FUZZY LOGIC FOR SEARCH ENGINE RANKING

Longzhuang Li, Yi Shang, and Wei Zhang
Dept. of Computer Engineering and Computer Science
University of Missouri-Columbia
Columbia, MO 65211

Abstract

In this paper, we present two fuzzy logic methods to evaluate the performance of search engines, e.g. precision, on the Web. A search engine's precision is evaluated in two steps: (a) compute the crisp relevance scores of hits from each search engine, and (b) rank the search engines using the fuzzy inference system or the extended version of Choquet fuzzy integral. In computing the crisp relevance scores of hits, we use four relevance scoring algorithms: vector space model, Okapi similarity measurement, cover density ranking, and three-level scoring method. For the fuzzy inference system, the rankings of search engines are based on the number of documents that are inferred as relevant or slightly relevant. For the extended version of Choquet fuzzy integral, we rank the search engines based on the statistical comparison of the *probability of win*. In our experiments, six popular search engines, *AltaVista*, *Fast*, *Google*, *Go*, *iWon*, and *NorthernLight*, are evaluated based on two sets of short queries, (a) 1000 queries on *parallel and distributed computing* (PDC) and (b) 1000 queries on *knowledge and data engineering* (KDE), under the default search mode of each search engine. Overall, *Google* is the best, and *NorthernLight* is second best for both fuzzy logic methods.

1 INTRODUCTION

Search engines are designed to help users to quickly find useful information on the Web. With thousands of search engines available and each with different ranking methods and different coverage, finding the one that gives the best results for a query becomes a difficult task.

In previous studies, both manual and automatic methods have been developed to evaluate the precision of search engines given a set of queries. Due to the fast changing nature of both the Web and the search engines, as well as the large amount of information on the Web, automatic relevance evaluation is a key. Usually, the automatic evaluation methods adapt similarity measures used in the traditional information retrieval (IR) area. It is found that some IR relevance scoring methods, such as Okapi similarity measurement, work well on both traditional text data and Web documents at the TREC conference [3]. In this study, we evaluate the performance of search engines only based on the content information because it is found at the TREC conference that the link information may not result in better rankings [4].

The relevance of a Web page to a query is a fuzzy concept and is hard to define precisely. The use of fuzzy set theory in IR significantly fits this requirement: express the similarity between a Web page and a query as a fuzzy relation. To apply the fuzzy set theory to evaluate the performance of search engines, we need to use the appropriate scoring methods to calculate the crisp similarity values at first. Previous studies show that the performance of search engines depends on the performance measures used and query domains [7]. Due to the lack of criteria to judge which scoring method is the best, we can combine the relevance scores from different methods into one score, which may better represent the relevance of the document to a query.

In the paper, we propose two ways to combine the relevance scores from different scoring methods, (i) we develop a fuzzy inference system, which combines scores from two methods: Okapi similarity measurement and three-level scoring method. (ii) we treat the relevance score from each scoring method as a fuzzy number, and adapt the extended version of the Choquet fuzzy integral [2] to fuse the similarity information from different scoring methods.

2 RELEVANCE EVALUATION METHODS

To calculate the similarity between a query q and a document x_i, let N be the the total number of documents in the collection, m the number of unique terms in the document collection, f_{ij} the number of occurrence of a term y_j in the document x_i, d_j the number of documents containing the term y_j, and v_j the frequency of the term y_j in the query q. In our experiments, we use four relevance scoring methods, they are vector space model, Okapi similarity measurement, cover density ranking, and three-level scoring method.

Vector Space Model (VSM) [8]. The similarity is

$$sim_{vs}(q, x_i) = \sum_{j=1}^{m} f_{ij} (\log(N/d_j))^2 \tag{1}$$

Okapi Similarity Measurement (Okapi) [3]. The similarity is

$$sim_o(q, x_i) = \sum_{j=1}^{m} v_j \cdot \frac{f_{ij} \cdot log(\frac{N-d_j+0.5}{d_j+0.5})}{2 \cdot (0.25 + 0.75 \cdot \frac{dl}{avdl}) + f_{ij}} \tag{2}$$

where dl is the length of the document (in bytes), and $avdl$ is the average document length in the collection (in bytes).

Cover Density Ranking (CDR) [1]. It contains two steps: (a) Documents are ranked by coordination level. (b) The documents at each coordination level are ranked to produce the overall ranking.

$$Sim_c(q, x_i) = \sum_{j=1}^{n} I(p_j, q_j) \quad I(p_j, q_j) = \begin{cases} \frac{\lambda}{q_j - p_j + 1} & \text{if } q_j - p_j + 1 > \lambda \\ 1 & \text{otherwise} \end{cases} \tag{3}$$

where p_j and q_j are the positions of terms, and λ is set to 16.

