Evaluation of a Belief-Network–Based Reminder System that Learns from Utility Feedback

Michael M. Wagner, Gregory F. Cooper, Section of Medical Informatics, University of Pittsburgh

PRETRIEVE is a belief-network–based, unsolicited information-retrieval system that performs machine learning based on user feedback. We report here on the document-ordering and document-retrieval performance of PRETRIEVE.

We developed a test collection of 410 judgments of document utility in a simulated medical order-entry context. We characterized the validity of these judgments, which were elicited from domain experts, by measuring interrater and intrarater reproducibility. We developed a measure of the quality of document orderings similar to the ROC-curve analysis used to evaluate document-retrieval systems. We found that the ordering performance of the PRETRIEVE system was (1) substantially better than random, (2) somewhat less than ideal, and (3) superior to that of versions of the PRETRIEVE system that used relevance feedback instead of utility feedback. Under a set of assumptions, which we make explicit, we found that the documents retrieved by a version of PRETRIEVE that modeled time cost were of higher utility than those retrieved by a similar rule-based system.

INTRODUCTION

A reminder system is an example of an **unsolicited information-retrieval (UIR) system,** which we define as a system that retrieves information for a person without requiring him to formulate a query or initiate a search. UIR systems are of increasing importance in medicine because they relieve clinicians of the burden of formulating queries, knowing *where* and *how* to look for information, and even knowing *whether* to look for information (a clinician may be unaware of a recent change in medical practice and therefore not look for information about it). UIR systems are highly acceptable to physicians [1], and they have significant potential as solutions to the problem of unmet information needs [2] and as a means for the dissemination of practice guideline [3].

To understand UIR systems, we formulated a decision-theoretic model of UIR (DT-UIR) [4]. We chose the decision-theoretic formalism because we wanted to model explicitly the tradeoff between the value of the information in a document and the cost of the physician time spent in *extracting* that information from the document. We also wanted to model the uncertainty inherent in the inference that a UIR system must make, namely, given the evidence available to the UIR system in computer-readable form, what is the expected utility of a document to a particular patient being treated by a particular physician.

In this paper, we briefly review our DT-UIR model. We then describe PRETRIEVE—a belief-network–based version of the model—and we present data on (1) the document-ordering performance of PRETRIEVE (and variants), (2) the document-ordering performance of a hybrid rule-based/decision-theoretic system, and (3) the document-retrieval performance of a pure decision-theoretic reminder system that models time cost and value of information.

DT-UIR MODEL

A computer-based UIR system selects *information objects,* which we refer to as **documents,** for a person based on **evidence** of that person's information needs. A decision-theoretic UIR system models the probability that documents will be useful to that individual, and the degree to which they may be useful (their utility). It assembles an *optimal* set of documents by computing the set of documents with the greatest expected utility given available evidence of information need.

Our DT-UIR model is based on a **document-independence assumption**, that is, we assume that the utility of a set of documents is equal to the sum of the utilities of the individual documents. This assumption, which is made by all document-retrieval systems, permits a system to find the optimal *set* of documents by locating just those whose expected utility (EU) is greater than zero. Under this assumption, a system can distribute this computation among d independent processes, each computing the EU of a single document (Fig. 1). We refer to these processes as **automated document experts**.

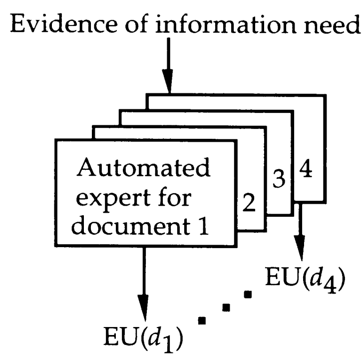

Fig. 1. DT-UIR model

0195–4210/95/$5.00

THE PRETRIEVE SYSTEM

We implemented the PRETRIEVE system in a setting that simulates the point at which a physician has just finished writing admission orders for a patient. We shall explain the operation of PRETRIEVE in this setting and we shall refer to it as a *reminder* system, although it sends a broader class of documents to physicians than are sent by a typical reminder system.

