retrieval system to provide better resulting images. Unfortunately, image search engines have very limited usefulness since it is still difficult to provide different users with what they are searching for. Usually different people issuing the same query are looking for different images. A good image search engine should not produce top results in the ranked list that contain only relevant items of a single theme, but rather diverse items representing sub-topics within the results, yet keeping high level of relevancy.

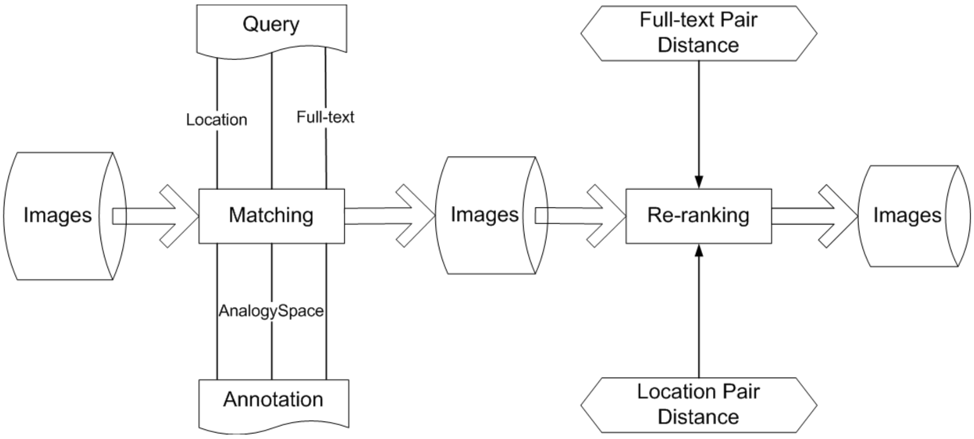

Figure 1: Flow Diagram of System Architecture

This paper describes our participation for the photographic retrieval task of ImageCLEF 2008. ImageCLEF 2008 is a track running as part of the CLEF (Cross Language Evaluation Forum) campaign. It comprises five tasks related to image retrieval and annotation techniques, namely, photographic retrieval, medical retrieval, general photographic concept detection, medical automatic image annotation, and image retrieval task from a collection of Wikipedia images. We present our development and contributions to the first task of which the goal is to promote diversity in the top ranked list of resulting images.

2 Approach and Implementation

Using surrounding text of the images or annotation as a means to interpret them is a classic research methodology. To date, however, most research efforts have only been concentrating on relevancy than diversity. The latter is also a quite important factor since the search engine usually knows nothing about the user. Furthermore, most of the time, people solve the problem through selecting some keywords and features of images to represent the photograph rather than trying to understand the semantic nature of annotation and the query. In this paper, we approach these problems as follows:

- To enable diversity, we use commonsense knowledge as a tool for term expansion. We consider ConceptNet [8] as our commonsense knowledge database. ConceptNet is made up of a network of everyday concepts that have been automatically generated from English sentences of the Open Mind Common Sense corpus. The corpus has been handcrafted by the general public since 2000 [9]. Those concepts are connected by one of about twenty relationships such as “IsA”, “PartOf”, “locationAt”, “Desires”, “CapableOf”, “UsedFor”, etc. We use ConceptNet for diversity purposes because a term can be expanded to its contextually related concepts that are not necessarily its synonyms. Furthermore, those related concepts reflect the commonsense way of people’s thinking and how they relate concepts since they are input by human beings with a specific purpose. For instance, “drink coffee” relates to “wake up”, “yawn”, “read newspaper”, etc. However, diversity should not come as a compensation of relevancy. Therefore, we also try to maintain the level of precision by combining the former with both full-text and location matching.
- Re-ranking technique is performed afterward to re-rank the results of the previous step by trying to eliminate duplicate and near duplicate results.

Figure 1 illustrates the process of our proposed approach. The flow can be divided into two major steps, namely, matching and re-ranking.

2.1 Matching

As shown in Figure 1, we introduce three kinds of matching between query and annotation of the image, namely *location*, *AnalogySpace*, and *full-text*.

