Evaluating the Trade-offs Between Diversity and Precision for Web Image Search using Concept-Based Query Expansion

Enamul Hoque, Orland Hoeber, and Minglun Gong
Department of Computer Science
Memorial University of Newfoundland
St. John's, NL, Canada A1B 3X5
{*enamulp, hoeber, gong*}*@mun.ca*

***Abstract*—Even though Web image search queries are often ambiguous, traditional search engines retrieve and present results solely based on relevance ranking, where only the most common and popular interpretations of the query are considered. Rather than assuming that all users are interested in the most common meaning of the query, a more sensible approach may be to produce a diversified set of images that cover the various aspects of the query, under the expectation that at least one of these interpretations will match the searcher's needs. However, such a promotion of diversity in the search results has the side-effect of decreasing the precision of the most common sense. In this paper, we evaluate this trade-off in the context of a method for explicitly diversifying image search results via concept-based query expansion using Wikipedia. Experiments with controlling the degree of diversification illustrate this balance between diversity and precision for both ambiguous and specific queries. Our ultimate goal of this research is to propose an automatic method for tuning the diversification parameter based on degree of ambiguity of the original query.**

***Keywords*-Image search results diversification; concept-based query expansion; diversity-precision trade-off.**

I. Introduction

Image search queries on Web are often are often very short and ambiguous [1]. The difficulty with ambiguous queries is that they can be open to many different interpretations. It is possible that different searchers may enter the same query, but their intentions and needs may vary significantly from one another. In situations such as this, the matching algorithms used by image search engines promote the interpretation that is most common and popular. In our research, we explore an alternate approach that generates a diversified set of images covering the various aspects of the query, under the expectation that at least one of the interpretations matches the searcher's intent. Providing interactive zoom and focus operation can allowing users to visually explore the image space as they seek relevant images [2].

Moving beyond independent relevance ranking, diversification approaches aim to improve the coverage of the search results set with respect to the different senses of the original query. A common method for diversification is query expansion, whereby additional terms are added to the query to generate a collection of new queries that when taken together, are more broad than the original [3]. However, in doing so there is a danger in broadening the query too much, resulting in a potentially significant decrease in precision. That is, the more broad and diverse the search results are, the less chance that a particular search result will be relevant to the searcher's information need. As such, this trade-off between diversity and precision must be studied in order to understand the situations where more or less diversification is beneficial.

Maintaining a good balance between diversity and precision requires an automatic modelling of the searcher's query to determine an appropriate degree of diversification to promote. In many cases, image search queries are inherently ambiguous. For example, "Washington" might be interpreted as "Washington (state)", "Washington D.C.", or "George Washington". Even within a more specific query, searchers might have an interest in seeing images that are related to but not explicitly identified in the query. For example, if a searcher submits a query such as "Hong Kong", they may wish to see some representative images of different landmarks in the Hong Kong area. In other cases, a query might be very specific, and the scope for broadening the query is limited. For example, queries for a particular landmark within a specific setting like "Eiffel Tower Bastille Day" may leave little room for diversifying the search results.

In previous research, we have presented a method for diversifying image search results using concept-based query expansion based on information derived from Wikipedia [2]. In this work, we have modified our approach such that the degree of diversification can be controlled, allowing us to evaluate the trade-offs between diversification and precision among the image search results. The experimental results can provide the basis for automatically determining the degree of diversification based on the level of ambiguity of the query.

II. Related Work

The diversity problem is a challenging topic in Web image search. The problem may be approached from the perspective of promoting semantic diversity [4] or visual diversity [5] among the search results. Although there may be some benefit to promoting both types of diversity, the focus of our research is on semantic diversity. Moreover, in the document centric retrieval methods, the documents themselves provide some useful ways to allow direct comparisons to promote diversification in search results [6]. But in image retrieval, the limited amount of textual metadata associated with the retrieved images are not reliable enough nor sufficient to allow for the computation of similarities between their associated images.

The diversification problem of image search results have been often addressed by semantic and/or visual clustering [4], [5]. However, there are challenges with clustering, including determining a suitable similarity measure, the efficiency of clustering algorithms, determining an appropriate number of clusters to use, and deciding how to order or organize the clusters.

Some have begun to study the trade-off between precision and diversification within the context of image search. For example, van Zwol et al. proposed a method for optimizing this trade-off by estimating a query model from the distribution of tags that favours the dominant sense of an image search query [7]. Our work takes a different approach, enhancing the diversity of image search results through concept-based query expansion [8], [2].

