Information Retrieval, 6, 75–97, 2003

Using the Information Structure Model to Compare Profile-Based Information Filtering Systems

YUVAL ELOVICI elovici@inter.net.il
BRACHA SHAPIRA bshapira@bgumail.bgu.ac.il
Department of Information Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel

PAUL B. KANTOR kantor@scils.rutgers.edu
Department of Library and Information Science, School of Communication Information and Library Studies (SCILS), RUTGERS—The State University of New-Jersey, 4 Huntington St., New Brunswick, NJ 08901-1071, USA

Received April 10, 2002; Revised September 24, 2002; Accepted October 7, 2002

Abstract. In the IR field it is clear that the value of a system depends on the cost and benefit profiles of its users. It would seem obvious that different users would prefer different systems. In the TREC-9 filtering track, systems are evaluated by a utility measure specifying a given cost and benefit. However, in the study of decision systems it is known that, in some cases, one system may be unconditionally better than another. In this paper we employ a decision theoretic approach to find conditions under which an Information Filtering (IF) system is unconditionally superior to another for all users regardless of their cost and benefit profiles.

It is well known that if two IF systems have equal precision the system with better recall will be preferred by all users. Similarly, with equal recall, better precision is universally preferred. We confirm these known results and discover an unexpected dominance relation in which a system with lower recall will be universally preferred provided its precision is sufficiently higher.

Keywords: performance, evaluation, formal model, user-profile, information economics

1. Introduction

Information Filtering (IF) systems aim to present to the user only the relevant part of an incoming stream of information. The user's information needs are expressed in their profiles that are matched against incoming documents. A good IF system can successfully estimate the relevance of incoming items, and thus protect the user from irrelevant information without overlooking that which is relevant (Hanani et al. 2001, Oard 1997, Belkin and Croft 1992). One major difference between IF and IR systems is that IF systems usually target long term users with stable information needs, represented by user profiles, while many IR systems target users with ad-hoc information needs.

Filtering was introduced as one of the Text Retrieval Conference (TREC) tasks in TREC-4, and required the development of new evaluation measures. The reason for this is that filtering systems make binary decisions (to accept or reject a document of an incoming string) and the outcome of the system is an unranked set of documents. Therefore, the standard evaluation measures (like average precision) that apply for ranked sets do not apply for filtering systems. One of the measures employed for evaluating IF systems is general linear utility (Robertson 2002, Hull and Robertson 2000).

The general utility measure depends on utility parameters that reflect the importance of detecting, and of missing the filtered topic to the user. For example, assuming a system that filters economic news, the cost of missed documents may be higher for a stock exchange broker than the cost to a student of economics. As evaluation based on utility reflects only specific user's cost and benefits, it is interesting to ask if there is an evaluation method that might find one system's superiority over another, regardless of its user's preferences (costs and benefits).

In this paper we employ a decision theoretic approach and use the information structure model (IS) to find conditions under which an IF system is unconditionally superior to another for all users, regardless of their cost and benefit. The IS model is frequently used in information economics to assess the value of information in terms of the payoff to a user of an imperfect information system. The IS model presumes an optimal decision strategy on the part of the user of the system, i.e., the user takes the most profitable action, given the system output (McGuire and Radner 1986).

In this paper an IF system is modeled by an IS, based on its precision and recall and one additional parameter explained below. This modeling permits the evaluation of the IF system in terms of benefit to the user, in a way similar to the TREC utility measure. The evaluation can be used to rank order the IF systems. However, since this rank ordering is based on the user utility parameters, it only reflects the ordering for a specific user.

We find that it is possible to partially rank order IF systems that are modeled by IS regardless of specific user utilities parameter. A system will be said to dominate another system when it is superior for all users for any utility parameters. The system comparison is done using the Blackwell theorem (McGuire and Radner 1986). If the modeled IF systems cannot be compared using the Blackwell theorem, they can still be ordered by their expected payoff *to a specific user utility parameters* as done in TREC.

There are some obvious dominance relations between systems that we confirm using the IS model:

1. System A is better than system B for all its users if they have equal precision, and system A has higher recall.
2. System A is better than system B for all its users if they have equal recall and system A has higher precision.

In this paper we show an unexpected dominance relation in which a system A with lower recall than system B is still better for all its users provided its precision is sufficiently higher.

The remainder of this paper is structured as follows: Section 2 reviews the IS model. Section 3 models a basic IF system as an IS. Section 4 presents the relation of IS parameters to precision and recall. Section 5 illustrates comparison of IF systems modeled by ISs, using the Blackwell theorem. Section 6 extends the model to support IF systems that maintain detailed user profiles, and Section 7 includes concluding remarks and future research directions.

2. The information structure model: A review

The Information Structure (IS) model developed by Marschak ((1971); is an information-economic approach to assess the value of information for a textbook see McGuire and Radner (1986)). The IS model represents an information system as a Markov (stochastic) matrix describing the stochastic transformation of states of nature, which are the inputs that the system receives, to signals, which are the system's outputs. The model assumes that the decision maker, the user of the system, is rational and, given its values of payoffs, can assess the benefits of using the system and choose the optimal usage strategy.

A brief review of the IS model is presented here. A more detailed introduction can be found in McGuire and Radner (1986) and the original argument appears in Blackwell and Girshick (1954, 1979 especially Ch. 5).

Let S be a finite set of Events: $S^t = \{s_1, \ldots, s_{ns}\}$. Let Y be a finite set of signals: $Y^t = \{y_1, \ldots, y_{nz}\}$. Let π be a vector of prior probabilities of the events: $\pi = \{\pi_1, \ldots, \pi_{ns}\}$ where $\sum_{i=1}^{ns} \pi_i = 1$, $\pi_i \geq 0$. An information structure (IS) is defined as a Markov (stochastic) matrix of conditional probabilities. Its elements define the probability for a signal to be displayed when an event occurs ($q_{i,j}$ indicates the probability of displaying signal y_j given event s_i).

$$\text{IS:}\quad Q = \begin{bmatrix} q_{1,1} & \cdot & \cdot & \cdot & q_{1,nz} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ q_{ns,1} & \cdot & \cdot & \cdot & q_{ns,nz} \end{bmatrix} \quad \sum_{j=1}^{nz} q_{i,j} = 1 \quad \forall\, i, \text{ and } 0 \leq q_{i,j} \leq 1 \quad \forall\, i, j \qquad (1)$$

Let A be a finite set of actions that can be taken by the decision maker (DM): $A^t = \{a_1, \ldots, a_{na}\}$. Let U be a cardinal payoff matrix of real numbers that associates payoffs to pairs of an action and an event: $U\colon A \times S$ ($u_{i,j}$ indicates the payoff when the DM takes action a_i and the event turns out to be s_j).