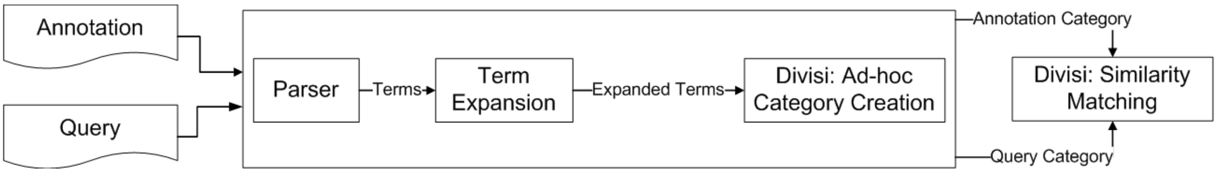

Figure 2: AnalogySpace Matching

2.1.1 location matching

We begin by parsing the annotation to get location named entities. GATE is used for this purpose [10]. Then, we establish a location hierarchy from the annotation before we perform the matching. For instance, *Lima* is expanded to *Lima >> Peru >> South America*. Location names found in image annotations and query topics are expressed as sets with prepositions found in the query as a matching condition. To do this, we simply create two sets of prepositions namely, *include set* and *exclude set*. Prepositions in include set are such as 'in', 'of', 'along', 'on', 'near', 'by', 'in', etc., while the other set includes prepositions such as 'out of', 'outside', etc. For example, in the query “Sport stadium outside Australia”, *outside* serves as an excluding condition.

2.1.2 AnalogySpace matching

AnalogySpace is a vector space representation of commonsense knowledge built on the top of ConceptNet using Principal Component Analysis [3]. This representation can be used as a reasoning tool as it reveals large-scale patterns in the data while smoothing over noise. In our case, we use an implementation of AnalogySpace called Divisi [11] to create ad-hoc category for each annotation and query. We then match the query against the annotation. The degree of similarity between the two ad-hoc categories is the dot product of matrices of the shared similar concepts and features.

Since ConceptNet depends on sentences contributed from human, it does not contain all the terms a dictionary has. To cope up with unknown terms, we use their synonym and hypernym. We create the set of expanded term for the unknown term using its Wordnet's synsets and hypernym regardless of its part of speech. However, we only choose one term as our replacement for the unknown term. The best term is the term that is most uniform to other terms of the annotation. This is achieved via dot product of the matrix of an ad-hoc category created from a combination of other terms of the annotation, against the ad-hoc categories created from each term from the expanded set if it exists in ConceptNet. We chose the term that has the highest similarity score. Figure 2 shows the process.

2.1.3 full-text matching

Vector Space Model is used to represent the annotations and query topics. Term frequency is used for our vector space model. Each document is represented as a vector, where each dimension corresponds to the frequency of a given term. In our case, terms are reduced to their stems.

Some terms from query topics might not be found in the index of the annotation documents. To cope up with this, we expand unknown query terms with their synsets and hypernym from WordNet. We select top three terms among the set of synonyms found. AnalogySpace is used to compute the similarity score between the unknown term and its synonyms.

The similarity distance between a document vector and a query vector is expressed as cosine distance. Figure 3 illustrates the technique.

Finally, we normalize each matching score according to its maximum and minimum value. The total matching score is expressed as the product of all the three matching scores. This is the simplest way to combine the scores and yet make the large differences count for even more.

2.2 Re-ranking

In this step, the results from the first step are re-ranked according to their semantic similarity by giving penalty to the ones with high similarity between each other.

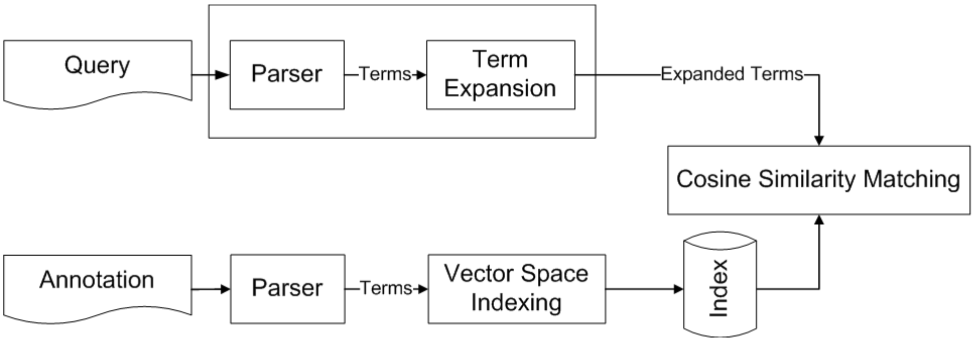

Figure 3: Full-text Matching

2.2.1 Pair distance similarity

We calculate full-text and location similarity. Same as in the matching process between query topic and photograph annotation, boolean logic is used for location similarity calculation, while vector space model is used for full-text similarity calculation. We compute the total pair distance of images as the product of both distance scores.

2.2.2 Re-rank

The similarity distance score obtained can be used to filter and re-rank the preliminary results. We use a method called Hill Climbing to find a threshold of similarity distance that can help optimize both the precision and diversity. We introduce a loop where Hill Climbing starts with a random threshold and looks for the set of solutions which are better from its neighbors. The loop goes on until we obtain the best compromise.