One of the primary challenges associated with concept-based query expansion is to find an appropriate source of knowledge required for the expansion process. It has been noted that many image search queries are associated with conceptual domains that include proper nouns (e.g., people's names and locations) [1]. As such, finding a suitable knowledge base that has sufficient coverage of a realistic conceptual domain is very important. Wikipedia is a good candidate for such a knowledge base since it includes a large number of articles describing people, places, landmarks, animals, and plants. The challenge in using Wikipedia is to design efficient and effective algorithms that can process the semi-structured knowledge to derive meaningful terms for use in the query expansion process.

III. Image Search Results Diversification

In our diversification framework, image search results are explicitly diversified based on concept-based query expansion. For the short and ambiguous queries that are common in image search, query expansion attempts to capture the various different aspects of the query. We model the original query by first extracting the different possible senses of the original query; for each sense, the set of concepts pertaining to the query are extracted from Wikipedia. These concepts are ranked according to their semantic relatedness to the original query, and only the top-N most related concepts are used within the query expansion process to retrieve a diverse range of images.

Within this process, the value of N is an explicit indicator of the degree of diversification. With a smaller value of N, fewer concepts will be used, and the search results will remain more focused. If we increase N, then more concepts will be used for query expansion, and in turn the search results will be more diversified covering more aspects pertaining the query. The fundamental trade-off between diversity and precision is based on the fact that as we increase N, there is a higher chance that a concept will be selected for the query expansion process that is not relevant to the searcher's information needs, resulting in images that are not relevant being included in the search results. In this context, we have developed our concept-based query expansion method such that the degree of diversification can be controlled through the diversification parameter N. This method for query expansion follows three steps: extracting the concepts, ranking the concepts, and retrieving the images.

A. Concept Extraction Using Wikipedia

Matching the user-supplied query Q to the knowledge-base is performed by selecting the best matching articles (referred as the home articles) using Wikipedia's search feature. In the case where the query is ambiguous and Wikipedia suggests multiple interpretations (senses), the ones with higher commonness values are used as the home articles. Here the commonness value of an article is calculated based on how often it is linked by other articles. Then, the candidate concepts for the query are extracted from the in-going links (articles having links to a home article), the out-going links (articles to which a home article links), and the the captions. The end result of this process is the selection of a set of home article(s) $\{h_s | 1 \leq s \leq q\}$ (for q senses of given query Q), along with a list of all the candidate concepts C_{h_s} for each home article h_s. These concepts provide the basis for the automatic query expansion process.

B. Ranking the Extracted Concepts

Due to the rich and interconnected nature of Wikipedia, the number of concepts obtained in the process described above may become very large. Thus, a filtering step is necessary to ensure the quality of the concepts that are extracted. Here, our objective is to select the top-N concepts from among all the candidate articles. Considering the difference in the importance of each of the senses, we distribute these top-N concepts among the candidate concepts C_{h_s} of each home article h_s. As such, the number of related concepts N_{h_s} that are to be selected for a particular home article h_s is determined as follows:

$$N_{h_s} = \frac{|C_{h_s}| \times N}{\sum_{j=1}^{q} |C_{h_j}|}$$

Note that the sum of all N_{h_s} values equals N.

To select N_{h_s} concepts for each home article h_s, it is necessary to rank the candidate concepts C_{h_s} based on their relevance. To achieve this objective, a semantic relatedness measure based on WLM [9] is applied between the home article h_s and each of the candidate articles $c_i \in C_{h_s}$. The outcome of this process is that the top-N concepts are selected from among the candidate articles. These concepts are used as the source for the query expansion. How diversification parameter N affects the precision of search results for each of the senses of the query are discuss in Section IV.

C. Concept-Based Query Expansion and Image Retrieval

In order to ensure that the expanded queries remain focused on the topic of the query itself, the original query Q is pre-pended to each of the top-N related concepts $\{c_r|0 \leq r \leq N\}$ as $< Q, c_r >$. We define c_0 to be null, producing the original query plus N expanded queries.