Three-Level Scoring Method (TLS) [7]. It also has two steps: (a) a raw score is calculated as:

$$A(q, x_i) = \frac{t_n \cdot k^{n-1} + t_{n-1} \cdot k^{n-2} + \cdots + t_1}{k^{n-1}} \tag{4}$$

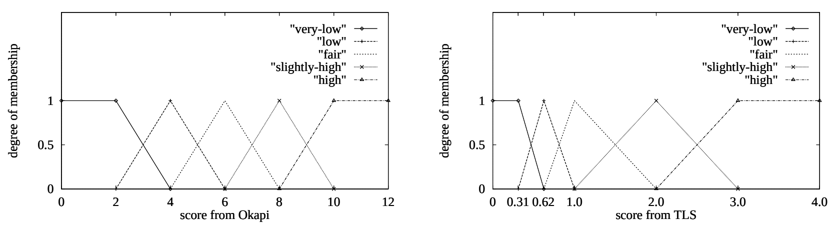
where k is a constant, t_i is the number of occurrence of the sub-phrases of length i, and n is the number of terms in the query q. (b) Convert $A(q, x_i)$ to a three-level similarity score through thresholding, with value 2 for relevant, 1 for partially relevant, and 0 for irrelevant.

In the experiments, we use the raw score $A(q, x)$ instead of the thresholded similarity score. k is set to 10.

3 A FUZZY INFERENCE SYSTEM

The fuzzy relevance inference system consists of four steps: fuzzification, inference, composition, and defuzzification [6].

Fuzzification. The crisp values are fuzzified by the five membership functions describing the linguistic term. The crisp values of scores from Okapi and TLS can be transformed into the linguistic term sets O_k(score) = {very low, low, fair, slightly high, high} and $T_l(score)$ ={very low, low, fair, slightly high, high}, respectively, see figure 1. These settings come from our empirical observance.



(a) (b)

Figure 1: Membership function of the fuzzy scores for (a) Okapi method, and (b) TLS method.

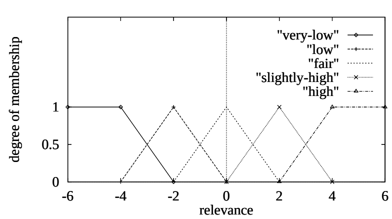


Figure 2: Membership function of the fuzzy relevance.

Inference Engine. A fuzzy proposition can be constructed as an "IF-THEN" statement such as "if the TLS score T_l is high and the Okapi score O_k is high then the document is of high relevance". Basically 25 rules are used to do *MIN* inferencing. The conclusion parts of the 25 rules have five linguistic terms which describe Web documents' relevance: R={very low, low, fair, slightly high, high}. The 25 inferencing rules are shown in table 1, and the membership functions for the output are described in figure 2.

Table 1: Fuzzy inference rules for predicting the relevance of Web documents. *s.* stands for slightly and *S.* for Slightly.

		Okapi Fuzzy Score				
		Very Low	Low	Fair	S. High	High
TLS Fuzzy Score	Very Low	very low	low	low	low	low
	Low	low	low	low	fair	fair
	Fair	low	low	fair	fair	fair
	S. Hight	fair	fair	fair	s. high	s. high
	High	fair	fair	s. hight	s. high	high

Composition. All of the fuzzy sets assigned to the conclusion part of the rules under inference are combined together to generate a single fuzzy set for each output variable. The maximum composition operator (MAX) is used to construct the combined output fuzzy set.

Defuzzification. The value from composition sub-procedure is converted to a crisp value of relevance. The center of maxima method is used to for this defuzzification.

4 FUZZY INTEGRAL METHOD

The performance evaluation results of search engines depend on the the relevance scoring methods used, and usually it is hard to judge which scoring method is the best. Hence, we use the extended version of Choquet fuzzy integral to fuse the scores from different relevance scoring methods because it assumes that the confidence function is a fuzzy number which is not necessarily bounded in [0, 1].

Let $\mathbf{h} : \mathbf{X} \rightarrow \Im\,[0, 1]$ where $\Im([0, 1])$ is a fuzzy power set of [0, 1]. suppose $\mathbf{h}_1, \ldots, \mathbf{h}_n$ are confidence fuzzy values of the scoring methods $\mathbf{x}_1, \ldots, \mathbf{x}_n$ respectively, and let $[\mathbf{h}_i]_\alpha = \lfloor \underline{[\mathbf{h}_i]_\alpha}, \overline{[\mathbf{h}_i]_\alpha} \rfloor$ for $1 \le i \le n$ and $0 \le \alpha \le 1$. The extended version of Choquet fuzzy integral with respect to a fuzzy measure g is defined as [2]:

$$[\int_c h \circ g]_\alpha = [\int_c \underline{[\mathbf{h}]_\alpha} \circ g, \overline{[\mathbf{h}]_\alpha} \circ g] \quad \text{for } 0 \le \alpha \le 1 \tag{5}$$

We use α-cut to compute the fuzzy integral because the measure is still defined in the standard sense. In our experiments, λ-fuzzy measures are used, and the fuzzy density computation is modeled after U-uncertainty [6].

