Decision Support Information Gathering System

Chiu-Che Tseng   Piotr J. Gmytrasiewicz

Department of Computer Science Engineering
University of Texas at Arlington
Arlington, Texas 76011
tseng, piotr@cse.uta.edu

Abstract[1]

The Decision Support Information Gathering System, Digs, uses influence diagrams to model user's decisions and to calculate the value of imperfect information for each available information source. The system then plans and executes the information gathering process providing the most valuable information to the user. Thus, the system saves time and cost of, sometimes random, search for information performed by using keyword search. Furthermore, it can suggest the proper decision to the user based on the newly retrieved additional information.

Introduction

We describe a Decision Support Information Gathering System (Digs) that suggests to the user how best to retrieve information related to the user's decision situation. Thus, we assume that a user is engaged in making a decision, and that there are many alternative information sources that can be used to aid this process. Since the information may not be available for free, and it may take substantial time to deliver the information to the user, our system evaluates beforehand whether consulting an information source, such as a WWW page, is worthwhile.

The fundamental approach of our Digs system is to endow the system with the model of the user's decision situation. The system then uses this model to execute the information gathering process that best serves the needs of the particular situation at hand using the principled and well-defined notion of value of information. This approach contrasts one taken in a number of information retrieving engines available on the market, such as Lycos, Infoseek, etc. In these systems, the information retrieval is based on keyword search or pattern matching without considering the actual relevance of the information to the user's decision making process.

To accomplish its task, Digs has to consider the following factors.

Search costs. Some information providers may charge for the information provided either per line or per tuple. The system has to calculate the cost of the information gathering and keep it within the user's budget.

Time constraint. In realistic situations, say financial or defense-related, decisions' quality deteriorates with time delay. Once an opportunity is missed, one may not be able to do it again. The system has to be able to perform its duty under the time constraint, which means that the system has to gather the information and present it to the human user before it is too late.

Quality of the information. The information providers seldom provide the perfect information regarding the situation at hand. The information often contains some degree of uncertainty and inaccuracy. The system has to be able to assess and reason with the reliability of the information it can expect from a source.

Digs uses the notion of value of information, as defined in decision theory, to guide the information gathering process and possibly to provide the user with the decision suggestion. The system's four major modules shown in Figure 1.

- **Knowledge base** – contains the information about the information sources.
- **User model –** stores a library of influence diagrams that represent the human decision models.
- **Executor** - perform actual information gathering actions based on the source that the system decides.
- **Interface** - provides communication with the human user.

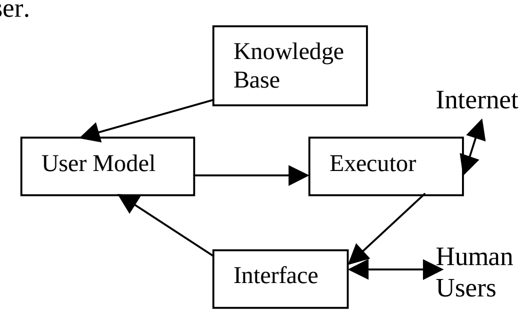

Figure 1. Decision Support Information Gathering System

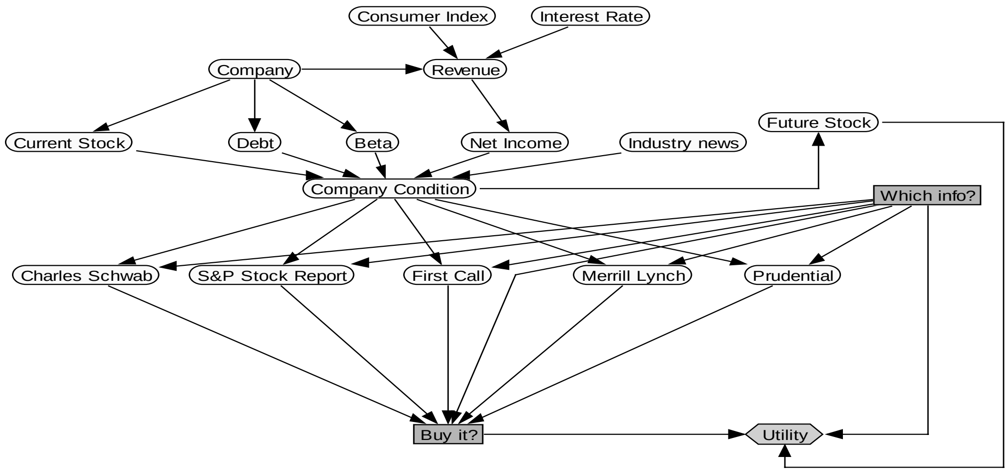

Figure 2. User model of the stockbroker

The remaining of this paper describes the components of the Digs system in more detail, presents our implementation and sample runs, and compares it to existing approaches.

Computing Value of Information

Digs uses the values of information to decide which information is worth retrieving. The system uses an influence diagram, described in Section 3, to calculate the value of imperfect information for each information source in the diagram.