The automated document experts in the PRETRIEVE system are belief networks (Fig. 2). Each belief network comprises a single evidence node and a single utility node. The **evidence node** represents a logical statement about patient characteristics and physician orders that might predict whether the document would have utility for a patient, if it were to be sent to the ordering physician. As in a rule-based reminder system, this statement is formulated by a domain expert. We denote the node that represents this statement as **LHS,** for *left-hand side* because of this similarity to rules, and because we also used these statements as the left-hand sides of rules in a rule-based reminder system that we built to compare with PRETRIEVE. (Although we use a single evidence node in most experiments, PRETRIEVE places no restriction on the number of evidence variables.)

The **utility node** represents the value of the information in the document to the recipient of the document. We call this node **VDI,** for *value of document information*. In this medical application, we define document utility as the expected benefit (or damage) to the patient of any changes in physician orders or physician management plans that were caused by the information in the document. If there are no changes in management, VDI is zero. We measure VDI in units of quality-adjusted life days (QALDs).

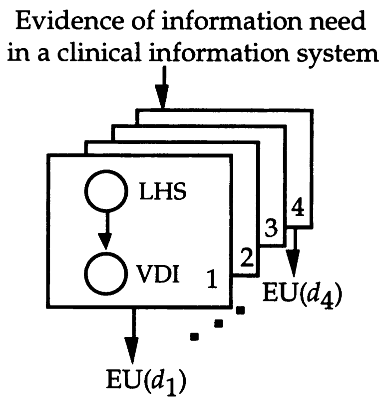

Fig. 2. The PRETRIEVE system

PRETRIEVE uses the K2 machine-learning algorithm [5] to learn its belief-network models from document-utility data, as we shall discuss in *Methods*. In this two-variable implementation, PRETRIEVE, may learn that the evidence is irrelevant to utility (a belief-network structure in which the LHS node is not connected to the VDI node by an arc), or it may learn that the two variables are probabilistically dependent. With multiple evidence variables, more complex models are possible.

To impose an ordering on a set of documents (or to select the optimal subset as in the full DT-UIR model), PRETRIEVE (1) determines the truth-value of the LHS node (for each automated document expert) by checking a database representation of the patient, (2) sets the value of the LHS node, (3) queries the belief network to compute the expected VDI, which we refer to as EVDI, and (4) orders the documents by the EVDIs thus computed. In the case of the complete DT-UIR model, PRETRIEVE finds the optimal *set* of documents by eliminating documents for which the expected time cost to realize the benefit of information (i.e., the cost of the time to read and act on the information in the document) exceeds the EVDI.

METHODS

To evaluate PRETRIEVE, we created an experimental situation that retains the essential features of the *reminding* situation, while controlling selected experimental conditions.

Test collection development

We created a **test collection** of 410 document-utility assessments in the following manner. We recruited two experts on the treatment of HIV-infected patients to serve as domain experts. With the help of the HIV experts, we assembled (and sometimes wrote) 10 *documents* about the management of HIV-infected patients. This set included journal articles (2), American College of Physician Journal Club reviews (3), clinical reminders (5) and a published practice guideline (1). We assembled a set of 41 admission history and physical reports for HIV-infected patients that we drew at random from a hospital information system, edited to remove identifying data, and formatted to improve readability. We also recruited nine geriatricians with limited experience in treating HIV-infected patients to serve as treating physicians.

We assigned three to five patient cases to each of the nine geriatricians (without overlap). Each geriatrician wrote admitting orders and described his management plans for his assigned cases. After writing orders for all cases, a geriatrician was given the first of the 10 documents and instructed that he should use it to try to improve upon the management plan and orders that he had just written. We recorded plan changes *resulting from* the first document. We presented the remaining documents, instructing the geriatrician to ignore the previous documents, and recorded their effects on management plans. We recorded the time that each physician spent reading and deciding whether to change his plan for the document on its first presentation.

For each of the 410 patient–document pairs, we determined the *utility* of the document as follows. If a document caused no change in management plan, the utility was zero (327/410 instances). The utility of the 83 instances in which a document caused a plan change, measured in QALDs, were determined by an HIV expert and author *MMW* in sessions in which *MMW* served as decision analyst. In these sessions, we reviewed published cost–utility and effectiveness data about the management-plan changes of the geriatricians. We divided these 83 assessment tasks randomly between the two experts by document. To partially validate the assessment method, we repeated 114 utility assessments; 65 in which the same expert did the same assessment (same patient, same document) after a delay of at least 3.5 months, and 49 in which the other expert repeated the assessment.

