UNT at ImageCLEF 2011: Relevance Models and Salient Semantic Analysis for Image Retrieval

Miguel E. Ruiz[1], Chee Wee Leong[2] and Samer Hassan1[2]

University of North Texas,
[1] Department of Library and Information Sciences,
[2] Department of Computer Science and Engineering,
1155 Union Circle 311068,
Denton, Texas 76203, USA

meruiz@gmail.com, cheeweeleong@my.unt.edu , samer@unt.edu

Abstract. This paper presents the result of the team of the University of North Texas in the ImageCLEF 2011 Wikipedia and Medical Image Retrieval tasks. For Wikipedia image retrieval we compare the two query expansion methods: relevance models and query expansion using Wikipedia and flicker as external sources. The relevance models use a classic relevance feedback mechanism for Language models as proposed by Levrenko. The external query expansion mechanism uses an unsupervised two steps method that takes advantage of Salient Semantic Analysis (SSA) using Wikipedia and estimates the "picturability" of terms using Flicker tags. Our results show that SSA and Flickr picturability can be used effectively to create very competitive runs that capture the semantic context of the original query. For Medical Image Retrieval we also use relevance models and query expansion using terms generated by MetaMap.

Keywords: Image Retrieval, Query Expansion, Salient Semantic Analysis, Language Models, Relevance Models.

1 Introduction

Image retrieval is becoming an common user activity on the web as well as in the medical domain. Despite the many advances in image retrieval research there are still some serious problems that need to be further explored such as the well known "semantic gap". This makes the ImageCLEF initiative very relevant to foster research in this area. In this paper we present the results of the University of North Texas (UNT) team in the Wikipedia image retrieval and the adhoc medical image retrieval tasks. This year we were inspired by the idea of exploring unsupervised techniques that can help to close the semantic gap in image retrieval. Our work focused on using relevance models and query expansion using semantic salient analysis and trying to predict the picturability of terms in the query using Filckr tags.

2 Wikipedia Image Retrieval Task

For our participation in the Wikipedia image retrieval task we used the standard 2011 Wikipedia image collection, which is described in more detail in the overview paper of this task [1]. Our goal for this year focused on using corpus based methods to build a query expansion that could capture semantic meaning and identify terms that are more likely to describe images. For this purpose we use Semantic Salient Analysis and Flickr picturability.

Salient Semantic Analysis (SSA) [2] is a method that computes semantic similarity between words based on salient content links from a corpus such as Wikipedia. The meaning of each word is represented by links of salient concepts defined in wikipedia. For example, given the following text from Wikipedia:

"Plants are living organisms belonging to the kingdom **Plantae**. Precise definitions of the kingdom vary, but as the term is used here, plants include familiar organisms such as trees, flowers, herbs, bushes, grasses, vines, ferns, mosses, and green algae."

The semantic meaning for the word "*plants*" is represented by a weighted vector of the salient links: living organisms, kingdom, trees, flowers, herbs, bushes, grasses, vines, ferns, mosses, and green algae.

To measure the semantic association between two terms or between two pieces of text SSA uses a similarity value computed on the co-occurrence with in a window of size k in a given corpus. The similarity value is controlled by a parameter (λ) that represents a threshold of the semantic gap between terms that are perfect synonyms (e,g, tiger-tiger) and near synonyms (e.g., tiger-feline) .

Flickr picturability is a method based on rewarding terms that match tags assigned to images in Flickr. For this purpose the method builds a corpus with the top Flickr tags most related to the query terms and weights them according to the co-occurrence of the term in the contexts of other query terms.

The reader can find the complete details about these methods in our short paper in CLEF 2011 which is available in the conference proceedings [3].

2.1 Collection Preparation

We decided to solve the cross language issue by translating all documents (image captions) from French and German into English using the Bing Translation service. This basically converts the cross-language retrieval problem into an English monolingual retrieval problem. We used not only the captions but also the full text of

the Wikipedia article that includes the image. The collection was indexed using Indri (available at http://www.lemurproject.org/).

2.2 Query Expansion

Query expansion was performed using a two step process that first generates a list of possible candidate words by retrieving the top *m* Wikipedia articles that are more relevant to the original query Q. Then the process adds the SSA scores over all individual concept vectors of each term in Q. After discarding stop words in the titles, remaining words are ranked using a fusion formula that incorporates the Flickr pictutability :

$$Weight(w_i) = tf(w_i) * 1/rank(w_i) * flickr(w_i) \quad (1)$$

Where $tf(w_i)$ is the term frequency w_i across all m Wikipedia titles, $rank(w_i)$ is the highest rank of the title (in descending order) that contains the word w_i, and $flickr(w_i)$ is the Flickr picturability score computed on the corpus.

The second step is the candidate selection. In this step the top W words (ranked in reversed order) by the weight computed using (1) are used as a working set. If the SSA similarity $Sim(Q,w) \geq \alpha$, then the word is added to the expanded query.

This approach basically tries to add picturable terms that are semantically related to the original query.

For our experiments we used the following parameter values that were determined empirically using the ImageCLEF 2010 queries. The number of tokens in the corpus used to compute SSA was set to m=1000, two values for the number of top Wikipedia articles retrieved for candidate selection W=50 and W=150. When measuring similarity, our SSA model was set to $\gamma = 1.2$ and $\lambda = 0.02$

2.3 Retrieval model

We use a standard unigram language model with Dirichlet smoothing, Krovetz stemming and a list of English stopwords. We also use a weighted query with parameter β such that

$$Weighted\_query = \beta\ Q\_original + (1-\beta)\ Q\_expansion \quad (2)$$

In our experiments the value was set to β=0.5 based on the parameters set for Lavrenko's relevance model [4].

