Information Retrieval Based on Axiomatic Decision Theory

S. K. M. Wong
Y. Y. Yao

TR 89-985
April 1989

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

Information Retrieval Based on Axiomatic Decision Theory

S.K.M. Wong, P. Bollmann, Y.Y. Yao

Abstract

The main objective of this paper is to establish a coherent framework for information retrieval based on the axiomatic decision theory. In information retrieval one has to deal with two difficult problems (knowledge representation and query formulation), both of which are absent in conventional database systems. It is argued that the axiomatic decision theory provides a useful framework to study these complex issues. Two quantitative representation systems are introduced. One is developed from the expected utility model and the other is derived from the concepts of evidential reasoning. An inductive learning algorithm is suggested for constructing a user query. The experimental results seem to provide some support for the theoretical arguments presented here.

Although the focus in this paper is mainly on information retrieval, the current work may be viewed as a preliminary effort towards unifying symbolic and numeric reasoning with incomplete or uncertain information.

1. Introduction

Consider a finite set of information items stored in the computer. One of the main design objectives in a database management system is to provide both efficient and effective means for a user to retrieve the relevant information. In the relational database model, for example, we have a set of simple objects called tuples. Each tuple is simply a record consisting of several attribute values. In database systems an information request is an *objective* notion and it can therefore be expressed unambiguously in terms of some well defined query languages such as those based on tuple or domain calculus. However, when we are dealing with a set of more complex objects, say text documents as being the case in information retrieval, new problems immediately arise. First of all, we have to find a way to represent the information content of a document. In contrast, one needs not be concerned with the problems of knowledge representation in database design at all. The second problem is a more complex user preference or choice that we may have to contend with in information retrieval. This problem is further complicated by the fact that a preference is a *subjective* notion. For this reason, it is much harder to design a language with sufficient expressive power to describe a variety of user preferences. For example, imagine the difficulty in trying to give an accurate instruction to someone to rank a group of people in accordance with your preference. Even if one may succeed developing such a powerful query language, it would probably be too difficult for anyone to use in practice.

The main differences between a database and an information retrieval system are summarized in Table 1. Let us reiterate the two design problems in information retrieval:

- (1) representation of complex objects (documents representation),
- (2) description of user preference (query formulation).

Before going into the detailed discussion of these two problems, first let us define more precisely the objectives of an information retrieval system. The discussion here is centered on the notion of user preference [1, 18, 20].

	Database	Information Retrieval
1 Objects	simple objects: tuples	complex objects: documents
2 Information request	an <i>objective</i> structure	a <i>subjective</i> structure
3 Query language	predicate calculus	how ?
4 Knowledge representation	transparent	a complex process

Table 1. Database vs. information retrieval

Let D denote a set of documents. A user preference can be represented by a binary relation $\bullet>$ on D : for $d, d' \in D$,

$$d \bullet> d' \Leftrightarrow \text{the user prefers } d \text{ to } d'.$$

It is understood that the preference relationship may not hold for some document pairs. For instance, a user may judge two documents being equally useful or he may judge both being irrelevant to his information needs. Thus, an indifference relation $\sim$ on D can be defined as:

$$d \sim d' \Leftrightarrow \neg(d \bullet> d') \ \& \ \neg(d' \bullet> d).$$

In information retrieval, our primary aim is to find a real-valued function f on D such that

$$d \bullet> d' \Leftrightarrow f(d) > f(d').$$

A function satisfying this condition is called a *perfect* order-preserving or ranking function (see Figure 1). That is, the numbers $f(d), f(d'), \dots$ as ordered by $>$ faithfully reflect the order of the documents $d, d', \dots$ under the user preference relation $\bullet>$.

Does a perfect order-preserving function exist? The answer to this question depends on the preference structure. We may assume that the preference relation $\bullet>$ satisfies the following axioms:

- (i) *Asymmetry* : $d \bullet > d' \Rightarrow \neg(d' \bullet > d)$

This axiom says that if the user prefers d to d' , he does not prefer d' to d .