LHS (evidence) and _VDI_ (utility) variables

To create the evidence variables for the automated document experts, we asked one of the experts to state 10 *rules* (as in a rule-based reminder system rule) describing when a hypothetical computer-based reminder system should send each document to a physician.

We discretized the *VDI* variable in each automated document expert into three categories—negative, zero, and positive utility. We set the value of each category equal to the mean of the utility determinations for that category. For example, the three categories for document 4 (a reminder about steroids for *Pneumocystis carinii* pneumonia) were set to -7.01, 0, and 38.57 QALDs; the -7.01 value was the mean document utility of the three cases in which this document had a negative utility, as determined by the HIV experts.

Patient database

Because no coded representation of the patient cases was available to us, we created one by representing the age, symptoms, allergies, medications, diagnoses, and laboratory results for the 41 patients in the form of object-attribute-value triples (e.g., 'patient 1', 'diagnosis', 'Pneumocystis pneumonia'). We created these triples before conducting any sessions with the experts or the geriatricians (a geriatrician's initial orders and plans were, of necessity, added to the representation after the session with that physician).

Experimental design

PRETRIEVE used the test-collection data as the source of training data for the K2 machine-learning algorithm. We used the same patient cases to test PRETRIEVE. To avoid overfitting bias, we employed a leave-one-out cross-validation experimental design in which the automated document expert that predicted EVDI for a case, *C*, was learned by K2 from training data that excluded case *C*.

For the ordering experiments, PRETRIEVE (and tested variants) ordered the documents for each of the 41 patients by (1) computing the EVDI for the documents, given the evidence in the database about the patient, and (2) then sorting documents by EVDI.

For the hybrid-system evaluation, the hybrid system (1) retrieved only those documents whose LHS proposition (rule) was satisfied for a patient, and (2) then sorted those documents by EVDI.

For the document-retrieval experiments, the system retrieved just those documents whose EU—computed as the difference between EVDI and expected time cost—was greater than zero.

Statistical analysis

The statistical analyses of the ordering data evaluated whether PRETRIEVE could order the documents relevant to a patient encounter according to the utility of the documents as judged by the experts.

For each patient encounter, there were 10 documents of varying expert-judged utility (VDI) that were ordered by PRETRIEVE. We computed 10 partial sums for the expert-judged utilities of this document order. The first sum was just the utility of the first document. The second sum was the utility of the first and second documents, and the tenth sum included all 10 utilities. If, for each patient case, we graphed these partial sums against the number of terms in the partial sums, we would have an ascending curve. If we graphed the means for these partial sums, which is a summary over all patient cases, we would have a **system curve** like the curves in Fig. 3.

The system curves could be compared against other system curves. The **ideal curves** were the partial sums of the best possible orderings as judged by the experts. The **random curves** were the partial sums of the expectation over all randomly generated document orderings for each patient case. All curves, by definition, agreed at positions zero and 10. For a better ordering, however, the curve rose faster at the initial positions and dominated inferior orderings. Hence (as in an ROC curve), the quality of the ordering by each method could be summarized by the **area under the curve (AUC)** for that method. We give a decision-theoretic interpretation of AUC in [7].

The difference between the AUC of PRETRIEVE and the AUC of a random list was calculated for each patient case. To determine whether PRETRIEVE produced a better than random order, we tested the null hypothesis that the mean difference score was zero against the alternative that it was positive. Since the difference scores were not normally distributed, we used the bootstrap algorithm [6] to perform a nonparametric equivalent to the *t*-test. We expected the difference between the AUC of PRETRIEVE and the AUC of the ideal order to be less than zero, hence we did not expect a statistical test of a null hypothetical comparison to be informative. We anticipated, however, that graphing the mean curves (as in Fig. 3) would assist in judgments of whether the difference between PRETRIEVE and the ideal order was of clinical significance.

RESULTS

Validity of utility assessments

The correlation coefficient for the interrater utility assessments was 0.5. The intrarater correlation coefficient was 0.73.

Quality of the evidence

Five of 10 *learned* automated document experts incorporated the influence of the evidence variable (i.e., there was an arc from the LHS to the VDI node in the belief network). In the other five, the expert's rule did not predict the VDI of the document.