A payoff matrix U is written as follows:

$$U = \begin{bmatrix} u_{1,1} & \cdot & \cdot & \cdot & u_{1,ns} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ u_{na,1} & \cdot & \cdot & \cdot & u_{na,ns} \end{bmatrix} \qquad (2)$$

The DM observes the signals and chooses actions accordingly. The DM decision rule can be described by a Markov matrix D where each element of the matrix represents the probability that the DM will take a certain action when observing a certain signal ($d_{i,j}$ indicates the probability that the DM will take action a_j given the signal y_i). Note that we are admitting mixed strategies. However, when there are no resource limitations, there will always be an optimal pure strategy. The DM wishes to maximize the expected payoff by choosing the optimal decision rule.

$$D = \begin{bmatrix} d_{1,1} & \cdot & \cdot & \cdot & d_{1,na} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ d_{nz,1} & \cdot & \cdot & \cdot & d_{nz,na} \end{bmatrix} \quad \sum_{j=1}^{na} d_{i,j} = 1 \quad \forall\, i, \text{ and } 0 \le d_{i,j} \le 1 \quad \forall\, i, j \tag{3}$$

Finally, the matrix Π is a square matrix with the vector π of a priori probabilities on its main diagonal and zeros elsewhere:

$$\Pi = \begin{bmatrix} \pi_1 & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & \pi_{ns} \end{bmatrix} \tag{4}$$

The expected payoff (utility) for the DM is $EU = tr(QDU\Pi)$ where tr is the matrix trace operator (i.e., the sum of the elements on the main diagonal). The DM maximizes EU by selecting the optimal decision rule D. Maximization is obtained by solving a linear programming problem. Because the optimum for a linear problem occurs on the boundary, at least one of the optimal decision rules will be pure, i.e., a matrix whose elements are only ones and zeros (McGuire and Radner 1986, p. 102).

Given two ISs Q and T operating on the same set of events S, Q is said to be *generally more informative* than T if the maximal expected payoff yielded by T is not larger than that yielded by Q for *all* payoff matrices U and *all* probability vectors π. A partial ordering of ISs is provided by the Blackwell theorem (Blackwell 1951, Blackwell and Girschick 1954, 1979, McGuire and Radner 1986) stating that

Theorem (*Blackewell*). *Q is generally more informative than T if and only if there exists a Markov matrix M with appropriate dimensions such that $Q \cdot M = T$; M is sometimes called the garbling matrix.*

Hereafter we will use the term "informativity" for the concept "generally more informative". The informativity relation imposes a partial ordering over the set of Information Structures. (Usually in information economics the term "informativeness" is used for what we describe as "informativity", however, in IR the term "informativeness" was used in a different context (Tauge-Sutcliffe 1992), and we have introduced a new term to avoid confusion.

The IS model was expanded by Demski (1972), Ahituv (1981), Ahituv and Wand (1984), Ahituv and Ronen (1988), Marshall and Narashimhan (1989), Sinchcombe (1990), Ronen and Spector (1995), and others. Carmi and Ronen (1996) applied the IS model to quality control attribute sampling, by analyzing a real life situation in the Israel Postal Authority. Another discussion of the necessary and sufficient conditions of the Blackwell theorem can be found in Hilton (1990).

In this paper we use the IS model to describe IF systems. Calculation of the user's expected payoff from the system supports comparison of the performance of IF systems. The next section introduces modeling of IF systems with ISs.

3. Modeling filtering systems as information structures

IF systems receive incoming information for their users. Their goal is to discriminate between relevant and irrelevant documents so as to expose their users only to information that is relevant, while not missing important relevant information (Hanani et al. 2001, Oard 1997, Belkin and Croft 1992). IF systems maintain users profiles that represent users' long term information preferences. Every piece of incoming information to the IF system is matched against a specific user's profile to determine if it is should be routed to him according to his profile. Many profile representation methods are described in IF related literature. One popular implementation of a user profile consists of a vector of weighted terms whose weights represent their significance to the user. The IF systems goal is, on the one hand, to maximize the fraction of the relevant incoming information that is presented to the user, and on the other hand to minimize the fraction of the non-relevant information that it mistakenly presented to the user. In this section modeling of IF systems that hold *basic* user profiles is presented. In Section 6 an extension is presented that includes modeling of IF system that maintain *detailed* user profiles.

There is an isomorphism between the filtering process and the information structure model. The fractions of relevant and non-relevant information entering the IF system can be thought of as the a priori probabilities of relevant and non-relevant information flowing into the IF system (the events in the IS model terminology are "relevant" and "non-relevant"). Van Rijsbergen (1979) defines for IR systems the *Generality* parameter (G) as the measure of the density of relevant documents in the collection (i.e., the number of relevant documents in a collection divided by the total number of documents in the collection). We slightly alter the definition of G for IF systems to be the average density of relevant documents in an incoming stream of information, i.e., the average chance that an incoming item is relevant. Hereafter G will denote the generality (density). The a priori probability of a relevant document will be G while the a priori probability of a non-relevant document will be $(1 - G)$.

The IF system's output can be considered as signals indicating whether a piece of information is estimated to be relevant or not. In order to distinguish between the events named "relevant" and "non-relevant" and the signals, the signals will be called "flagged" and "non-flagged". A "flagged" signal means that the system suggests that the user read the piece of information while a "non-flagged" signal means that the system suggests that the user ignores the piece of information.

The filtering process can be represented by a 2 by 2 IS, denoted by Q. We refer to the elements of the matrix Q by what they represent, where R means "relevant", R' means "non-relevant", F means "flagged" and F' means "non-flagged".

$Q[R, F]$ represents the fraction of the relevant information from the incoming stream of information that is flagged. $Q[R, F']$ represents the fraction of the relevant information that is not flagged. $Q[R', F]$represents the fraction of the non-relevant information that the system mistakenly flags as relevant. $Q[R', F']$ represents the fraction of the non-relevant information that the system flags as not relevant. Based on this filtering process the user can choose whether to follow the system's recommendation (i.e., whether to examine the pieces of information indicated by the system as relevant) or not.

In the IS model terminology, the IF users' activities are translated into two actions: read the information (action named "read"), and disregard the information (action named "disregard").

The following example summarizes and illustrates the modeling of IF system with the IS model:

- Events: $S = \{$Relevant, Non-relevant$\}$
- A priori probabilities: $\pi = \{G$ for Relevant, $(1 - G)$ for Non-relevant. In this example we take $G = 0.2\}$
- Signals: $Y = \{$Flagged, Non-flagged$\}$
- Actions: $A = \{$Read, Disregard$\}$.

Suppose the actual IS, Q, modeling an IF system is as follows:

	Signals	
IS Q events	Flagged	Non-flagged
Relevant	0.9	0.1
Non-relevant	0.2	0.8

The user profile that typically describes users preferences, is represented in the IS model as the user's payoff matrix. A basic profile should define the importance of the relevant information for the user, and is denoted in the IS model by his or her cost and benefit from reading and disregarding relevant and non-relevant information. (The representation of a more detailed profile in the IS model is described in Section 6).

Suppose that the user has the following payoff matrix U:

	Events	
U actions	Relevant	Non-relevant
Read	20	−5
Disregard	−10	0

In the above payoff matrix, the user payoff from examining relevant information is 20. The user loss caused by examining non-relevant information is −5. The user loss caused by not examining an item of relevant information is −10, which is higher than the loss caused by using irrelevant information. The payoff from not using irrelevant information is set to 0 (In fact, both the zero point and the scale of the utility matrix may be set arbitrarily). The expected payoff to the user can be expressed in terms of the trace of a matrix product

$$\begin{aligned} EU &= tr(QDU\Pi) \\ &= G(d_{11}\, q_{11}\, u_{11} + d_{21}\, q_{12}\, u_{11} + d_{12}\, q_{11}\, u_{21} + d_{22}\, q_{12}\, u_{21}) \\ &\quad - (G-1)\,(d_{11}\, q_{21}\, u_{12} + d_{21}\, q_{22}\, u_{12} + d_{12}\, q_{21}\, u_{22} + d_{22}\, q_{22}\, u_{22}) \\ &= 2.8 \cdot d_{11} - 1.8 \cdot d_{12} - 2.8 \cdot d_{21} - 0.2 \cdot d_{22} \end{aligned} \quad (5)$$

For that EU, the optimal decision strategy D is:

	Actions	
D signals	Read	Disregard
Flagged	1	0
Non-flagged	0	1

In this case the optimal decision rule is to read the flagged information and disregard the non-flagged information to maximize expected payoff. This decision strategy is a direct result of the user payoff matrix, i.e. his profile. Applying the optimal decision rule to the *EU* equation yields the maximal *EU* (in this example $EU = 2.6$).

