

Information Retrieval with Commonsense Knowledge

Ming-Hung Hsu and Hsin-Hsi Chen

Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan

mhhsu@nlg.csie.ntu.edu.tw
hhchen@csie.ntu.edu.tw

Abstract

This paper employs ConceptNet, which covers a rich set of commonsense concepts, to retrieve images with text descriptions by focusing on spatial relationships. Evaluation on test data of the 2005 ImageCLEF shows that integrating commonsense knowledge in information retrieval is feasible.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Retrieval Models

General terms

Experimentation, Performance

Keywords

Commonsense Knowledge, ConceptNet, Concept Expansion

1. INTRODUCTION

External resources like WordNet have been introduced to deal with the issue of divergence in vocabulary between queries and documents [2,5]. However, a query sometimes diverges from its relevant documents not only in vocabulary, but also in “concept”. For example, a query “groom and bride” may be relevant to a document containing the word “wedding”. As “groom and bride” and “wedding” are related concepts in common sense, using common sense in IR may make progress for the state-of-the-art IR systems.

Commonsense knowledge has been explored in retrieval using Cyc [3] and ConceptNet [6]. Liu and Lieberman [4] used commonsense knowledge to expand the query with the related concepts. However, the above works did not make formal evaluation, so that we were not sure if the effects of introducing common sense are positive or negative. This paper investigates the usefulness of commonsense knowledge for image retrieval. The basic idea is to find concepts of objects that have spatial relationship (i.e., occur nearly in space) with each other, based on text descriptions of images.

2. IR WITH COMMON SENSE

2.1 ConceptNet

ConceptNet [6] is presently the largest commonsense knowledgebase. Its framework is a semantic network. Nodes in ConceptNet are semi-structured natural language fragments, e.g., “food”, “grocery store”, “buy food”, “at home”, etc. Each node represents one concept in real world. An edge between two nodes represents one relationship between two concepts. Nowadays ConceptNet covers relationships of twenty types such as causal, spatial, functional, etc.

Copyright is held by the author/owner(s).
SIGIR '06, August 6-11, 2006, Seattle, Washington, USA.
ACM 1-59593-369-7/06/0008.

ConceptNet provides a set of NLP tools for reasoning over text. Here we only adopt *project_spatial(concepts)* to consider spatial relationship between concepts. This function carries out spreading activation [7] through spatial relationships in ConceptNet. It helps IR systems find those concepts which usually coexist with the concepts of parameters in the space of real world. For example, the result of *project_spatial(“groom”, “bride”)* is {“tuxedo (15%)”, “bouquet (15%)”, “veil (15%)”, “wedding dress (14%)”, “gown (13%)”, “wedding cake (12%)”, ...}, which is a set of concepts and their strengths of relationships to “groom” and “bride”.

2.2 Commonsense Concept Expansion

For image collections with text annotations, text descriptions can be viewed as “documents”. We use ConceptNet to expand concepts of objects in a text description T as follows.

- I. Lemmatizing and POS-tagging T by using MontyLingua [6].
- II. Choosing concepts in T to be expanded. These concepts are called *expanded-concepts*.
- III. Utilizing spatial relationships in ConceptNet to find concepts that are spatially related to the expanded-concepts chosen at stage II. These concepts are called *projection-concepts*.
- IV. POS-tagging the projection-concepts found at stage III and filtering those concepts that are not nouns or noun phrases.

At stage II, the noun phrases that exist in ConceptNet are chosen, and nouns not part of a noun phrase chosen are selected too. For example, given a text description “*studio portrait of woman in patterned blouse*”, the concepts chosen for expansion, are “*studio portrait*”, “*woman*”, and “*blouse*”. At stage III, each of the projection-concepts returned by ConceptNet has its relationship related to the expanded-concept with some strength. A projection-concept with strength lower than a threshold will not be used later. We also limit the number of projection-concepts no more than half of the number of expanded-concepts.

2.3 IR with Concept Expansion

The original text descriptions and the projection-concepts of images are indexed separately. At query time, the IR system searches the two document collections and linearly combines the results as follows.

$$S(i) = c_1 S_1(i) + c_2 S_2(i) \quad (1)$$

where $S_1(i)$ and $S_2(i)$ are the scores of image i in the collection of original text descriptions and in the collections of projection-concepts, respectively, and $c_1 + c_2 = 1$.

3. EXPERIMENTS AND DISCUSSIONS

The ImageCLEF test collection in CLEF 2005 [1], i.e., the St. Andrews image collection and 28 ad hoc topics in English, is adopted. There are 28,133 images each of which is annotated with

a text description including three fields: headline (H), caption (P), and category (T). While we explore different fields for concept expansion, the corresponding baseline performance is obtained using the same fields. Okapi with BM25 is employed. Three evaluation criteria are considered: mean average precision (MAP), precision at top 20 documents (P@20), and R-Precision.