Ordering performance of PRETRIEVE

We defined the accuracy of utility prediction as the difference between the expert-judged utility and the predicted EVDI. This accuracy, which was not high in absolute terms [7], was sufficient for the system to impose a high-quality ordering on the documents. Fig. 3 compares the cumulative utility curve of the PRETRIEVE system with (1) the expected random curve, (2) the ordering by expert-judged utility (Ideal), and (3) the reverse ordering by expert-judged utility (Worst case). Although the AUC for the PRETRIEVE curve (416.6 QALDs) was less than ideal (462.6 QALDs; difference = 46.0, 95% C.I. 20.2, 78.4), it was substantially better than random (248.8 QALDs; difference = 167.8, 95% C.I. 74.4, 271.5).

Additional ordering experiments

We built many variants of the PRETRIEVE system and we used the test collection to measure and compare their document-ordering abilities. We summarize selected results from these experiments in Fig. 4. in which the height of each column represents the improvement in AUC of a system relative to the random expectation (which is represented by the floor of the graph). The AUC of the *ideal* ordering, positioned rear-right, represents an upper bound on performance.

We investigated the effect of *evidence* on document-ordering performance. We built variants of PRETRIEVE that used **no evidence** (i.e., the automated document experts had only VDI nodes) and **perfect evidence** (we set the value of the LHS node to predict perfectly the value of the VDI node for each case in the test collection). We found that the *no evidence* system was inferior to the PRETRIEVE system (difference = 34.8 QALDs, 95% C.I. 4.7, 71.1). The *perfect evidence* system was superior to the PRETRIEVE system (difference = 35.5 QALDs, 95% C.I. 14.8, 65.1). We conclude that PRETRIEVE used the evidence to improve the orderings, and still better performance can be expected if we improve on the simple evidence that we used. We built two systems that used multiple evidence variables. One system (VDI, LHS & MD in Fig. 4) added a second evidence variable—representing a physician's HIV-knowledge—to the *LHS* variable. Although three of the automated document experts incorporated this evidence variable into their belief-network models, the performance improvement was slight, and not significant. To build the second multiple evidence-variable system, we created a separate variable for many of the database elements that were referenced by the expert's rule (in effect, we were allowing the system to learn the rule). The performance of this system (VDI, LHS comps) was not significantly different than that of PRETRIEVE.

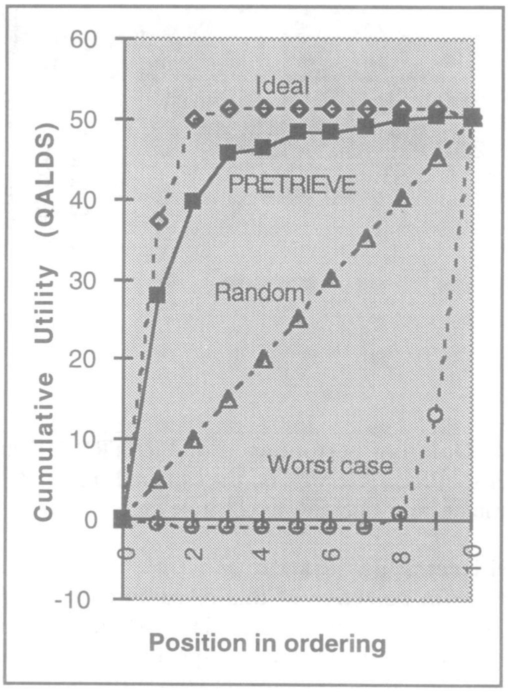

Fig. 3. PRETRIEVE document ordering.

There is an existing decision-theoretic model of document retrieval [8] that assigns all relevant documents utility = 1 and non-relevant documents utility = 0. We hypothesized that our three-valued utility measure would support better document ordering performance than would a 0/1 utility definition. To test this conjecture, we defined a **geriatrician-relevant** document as one that caused the physician to change his management plan (83 instances) and we assigned such documents utility =1. We also defined an **expert-relevant** document as a geriatrician-relevant document for which an expert agreed that the change was beneficial (60 instances). We created versions of PRETRIEVE based on each of these definitions. In these versions, the VDI nodes took the values 0 and 1, and the systems learned their automated document experts from such 0/1 relevance feedback, not utility feedback. We found that utility feedback produced better ordering performance than relevance feedback: The AUC difference between PRETRIEVE and the expert-relevance version was 18.0 QALDs (95% C.I. -1.3, 44.2), close to a statistically significant difference. The AUC difference with the geriatrician-relevance version was 22.3 QALDs (95% C.I. 2.5, 48.8).