- (ii) *Negative transitivity* : $\neg(d \bullet > d') \ \& \ \neg(d' \bullet > d'') \Rightarrow \neg(d \bullet > d'')$.

This means that if the user does not prefer d to d' and does not prefer d' to d'' , then he does not prefer d to d'' .

Such axioms can be looked at in two ways. The *prescriptive* or *normative* view interprets these axioms as conditions of rationality in making choices. A second interpretation is that these axioms are *descriptive*, in which an axiom is viewed as a testable condition.

Based on these two axioms, one arrives at the following existence theorem [3, 11].

Theorem 1

Suppose D is a finite set and $\bullet >$ a (preference) relation on D . Then there is a real-valued function f on D such that

$$d \bullet > d' \Leftrightarrow f(d) > f(d')$$

if and only if $(D, \bullet >)$ is a weak order, i.e., $\bullet >$ is asymmetric and negatively transitive.

Figure 2 depicts a graph which represents a preference relation with $(D, \bullet >)$ being a weak order. There is an edge connecting d and d' in the graph for $d \bullet > d'$. By default, the document in the upper end of an edge is more preferred. Note that if a preference relation is asymmetric and negatively transitive, then the relation is also transitive, i.e., $(d \bullet > d') \ \& \ (d' \bullet > d'') \Rightarrow (d \bullet > d'')$.

In information retrieval one is primarily interested in ensuring that the more preferred documents are listed in front of the less preferred. In this case, it is sufficient to consider a function f satisfying a weaker condition than that of Theorem 1, namely,

$$d \bullet > d' \Rightarrow f(d) > f(d') ,$$

which implies:

$$f(d') \geq f(d) \Rightarrow \neg(d \bullet > d') .$$

A function satisfying the above condition is called an *acceptable* ranking function (see Figure 3). We will consider only this type of ranking functions in subsequent discussions. The corresponding existence theorem is given below [3, 11].

Theorem 2

Suppose D is a finite set and $\bullet >$ a (preference) relation on D . Then there is a real-valued function f on D such that

$$d \bullet > d' \Rightarrow f(d) > f(d')$$

if $(D, \bullet >)$ is a strict partial order, i.e., $\bullet >$ is asymmetric and transitive.

A graph of a preference relation with $(D, \bullet >)$ being a strict partial order is shown in Figure 4. Comparing Figures 1 and 2 with Figures 3 and 4, it can be seen that one cannot find a perfect order-preserving function for a relation $\bullet >$ if $(D, \bullet >)$ is a strict partial order but not a weak order.

Theorems 1 and 2 are important because they provide a basis for a quantitative representation of the user preference structure.

2. Quantitative Representation of Preferences

In the following two sections, we discuss how one may construct a quantitative system to represent a user preference. First, a document representation scheme is chosen. Then the structure of a user preference of interest is defined by an appropriate set of axioms. This set of axioms provides a qualitative representation of the preference structure. In order to transform a qualitative representation system into a quantitative representation system, one must choose a ranking function consistent with the required axioms. Finally, a systematic method is required for the determination of the necessary parameters in the ranking function, which will be discussed in Section 3.

The quantitative representation systems of two plausible preference structure are discussed. One is based on the expected utility (the axiomatic decision theory [3, 4, 8,

11]) and the other is derived from the notion of belief functions (the mathematical theory of evidence [14-16]).