Thus, given the properties of an IF system and his or her payoff matrix, a user can, determine whether using the IF system will increase expected payoff. In the next section we will demonstrate how IF system modeled by ISs can be compared using the Blackwell theorem.

4. Relating IS parameters to precision and recall evaluation measures

The two most commonly used measures for performance of information retrieval and information-filtering systems are precision and recall (Van Rijsbergen 1979, Frakes and Baeza-Yates 1992, Baeza-Yates and Ribeiro-Neto 1999). "Precision is the fraction of those documents that have been retrieved that are relevant. Recall is the fraction of all relevant documents that have been retrieved" (Frakes and Baeza-Yates 1992). A precision and recall curve usually describes the performance of a ranking system for one query. The curve represents the precision for different recall points. When used as a system's performance measure, the curve represents some average of these precision and recall curves over some distinct queries. Information retrieval researchers have defined several single-number measures to combine the precision and recall measure in an effort to avoid the essential complexity of comparison of precision-recall curves (Frakes and Baeza-Yates 1992, Van-Rijsbergen 1979). When the user utilities and the a priori probabilities are specified, modeling IF systems with IS will also produce such a single measure, the payoff value for the user. In this section the relation between IS parameters and precision-recall measure is given. In Section 4.1 the mathematical derivation of precision and recall from IS modeling IF system is presented, and in Section 4.2 the derivation of IS parameters for an IF system is illustrated based on the system's precision and recall.

4.1. *Precision and recall derivation from an IS modeling an IF system*

Assume the following IS matrix Q models an IF system:

	Signals	
IS Q events	Flagged	Non-flagged
Relevant	0.9	0.1
Non-relevant	0.2	0.8

Assume as before that the a priori probabilities of the events ("relevant", "Non-relevant") are: Pr(*Relevant*) = 0.2, Pr(*Non-relevant*) = 0.8. The relevant density in the IF model is $G = \Pr(\mathit{Relevant}) = 0.2$.

The *recall* measure (in probabilistic terms) is the probability to flag a relevant piece of information, given the set of relevant documents in the system (hereafter denoted by R). We can derive the recall for the above system Q as follows

$$R = \Pr(\mathit{Flagged}/\mathit{Relevant}) = Q[1, 1] = 0.9 \tag{6}$$

The *precision* measure (in probabilistic terms) is the probability of relevance for a piece of information, given that the system has flagged it to be relevant (hereafter denoted by P). Using Bayes' theorem we can derive the precision for above system Q, as follows:

$$\begin{aligned} P &= \Pr(\mathit{Relevant}/\mathit{Flagged}) \\ &= \frac{\Pr(\mathit{Flagged}/\mathit{Relevant}) \cdot \Pr(\mathit{Relevant})}{\Pr(\mathit{Flagged}/\mathit{Relevant}) \cdot \Pr(\mathit{Relevant}) + \Pr(\mathit{Flagged}/\mathit{Non\text{-}relevant}) \cdot \Pr(\mathit{Non\text{-}relevant})} \\ &= \frac{R \cdot G}{R \cdot G + Q[2, 1] \cdot (1 - G)} = \frac{0.9 \cdot 0.2}{0.9 \cdot 0.2 + 0.2 \cdot 0.8} = 0.529 \end{aligned} \tag{7}$$

4.2. *Derivation of an IS model of IF system based on its precision and recall*

Assume an IF system with precision 0.8 and recall 0.9 and assume that the known generality (density) G is 0.2. In the IS model of that system, the a priori probabilities are: (Pr(*Relevant*) = G = 0.2, Pr(*Non-relevant*) = $(1 - G)$ = 0.8).

It is possible to "translate" the *recall* measure to IS terminology as follows, and to calculate the probabilities of the signals given the relevant event:

$$\begin{aligned} &\Pr(\mathit{Flagged}/\mathit{Relevant}) = R = 0.9 \\ &\Pr(\mathit{Non\text{-}flagged}/\mathit{Relevant}) = 1 - R = 0.1 \end{aligned} \tag{8}$$

When transforming the *precision* measure to IS terminology, it is possible to calculate the probabilities of the signals (“flagged”, “non-flagged”) given the not-relevant event:

$$P = \frac{R \cdot G}{R \cdot G + \Pr(\mathit{Flagged}/\mathit{Non\text{-}relevant}) \cdot (1 - G)}$$
$$\Downarrow$$
$$\Pr(\mathit{Flagged}/\mathit{Non\text{-}relevant}) = \frac{R \cdot G - R \cdot G \cdot P}{(1 - G) \cdot P} \tag{9}$$
$$= \frac{0.9 \cdot 0.2 - 0.9 \cdot 0.2 \cdot 0.8}{0.2 \cdot 0.8} = 0.225$$
$$\Pr(\mathit{Non\text{-}flagged}/\mathit{Non\text{-}relevant}) = 1 - \Pr(\mathit{Flagged}/\mathit{Non\text{-}relevant}) = 0.775$$

The IS thus derived for the above IF system is

	Signals	
Events	Flagged	Non-flagged
Relevant	0.9	0.1
Non-relevant	0.225	0.775

5. A comparison of IF systems modeled by IS

The motivation for IF modeling as an IS is to evaluate and compare IF systems using a single measure, without the necessity of comparing precision-recall curves. This is not, in general, unconditionally possible. However, once we represent the IF systems using ISs, we may compare them using the Blackwell theorem. Since the Blackwell theorem offers only a partial ordering of ISs it is not guaranteed that each pair of IF systems will be comparable. But, sometimes, one dominates another. In this section we characterize pairs of IF systems that can be compared by the Blackwell theorem based on the precision and recall measures of the systems. Thus we define the partial order of the Blackwell theorem in precision-recall terms.

Let Q and T be two ISs modeling two different IF systems. A partial rank ordering of ISs is provided by the Blackwell theorem quoted above. It can be shown that a more informative IF system never yields lower expected payoff for any choice of a priori probabilities and payoff matrix.

For example, assume two ISs Q and T, modeling IF systems, have the following matrices

	Signals	
IS Q events	Flagged	Non-flagged
Relevant	0.9	0.1
Non-relevant	0.2	0.8

	Signals	
IS *T* events	Flagged	Non-flagged
Relevant	0.8	0.2
Non-relevant	0.2	0.8

There exists a garbling matrix M that $Q \cdot M = T$:

	Signals	
M signals	Flagged	Non-flagged
Flagged	0.8857140	0.114286
Non-flagged	0.0285714	0.971429

In this case, Q is preferred to T by *all* users. A detailed description of how to find the matrix M, using linear programming, is given in Appendix A.