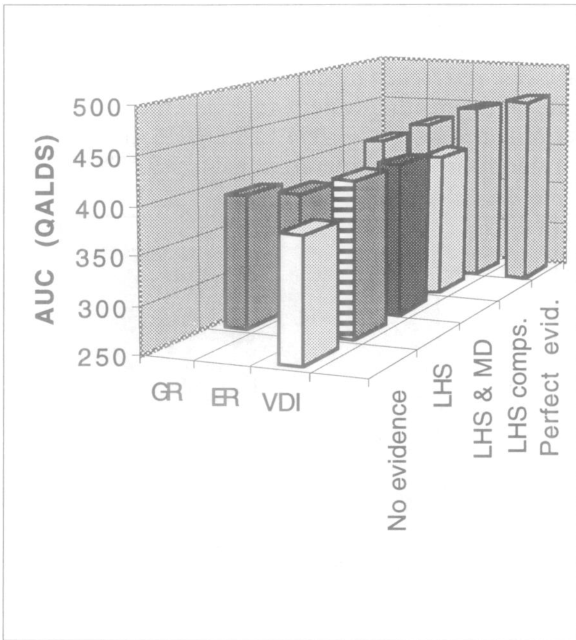

Fig. 4. Ordering performance of PRETRIEVE and variants. PRETRIEVE is the striped bar. GR, geriatrician-relevant; ER, expert-relevant.

Hybrid reminder system

A long-term objective of this research is to build a pure decision-theoretic reminder system (one that computes EU as EVDI minus expected time cost and retrieves only those documents whose EU > 0). We present data about the status and current performance of this model in the next section. Here, we show how the ordering ability of the EVDI-predicting version may be used in a rule-based reminder system.

In Fig. 5, we plot the mean cumulative utility of documents that were retrieved by a rule-based reminder system (labeled *Rule-based*), and we plot the curve for a **hybrid system** that orders the documents selected by the rule-based system according to their EVDI. (We assume that a rule-based system has no ordering capability and therefore in Fig. 5 for the rule-based curve we plot the *expected* AUC over all possible orderings of each set of documents retrieved by the rule-based system.) The ordering performance of the hybrid system, judged by our AUC metric, was significantly better (difference = 22.6 QALDs; 95% C.I. 7.7, 42.6) than that of the rule-based system.

Incidentally, this graph allows us to compare the **retrieval performance**—which we define as the mean utility of *the* set of documents sent to a physician—of the rule-based system to the maximum potential retrieval performance. If we compare the rule-based curve with the ideal curve in Fig. 5, we can see that the ideal curve plateaus after the third position in the ordering at 51.4 QALDs. This value is a measure of the average improvement in patient management attributable to the information in the (beneficial) documents. (The ideal curve drops slightly in the ninth and tenth positions to 50.3 QALDs because of the effect of negatively-valued documents that appear last in an ideal system's orderings). The rule-based system plateaus at 42.6 QALDs because it misses some documents that (would have) changed management plans in ways that the experts judged beneficial to the patients. We could characterize the performance of this reminder system by saying that it is achieving 83% (42.6/51.4) of its potential benefit.

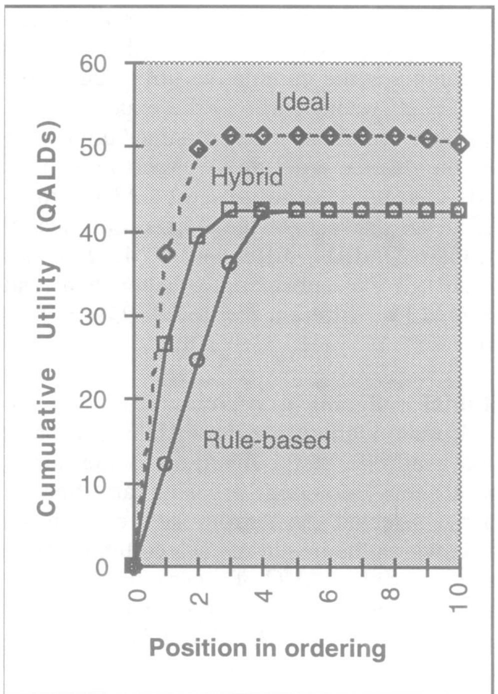

Fig. 5. Rule-based and hybrid reminder systems

PRETRIEVE-TC

To examine the effect of adding time cost (TC) to the model, we built systems based on a wide range of economic assumptions. In particular, we varied the economic value of one QALY (quality–adjusted life year) over the range \$10,000 to \$100,000 and the annual cost of a physician's time from \$100,000 to \$400,000. Using these assumptions, and additional assumptions about the time–utility function [7], we built systems whose automated document experts computed EU as EVDI minus expected time cost. These systems retrieved only those documents with EU > 0.