2.1. The expected utility model

Let $\Theta = \{t_1, t_2, \dots, t_n\}$ be a finite set of mutually exclusive elementary concepts. In the expected utility model [3, 19], it is assumed that each document $d \in D$ can be uniquely characterized (represented) by a simple probability measure P_d on Θ with

$$\sum_{j=1}^n P_d(\{t_j\}) = 1, \text{ and } P_d(\{t_j\}) \geq 0, \text{ for all } d \in D.$$

Let Ξ be the set of all simple probability measures on Θ (see Figure 5) and let $\bullet>$ denote a preference relation on Ξ . In terms of the above choice of document representation, the preference structure of interest can be explicitly characterized by the following set of axioms: for all $P_{d_1}, P_{d_2}, P_{d_3}, P_{d_4} \in \Xi$,

- (1) $(\Xi, \bullet>)$ is a strict partial order,
- (2) If $0 < \alpha < 1$, then
$$P_{d_1} \bullet> P_{d_2} \iff (\alpha P_{d_1} + (1 - \alpha)P_{d_3}) \bullet> (\alpha P_{d_2} + (1 - \alpha)P_{d_3}),$$
- (3) If $(\alpha P_{d_1} + (1 - \alpha)P_{d_3}) \bullet> (\alpha P_{d_2} + (1 - \alpha)P_{d_4})$ for all $0 < \alpha \leq 1$, then $\neg(P_{d_4} \bullet> P_{d_3})$.

Axiom (2), a form of independence axiom, is regarded by many as the core of expected utility theory, for without it the *expectation* part of expected utility vanishes. This condition is often regarded as the principal nominative criterion of the theory, along with transitivity of $\bullet>$. Axiom (3) is sometimes referred to as the Archimedian axiom. A more detailed discussion of these axioms can be found in the literature of axiomatic decision theories [3, 4, 8, 11].

Theorem 3

If axioms (1), (2) and (3) hold for a preference relation $\bullet>$ on Ξ , then there is a real-valued function u on Θ such that for all $P_d, P_{d'} \in \Xi$,

$$P_d \bullet> P_{d'} \Rightarrow \sum_{j=1}^n P_d(\{t_j\})u(t_j) < \sum_{j=1}^n P_{d'}(\{t_j\})u(t_j) .$$

Theorem 3 enables us to proceed from a qualitative system to a quantitative system. Although at this point we know the quantitative representation of the preference structure explicitly defined by the expected utility function $f(d) = E(P_d, u)$:

$$E(P_d, u) = \sum_{j=1}^n P_d(\{t_j\})u(t_j) .$$

the utility $u(t_j)$ for the individual elementary concept t_j is not yet known. We will demonstrate later how one can compute $u(t_j)$ by an inductive learning method.

The ranking function $E(P_d, u)$ has the form reminiscent of the similarity measure in the vector space model introduced by Salton [12] some years ago. The tuple $\mathbf{d} = (P_d(\{t_1\}), P_d(\{t_2\}), \dots, P_d(\{t_n\}))$ can be viewed as a *document vector*, while the tuple $\mathbf{q} = (u(t_1), u(t_2), \dots, u(t_n))$ represents a *query vector*. If the term concepts derived from autoindexing are assumed to be mutually exclusive, the probability $P_d(\{t_j\})$ can be interpreted as the relative occurrence frequency of term concept t_j in the document [19].

2.2. The belief function model

Here we consider the quantitative representation of a user preference structure that conforms to the properties of a belief function [14-16].

Let $\Theta = \{t_1, t_2, \dots, t_n\}$ be a finite set of mutually exclusive elementary concepts. Suppose that each document can be uniquely represented by a subset of Θ . That is, a document $d \in D$ is characterized by an element S_d in the power set 2^Θ of Θ (see Figure 6). With such a document representation scheme, we are interested in a preference relation $\bullet>$ satisfying the following axioms:

- (a) $(2^\Theta, \bullet>)$ is a strict partial order,
- (b) For all $S_d, S_{d'} \in 2^\Theta$, $S_d \supseteq S_{d'} \Rightarrow S_d \bullet> S_{d'}$ or $S_d \sim S_{d'}$,
where $S_d \sim S_{d'} \Leftrightarrow (\neg(S_d \bullet> S_{d'}) \ \& \ \neg(S_{d'} \bullet> S_d))$.

Axiom (b) is referred to as the *dominance* axiom. If a preference relation satisfies these two axioms, it can be quantified by a belief function as indicated by the following theorem.