3 EVALUATION

3.1 Process

Organizers of ImageCLEF 2008 provide participants with a collection of annotated images, together with query topics. Participants use these resources with their retrieval systems and submit to the organizers the identifiers of the relevant documents for each query topic. Then, the organizers evaluate the result set of each submission from every participant and rank submissions according to standard evaluation measures.

3.2 Dataset

The collection of images used for ImageCLEF 2008 is the IAPR TC-12 photo collection consisting of 20,000 natural images taken from locations around the world [2]. The collection includes images of various sports and actions, photos of people, animals, cities, landscapes and many other aspects of contemporary life. Each image is also associated with an alphanumeric caption stored in a semi-structured format. These captions include the title of the image, its creation date, the location at which the photograph was taken, a semantic description of the contents of the image by the photographer and some additional notes. Table 1 shows the example of a photograph and its metadata. In our system, we use only the *title*, *description,* and *location* parts of the metadata.

3.3 Query

There are a total of 39 queries used in this study ranging from the very specific to the very abstract ones with different levels of difficulty. Here are some of the query topics: "animal swimming", "destinations

Table 1: Example of a photograph of the collection and its attached metadata

	DocNo: annotations/37/37394.eng Title: The Saint Basil's Cathedral Description: a cathedral with crosses on many onion domes; people, trees, a statue and snow on a square in front of it; a grey sky in the background; location: Moscow, Russia Date: February 2001 Image: images/37/37394.jpg Thumbnail: thumbnails/37/37394.jpg

DocNo	annotations/37/37394.eng
Title	The Saint Basil's Cathedral
Description	a cathedral with crosses on many onion domes; people, trees, a statue and snow on a square in front of it; a grey sky in the background;
location	Moscow, Russia
Date	February 2001
Image	images/37/37394.jpg
Thumbnail	thumbnails/37/37394.jpg

Table 2: Example of a query topic

Num	2
Title	Church with more than two towers
Cluster	City
Narr	Relevant images will show a church, cathedral or a mosque with three or more towers. Churches with only one or two towers are not relevant. Buildings that are not churches, cathedrals or mosques are not relevant even if they have more than two towers.
Image	images/16/16432.jpg
Image	images/37/37395.jpg
Image	images/40/40498.jpg

in Venezuela", "church with more than two towers", "sunset over water", etc. Query topics are provided as a structured information. It is composed of the query title, cluster, narration of how relevant images should be, and some examples of relevant image files. Table 2 shows the example of a query topic. In our system, we use only the *topic title*.

3.4 Measurement techniques

To ensure both relevancy and diversity, the evaluation is based principally on two measures, namely, precision at 20, and instance recall at rank 20 [4]. The technique is a relatively new evaluation methodology that considers results of a query as interdependence rather than a standalone. A good engine will produce results that maximize the two measurements.

4 Results and Discussions

We present below the results of the four runs that we submitted to ImageCLEF photographic task 2008.

1. AnalogySpace: In this run, we combine location matching and AnalogySpace.
2. Full-text: In this run, we simply use location matching and full-text search.
3. Full-text (no query expansion) + AnalogySpace: In this run, we combine location matching, full-text matching, and AnalogySpace matching.
4. Full-text (with query expansion) + AnalogySpace: The same as the previous one, we combine the three matching. We further expand the terms of query topics in full-text matching with their synsets and hypernym.

Table 3: Precision (P) at the top n results

Runs	P5	P10	P15	**P20**	P30	P100
AnalogySpace	0.24	0.23	0.22	**0.22**	0.2	0.11
Full-text	0.32	0.3	0.29	**0.27**	0.25	0.16
Full-text (no query expansion) + AnalogySpace	0.3	0.28	0.27	**0.27**	0.25	0.16
Full-text (with query expansion) + AnalogySpace	0.33	0.3	0.28	**0.26**	0.24	0.16

Table 4: Cluster Recall (CR) at the top n results

Run	CR5	CR10	CR15	**CR20**	CR30	CR50	CR100	CR1000
AnalogySpace	0.09	0.13	0.16	**0.21**	0.24	0.27	0.35	0.67
Full-text	0.14	0.21	0.25	**0.28**	0.35	0.46	0.55	0.81
Full-text (no query expansion) + AnalogySpace	0.14	0.21	0.25	**0.31**	0.38	0.46	0.54	0.82
Full-text (with query expansion) + AnalogySpace	0.13	0.19	0.22	**0.25**	0.33	0.42	0.52	0.8

Table 5: Other measurements: Number of Relevant Retrieved images (NumRelRet), Number of Relevant images (NumRel), Mean Average Precision (MAP), Geometric Mean Average Precision (GMAP), Blind REelevance Feedback (BREF)

