Web metasearch as belief aggregation

Sergio A. Alvarez
Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA 01609
alvarez@cs.wpi.edu

Abstract

Web metasearch requires a mechanism for combining rank-ordered lists of ratings returned by multiple search engines in response to a given user query. We view this as being analogous to the need for combining degrees of belief in probabilistic and uncertain reasoning in artificial intelligence. This paper describes a practical method for performing web metasearch based on a novel transformation-based theory of belief aggregation. The consensus ratings produced by this method take into account the item ratings/rankings output by individual search engines as well as the user's preferences.

Introduction

Web search engines (WSE) use tools ranging from simple text-based search to more sophisticated methods that attempt to understand the intended meanings of both queries and data items. There has been much work in this area in recent years. The link structure of the web has been used to understand the relationships between documents (Chakrabarti *et al.* 1999). Machine learning techniques have been applied to web search (McCallum *et al.* 1999), (Boyan, Freitag, & Joachims 1996). Specialized agents that mine the web have been described (Doorenbos, Etzioni, & Weld 1997). Light is shed on web search from a different perspective by work on human behavior (Macskassy *et al.* 1998). Related problems include those of intelligently recommending scientific papers (Basu *et al.* 1999) and creating digital libraries for efficient indexing and retrieval of scientific documents (Lawrence, Bollacker, & Giles 1999). Reviews of work in web searching include (Lawrence & Giles 1999), (Filman & (guest editors) 1998), and (Lawrence & Giles 1998).

We are interested in web metasearch engines (MSE) (Selberg & Etzioni 1995), (Glover *et al.* 1999), which dispatch user queries to several available WSE; each WSE produces an ordered list of data items in response to the query, and the MSE combines these lists into a single summary list that is then passed on to the user. In the present paper we present a new approach to web metasearching. Numerical relevance ratings are provided as part of our method's output. A useful feature of our approach is that it allows the user to give subjective confidence values for the particular WSE being employed. These confidence values determine the relative importance accorded to the different WSE when producing the final search summary that the user receives as output. Our approach is based on a framework (Alvarez 1997), (Alvarez 2000) that provides a set of tools with which to systematically construct combination operators for belief aggregation, each determined by a different choice of geometric transformation or reference frame in an abstract space (in the present case this space is the space of relevance ratings). Combination operators allow one to uniformly assign numerical relevance ratings to the items found by the WSE being polled by the system, and thus ultimately to produce the final summary list. Our approach assumes that the WSE return numerical ratings in addition to rank-ordered lists of hits. If this is not the case, ratings may be assigned to rankings in some way before combination is to be performed. Algorithms for combining rankings when numerical ratings are not available have been studied previously, e.g. (Freund *et al.* 1998). High flexibility and configurability are two properties that our approach inherits from the theoretical framework of (Alvarez 2000). Our framework provides a natural mechanism to vary the sensitivities of the resulting combination operators to their various inputs. This allows a system based on this approach to adapt according to the user's preferences.

Contents of the paper

The paper begins with a brief discussion of our approach to metasearch based on the new framework for combination operators described in (Alvarez 2000). This is followed by a brief treatment of the incorporation of user preferences. Preliminary experimental evaluation of our approach is provided through two examples included in the above mentioned sections. The conclusions section summarizes our contributions to date and describes work in progress.

Combination operator framework for metasearching

A block diagram for a generic web metasearch system appears below in Figure 1. The user presents a query to the system through the user interface. The query is passed on to the dispatcher, which decides which of the available web search engines (WSE) to submit the query to. Once the polled WSE have returned hit lists in response to the query, the combination mechanism produces a consensus list of hits (and ratings) which is presented to the user.

We focus in the present paper on the combination mechanism shown in the Figure. As described above, the purpose of this mechanism is to produce a consensus list of hits (and ratings) from the individual lists returned by the search engines polled by the dispatcher. We view this process as being analogous to the combination of degrees of belief in probabilistic and uncertain reasoning in artificial intelligence. In the latter context, one needs to combine subjective probabilities or similar numerical measures associated with different sources of information into a single summary measure that reflects their individual contributions. Our approach to web metasearch is to first combine the numerical relevance ratings for the hits produced by different search engines into a single summary rating. This provides an objective basis for the final ordering produced by the metasearch system. Specifically, the combined summary rating is used as the dimension along which the results are ordered in the combined output of the metasearch system.