Theorem 4

Suppose $(2^\Theta, \bullet>)$ is a strict partial order. Then there exists a belief function $Bel: 2^\Theta \rightarrow [0,1]$ satisfying:

$$S_d \bullet> S_{d'} \Rightarrow Bel(S_d) > Bel(S_{d'})$$

if and only if the dominance axiom holds.

This theorem is interesting by itself because it establishes a link between symbolic and numeric reasoning with uncertainty. For completeness, some key properties of a belief function Bel are summarized below:

- (i) $Bel(\emptyset) = 0$,
- (ii) $Bel(\Theta) = 1$,
- (iii) $Bel(S_{d_1} \cup \dots \cup S_{d_M}) \geq$

$$\sum_i Bel(S_{d_i}) - \sum_{i < j} Bel(S_{d_i} \cap S_{d_j}) \pm \dots + (-1)^{M+1} Bel(S_{d_1} \cap \dots \cap S_{d_M}) .$$

Property (iii) indicates that a belief function is not necessarily additive, which is in fact a generalization of the additive Bayesian probability function. A belief function can be expressed by using another function, $m: 2^\Theta \rightarrow [0, 1]$, which is called the basic probability assignment satisfying the following properties:

$$m(\emptyset) = 0, \quad \sum_{F_j \in 2^\Theta} m(F_j) = 1 .$$

In terms of the basic probability assignment, $Bel(S_d)$ can be computed by:

$$Bel(S_d) = \sum_{F_j \subseteq S_d} m(F_j) ,$$

where the summation is over all elements $F_j \in 2^\Theta$.

Corresponding to each S_d , we define a tuple $\mathbf{d} = (c_1^d, c_2^d, \dots, c_N^d)$ with $N = |2^\Theta|$ components as follows:

$$c_j^d = \begin{cases} 1 & \text{if } S_d \supseteq F_j , \\ 0 & \text{if } S_d \supset F_j . \end{cases}$$

With this numeric document representation, the belief function $Bel(S_d)$ can be conveniently expressed as:

$$Bel(S_d) = \sum_{j=1}^N c_j^d m(F_j) .$$

In contrast to the *weighted* document representation in the expected utility model, the tuple $\mathbf{d}$ defined here is a *binary* vector.

Thus far we have seen the quantitative representation systems of two different preference structure. The first quantitative representation is realized by an expected utility function with weighted document representation, whereas the second is obtained by a belief function with binary document representation. Although these two quantitative representation functions are both linear, they have rather different physical interpretations. It is perhaps worth mentioning that a majority of existing information retrieval models use linear ranking functions defined for either weighted or binary document representation. The arguments presented here seem to provide a theoretical justification for these approaches.

3. Query Formulation

The ranking functions derived from the expected utility and the belief function models can both be written in the form:

$$\mathbf{d} \cdot \mathbf{q} = \sum_{j=1} d_j q_j ,$$

where $\mathbf{d} = (d_1, d_2, \dots)$ represents a document d with $d_j = P_d(\{t_j\})$ or c_j^d , and the tuple $\mathbf{q} = (q_1, q_2, \dots)$ represents a user preference with $q_j = u(t_j)$ or $m(F_j)$.

Now assume that we have chosen a document representation based on some indexing scheme. The theorems given in Section 2 identify the necessary or/and sufficient conditions for the quantitative representation of a preference structure by a linear function. However, we are still faced with the problem of formulating a tuple $\mathbf{q}$ such that

$$\mathbf{d} \cdot \mathbf{q} > \mathbf{d}' \cdot \mathbf{q}$$

or

$$(\mathbf{d} - \mathbf{d}') \cdot \mathbf{q} = \mathbf{b} \cdot \mathbf{q} > 0 .$$

for all document pairs (d, d') in $D \times D$ satisfying $d \bullet > d'$. In practice, the user preference relation $\bullet >$ is not known *a priori* for the entire document collection, this makes the problem even more difficult.