Table 1 shows the experiment results. Baseline (BL) is obtained with $c_1=1$ and $c_2=0$, i.e., concept expansion is not used. It shows that IR with commonsense knowledge (OURS) is better than the baseline (BL). When using headline and caption, the improvement is verified as significant by a t-test with a confidence level of 95% (marked by an asterisk in Table 1). It indicates that our approach is more suitable for precision-oriented tasks. The following examples explain how common sense improves P@20. Several relevant images of topic 18 “*woman in white dress*” are ranked low in the baseline because their captions describe occasions about wedding, but do not contain the critical words such as “white dress”. These images are expanded with projection-concepts like “white dress”, so their ranks are moved to the top 20 in our approach. Another example is topic 24 “*close-up picture of bird*”. While captions of some relevant images mention bird-related concepts, such as “eagle”, “nest”, etc, our approach can enhance the concept of “bird” for those images highly related to bird, so that the performance gets improved. Another interesting example is topic 9 “*Horse pulling cart or carriage*”. Some of the relevant images describe the royal in a carriage. In the descriptions of these images, the concept “carriage” doesn’t occur but “royal” does. Our approach expands these images with “carriage” since “carriage” and “royal” have spatial relationship with each other in ConceptNet. This kind of common sense is culture-specific.

Table 1. Comparison of our approach with the baseline

Field(s) used	MAP		P@20		R-Precision	
	BL	OURS	BL	OURS	BL	OURS
H ($c_1=.8, c_2=.2$)	.0799	.0824 (3.13%↑)	.1875	.1911 (1.92%↑)	.1230	.1238 (0.65%↑)
H+P ($c_1=.7, c_2=.3$)	.2693	.2697 (0.15%↑)	.3821	.3929* (2.83%↑)	.2951	.3082 (4.44%↑)
H+P+T ($c_1=.95, c_2=.05$)	.3544	.3545 (0.03%↑)	.4536	.4554 (0.40%↑)	.3605	.3635 (0.83%↑)

By comparing the performance improvements in P@20, we observe how the effects of commonsense knowledge vary when different fields are used. The improvement of using headline only is 1.92% and not significant. In contrast, when both headline and caption are used, the improvement in P@20 (i.e., 2.83%) is significant. It confirms that commonsense knowledge is deeply context-sensitive. When we use headline only for concept expansion, the context information is so little that introducing commonsense knowledge has comparatively less benefit. When both headline and caption are used, the context information is rich enough to introduce useful commonsense knowledge. However, when all the three fields are used, it shows that the context information increases but the improvement of our approach (0.4%) is much smaller than that of using headline and caption. The reason is that the category field has covered much commonsense knowledge needed for this task. Note that this field consists of tags of multiple categories annotated by librarians working at St. Andrews Library. It also reflects that humans always employ their commonsense knowledge when dealing with document (image) classification. Even we can ascertain

commonsense knowledge is useful in IR because the baseline performance of using the category field is obviously superior to that without using the category field.

When headline and caption are used and the evaluation criterion is P@20, five topics get improvements without lowering the performance of the other topics. Table 2 shows the five topic numbers and their performances. We find that four of the five topics have baseline performance lower than the average (0.3821). This phenomenon indicates that our approach mainly takes effect on “difficult” topics. In other words, it illustrates that the use of commonsense knowledge is necessary for the tasks that are “difficult” for current IR systems.

Table 2. P@20 of 5 critical topics

Topic No.	<5>	<9>	<11>	<18>	<24>
BL	0.20	0.40	0.05	0.00	0.30
OURS	0.25	0.50	0.10	0.05	0.40

4. CONCLUSIONS AND FUTURE WORK

We introduce commonsense knowledge into IR by expanding concepts in image descriptions with spatially related concepts. The experiment results show that our approach is more suitable for precision-oriented tasks and for “difficult” topics. In future work, we will investigate a dynamic weighting scheme to combine scores in a way of query by query. We will also investigate how to apply other kinds of relationships in ConceptNet to other IR tasks.

5. ACKNOWLEDGMENTS

Research of this paper was partially supported by National Science Council, Taiwan, under the contracts NSC94-2752-E001-001-PAE and NSC95-2752-E001-001-PAE.

6. REFERENCES

- [1] Clough, P., Müller, H., Deselaers, T., Grubinger, M., Lehmann, T.M., Jensen, J., and Hersh, W. The CLEF 2005 cross-language image retrieval track. In *Proceedings of the 2005 Cross Language Evaluation Forum*, LNCS, 2006.
- [2] Kim, S.B., Seo, H.C., and Rim, H.C. Information retrieval using word senses: root sense tagging approach. In *Proceedings of the 27th Annual International ACM SIGIR Conference*, 2004, 258-265.
- [3] Lenat, D.B. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 1995, 33-38.
- [4] Liu, H. and Lieberman, H. Robust photo retrieval using world semantics. In *Proceedings of the LREC 2002 Workshop on Creating and Using Semantics for Information Retrieval and Filtering*, 2002, Canary Islands.
- [5] Liu, S., Liu, F., Yu, C.T., and Meng, W. An effective approach to document retrieval via utilizing WordNet and recognizing phrases. In *Proceedings of the 27th Annual International ACM SIGIR Conference*, 2004, 266-272.
- [6] Liu, H. and Singh, P. ConceptNet: a practical commonsense reasoning toolkit. *BT Technology Journal*, 22(4), 2004, 211-226.
- [7] Salton, G. and Buckley, C. On the use of spreading activation methods in automatic information retrieval. In *Proceedings of the 11th ACM-SIGIR Conference*, 1988, 147-160.