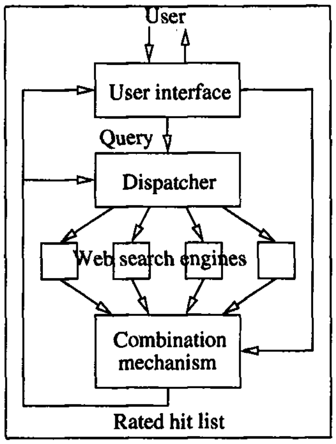

Figure 1: A generic web metasearch engine

Axioms for combination operators

The process of obtaining a consensus rating as described above may be achieved through a *combination operator*. Such operators are important in a variety of applications ranging from knowledge-based systems to experimental psychology. The following definition (see (Alvarez 2000)) makes the notion of combination operator precise. Although only binary operators are mentioned explicitly in the definition below, n-ary operators for $n > 2$ can of course be constructed simply by composing binary operators with themselves.

Definition 0.1. A function $\oplus : [0,+1] \times [0,+1] \rightarrow [0,+1]$ is an *admissible combination operator* if and only if it satisfies the following axioms:

Commutativity $p \oplus q = q \oplus p$

Monotonicity $(\cdot) \oplus q$ is an increasing function for each q

Boundary values $0 \oplus q = q, \;\; 1 \oplus q = 1$

Notice that associativity is not required. This allows combination operators that are sensitive to the order in which ratings are presented, which can be desirable in the present context of rank-ordered rating lists.

Frame transformations

In (Alvarez 2000), a framework is developed that allows the systematic generation and manipulation of admissible combination operators. The framework starts from a single well-chosen canonical form and uses geometric transformations to generate multiple combination operators from this canonical form. The resulting combination operators may be expressed as follows:

$$a \oplus_\beta b = \beta^{-1}(\beta(a) + \beta(b)) \tag{1}$$

Here, a and b represent two relevance values between 0 and 1; $\oplus_\beta$ is the combination operator associated with a given choice of the *frame transformation* β that appears on the right-hand side of Eq. 1. It may be shown (Alvarez 2000) that the combination operator of Eq. 1 is admissible in the sense of Def. 0.1 if and only if the frame transformation $\beta : [-1, +1] \to [-\infty, +\infty]$ is increasing and satisfies the boundary conditions $\beta(0) = 0$, $\beta(+1) = +\infty$.

Nonlinear scaling

In the frame transformation framework, combination operators occur naturally in parametrized families; for each *steepness* value $t > 0$ one has a combination operator obtained from the standard $t = 1$ version by "nonlinear scaling" as follows:

$$p \oplus_t q = \beta^{-1}\left(t(\beta(p) + \beta(q))\right) \tag{2}$$

The scaling mechanism given here is extremely useful as a means of adjusting the sensitivity of the resulting combination operator. In the present context one seeks a nonlinear average of the available relevance ratings, and so the most natural choice for the scaling parameter t is $1/n$, where n is the number of WSE being polled. It is also possible to specify nonlinear weighted averages by choosing different scaling parameters for different WSE. This allows fine tuning of the operator to match a given user's rating preferences, and to compensate for differences among rating scales for different WSE.

An admissible combination operator

We now give an example of a combination operator obtained via the frame transformation framework described above. Further examples may be found in (Alvarez 1997). By choosing the frame transformation β to be the function $\tanh^{-1}$ in Eq. 1, we obtain the following very simple expression for the associated combination operator:

$$p \oplus q = \tanh\left(\tanh^{-1} p + \tanh^{-1} q\right) = \frac{p+q}{1+pq} \tag{3}$$

This operator may be shown to have interesting interpretations in terms of probability theory, Dempster-Shafer evidence theory, and even Minkowski spacetime geometry (Alvarez 1997). Nonlinear scaling by t as in Eq. 2 generalizes the combination operator of Eq. 3 to the following family:

$$p \oplus_t q = \frac{\left(\frac{1+p}{1-p}\right)^t - \left(\frac{1-q}{1+q}\right)^t}{\left(\frac{1+p}{1-p}\right)^t + \left(\frac{1-q}{1+q}\right)^t} \tag{4}$$

We will use this combination operator with $t = 0.5$ below to carry out a metasearch example. The above operator may be extended to an n-ary operator by considering n operands instead of 2 in Eq. 2.