For simplicity, one may use a non-inductive method to construct a $\mathbf{q}$ directly from the query text submitted by the user. This is in fact one of the methods used in the vector space model [12]. As mentioned before, it is rather difficult to describe a user preference. It is therefore not surprising that the performance results obtained by non-inductive methods are not very satisfactory. Instead of using the non-inductive methods, we suggest a learning procedure to formulate $\mathbf{q}$ from a set S of sample documents selected from D .

Suppose the preference relation $\bullet >$ on S is given by the user. Let

$$\mathbf{B} = \{ \mathbf{b} = \mathbf{d} - \mathbf{d}' \mid d, d' \in S, d \bullet > d' \} .$$

The following iterative algorithm, a modification of the perceptron learning algorithm in the neural network [2, 9, 10, 18], may be used to construct a $\mathbf{q}$ satisfying $\mathbf{b} \cdot \mathbf{q} > 0$ for all $\mathbf{b} \in \mathbf{B}$:

- (i) Let $k = 0$; choose an initial $\mathbf{q}^0$;
- (ii) Let $\mathbf{q}^k$ denote the tuple $(q_1^k, q_2^k, \dots)$ in the $(k + 1)$ th iteration;
Compute a set $\Gamma(\mathbf{q}^k)$ defined as follows:

$$\Gamma(\mathbf{q}^k) = \{ \mathbf{b} \in \mathbf{B} \mid \mathbf{b} \cdot \mathbf{q}^k \leq 0 \} ;$$

If $\Gamma(\mathbf{q}^k) = \emptyset$, terminate the procedure; otherwise, let

$$\mathbf{q}^{k+1} = \mathbf{q}^k + \sum_{\mathbf{b} \in \Gamma(\mathbf{q}^k)} \mathbf{b} ;$$

- (iii) Let $k = k + 1$; go back to step (ii);

It has been proved [18] that if the required axioms given in Section 2 hold for a preference relation, the above procedure indeed converges to a query vector $\mathbf{q}$ satisfying $\mathbf{b} \cdot \mathbf{q} > 0$ for all $\mathbf{d} \in \mathbf{B}$.

To verify our theoretical arguments, we examined the expected utility model with five standard test document collections [13, 17], which indicate that all user preferences in these test collections have a linear structure. Experiments were also carried out to test the effectiveness of our inductive learning method [17]. These experimental results clearly show that the performance of the proposed method is substantially better than that of a non-inductive method. However, the point we want to stress here is not so much on the improvement in performance but rather on the possibility that we may have taken a step in the right direction.

4. Conclusion

A number of problems in document indexing remain to be resolved. For example, in the expected utility model the elementary concepts are assumed to be mutually

exclusive. In practice, many indexing schemes actually violate this assumption. On the other hand, in the belief function model a document is represented as a set of elementary concepts or a binary vector. Again, this condition may not necessarily be fulfilled by an indexing scheme.

We believe it is worthwhile to look into more closely the connection between information retrieval and axiomatic decision theories. The results of such an investigation may enhance our understanding about information retrieval. As for the future one should perhaps look beyond the problems dealing with the retrieval process. It may be beneficial to find a way to make better use of the vast quantity of information available for decision making. The progress made in information retrieval may have a significant impact on the future design of intelligent information systems.

Although the focus in this paper is on information retrieval, many issues discussed are actually related to uncertainty management. Reasoning with incomplete or uncertain information has received much attention lately [5-7]. There are two general approaches to uncertainty management. One is based on the idea of extending ordinary truth-functional logic to allow the incorporation of defaults or assumptions. A system based on such a logic can set default values for uncertain propositions and reason as if these values were known, but revise its defaults when they give rise to a contradiction. This capability mimics the non-monotonic character of intelligent human reasoning. On the other hand, in the quantitative approach uncertainty about truth value of a proposition is expressed by a number or numbers typically in the range between 0 and 1. Inference rules also have numbers associated with them indicating how strongly belief in the premises warrants belief in the conclusions. Intelligent reasoning should combine aspects of both these viewpoints. The current work may be viewed as a preliminary effort towards unifying symbolic and numeric reasoning with incomplete or uncertain information.