Given that individual expanded queries have differing degrees of relevance to the original query, we dynamically determine how many images to retrieve for each expanded query based on their relatedness score to the home articles, using the following formula:

$$I_r = \frac{r(c_r, h_s) \times I_t}{\sum_{k=0}^{N} r(c_k, h_s)}$$

Here, r is the same function used to generate the relatedness score in the concept ranking process, and I_t is the total number of images to be retrieved by all of the queries. We set $I_t = 60$ for the purposes of performing the evaluation within this paper, but it can be set to any reasonable number of images. Since the null expanded query (c_0) is the original query, we define $r(c_0, h_s) = 1$ in the above calculation. This dynamic method for determining how many images to retrieve ensures that more images are provided for concepts that are most similar to the original query (even when the original query has multiple meanings). All of the queries are then sent to the Google AJAX Search, and the desired number of images are retrieved.

IV. Evaluation

In this section, we wish to evaluate the inherent trade-off between precision and diversity in detail. In particular, for a set of queries, we explore how the precision changes as diversity in the image search results is promoted.

A. Experimental Setup

For these experiments, we chose 6 query topics, split between those we deemed to be highly ambiguous (having three or four senses) and those we deemed to be slightly ambiguous (having one or two senses). This distribution of different degrees of ambiguity allowed us to examine the effect of the experimental condition (i.e., the varying of the degree of diversification) in the context of ambiguity.

To evaluate the effect of diversification on precision, we retrieved the top 60 search results using our concept-based query expansion method with different values of N, ranging from 0 to 50. Here, $N = 0$ implies that no query expansion has occurred (i.e., the search results are not diversified, and are simply the results provided by the underlying image search engine). At the other extreme, $N = 50$ causes the system to return a highly diversified set of image search results from 50 different associated concepts chosen in the query expansion procedure.

For each of the different senses of the query, assessors were asked to judge the relevance of each image. This assessment of relevance was used as ground truth information in the calculation of the precision scores.

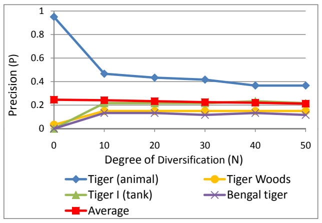
B. Results

In these experiments we measured the precision for each of the test queries as the diversification parameter N was varied from 0 to 50 in increments of 10. Our hypothesis was that as N increased, the distribution of the senses would become more balanced across all of the meanings of the query. This would result in a reduction in the precision for the most common senses of the query, and an increase in the precision for the less common senses. This feature can be readily identified in the graphs in Figure 1. To further understand this effect, we also plotted the average precision across all of the senses (the red lines in the graphs).

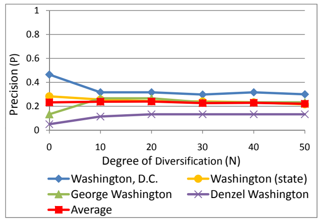
Figure 1 (a - d) shows the results from the highly ambiguous queries. “Tiger” and “Washington” had four different senses; while “Beetle” and “Fuji” had three different senses. The general effect that can be seen here is that the precision for the most common sense (i.e., the blue line in each graph) was automatically reduced as a result of the diversification, even when N was as small as 10. At the same time, the precision for all other senses increased.

Our expectation when designing these experiments was that for queries that have a high degree of ambiguity, it would be necessary to set the diversification parameter rather high in order to capture enough information on all of the different senses. However, it is clear that even with a diversification parameter set at $N = 10$, the desired effect appears. By carefully inspecting the least common senses in these queries (i.e., the curves that have the lowest precision), we can see that in some cases setting $N = 20$ allowed more information on these senses to also be included. Clearly, setting N in the 30 - 50 range did not result in much change the precision across the different senses; on the contrary, it seemed to have had the effect of starting to include some less relevant concepts resulting in slightly reduced precision values across many of the senses. This can be seen not only for individual senses, but also in the gradual dip in the average precision across all senses for each query.