Time measurements. During the sessions with the geriatricians, time data were collected as described in *Methods*. We found that the mean time that geriatricians spent reading a document and considering whether to change the management of their first patient varied by document, ranging from 29.1 to 219.1 seconds. We used these values as fixed estimates of the expected time that a physician would take to read a document and decide whether to make a change in orders or management plan.

Time cost. We discuss the range of time-cost functions examined in [7]. Here we present results for a system based on one set of assumptions. Specifically, we assumed an annual provider cost of \$200,000, and value of one QALY of \$5,000 (which is set this low for purposes of illustration). We also assumed that the form of the time–cost function was linear. Under these assumptions, one minute of a physician's time is worth \$1.39, 1 QALD is valued at \$13.50, and the time–cost function relating a physician's time (in mins.) to a patient's benefits is

$$TC = time \cdot 0.101 \text{ QALDs / minute.} \quad \text{(Eq. 1)}$$

Comparison with a rule-based system. Since the PRETRIEVE-TC system is an implementation of the complete model, that is, it retrieves the optimal set of documents, we compared *retrieval performance,* as defined in the hybrid-system section, not ordering performance.

We found notable differences in the behavior of the decision-theoretic reminder system and the rule-based reminder system (Table 1). The rule-based system retrieved, on average, 3.1 documents per patient, took 175.6 (expected) seconds of provider time, and yielded a mean improvement in plans for a patient of 42.6 QALDs. The decision-theoretic system retrieved 7.9 documents, took 739 seconds of provider-time, and yielded a mean improvement of 50.1 QALDs. A non-discriminating system (*All 10* in Table 1) would have retrieved all 10 documents, achieved a benefit of 50.3 QALDs, and required 1007 seconds of provider time.

The tradeoff between the PRETRIEVE-TC and the rule-based system can be understood by converting the increment in physician time into QALDs. Under our economic assumptions, the utility of 739-175.6= 563.4 seconds of additional physician time can be computed using Eq. 1 as 0.95 QALDs, which is small compared to the benefit of 50.1-42.6=7.5 QALDs. This comparison suggests that the marginal time to read the additional PRETRIEVE-TC documents was warranted. The 0.2 QALDs gained by the *All 10* system relative to the PRETRIEVE-DT system were gained at the expense of a calculated TC of 0.45 QALDs. We reported here the results of our analysis using the TC assumptions that produced the *highest* TC penalties, which were, nevertheless, small relative to document utility.

Under the assumptions and the conditions of this experiment, we conclude that *this* rule-based system in *this* domain for *these* users is overly restrictive. Given the discriminating ability of the rule-based system, that of the decision-theoretic system, and the value of the documents (for these patients and these physicians), it is better to use the decision-theoretic approach than the rule-based approach. Although this conclusion may, at first, seem counterintuitive to clinicians because of the number of documents that would be retrieved, we should remember that the treating physicians in this experiment were geriatricians who rarely care for HIV patients. For such physicians, sending most of the 10 documents on the rare occasion that they see an HIV patient may very well be the optimal policy.

Table 1. Retrieval performance of PRETRIEVE-TC and a similar rule-based system. *R-based,* rule-based reminder system; *PR-TC,* the time-cost version of PRETRIEVE; *All 10,* the effect of sending all 10 documents for each patient case. Time is in seconds.

System	N	SD	Mean VDI	SD	Mean time	SD
R-Based	3.1	1.4	42.6	83.8	175.6	145.0
PR-TC	7.9	0.2	50.1	86.0	739	20.5
All 10	10	0	50.3	86.0	1007	0

DISCUSSION

The ability of PRETRIEVE to order documents in order of expert-judged utility was substantially better than random. This capability may be of practical use in a UIR application in which documents with the highest expected utility relative to a particular clinical situation are displayed automatically for a clinician.