Acknowledgment

This paper was prepared while the authors S.K.M. Wong and Y.Y. Yao visiting the Department of Computer Science, Cornell University. They wish to thank J. Hopcroft for providing the opportunity to work at Cornell and they are particularly indebted to G. Salton for his encouragement and hospitality.

Reference

- [1] Bollmann, P.; Wong, S.K.M. "Adaptive linear information retrieval models," *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 157-163, 1987.
- [2] Duda, R.O.; Hart, P.E. *Pattern Classification And Scene Analysis*, New York: Wiley, 1973.
- [3] Fishburn, P.C. *Utility Theory for Decision Making*, New York: Wiley, 1970.
- [4] Fishburn, P.C. *Mathematics of Decision Theory*, Mouton: UNESCO, 1972.
- [5] Kanal, L.N.; Lemmer, J.F. (ed.) *Uncertainty in Artificial Intelligence*, Amsterdam: North-Holland, 1986,
- [6] Laskey, K.B.; Lehner, P.E. "Belief maintenance: an intergrated approach to uncertainty management," *Proceedings AAAI-88, Seventh National Conference on Artificial Intelligence*, Vol. 1, 210-214, 1988.
- [7] Lemmer, J.F.; Kanal, L.N. (ed.) *Uncertainty in Artificial Intelligence 2*, Amsterdam: North-Holland, 1988,
- [8] Menges, G. (ed.) *Information, Inference and Decision*, Boston: D. Reidel Publishing Company, 1974.
- [9] Minsky, M.; Papert, S. *Perceptrons -- An Introduction to Computational Geometry*, Massachusetts: MIT Press, expanded edition, 1988.
- [10] Nilsson, N.J. *Learning Machine -- Foundations of Trainable Pattern Classifying Systems*, New York: McGraw-Hall, 1965.
- [11] Roberts, F.S. *Measurement Theory*, New York: Academic Press, 1976.
- [12] Salton, G. *Automatic Text Processing*, Reading, MA: Addison-Wesley Publishing Company, 1989.
- [13] Salton, G.; Buckley, C. "Improving Retrieval Performance by Relevance Feedback," *Technical Report*, Department of Computer Science, Cornell University, No. 88-898, 1988.
- [14] Shafer, G. *A Mathematical Theory of Evidence*, Princeton: Princeton University Press, 1976.
- [15] Shafer, G. "Belief functions and possibility measures," in Bezdek, J.C. (ed.) *Analysis of Fuzzy Information*, Vol. I, 51-84, CRC Press, 1986.

- [16] Shafer, G. "Probability judgment in artificial intelligence and expert systems," *Statistical Science*, 2: 3-16, 1987.
- [17] Wong, S.K.M; Salton, G.; Yao, Y.Y. "Relevance feedback based on an adaptive linear model," *Technical Report*, Department of Computer Science, Cornell University, 1989.
- [18] Wong, S.K.M; Yao, Y.Y. "Query formulation in linear retrieval models," To appear in *Journal of the American Society for Information Science*, 1988.
- [19] Wong, S.K.M.; Yao, Y.Y. "A probability distribution model for information retrieval," To appear in *Information Processing and Management*, 1989.
- [20] Wong, S.K.M.; Yao, Y.Y.; Bollmann, P. "Linear structure in information retrieval," *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 219-232; 1988.

Figure 1. A perfect ranking function

Figure 2. An asymmetric and negatively transitive preference relation $\bullet >$

Figure 3. An acceptable ranking function

Figure 4. An asymmetric and transitive preference relation $\bullet \succ$

Figure 5. A geometric illustration of simple probability measures with $n = 2$

Figure 6. Document representations with $\Theta = \{t_1, t_2, t_3\}$ in the belief function model