

User-Centric Multi-Criteria Information Retrieval

Shawn R. Wolfe^{†‡}, Yi Zhang[†]

[†]School of Engineering, University of California Santa Cruz, Santa Cruz, CA, USA

[‡]Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA, USA
{srwolfe, yiz}@soe.ucsc.edu

ABSTRACT

Information retrieval models usually represent content only, and not other considerations, such as authority, cost, and recency. How could multiple criteria be utilized in information retrieval, and how would it effect the results? In our experiments, using multiple user-centric criteria always produced better results than a single criteria.

Categories and Subject Descriptors:

H.3.3 [Information Search and Retrieval]: retrieval models

General Terms:

Algorithms, Performance, Experimentation

Keywords:

multi-criteria decision making

1. INTRODUCTION

The goal of an information retrieval system is to help the user find good documents. Creating a clear definition of what is a good document remains a challenging problem, however, so often the utility of the document is used as an approximation of the user's criteria. In information retrieval research, utility is usually reduced to a narrow definition of "topical relevance" or "related to the matter at hand (i.e., aboutness)". However, prior research has found that a wide range of factors (such as personal knowledge, topicality, quality, novelty, recency, and authority) affect human judgments of relevance. Information novelty is one specific example of an additional implicit criterion that has been studied in the context of search, summarization, filtering and topic detection and tracking. Multiple criteria have also been used in some operational aspect of several recommender systems [3], and more complex rank-based methods have used multiple criteria to support search [2].

This motivates us to explore a more complex representation of utility, using multi-criteria decision theory, to explicitly incorporate multiple criteria in hope of better representing the user's need. Examples of user preferences that go beyond content include: preferring a less relevant article on appendicitis symptoms from the Mayo clinic than a more relevant article on a less authoritative personal homepage; preferring a less relevant article on learning to rank methods from Wikipedia than a more relevant one that incurs a fee; or preferring a less relevant but more recent article mentioning an election recount result over a more relevant but out-of-date article from the USA Today. Unlike much

of the related literature, we are interested in problems that have multiple user-stated criteria, rather than techniques that combine multiple features or the output of multiple methods. There are two major potential advantages of our approach. First, it gives the system an ability to explicitly optimize the user-specified multi-criteria utility. Second, the user can better understand how options were ranked.

The operations research community has extensively studied the use of multiple criteria in multi-criteria decision making (MCDM), also known as multi-criteria decision analysis, which aids decision makers in making difficult choices evaluated under potentially conflicting criteria. A variety of methods have been developed for MCDM, ranging from straightforward single formula methods to more complex methods that use multiple stages to induce a ranking. Though these techniques are designed for decision analysis, it is worth exploring how can they be adapted to the ranking problem of information retrieval. As a starting point, we have applied MCDM techniques to two different information retrieval applications: air travel booking, which has no dominant criterion (e.g., content); and information filtering (of news articles), which has no explicit query. Airline ticket booking is a particularly interesting search problem, because it differs significantly from other commonly studied information retrieval problems (such as web document retrieval). It lacks a single criterion (e.g. content) that is overwhelmingly vital to search results, and in general is likely to have multiple criteria that are important. To contrast with the airline ticket task, we examined the news filtering problem as a task more aligned with traditional information retrieval. There are many criteria a news filtering user might use to judge a new item. In practice, the ratings for a news item on each criteria will be unknown and must be estimated by the filtering system. Based on the estimation of these criteria, the filtering system can further predict whether a user would like the news or not and make filtering decision accordingly.

To evaluate the potential of MCDM in information retrieval, we adapted two MCDM algorithms [4] and compared them to a single-criterion baseline. For the MCDM algorithms, each criterion is given a weight in advance, with the sum of these weights equal to one. We translated a user-specified priority ranking of the criteria into a weights for simplicity, though direct weighting by users is also possible. The simpler of the two algorithms is the weighted sum: the score for each option (e.g., document) is simply the sum of each rating by each criterion multiplied by its corresponding weight. The second method, ELECTRE II, is an outranking method which orders the different options directly by com-

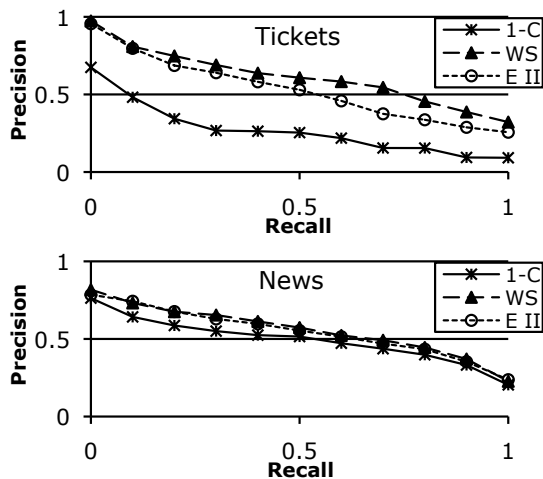


Figure 1: Interpolated precision and recall for single-criterion (1-C), WeightedSum (WS) and ELECTRE II (E II) methods.