Preliminary evaluation: an example of ratings combination

The query "web metasearch" was presented to two WSE: Excite and WebCrawler. The top 5 hits from the resulting lists are shown below. A consensus list was then computed through the belief aggregation approach of the present paper, using the combination operator with frame transformation $\beta(x) = \tanh x$ described above in Eq. 4. The default value $t = 0.5$ was used for the nonlinear scaling parameter.

```
Excite Results

    67% MetaSearch
    http://metasearch.langenberg.com/

    65% MetaSearch inc.
    http://www.metasearchinc.com/

    64% Directory of MetaSearch Engines
    http://www.searchiq.com/directory/multi.htm

    63% Metasearch
    http://www.metasearch.com/

    63% Verio Metasearch
    http://search.verio.net/


WebCrawler Results

    64%     W3 Search Engines
    http://cuiwww.unige.ch/meta-index.html

    61%     Directory of MetaSearch Engines
    http://www.searchiq.com/directory/multi.htm

    60%     MetaSearch
    http://metasearch.langenberg.com/

    59%     SavvySearch
    http://www.savvysearch.com/

    58%     Verio Metasearch
    http://search.verio.net/
```

```
Consensus Ratings/Ranking
(using tanh combination with steepness 0.5)

  .6363 MetaSearch
   http://metasearch.langenberg.com/

  .6252 Directory of MetaSearch Engines
   http://www.searchiq.com/directory/multi.htm

  .6056 Verio Metasearch
   http://search.verio.net/

  .3693 MetaSearch inc.
   http://www.metasearchinc.com/

  .3619 W3 Search Engines
   http://cuiwww.unige.ch/meta-index.html

  .3546 Metasearch
   http://www.metasearch.com/

  .3264 SavvySearch
   http://www.savvysearch.com/
```

The following two observations may be extracted from these results.

1. A major factor in determining the consensus ranking of a site is the number of search engines that retrieved it. For example, SavvySearch, the fourth-ranked site in the WebCrawler list, appears at the very bottom in the consensus list in part because this site is not among Excite's top 5 hits. Note that this phenomenon is dependent on the total number of hits considered in computing the consensus ratings. If the top 100 hits were used instead of the top 5, the relative positions in the final list would be different. Indeed, for the query used in the present example, Excite lists SavvySearch among its top 100 hits and gives it a rating of .59; this would lead to moving SavvySearch above Excite's top ranked site in the resulting consensus list. This sensitivity to the number of WSE that retrieved a given site decreases as the total number of WSE polled increases.
2. The relative ordering of two sites in the final list indeed depends on the numerical ratings attributed to these sites by the WSE considered. In the above example, the final ranking of the top three sites is consistent with the first WSE's ranking but not with that of the second WSE; this can be traced to the numerical relevance ratings for the two WSE.

Modeling user preferences

Our approach addresses user preferences in two ways. First, it allows the user to specify his/her relative confidence in different web search engines (WSE) by providing a confidence value between 0 and 1 for each; these confidence values are then used by the system to weigh the ratings output by the corresponding WSE. Second, our approach allows fine tuning of the combination mechanism so that consensus values that accurately reflect the user's subjective ratings are produced. In both cases, it is possible for a specific system based on this approach to learn the appropriate settings through passive user feedback based on the user's behavior while using the system.

Specifying confidence in a WSE

The user's *confidence* in a given WSE E is a number $c(E)$ between 0 and 1 that reflects the degree to which the user trusts results returned by E. The higher the confidence, the more trustworthy the user considers this WSE to be for the given type of query. Our approach allows confidences to be incorporated quite easily. Given a frame transformation β, and given confidences c_i for the available WSE E_i, $i = 1..n$, one aggregates ratings r_i according to the following combination operator:

$$\oplus r_i = \beta^{-1}\left(t\sum_{i=i}^{n}\frac{c_i}{\frac{1}{n}\sum_{i=1}^{n}c_i}\beta(r_i)\right) \qquad (5)$$

WSE with higher confidence values are given greater weight in producing the combined rating through Eq. 5. The denominator of the fractions in Eq. 5 is needed to correctly scale the resulting values. Adaptation to user preferences becomes possible by experimentally estimating the user's ratings and viewing Eq. 5 as a nonlinear regression equation for the confidence values c_i and the steepness parameter t. These parameters may also be adjusted to compensate for differences in rating steepness among different WSE.