We demonstrated a second use of EVDI: to order a set of reminders that are selected by a rule-based reminder system. We showed that a hybrid rule-based/EVDI system produced a better ordering than the random ordering typical in current reminder systems.

There is an existing approach to ordering multiple reminders in which an ad hoc priority number is assigned to a reminder. If multiple reminders are triggered by a rule-based system, they are presented to the user in the order of their priority numbers. The priority number is set by the reminder author as a context-*insensitive* property of the reminder [9]. A theoretically-sound definition of the priority number is suggested by our *no evidence* system. The *no evidence* type of system does not use information about a particular patient in its EVDI predictions: it uses the mean utility of a document over all patients. Developers of rule-based reminder systems could measure the mean utility of reminders (without the machine-learning apparatus described here) and use it as a priority number.

Under strong assumptions about time and time cost, the set of reminders retrieved by the PRETRIEVE-DT system were of higher time–cost–adjusted utility than those retrieved by a similar rule-based system. This finding raises the possibility that, contrary to current belief, existing rule-based system may not send enough reminders. The observed results depend heavily on the conditions of the experiment, however, and alternative explanations include (1) that our time-cost estimates are low, and (2) that our method for assessing document utility produces overestimates of document utility. If these alternative explanations are not borne out by future experimental work, however, our results suggest that the answer to current concerns that reminder systems send too many reminders (and therefore risk being ignored by physicians) may not be to reduce the number of false alarms (because too many true alarms go with them), but rather to build time into clinicians' schedules to check alarms, and to educate clinicians to accept the additional work imposed by fail-safe systems.

The measurement of document utility is important in our approach. We showed moderately good interrater and intrarater stability, partly validating our method for obtaining such assessments from experts. We did not attempt to elicit utility directly from treating physicians because we wanted to isolate the evaluation of PRETRIEVE from this research problem. To increase the practicality of our approach, however, additional research on eliciting such assessments directly from ordering physicians is needed.

The AUC metric described here is a new method for measuring information-retrieval system performance. This metric has potential application for measuring the quality of any ordered list of information items, such as a differential diagnosis list output by a diagnostic expert system. Our use of a test collection to evaluate, in effect, a reminder system is another methodological innovation of this research. Test collections have not been used in reminder systems research to compare different algorithms (developers test new reminders against historical patient cases, but these historical collections do not contain data about the effects of reminding). Test collections have facilitated the development of document-retrieval algorithms and may have a similar facilitating effect on reminder systems. A *complete* test collection, that is, one with a utility or relevance determinations for every document for every patient, also enables one to measure the performance of a system relative to a theoretical maximum.

The prediction accuracy of our learning approach was not discussed in detail here. In the experiments described, the utility predictions were noisy due to insufficient data and to the use of simple evidence models that did not represent many of the determinants of document utility. We interpret the fact that we could demonstrate some improvements in document-ordering performance and document-retrieval performance to suggest that very good performance may be possible with better evidence and more training data.

ACKNOWLEDGMENTS

We thank Bob Arnold, Deborah McMahon, and our colleagues in geriatrics for their participation. This work was supported in part by grant BES-9315428 from the National Science Foundation and by grant LM05291-02 from the National Library of Medicine.

References

1. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. N Engl J Med 1976;295:1351-5.
2. Gorman PN, Helfand M. Information seeking in primary care: How physicians choose which clinical questions to pursue and which to leave unanswered. Med Dec Mak 1995;15(2):113-9.
3. Field M, Lohr K, ed. Guidelines for Clinical Practice: From Development to Use. Washington, DC: National Academy Press, 1992.
4. Wagner MM, Cooper GF. Decision-theoretic information pretrieval: a generalization of reminding. Proc. 17th Ann SCAMC, 1993: 512-516.
5. Cooper G, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Mach Learn 1992:309-47.
6. Efron B. The Jacknife, the Bootstrap, and other Resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics. 1982
7. Wagner MM. Decision-Theoretic Reminder Systems that Learn. University of Pittsburgh Dissertation, August 1995.
8. van Rijsbergen C. Information Retrieval. London: Butterworths, 1986.
9. Working group E31.15 AS. Standard specification for defining and sharing modular health knowledge bases (Arden syntax for medical logic systems). ASTM, Report E1460-92, 1992.