binning combining partial orderings with progressively more relaxed consistency conditions. Due to its relative complexity, the details of the ELECTRE II method are beyond the scope of this paper. For our baseline method, the most appropriate single criterion for the given task was selected, and the options were ranked directly by the ratings on this single criterion.

2. EXPERIMENTS

For the ticketing application, we culled information from several online databases [1] to develop a representative set of ticketing options and expected delay profiles. We created five ticketing tasks using different criteria for each, and asked three subjects to mark tickets as relevant or not relevant, according to the task. The following criteria were identified: 1) the desired origin of the flight; 2) the desired destination of the flight; 3) the desired quarter (temporal) of the flight; 4) the price of the fare; 5) the expected flight time; 6) the number of connections; 7) the expected delay and standard deviation; 8) popularity, defined as the number of tickets sold for this final destination 9) stopover popularity, defined as above, but for the connection airports (presumes sightseeing at the connection is possible).

For the news filtering application, we used a data collected from a previous study: more information on the dataset is provided by Zhang [5]. In that study, approximately 20 users rated news articles on several criteria from a corpus of almost 9000 articles. On this data set, the following criteria are included: novelty, authority, readability, and relevancy (to the category assigned to the news article).

Figure 1 shows the average interpolated precision and recall averaged over all subjects and tasks. On the ticketing application, the MAP (mean average precision) averaged over all tasks and subjects was 0.250, 0.586, and 0.511 for the baseline, weighted sum and ELECTRE II, respectively; for news filtering, the MAP averaged over all subjects was 0.463, 0.544, and 0.534 for the baseline, weighted sum and ELECTRE II, respectively. Though the MCDM methods

performed better than the single-criterion baseline in both IR applications, the gain was slight for the news filtering application. This is likely due to the correlation of the criteria: in the news filtering application, there was a high degree of correlation with the criteria and the target attribute (ranging from 0.47 to 0.74) and between criteria, whereas the correlation was much lower in the ticketing application. Nonetheless, both MCDM methods were able to slightly increase performance even with what little additional information was available in the additional criteria, and never hurt performance in our experiments. ELECTRE II did not perform as well as the simpler weighted sum algorithm. It may be the the domains chosen were not suited to this algorithm; ELECTRE II is designed to find compromise solutions in the presence of conflicting criteria, which was not particularly problematic in these applications.

3. CONCLUSIONS AND FUTURE WORK

This paper explores how to apply MCDM algorithms to search or filtering tasks that have multiple user criteria. A major potential advantage of multi-criteria utility measures is that the system explicitly models multiple user criteria and estimates the separate components of document utility using different sub-utility measures. We expect it would be easier for the system to predict the overall utility of an document based on the estimation of the utility components, compared with predicting the inherently complex user utility directly using standard machine learning or IR models. Our experimental results are consistent with this conjecture.

Given the limited scope of our study, the suitability of MCDM methods for any information retrieval problem remains an open question. However, the fact that simple un-tuned MCDM methods performed well in both experiments is encouraging. Our future work is to have a larger scale evaluation with more users that will help us better understand how the conclusions may generalize to the larger population, and characterize any exceptional situations that contradict our current conclusions.

4. ACKNOWLEDGMENTS

We thank Francis Enomoto and James Davis for valuable discussion about the research. This research was supported in part by National Science Foundation IIS-0713111.

5. REFERENCES

- [1] Bureau of Transportation Statistics. Data library: Aviation. http://www.transtats.bts.gov/databases.asp?Mode_ID=1, accessed on April 15, 2007.
- [2] M. Farah and D. Vanderpooten. An outranking approach for rank aggregation in information retrieval. In *30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 591–598, 2007.
- [3] N. Manouselis and C. Costopoulou. Analysis and classification of multi-criteria recommender systems. In *World Wide Web*, volume 10, pages 415–441, 2007.
- [4] E. Triantaphyllou. *Mutli-Criteria Decision Making Methods: A Comparative Study*. Kluwer Academic Publishers, 2000.
- [5] Y. Zhang. Yow user study data. <http://www.soe.ucsc.edu/~yiz/papers/data/YOWStudy/>, last visited in Jan. 2009.