Note that the Blackwell theorem offers only partial ordering, meaning that only some pairs of ISs can be compared unconditionally. If two ISs cannot be compared using the theorem, the expected payoff of each of the systems must be evaluated using the model for a given user profile (called the payoff matrix in the IS model terminology) and the relevance density G (a priori probabilities in the IS terminology). The system which delivers higher expected payoff will be considered "more informative", and the choice is conditioned on the specific details of the situation.

To be concrete, assume the user has the following payoff matrix

	Events	
U actions	Relevant	Non-relevant
Read	20	−5
Disregard	−10	0

Assume the a priori probabilities are

	Events	
	Relevant	Non-relevant
Probability	0.2	0.8

The expected payoff using the IF system modeled by Q is 2.6 while the expected payoff using the IF system modeled by T is 2.0. The IF system modeled by Q delivers higher expected payoff to the user. This result is not surprising, since the IS Q is more informative than T according to the Blackwell theorem.

By straightforward arithmetic we can express the information structure in terms of precision, generality (or density or relevant documents), and recall. With precision (P), recall (R) and density (G): the result is

	Signals	
IS Q events	Useful	Trash
Relevant	R	$1 - R$
Non-relevant	$\frac{R \cdot G - R \cdot G \cdot P}{(1-G) \cdot P}$	$1 - \frac{R \cdot G - R \cdot G \cdot P}{(1-G) \cdot P}$

Theorem 1. *Assume an IF system A with recall R and Precision P and consider a bounded region in the p-R coordinates defined by the following parameterized curves.*

1. $0 \leq \alpha \leq 1 \begin{cases} R^\alpha = \alpha \cdot R \\ P^\alpha = P \end{cases}$

2. $0 \leq \beta \leq 1 \begin{cases} R^\beta = \beta \\ P^\beta = G \end{cases}$

3. $0 \leq \gamma \leq 1 \begin{cases} R^\gamma = \gamma R + (1 - \gamma) \\ P^\gamma = \dfrac{R^\gamma P}{R^\gamma \cdot P + \left(\gamma R(1 - P) + \frac{P}{G}(1 - \gamma)(1 - G)\right)} \end{cases}$

Any IF system with precision and recall within the bounded region is inferior to system A according to the Blackwell theorem.

Proof: Appendix B. □

The key to the proof is that we transform a simple convex set, defined in the space of the elements of the IS, into the more complex curves shown here.

Two known relations can be derived from Theorem 1 and are expressed in Corollary 1 and 2:

Corollary 1. *Any set of IF systems modeled by ISs can be rank ordered using the Blackwell theorem if their precision measure is the same.*

Proof: Appendix C. □

In Corollary 1 it is shown that any group of IF systems whose representative precision is equal can be ranked ordered using the Blackwell theorem. This ranking is true for every payoff matrix of the user. An IF system that has a higher recall will be a more informative IS.

Corollary 2. *IF systems modeled by ISs can be rank ordered using the Blackwell theorem if their recall measure is the same.*

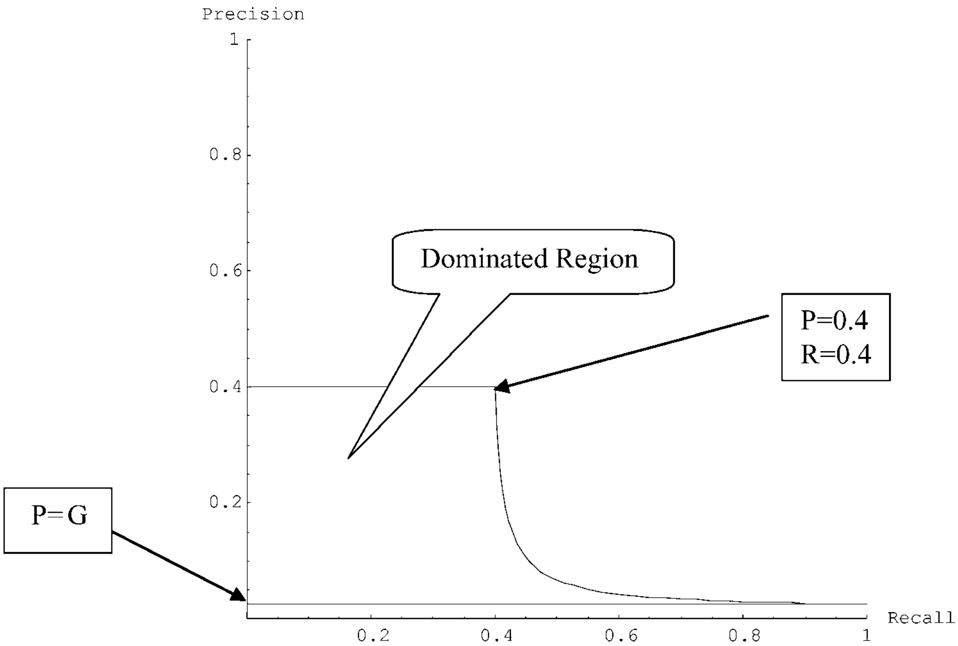

Figure 1. Dominance in *p-R* coordinates for a system with $P = 0.4$, $R = 0.4$, $G = 0.025$.

Proof: Appendix D. □

In Corollary 2 it is shown that any group of IF systems whose representative recall points are equal can be ranked ordered using the Blackwell theorem. This ranking is true for every payoff matrix of the user. An IF system that has a higher precision has a more informative IS.

Figure 1 is a graphical presentation of the bounded dominance region defined in Theorem 1 for a system with $R = 0.4$, $P = 0.4$, and $G = 0.025$. Any other IF systems whose representative points in the *p-R* place fall inside a bounded dominance region are inferior to the system according to the Blackwell theorem.

Assume an IF system A whose dominance region is illustrated in figure 2. It is possible to characterize the IF systems that are dominated by system A based on their precision and recall representative points:.

1. It is known that if two systems have equal precision, the system with better recall dominates the other. This known relation is stated in Corollary 1, and is graphically pointed to on figure 2 as the "Corollary 1" line, which presents the domination over systems with lower recall when equal precision is observed.
2. It is also known that if two systems have equal recall, the system with better precision dominates the other. This known relation is stated in Corollary 2, and is graphically pointed to on figure 2 as the "Corollary 2" line which presents the domination over systems with lower precision when equal recall is observed.

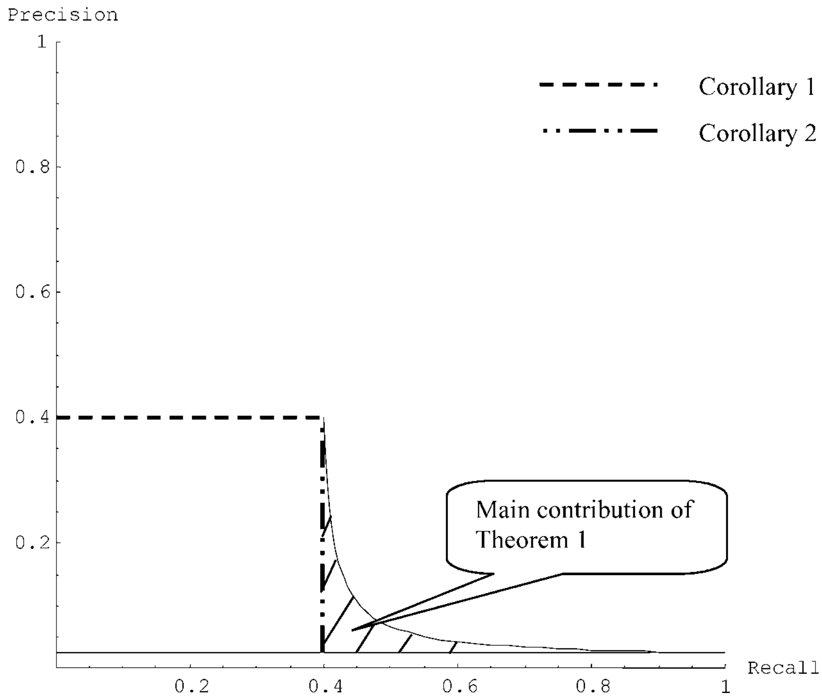

Figure 2. Three areas of dominance in *p-R* coordinates.