2.4 Results

We submitted 7 official runs which are presented on Table 1. All our runs used textual features only. French and German text in the documents was translated to English and use only the English queries.

The first run listed is an unofficial baseline run using language models on the original query terms with no expansion. Surprisingly this baseline achieves a quite competitive MAP of 0.2621 and quite high values for P10 and P20. All runs labeled with SSA use the expanded queries with terms selected based on Salient Semantic Analysis and Flickr picturability scores. The label W indicates whether the run uses the weighted query scheme. Runs labeled with rf use a relevance model to perform pseudo relevance feedback.

From Table 1 we can see that just using the top 50 expanded query terms selected with SSA and Flickr picturability is our lowest performing run (2011_SSA50) even significantly below our baseline. However, when we use the weighted query and retrieval feedback (UNTESU_SSA50Wrf) the performance improves to 0.2794 (6.6% above the baseline). This indicates that to improve retrieval performance with the query expansion method we must use an appropriate weighted query. Selecting the top 150 terms with a weighted query and relevance feedback shows the highest performance of our SSA runs with 0.2820 (7.6%) above the our baseline.

Table 1. Official results in the Wikipedia Image Retrieval task.

Run name	FB/QE	MAP	P10	P20	Rprec	Bpref
Baseline (unofficial)		0.2621	0.5493	0.4434	0.2900	0.2522
2011_SSA50	QE	0.2143	0.3260	0.2900	0.2438	0.2027
UNTESU_SSA150rf	QEFB	0.2292	0.3120	0.2810	0.2476	0.2050
2011_SSA50_FB	FB	0.2327	0.3160	0.2860	0.2543	0.2113
UNTESU_SSA150W	QE	0.2577	0.4060	**0.3510**	0.2835	0.2401
UNTESU_SSA50Wrf	QEFB	0.2794	0.4240	0.3630	0.3107	0.2647
UNTESU_SSA150Wrf	FB	0.2820	0.4200	0.3610	0.3190	0.2679
UNTESU_BLRF	FB	**0.2866**	**0.4220**	0.3650	**0.3276**	**0.2821**

We also submitted a run that uses Lavrenko's relevance model [4] for query expansion (which performs pseudo-relevance feedback on language models). This run

(UNTESU_BLRF) was our highest performing run with MAP=0.2866 (9.3% above our baseline). Overall this run was ranked as the 3rd best textual run in the Wikipedia task. However it seems clear from the results of other teams that a mixed approach using textual and visual features could yield a much higher performance for the task.

2 Medical Image Retrieval Task

For the medical image retrieval task we participated only in the adhoc retrieval task [5]. We indexed the data using Indri with standard parameters. We used an approach that expands queries using MetaMap [6] and identifies whether a specific image modality (e.g. x-rays image) is requested in the query. We added a field to each document with the image modality predicted type provided by the organizers of the medical image classification task [5].

We created structured queries that used the phrase operator to ensure that the multiword terms generated by MetaMap where matched as a single term instead of individual words. We also use a weighted formulation that included three components: the original query terms, the image modality requested in the query, and the expanded terms generated by MetaMap.

Our official results for the ad-hoc medical retrieval runs are presented in Table 2. The results show clearly that query expansion using the structured queries actually decreased performance slightly. Relevance feedback models have the same effect of decreasing performance slightly. Although we still have to do a more thorough analysis of the results it seems that the large number of expansion terms generated by MetaMap is affecting the focus of the query. We probably will need to use a similar technique like the SSA presented in our Wikipedia retrieval task to do a more focused term expansion and rank the terms generated by MetaMap.

Table 2. Official results in the Wikipedia Image Retrieval task.

Run name	FB/QE	MAP	P10	P20	Rprec	Bpref
ESU_Ib_bl		0.1590	0.2670	0.2070	0.1940	0.1890
ESU-Ib_blRF	FB	0.1560	0.2430	0.2100	0.1760	0.1870
ESU_Ib_Struc		0.1540	0.2800	0.2300	0.1870	0.1910
ESU_Ib_StrucRF	FB	0.1350	0.2300	0.2000	0.1610	0.1870

3 Conclusions

We presented in these paper experiments that focus on query expansion methods using external resources as well as using collection based query expansion with pseudo relevance feedback.

In the case of Wikipedia retrieval our results show that query expansion using SSA and Flickr picturability are equivalent to using traditional collection based relevance model for relevance feedback.

For the Medical image retrieval our results are mixed and do not show improvements when using structured queries and query expansion using MetaMap. However, we still need to do a more thorough analysis to try to understand what is hurting performance in these runs.

References

1. Tsikrika, T., Popescu, A, and Kludas, J. Overview of the wikipedia image retrieval task at ImageCLEF 2011. In: CLEF 2011 Working Notes, Amsterdam, The Netherlands, (2011)
2. Hassan, H., Mihalcea, R.: Semantic relatedness using salient semantic analysis. In : Proceedings of AAAI Conference on Artificial Intelligence (2011)
3. Leong, C., Hassan, S., Ruiz, M., Mihalcea, R.: Improving query expansion for image retrieval via saliency and picturability. In: CLEF 2011 Conference on Multilingual and Multimodal Information Access Evaluation , Amsterdam (2011)
4. Lavrenko, V., Croft, B.: Relevance-Based Language Models. In : Proceedings of the ACM SIGIR 2001 Conference on Research and Development in Information Retrieval, New Orleans, LA (2001)
5. Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., de Herrera, A., Tsikrika, T. The CLEF 2011 medical image retrieval and classification tasks. In: CLEF 2011 Working Notes, Amsterdam, The Netherlands, (2011)
6. Aronson, A.: Effective Mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. In : AMIA Annual Symposium, Washington, DC (2001)