Run	NumRelRet	NumRel	**MAP**	GMAP	BREF
AnalogySpace	1247	2401	**0.14**	0.01	0.51
Full-text	1420	2401	**0.21**	0.06	0.64
Full-text (no query expansion) + AnalogySpace	1451	2401	**0.2**	0.06	0.65
Full-text (with query expansion) + AnalogySpace	1462	2401	**0.2**	0.04	0.65

Table 3, 4, and 5 show the precision, cluster recall, and other measures respectively. From the results, we notice that there is only a slight improvement in *recall* when introducing AnalogySpace. Table 4 shows that AnalogySpace helps to gain a little bit better cluster recall at 20 over the conventional full-text vector space model. The number of relevant images retrieved also increases as shown in Table 5. However, Table 3 and 5 show that the precision at 20 and the Mean Average Precision (MAP) which is the summary of recall and precision do not produce better result with AnalogySpace. Therefore, the results are not yet conclusive. We also notice that the improvement happens only when there is no query expansion in the full-text matching.

We still believe that ConceptNet could help enriching diversity in the resulting images. To our understanding, the reason why we could not achieve better results is because of the fact that there are lots of terms that ConceptNet does not cover. When we try to expand those unknown terms using WordNet, we only introduce noise. That is because WordNet's synsets contain all the synonyms of the word from all its possible senses. We did not implement any sense disambiguation. We did not even check the part of speech. Therefore, most of the time, the replacement only twists the meaning of the original word since we do not select the most appropriate sense of the word. Moreover, we limit the number of selected synonym to only one in AnalogySpace term expansion, and only up to three in our full-text query expansion. This reduces the coverage of the meanings. Due to limited time, content-based technology was not taken into consideration. Should we have incorporated another content-based pair similarity distance in the re-ranking step, we might be able to get better resulting images. Hence, we are planning to tackle these issues in our future works.

5 Related Works

Image search at the major search engines today largely relies on looking at words that are used around images – on the pages that host them, in image file names, and in ALT text associated with them. No real image recognition is done by any of the majors. Datta et al. have recently produced a complete survey of the current image related techniques [6]. Hsu et al [1] have used ConceptNet as tool for query and document expansion in image retrieval task. Nevertheless, in doing this, the authors only use spatial relationship function to find the concepts that co-exist in space of the real world. Google recently introduced VisualRank – a method that guesses how the images would be linked together, with those being most similar having more virtual links to each other. As a result, the most "linked to" images are calculated to rank first [7].

6 Conclusion

User satisfaction is not solely a function of relevancy. When nothing is known about the user, diversity plays an important role in getting the results that user would like to see. We present a novel approach to enable rich diversity in the results by incorporating commonsense knowledge expansion and result re-ranking through elimination of duplicate and near duplicate results. The presented results are just our preliminary ones. Even they are not conclusive yet, they pave the way to help us to improve our current system. We are now working to address the weak points that we have discussed earlier.

References

[1] Ming-Hung Hsu and Hsin-Hsi Chen, 2006. Information retrieval with commonsense knowledge. Proceedings of the 29th ACM SIGIR '06, 651–652.

[2] Grubinger, M., Clough, P., Müller, H. and Deselaers, T., 2006. The IAPR TC-12 Benchmark: A New Evaluation Resource for Visual Information Systems, In Proceedings of International Workshop OntoImage'2006 Language Resources for Content-Based Image Retrieval, held in conjunction with LREC'06, pages 13-23, Genoa, Italy, 22 May 2006.

[3] Robert Speer, Catherine Havasi, and Henry Lieberman. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge, Chicago, Illinois, AAAI 2008

[4] Zhai, C. X., Cohen, W. W., and Lafferty, J. 2003. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval (Toronto, Canada, July 28 - August 01, 2003). SIGIR '03.

[5] Catherine Havasi, Robert Speer, and Jason Alonso. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge, RANLP 2007

[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image Retrieval: Ideas, Influence, and Trends of the New Age. ACM Computing Surveys, vol. 40, no. 2, article 5, 60 pages, 2008

[7] Yushi Jing and Shumeet Bajula. PageRank for Product Image Search. In Proceedings of the World Wide Web Conference. ACM WWW '08.

[8] Havasi, C., Speer, R. & Alonso, J. (2007) ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. Proceedings of Recent Advances in Natural Languages Processing 2007

[9] Singh P, Lin T, Mueller E T, Lim G, Perkins T and Zhu W L: 'Open mind commonsense: knowledge acquisition from the general public', Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems, Lecture Notes in Computer Science No 2519 Heidelberg, Springer (2002).

[10] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "GATE: a framework and graphical development environment for robust NLP tools and applications," in Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL '02), Philadelphia, Pa, USA, July 2002

[11] Divisi: a general-purpose tool for reasoning over semantic networks. Website: http://divisi.media.mit.edu/ (Last visit: August 11, 2008)