5 EXPERIMENTAL RESULTS

The experiments consist of the following major steps: (i) submit each query to each target search engine, and randomly pick one from the top 20 returned links, and compute the relevance score of the chosen link using four scoring methods: VSM, Okapi, CDR, and TLS; (ii) use the fuzzy logic inference system to combine scores from two scoring methods: Okapi and TLS, and rank all search engines based on the number of documents that are inferred as slightly relevant or relevant; (iii) use the extended version of the Choquet fuzzy integral to fuse the relevance scores from two, three, or four of the four scoring methods, and rank all search engines based on statistical comparison of the *probability of win* (P_{win}) [5].

In our experiments, queries from two query sets, KDE and PDC, are submitted to the default search mode of the six search engines, and the retrieved Web documents are converted to ASCII file with all HTML tags and scripts removed. Only lowercase queries are used since different engines may treat capitalized queries differently. To reduce the possibility that documents on the Web are changed between evaluations, the same query is submitted to the target search engines simultaneously through multi-threading in our search agent.

Figure 3 shows the rankings of six search engines from the fuzzy inference system. For both query sets, KDE and PDC. *Google* is the best, *NorthernLight* the second best, the rankings of other four search engines are *Go, Fast, AltaVista* (AV), and *iWon* in turn. Among the 1000 Web documents returned for each search engine for each query set, *Google* has the largest number of documents for both high relevant documents and slightly high relevant documents, *NorthernLight* (NL) comes at the next for two kinds of the relevant documents, and *iWon* is the search engine with the least number of the two kinds of the relevant documents.

Figure 4 shows the rankings of six search engines after the fusion of the relevance scores from two, three, or four of the four scoring methods using extended version of Choquet fuzzy integral and the *probability of win* based on two query sets: KDE and PDC. *Google* always is the best, *NorthernLight* the second best, and *Go* the third best mostly, while the rankings of other search engines vary with the query set and the scoring methods, which illustrates that the selection of scoring methods and the query sets may affect the rankings of search engines. *AltaVista, Fast* and *iWon* are the worst three ones. These results are very consistent with those drawn from the fuzzy inference system. During the statistical comparison of the *probability of win*, we found that most of these P_{win} values between two search engines are either near 1 or near 0, meaning the confidence that one search engine is better (or worse) than the other one is very high.

6 CONCLUSION

In this paper, we evaluate the precision performance of six search engines by using two fuzzy methods to combine the relevance scores from different scoring methods. Our experimental results show that these six search engines mostly perform consistently using two fuzzy methods, but may vary to some extent under different scoring methods and query sets. Overall, *Google* is the best. *NorthernLight* is second best under both fuzzy methods.

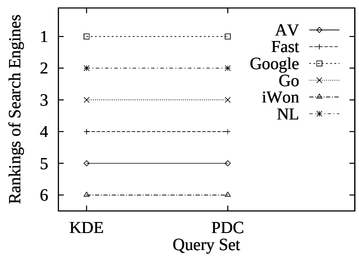


Figure 3: Rankings of six search engines from the fuzzy inference system based on two query sets KDE and PDC.

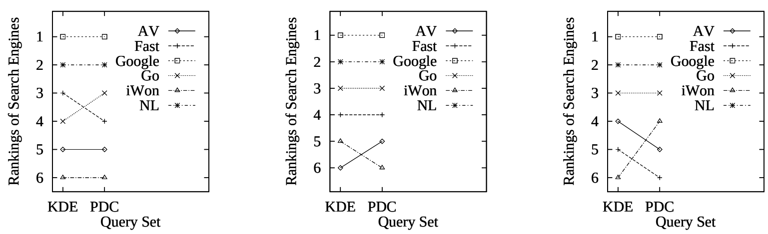


Figure 4: Rankings of six search engines from the extended version of Choquet fuzzy integral based on two query sets KDE and PDC, by fusing relevance scores from (a) Okapi and TLS; (b) Okapi, TLS, and CDR; (c) Okapi, TLS, CDR, and VSM.

References

[1] Charles L.A. Clarke, Gordon V. Cormack, and Elizabeth A. Tudhope. Relevance ranking for one to three term queries. *Information Processing & Management*, 36:291–311, 2000.

[2] M. Grabisch, H. T. Nguyen, and E. A. Walker. *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inderence*. London, Kluwer Academic Publishers, 1994.

[3] David Hawking, Peter Bailey, and Nick Craswell. Acsys trec-8 experiments. In *Proceedings of the TREC-8*, 1999.

[4] David Hawking, Ellen Voorhees, Nick Craswell, and Peter Bailey. Overview of the trec-8 web track. In *Proceedings of the TREC-8*, 1999.

[5] A. Ieumwananonthachai and B. W. Wah. Statistical generalization of performance-related heuristics for knowledge-lean applications. *Int'l J. of Artificial Intelligence Tools*, 5(1 2):61–79, June 1996.

[6] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice hall, 1995.

[7] L. Li and Y. Shang. A new statistical method for evaluating search engines. In *Proc. IEEE 12th Int'l Conf. on Tools with Artificial Intelligence*, Vancouver, British Columbia, Canada, 2000.

[8] G. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Addison-Wesley Series in Computer Science. Addison-Wesley Longman Publ. Co., Inc., 1989.