3. The area that is bordered by Corollaries 1 and 2 lines represents all the systems with lower recall and lower precision dominated by system A.
4. However, it is difficult to compare two systems where one system has higher recall and the other higher precision. We have found an unexpected dominance region, which is actually the main contribution of Theorem 1. This region represents all the systems that have higher recall and lower precision snd are none the less dominated by system A. The significance of this dominance region is that it is possible to rank ordered systems and to point up the universally superior system (by using the Blackwell theorem) when the superior system has lower recall but a sufficiently higher precision. This dominance area is illustrated in figure 2 as the "main contribution of Theorem 1".

6. Modeling IF systems with detailed user profiles

In this section the model is extended to support filtering systems that maintain detailed user profiles. The detailed profile represents the user long-term preferences regarding several specific areas of interest. Using IS model terminology, the user profile is defined as a payoff matrix reflecting the user's perceived payoff for the receipt of information items flagged for each area of interest included in the detailed user profile. The following is an illustration of modeling such IF systems by an IS. Two different users are represented by two distinct payoff matrixes which reflect their user profiles.

The IS model parameters in this example are:

- Events: $S = \{$News, Finance, Computers, Other subjects$\}$—representing the set of areas of interests (subjects), that the systems might receive documents about and is able to deal with.
- A priori probabilities: $\pi = \{$0.05 for News, 0.15 for Finance, 0.1 for Computers, 0.7 for other subjects$\}$—representing the probabilities of receiving documents relating to the known areas of interest (the events).
- Signals: $Y = \{$Flagged News, Flagged Finance, Flagged Computers, Non-flagged$\}$—representing the set of signals that the system can send notifying the user as to which area of interest the arriving document relates to.
- Actions: $A = \{$Read, Disregard$\}$—representing the actions that the user might take upon a receipt of a signal from the system.

The matrix of the IS Q is as follows:

	Signals			
IS Q events	Flagged News	Flagged Finance	Flagged Computers	Non-flagged
News	0.80	0.05	0.04	0.11
Finance	0.02	0.70	0.10	0.18
Computers	0.07	0.03	0.85	0.05
Other subjects	0.10	0.05	0.15	0.70

Each of the first three rows of the structure Q represents the ability of the IF system to "detect" incoming documents in the related subject that is represented by the column. Thus, for example, the cell ["Flagged Finance", "Finance"] presents the probability (0.70) that the system will indicate (correctly) that Finance related documents are received, in the event of receiving a Finance related document; while the cell ["Flagged Finance", "News"] presents the probability (0.05) that the system will indicate (incorrectly) that Finance related documents were received in the event of receiving News related documents. The three first rows of the table reflect the ability of the system to indicate the area (domain) of documents for all areas of interests that are part of the users' profile. In other words, it reflects the quality of the system filtering, i.e., the extent of the ability to match the user profile. The fourth row represents the "ability" of the IF system to flag information in subjects that are *not* of interest to the user. The fourth row ("Other subjects") reflects the extent of "under filtering" of the system. For example, the cell ["Flagged News", "Other Subjects"] presents the probability (0.10) that the system will (incorrectly) indicate that a News related document is received when actually a non-relevant document is received, i.e., a document not related to any of the areas of interests of the user's profile.

Consider the following payoff matrices U^1 *and* U^2 for two different users. The following matrices represent their profiles:

	Events			
U^1 actions	News	Finance	Computers	Other subjects
Read	15	10	30	−3
Disregard	−2	−5	−3	0

	Events			
U^2 actions	News	Finance	Computers	Other subjects
Read	15	10	7	−9
Disregard	−5	−15	−4	0

Each of the above payoff matrices (U^1, U^2), presents for one user the payoff for a received document for each area of interest, and the damage caused by disregarding relevant information for every area of interest. The payoff matrix represents the detailed user profile as it includes a declaration of the importance of each subject to the user.

The expected payoff for the first user (User1) is

$$EU^1 = tr(QD^1U^1\Pi) = 0.63 \cdot d_{11} - 0.116 \cdot d_{12} + 1.0725 \cdot d_{21} - 0.539 \cdot d_{22} + 2.415 \cdot d_{31} - 0.334 \cdot d_{32} - 0.9675 \cdot d_{41} - 0.161 \cdot d_{42} \quad (10)$$

For that expected payoff, the optimal decision strategy D^1 is:

	Actions	
D^1 signals	Read	Disregard
Flagged News	1	0
Flagged Finance	1	0
Flagged Computers	1	0
Non-flagged	0	1

The expected payoff for the second user (User2) is

$$EU^2 = tr(QD^2U^2\Pi) = 0.049 \cdot d_{11} - 0.273 \cdot d_{12} + 0.7935 \cdot d_{21} - 1.5995 \cdot d_{22} - 0.17 \cdot d_{31} - 0.575 \cdot d_{32} - 4.0225 \cdot d_{41} - 0.4525 \cdot d_{42} \quad (11)$$

For that utility, the optimal decision strategy D^2 is:

	Actions	
D^2 signals	Read	Disregard
Flagged News	1	0
Flagged Finance	1	0
Flagged Computers	1	0
Non-flagged	0	1

In this case the optimal decision rule suggests that both users should follow the system's recommendations (for all subjects on the profile) in order to maximize their expected payoff. When applying the optimal decision rule to the EU^1 *and* EU^2 equations the maximal expected payoff obtained in this example is: $Q\text{-}EU^1 = 3.9565$, $Q\text{-}EU^2 = 0.22$. This example demonstrates how different user profiles as expressed by different payoff matrices result in different expected payoff for the same IF system.

According to general utility theory, we do not attempt to compare utilities for different users so the fact that these numbers $Q\text{-}EU^1$ and $Q\text{-}EU^2$ are different has no implications for action. It does not, for example, mean that the system *would serve User1 better than it would serve User2*. Conclusions can only be made regarding the comparison of two systems for the same user.

The following example assumes that the same two users (User1, User2) are using another IF system modeled by the IS T, whose matrix is

	Signals			
IS *T* events	Flagged News	Flagged Finance	Flagged Computers	Non-flagged
News	0.7278	0.0873	0.0716	0.1133
Finance	0.0500	0.6042	0.1046	0.2412
Computers	0.0834	0.0395	0.8111	0.066
Other subjects	0.1295	0.0770	0.1540	0.6395

The expected payoff for the optimal strategy of the users (based on their above U1 and U2 Payoff matrices, i.e., their profiles) is found as above, yielding: $T\text{-}EU^1 = 3.63615$ $T\text{-}EU^2 = -0.41155$.