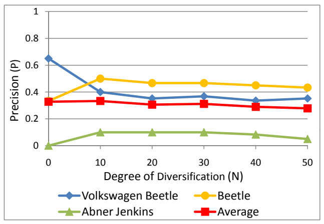
Figure 1 (e - f) shows the results from the slightly ambiguous queries. Query “Beckham” had two different senses; and “Eiffel Tower” had only one sense. As with the highly ambiguous queries, the effect of diversification with $N = 10$ when there are only two senses (“Beckham”) results



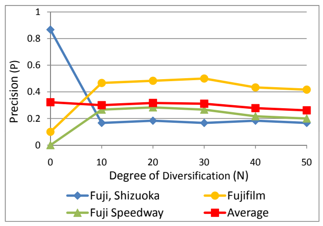
(a) Tiger



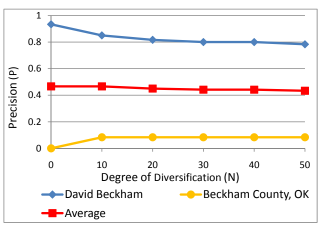
(b) Washington



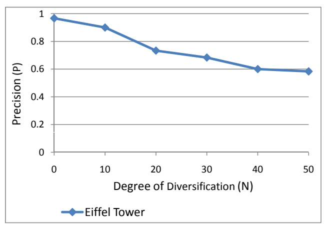
(c) Beetle



(d) Fuji



(e) Beckham



(f) Eiffel Tower

Figure 1. The effect of varying the degree of diversification (N) on precision (P) for highly ambiguous queries that had *four* (a, b) or *three* (c, d) different senses; and slightly ambiguous queries that had *two* (e) or *one* (f) sense(s).

in a balancing of the precision between the senses. However, we can see that even increasing N one step larger to 20 results in a reduced average precision, which is an indication that non-relevant concepts are starting to be included. For the case where there was only one sense ("Eiffel Tower"), it is clear that diversifying the search results can very quickly have a negative effect on the precision, although a small degree of diversification might be tolerable.

As a result of this analysis, we conclude that diversifying the image search results can be very useful for addressing the situation where an ambiguous query has multiple senses. Rather than relying on the search engine to choose the most common sense, we can diversify the image search results and let the user focus on those images that match their needs. The more senses that can be inferred from a query, the more diversification is necessary to sufficiently balance all of these senses in the search results. However, when there are few different senses, the degree of diversification should be limited to avoid including irrelevant concepts and their associated images in the search results.

V. Conclusions

In this paper, we present a novel approach for explicitly diversifying image search results using concept-based query expansion. We evaluated the trade-off between diversification and precision, using a set of test queries of varying ambiguity. The results show that the degree to which diversification is valuable depends on the level of ambiguity of the query; a highly ambiguous query can benefit more from diversification than a very specific query.

Instead of only satisfying the information needs for the most common interpretation of the query, our method provides a more balanced view of the different senses of the query. By choosing the degree of diversification based on the number of senses of the query, the negative impacts on the average precision across all of the senses can be avoided. Coupled with a visual interface for organizing the image search results both visually and conceptually and interaction methods that support dynamic filtering (such as the one described in [2]), search results diversification can be a powerful tool for enhancing image search experiences.

References

[1] P. André, E. Cutrell, D. S. Tan, and G. Smith, "Designing novel image search interfaces by understanding unique characteristics and usage," in *Proceedings of the IFIP Conference on Human-Computer Interaction*, 2009, pp. 340–353.

[2] E. Hoque, G. Strong, O. Hoeber, and M. Gong, "Conceptual query expansion and visual search results exploration for Web image retrieval," in *In Proceedings of the Atlantic Web Intelligence Conference*, 2011, pp. 73–82.

[3] C. J. van Rijsbergen, *Information Retrieval*. London: Butterworths, 1979.

[4] S. Wang, F. Jing, J. He, Q. Du, and L. Zhang, "IGroup: presenting Web image search results in semantic clusters," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2007, pp. 587–596.

[5] R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol, "Visual diversification of image search results," in *Proceedings of the International Conference on the World Wide Web*, 2009, pp. 341–350.

[6] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong, "Diversifying search results," in *Proceedings of the ACM International Conference on Web Search and Data Mining*, 2009, pp. 5–14.

[7] R. van Zwol, V. Murdock, L. Garcia Pueyo, and G. Ramirez, "Diversifying image search with user generated content," in *Proceeding of the ACM International Conference on Multimedia Information Retrieval*, 2008, pp. 67–74.

[8] D. Myoupo, A. Popescu, H. L. Borgne, and P.-A. Moëllic, "Multimodal image retrieval over a large database," in *Proceedings of the International Conference on Cross-Language Evaluation Forum: Multimedia Experiments*, 2009.

[9] D. Milne and I. H. Witten, "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links," in *Proceedings of the AAAI Workshop on Wikipedia and Artificial Intelligence*, 2008, pp. 25–30.