Example We revisit the example considered previously. We now somewhat arbitrarily attribute a confidence of 0.25 to the first WSE (Excite) and we retain the default confidence of 1.0 for the second WSE (WebCrawler). The resulting consensus ratings and ranking are shown below.

```
Consensus Ratings/Ranking
(tanh combination, t=0.5, c1=0.25, c2=1.0)

  .6161 Directory of MetaSearch Engines
   http://www.searchiq.com/directory/multi.htm

  .6148 MetaSearch
   http://metasearch.langenberg.com/

  .5904 Verio Metasearch
   http://search.verio.net/

  .5417 W3 Search Engines
   http://cuiwww.unige.ch/meta-index.html
```

```
.4946 SavvySearch
 http://www.savvysearch.com/

.1538 MetaSearch inc.
 http://www.metasearchinc.com/

.1472 Metasearch
 http://www.metasearch.com/
```

Comparing the above results with those from the example in the previous section, one sees that the ratings now agree more closely with those returned by WebCrawler, as expected. In fact, notice that the consensus ranking of the top three sites has actually changed from the Excite ranking to the WebCrawler ranking. This demonstrates that the confidence values have an impact on the final ranking, not just on the numerical values of the ratings. The purpose of the above example is merely to illustrate the effect of changing the WSE confidence values. We stress that our approach allows extracting the confidence values on the basis of feedback (either active or passive) from the user.

Conclusions

We have described an approach to web metasearch based on a new theory of belief aggregation in probabilistic reasoning. Our approach uses combination operators derived through transformations in the space of relevance ratings to produce consensus ratings for the set of web search engines (WSE) being polled. The consensus ratings lead to a rank-ordered consensus list that is output by the metasearch system. Our approach includes a mechanism that allows the use of confidence values for the WSE being used. These confidence values may either be specified by the user or else they may be learned by the system based on observations of the user's behavior. Our method is useful when numerical ratings are available for the rank-ordered lists returned by the WSE. We believe that our approach opens up interesting possibilities for interaction between probabilistic/uncertain reasoning and information retrieval techniques. Further experimental evaluation of this approach is needed.

References

Alvarez, S. 1997. Rational comparison of probabilities via a blow-up conjugacy. Research Report 97-NA-010, Center for Nonlinear Analysis, Carnegie Mellon University.

Alvarez, S. 2000. Belief combination by means of frame transformations. In *Sixth International Symposium on Artificial Intelligence and Mathematics (AMAI-2000).*

Basu, C.; Hirsh, H.; Cohen, W.; and Nevill-Manning, C. 1999. Recommending papers by mining the web. In *Working Notes of the Workshop on Learning about Users, and Working Notes of the Workshop on Machine Learning for Information Filtering, Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99).*

Boyan, J.; Freitag, D.; and Joachims, T. 1996. A machine learning architecture for optimizing web search engines. In *Proceedings of the AAAI workshop on Internet-Based Information Systems, AAAI Technical Report WS-96-06.*

Chakrabarti, S.; Dom, B.; Gibson, D.; Kleinberg, J.; Kumar, S.; Raghavan, P.; Rajagopalan, S.; and Tomkins, A. 1999. Mining the link structure of the world wide web. *IEEE Computer.*

Doorenbos, R.; Etzioni, O.; and Weld, D. 1997. A scalable comparison-shopping agent for the world-wide web. In *Autonomous Agents '97.*

Filman, R., and (guest editors), S. P. 1998. Special issue on internet search technologies. *IEEE Internet Computing.*

Freund, Y.; Iyer, R.; Schapire, R.; and Singer, Y. 1998. An efficient boosting algorithm for combining preferences. In Shavlik, J., ed., *Proc. Fifteenth Int. Conf. Machine Learning (ICML '98)*, 170–178.

Glover, E.; Lawrence, S.; Birmingham, W.; and Giles, C. 1999. Architecture of a metasearch engine that supports user information needs. In *Eighth International Conference on Information and Knowledge Management (CIKM '99).*

Lawrence, S., and Giles, C. 1998. Searching the world wide web. *Science* 280:98.

Lawrence, S., and Giles, L. 1999. Accessibility and distribution of information on the web. *Nature* 400:107–109.

Lawrence, S.; Bollacker, K.; and Giles, C. 1999. Indexing and retrieval of scientific literature. In *Eighth International Conference on Information and Knowledge Management (CIKM '99).*

Macskassy, S.; Banerjee, A.; Davison, B.; and Hirsh, H. 1998. Human performance on clustering web pages: A preliminary study. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD98)*, 95. AAAI Press/MIT Press.

McCallum, A.; Nigam, K.; Rennie, J.; and Seymore, K. 1999. Building domain-specific search engines with machine learning techniques. In *AAAI-99 Spring Symposium.*

Selberg, E., and Etzioni, O. 1995. Multi-service search and comparison using the metacrawler. In *Proc. 1995 World Wide Web Conference.*