The specific optimal decision strategies (D^1) and (D^2) of the User1 remains the same. For both users the expected payoff is lower when using system T compared to using system Q, for User1: $Q\text{-}EU^1 = 3.9565$ $T\text{-}EU^1 = 3.63615$, for User2: $Q\text{-}EU^2 = 0.22$ $T\text{-}EU^1 = -0.41155$. This suggests that for these two users the system modeled by T is inferior to the system modeled by Q. However, in order to conclude that the system modeled by T is *always* worse than system modeled by Q, i.e. for every user with every profile, there must exist a garbling matrix M such that $Q \cdot M = T$, and M is a Markov matrix. This will satisfy the Blackwell theorem conditions. The following matrix is the desired M for the above example:

$$M = \begin{bmatrix} 0.90 & 0.05 & 0.04 & 0.01 \\ 0.03 & 0.85 & 0.01 & 0.11 \\ 0.02 & 0.01 & 0.95 & 0.02 \\ 0.05 & 0.04 & 0.01 & 0.90 \end{bmatrix} \tag{12}$$

The existence of M for the above example allows us to conclude that for any user with *any* profile, system T is always worse than system Q. The ability to conclude such dominance between two such systems is an important extension to the model as it permits comparison of filtering systems that include *detailed profiles*.

7. Discussion and future research directions

We have shown that it is possible to take a deep theorem about information structures and transform it into the language familiar in IR, the language of precision and recall. The key ingredient is the specific value of G, the density of relevant items in the incoming stream. For different values of G, the region dominated by a system with specific (p, R) will vary. However, as long as p is greater than G (and if it is not, then the system is actually worse than random guessing) the dominated region will include the lines of constant p and constant R, as determined in Corollaries 1 and 2. Thus for all values of $G < p$, any system with smaller p and smaller R is dominated by the given system. This result, which is intuitively clear, is given a rigorous foundation in the theory of information structures, and the Blackwell theorem.

While the situation for the simple filtering case is easily represented in terms of a p-R graph, as shown here, the model also applies to more complex systems, which function as text classifiers. We have shown the basic model in Section 6. Those complex models are an interesting topic in their own right, applying to the general text classification problem, and they will be presented elsewhere. Classification of texts into multiple categories is of considerable importance in areas ranging from Customer Resource Management to intelligence work. We mention also that the complex shape of the curves in the precision recall plot can be translated into a less familiar plot whose axes represent $(1/p, 1/R)$. These variables were considered by Van Rijsbergen (1979) in the context of the so-called E and F measures. The dominance region, according to the Blackwell theorem can easily be seen on these plots. As might be expected, systems with higher values of E do not necessarily dominate those with lower values, or vice versa.

Real systems usually have a complex performance curve, with many possible values of the pair (p, R), depending on the setting of some threshold parameter. It can be shown that the Blackwell theorem applies here as well, with one system dominating another if and only if its precision-recall curve always lies above that of the other. This subsumes the approach using cumulated numbers of relevant documents applied to the evaluation of IR systems by Kantor and Voorhees (2000). In extreme cases, a *single* tuning of a more powerful system will be more informative than *any* tuning of the less informative system.

We believe that examining filtering systems from the point of view of information structures will provide a rigorous basis for the sometimes vexing issue of comparing the performance of those systems.

Appendix A: Using LP techniques to determine whether two systems can be ordered using the Blackwell theorem

It is a critical practical issue for systems comparison to be able to determine whether M exists for given Q and T and how to compute it when its existence is known. In this section we suggest converting the "existence of M" problem to an LP (Linear Programming) problem. In LP terminology the problem is defined as finding a matrix M, given matrices Q, T, to satisfy one of the following two sets of constraints:

1. $Q \cdot M - T = 0; \quad M_{ij} \geq 0; \quad \sum_j M_{ij} = 1$ in case that Q is better than T.
2. $T \cdot M - R = 0; \quad M_{ij} \geq 0; \quad \sum_j M_{ij} = 1$ in case that T is better than Q.

The second and the third constraints in each of the two sets stem from M being Markov. Any software that has Linear Programming capabilities (such as Mathematica) can apply an LP algorithm (such as SIMPLEX) to solve this problem. If any of the sets of constrains are not satisfied, then Q and T *cannot* be ordered by the Blackwell theorem.

For example: given the following Q T 2×2 matrices, and defining the above constraints would yield the following M when applying LP in Mathematica, meaning that Q and T are comparable, and that Q is always better than T.

$$
\begin{array}{ll}
Q = \begin{bmatrix} 0.94 & 0.06 \\ 0.11 & 0.89 \end{bmatrix} & \\
& \begin{cases} Q \cdot M - T = 0 \\ \sum_j M_{ij} = 1 \\ M_{ij} \geq 0; \end{cases} \Rightarrow M = \begin{bmatrix} 0.999880 & 0.000120482 \\ 0.168554 & 0.831446000 \end{bmatrix} \\
T = \begin{bmatrix} 0.95 & 0.05 \\ 0.26 & 0.74 \end{bmatrix} & \\
& \begin{cases} T \cdot M - Q = 0 \\ \sum_j M_{ij} = 1 \\ M_{ij} \geq 0; \end{cases} \Rightarrow \text{There is no } M \text{ that satisfies the contraints}
\end{array}
\tag{A1}
$$

However, applying LP to the following Q and T would result in a message indicating that M does not exist for either set of constraints, meaning that these two systems can not be compared under the Blackwell theorem.

$$Q = \begin{bmatrix} 0.91 & 0.09 \\ 0.21 & 0.79 \end{bmatrix} \qquad \begin{cases} Q \cdot M - T = 0 \\ \sum_j M_{ij} = 1 \\ M_{ij} \geq 0; \end{cases} \Rightarrow \text{There is no } M \text{ that satisfies the contraints}$$

$$T = \begin{bmatrix} 0.95 & 0.05 \\ 0.26 & 0.74 \end{bmatrix} \qquad \begin{cases} T \cdot M - Q = 0 \\ \sum_j M_{ij} = 1 \\ M_{ij} \geq 0; \end{cases} \Rightarrow \text{There is no } M \text{ that satisfies the contraints} \tag{A2}$$

Appendix B: Proof of Theorem 1

Theorem 1. *Assume an IF system A with recall R and Precision P and consider a bounded region in the p-R coordinates defined by the following parameterized curves*:

1. $0 \leq \alpha \leq 1 \quad \begin{cases} R^\alpha = \alpha \cdot R \\ P^\alpha = P \end{cases}$
2. $0 \leq \beta \leq 1 \quad \begin{cases} R^\beta = \beta \\ P^\beta = G \end{cases}$
3. $0 \leq \gamma \leq 1 \quad \begin{cases} R^\gamma = \gamma R + (1-\gamma) \\ P^\gamma = \dfrac{R^\gamma P}{R^\gamma \cdot P + \left(\gamma R(1-P) + \frac{P}{G}(1-\gamma)(1-G)\right)} \end{cases}$

Any IF system with precision and recall within the bounded region is inferior to system A according to the Blackwell theorem.

Proof: In this proof we show a graphical representation of the comparability condition of the Blackwell dominance relation and extend it to a precision-recall diagram. Thus, we graphically define a region (rather than points) for Blackwell comparable systems (i.e., systems that can be compared using Blackwell theorem).

The relation $Q = T \cdot M$ where T and M are Markov matrices admits a useful geometric interpretation, which will help us to extend it to the p-R diagram that is familiar to researchers in Information Retrieval. We note first that the IS T modeling an IF system has two characteristic probabilities $d = \Pr(\textit{Flagged/Relevant})$, and $f = \Pr(\textit{Flagged/Non-relevant})$, and can be represented in a plot whose axes are the conditional probabilities to flag a relevant (y-axis) and non-relevant (x-axis) document. The two numbers define a vector (f, d), and (trivially) a second vector, $(1 - f, 1 - d)$. Their sum, is $(1, 1)$. As shown in figure 3, they define a parallelogram. The point corresponding to the system T is marked by (f, d), and the reflected point marked by $(1 - f, 1 - d)$. The relation to the Blackwell

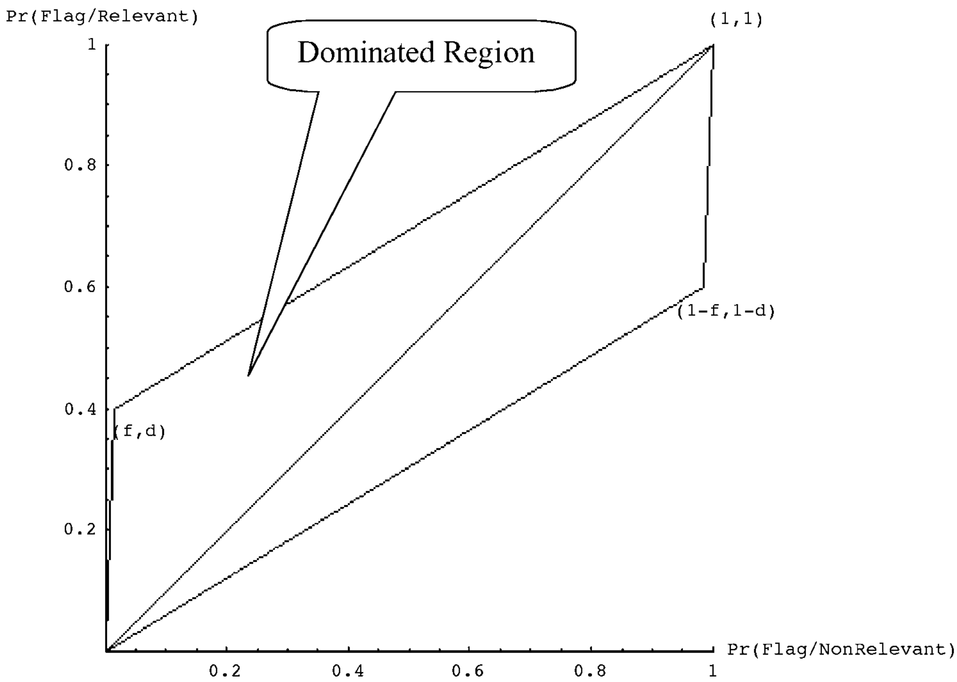

Figure 3. Dominance in f-d coordinates.

theorem is that a system whose representative pair of points lies within the parallelogram is exactly a system in the form $T \cdot M$, which is less informative than the system T. It is easy to see that the expression of any point within the triangle defined by (0, 0), (1, 1), and (f, d) as a convex combination of the three vertices can be transformed into an expression in terms of a Markov matrix such that the interior point is represented by $T \cdot M$. □

We must show how the Blackwell dominance region can be drawn in p-R coordinates based on the system's derived precision and recall on each edge.

The first edge of the triangle ((0, 0), (f, d)) represents all the systems $(\alpha \cdot f, \alpha \cdot d)$ where α is in the interval [0, 1]. The average precision of the systems on this line is constant and does not change with $\alpha(P^\alpha = P)$ whilst the recall changes from 0 for $\alpha = 0$ to the recall R of the system $T(R^\alpha = \alpha \cdot R)$. This edge represents all the systems based on the system that is represented by (f, d) and the trivial alternative "never flag". These systems retain (with probability α) documents that were flagged by the system (f, d), and throw the flagged documents with probability $(1 - \alpha)$ away. As α ranges from 0, 1, these systems sweep out the line from (0, 0) to (f, d).

The second edge of the triangle, ((0, 0), (1, 1)), represents all the systems flagging documents for retrieval entirely at random. Thus $d = f = \beta$ where β represents the fraction of all documents that are flagged by the system. Systems on this edge have a recall equal to $\beta(R^\beta = \beta)$ while the precision equals the density measure $G(P^\beta = G)$ (as we have defined G on the 3rd section to be average density of an incoming stream of information, i.e., the average chance that an incoming item is relevant). As β ranges from 0 to 1, it sweeps the line from (0, 0) to (1, 1).

The third edge of the triangle, $((f, d), (1, 1))$, is the most interesting one. It represents the systems that can be achieved by randomly mixing the system at hand, T, with the system that retrieves all the documents, which will be denoted by I (I has $d = f = 1$). For a particular mixture $(\gamma T + (1 - \gamma)I)$ we can calculate the actual values of the precision P^γ and recall R^γ. The recall equals a mixture of the recall of both systems ($R^\gamma = \gamma R + (1-\gamma)$). The precision of the mixed systems is derived in the following way:

The value of f of system T (based on Eq. 14) is:

$$f = \Pr(\mathit{Flagged}/\mathit{Non\text{-}relevant}) = \frac{R \cdot G - R \cdot G \cdot P}{(1 - G) \cdot P} \tag{B1}$$

The value of f of the mixed system (denoted by f^γ) is

$$f^\gamma = \gamma \cdot f + 1 - \gamma = \gamma \frac{R \cdot G - R \cdot G \cdot P}{(1 - G) \cdot P} + 1 - \gamma \tag{B2}$$

The precision of the mixed system (based on Eq. 12) is

$$\begin{aligned} P^\gamma &= \frac{R^\gamma \cdot G}{R^\gamma \cdot G + f^\gamma \cdot (1 - G)} = \frac{R^\gamma \cdot G}{R^\gamma \cdot G + \left(\gamma \frac{R \cdot G - R \cdot G \cdot P}{(1-G) \cdot P} + 1 - \gamma\right) \cdot (1 - G)} \\ &= \frac{R^\gamma \cdot G \cdot P}{R^\gamma \cdot G \cdot P + (\gamma R \cdot G - \gamma R \cdot G \cdot P + P(1 - \gamma)(1 - G))} \end{aligned} \tag{B3}$$

Each of the triangle edges in figure 3 is drawn in p-R coordinates in figure 1 based on the system's derived precision and recall on each edge (as summarized in Table 1). The line $((0, 0), (f, d))$ in f-d coordinates is transformed into the line $((0, P), (R, P))$ in p-R coordinates. The line $((0, 0), (1, 1))$ in f-d coordinates is transformed into the line $((0, G), (1, G))$ in p-R coordinates.

Table 1. Systems whose performance lies on the Triangle $((0, 0), (f, d), (1, 1))$.[a]

Triangle edge	Systems	Precision and recall
$((0, 0), (f, d))$	$\begin{bmatrix} \alpha d & 1 - \alpha d \\ \alpha f & 1 - \alpha f \end{bmatrix}$ for $0 \le \alpha \le 1$	$R^\alpha = \alpha \cdot R$ $P^\alpha = P$
$((0, 0), (1, 1))$	$\begin{bmatrix} \beta & 1 - \beta \\ \beta & 1 - \beta \end{bmatrix}$ for $0 \le \beta \le 1$	$R^\beta = \beta$ $P^\beta = G$
$((f, d), (1, 1))$	$\begin{bmatrix} \gamma d + 1 - \gamma & \gamma(1 - d) \\ \gamma f + 1 - \gamma & \gamma(1 - f) \end{bmatrix}$ for $0 \le \gamma \le 1$	$R^\gamma = \gamma R + (1 - \gamma)$ $P^\gamma = \frac{R^\gamma P}{R^\gamma \cdot P + \left(\gamma R(1 - P) + \frac{P}{G}(1 - \gamma)(1 - G)\right)}$

[a]We do not discuss systems with $P < G$ as these systems are irrelevant.

Appendix C: Proof of Corollary 1

Corollary 1. *Any set of IF systems modeled by ISs can be rank ordered using the Blackwell theorem if their precision measure is the same.*

Proof: Assume Q and T are ISs modeling two IF systems where R^Q, P^Q and R^T, P^T are the recall and precision of the systems respectively and $P^Q = P^T$.

The ISs Q and T have the following structure where $P = P^Q = P^T$:

$$Q = \begin{bmatrix} R^Q & 1 - R^Q \\ \dfrac{R^Q \cdot G - R^Q \cdot G \cdot P}{(1-G)\cdot P} & 1 - \dfrac{R^Q \cdot G - R^Q \cdot G \cdot P}{(1-G)\cdot P} \end{bmatrix}$$
$$T = \begin{bmatrix} R^T & 1 - R^T \\ \dfrac{R^T \cdot G - R^T \cdot G \cdot P}{(1-G)\cdot P} & 1 - \dfrac{R^T \cdot G - R^T \cdot G \cdot P}{(1-G)\cdot P} \end{bmatrix} \tag{C1}$$

Q is more informative than T according to Blackwell theorem if there exists a Markov matrix $M = Q^{-1} \cdot T$. In this case, Q is invertible, and:

$$M = Q^{-1} \cdot T = \begin{bmatrix} \dfrac{R^T}{R^Q} & 1 - \dfrac{R^T}{R^Q} \\ 0 & 1 \end{bmatrix} \tag{C2}$$

M will be Markovian if $R^Q > R^T$. Thus, any pair of IF systems with the same precision can be compared using the Blackwell theorem since M can be found for the IF with the higher recall. □

Appendix D: Proof of Corollary 2

Corollary 2. *IF systems modeled by ISs can be rank ordered using the Blackwell theorem if their recall measure is the same.*

Proof: Assume Q and T are ISs modeling two IF systems where R^Q, P^Q and R^T, P^T are the recall and precision of the systems respectively and $R = R^Q = R^T$.

Q is more informative than T according to Blackwell Theorem if there exists a Markov matrix $M = Q^{-1} \cdot T$. As Q is invertible we find directly:

$$M = Q^{-1} \cdot T$$
$$= \begin{bmatrix} \dfrac{G \cdot P^Q - P^Q \cdot P^T - G \cdot P^Q \cdot R + G \cdot P^T \cdot R}{(G - P^Q)\cdot P^T} & \dfrac{G \cdot (P^Q - P^T)(R-1)}{(G - P^Q)\cdot P^T} \\ \dfrac{G \cdot (P^T - P^Q)\cdot R}{(G - P^Q)\cdot P^T} & \dfrac{G \cdot P^T - P^Q \cdot P^T + G \cdot P^Q \cdot R - G \cdot P^T \cdot R}{(G - P^Q)\cdot P^T} \end{bmatrix} \tag{D1}$$

M will be Markovian if $P^Q > P^T$ since we assume $P^Q, P^T > G$ and these conditions make the off diagonal elements positive and less than 1. □

References

Ahituv NA (1981) Comparison of information structures for a "rigid decision rule" case. Decision Sciences, 12:399–416.

Ahituv N and Ronen B (1988) Orthogonal information structures—A model to evaluate the information provided by second opinion. Decision Sciences, 19:255–268.

Ahituv N and Wand Y (1984) Comparative estimation of information under two business objectives. Decision Sciences, 15:31–51.

Baeza-Yates R and Ribeiro-Neto B (1999) Modern Information Retrieval. ACM Press, Addison-Wesley, New York.

Belkin NJ and Croft WB (1992) Information filtering and information retrieval two sides of the same coin? Communications of the ACM, 35(12):29–38.

Blackwell D (1951) Comparison of experiments. In: Neyman J, Ed., Proceedings of the Second Berekeley Symposium on Mathematical Statistics and Probability, pp. 93–102.

Blackwell D and Girshick MA (1954, 1979) Theory of Games and Statistical Decisions. John Wiley and Sons (1954), Dover (1979), p. 353.

Carmi N and Ronen B (1996), An information value approach to quality control attribute sampling. European Journal of Operational Research, 92(3):615–627.

Demski JS (1972) Information Analysis. Addison-Wesley, Reading, MA.

Frakes WB and Baeza-Yates R (1992) Information Retrieval: Data Structures and Algorithms. Prentice-Hall, Englewood Cliffs, NJ.

Hanani U, Shapira B and Shoval P (2001) Information filtering: Overview of issues, research and systems. User Modeling and User-Adapted Interaction, 11:203–259.

Hilton RW (1990) Failure of Blackwell's Theorem under Machina's generalization of expected utility analysis without the independence axiom. Journal of Economics Behavior & Organization, 13(2):233–244.

Hull DA and Robertson S (2000) The TREC-8 filtering track final report. In: Voorhees EM and Harman DKm, Eds., The 8th Text Retrieval Conference (TREC-8), NIST SP 500–246, pp. 35–36.

Kantor PB and Voorhees EM (2000) The TREC-5 confusion track: Comparing retrieval methods for scanned text. Information Retrieval, v2:165–177.

Luce RD and Winterfeldt D (1994) What common ground exists for descriptive, prescriptive, and normative utility theories, Management Science, 40:263–279.

Marschak J (1971) Economics of information systems. Journal of the American Statistical Association, 192–219.

Marshall RM and Narashimhan R (1989) Risk constrained information choice. Decision Sciences, 20(4):677–684.

McGuire CB and Radner R (1986), Eds. Decision and Organization. North Holland, Amsterdam.

Oard WD (1997) The state of the art in text filtering. User Modeling and User Adapted Interaction (UMUAI), 7(3):141–178.

Robertson SE (2002) Overview of The TREC routing and filtering tasks. Information Retrieval, 5:127–137

Ronen B (1994) An information value approach to quality control attribute sampling, European Journal of Operational Research, 73:430–442.

Ronen B and Spector Y (1995) Evaluating sampling strategy under two criteria. European Journal of Operational Research, 80:59–67.

Sinchcombe MB (1990) Baysian information topologies. Journal of Mathematical Economics, 19(3):233–253.

Tauge-Sutcliffe J (1992) Measuring the informativeness of a retrieval process. In: Proceedings of the Fifteen Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 23–36.

Van Rijsbergen C (1979) Information Retrieval, 2nd edn. Butterworth, London (Especially chapter 7).