Suppose that the current information available is C, which can be represented as belief regarding the value of the different random variables (chance nodes) in the influence diagram. Let $Result_i(A)$ be the i-th possible outcome of the user's action A. The current best action, $\boldsymbol{a}$, is one with the highest expected utility computed with information C:

$$EU(\boldsymbol{a}|C) = \max_{A} \sum_{i} U(Result_i(A)) P(Result_i(A)|C, Do(A))$$

Now consider the situation in which the user can get additional information that will provides some evidence, X_n. The value of the best action $\boldsymbol{a}_{X_n}$ after obtaining the evidence X_n is:

$$EU(\boldsymbol{a}_{X_n}|C, X_n) = \max_{A} \sum_{i} U(Result_i(A)) P(Result_i(A)|C, X_n, Do(A)))$$

But X_n is a random variable whose value is currently unknown, so we must average over all possible value X_n^k that we might discover for X_n , using system's current belief about its value. The Value of Information (VI) on node n is:

$$VI(X_n) = (\sum_{k} P(X_n = X_n^k|C) EU(\boldsymbol{a}_{X_n}^k|C, X_n = X_n^k)) - EU(\boldsymbol{a}|C)$$

The VI is the value of the perfect information, but interestingly, it can be made to express the value originated from unreliable information sources. In order to account for imperfections of information sources we must represent them as separate nodes in the network, causally connected to the node. The strength of this connection is the representation of the faithfulness of the information source and correlates the actual value of the node to the values reported by the information source. We have included this effect in our implementations, as depicted in Section 4, in the domain of financial decision-making.

Another effect that we must consider is, as we mentioned, that querying some information sources may be costly. For example, to consult the Charles Schwab's SchwabNOW Online Investments one is charged $6 per report. We include these costs by linking the node specifying which information source to consult directly with the utility node of the influence diagram.

The information sources ranked according to the information value of each imperfect source. The top ranking one(s) is used to direct the information retrieval agents that reside in the executor module in our system to perform the information retrieval task.

System Modules

User Model

The Digs system uses the influence diagram as the representation for the human user's decision model. We built the influence diagrams for the decision-making of a stockbroker by consulting human expert of the field. The model can be stored in a library and reused whenever the similar situation occurs. The model reflects the decision criteria and the attitude toward risk of the stockbroker as in Figure 2. The model includes the information sources that could be valuable to the user and the external variables of the domain. By using the information value criteria, the system will return the information from the source(s) to the user to assist the broker to make the investment decisions.

In this version, the DSIGS will uses this model to choose among the information sources (e.g., Charles Schwab, etc.). It will also return the retrieved information (the recommendation from the financial experts on the web) to the user in order to assist the user in making the investment decision for a certain stock. The decision criteria of the system are discussed in the previous section.

Knowledge Base, Executor and Interface

The knowledge base contains the information about the alternative information sources not directly included in the influence diagram. For example, this includes data on the sources turnaround time, and their availability. It also lists the type of information the sources provided and the information needed to interact with the sources.

The executor module contains the retrieval agents that are used by Digs system to get the information from the sources. These agents are responsible for generating the visual reports from their information gathering results. The executor module then sends the report generated from the retrieval agents to the interface module.

The interface module handles the interaction between the human user and the system. It also allows the human user to enter the information to the user module, such as the currently available information. Further, the module displays the information that the executor module gathers and the decision suggestion from the system.

Implementation

Using the belief network library NeticaAPI provided by the Norsys Inc. and the internet building platform LiveAgent Pro from AgentSoft, we constructed two experimental prototype systems for the defense and financial domains. The Digs will return the best information source (or none) to retrieve the additional information. Based on the additional information, the system suggests to the user an appropriate action to take.

We tested the Digs system using the user models described in the Section 3. Below is a run from the financial domain prototype.

Stockbroker Scenario

A stockbroker is to decide which stock in a certain industry group he or she is going to invest in. He or she wants to gather valuable (cost sensitive) and useful information before making such decision.

The Digs model for the stock investor will be responsible for giving the best information source(s) to retrieve the information based on the evidence(s) inputs to the model (e.g., the interest rate, the specific company, etc.), and the decision suggestion on whether or not to invest the company's stock based on the additional information.

In this run, the investor is looking at a company, A, with the consumer index good and the interest rate low. Given this information, Digs calculated the information values for each of the additional source (see Table 1), and suggested obtaining the SP (S&P information source) information. It also suggested courses of action to take based on the information retrieved from the SP information source (see upper left window in figure 3). The Digs then triggered the executor module to obtain the information from the suggested source (see Figure 3).

INFORMATION SOURCES	VALUE OF IMPERFECT INFORMATION
CSW	7.855481
SP	9.951006
FC	7.239150
Zacks	4.010879
Argus	3.861681

Table 1. The Imperfect Information Value for each source

Related Work

The problem of data request has been formally treated in decision theory (Howard 1966) where expected utilities of the possible actions are the selecting criteria for the request decisions. Not until recently, this problem has become the focus of the researchers because of the rapid growth in the on-line information on the Internet such as digital libraries, information brokers, news, etc. Although much work has been done on both information gathering and decision support (Etzioni 1996) (Jensen and Liang 1994), little work has focused on developing a system that combines both of them.

Independently of our work, Grass and Zilberstein (Grass and Zilberstein 1997) have developed a framework that combines the information gathering and decision making processes together, which is similar to our work. In their framework, the value of information is based on the perfect information, while our system is based on the value of imperfect information, which is more realistic when applying real world problems. Also, their system does not provide the decision suggestions to the human.

Compared to other available commercial search engines such as Lycos, Infoseek, Webcrawer, etc., our Digs system provides the human with the most relevant information source based on the value calculations. This means that Digs' results are more accurate since they consider the way in which the acquired information will be used, as opposed to going by a keyword matching. This gain in accuracy, however, is achieved at the cost of increased sophistication and the computational burden the system requires.

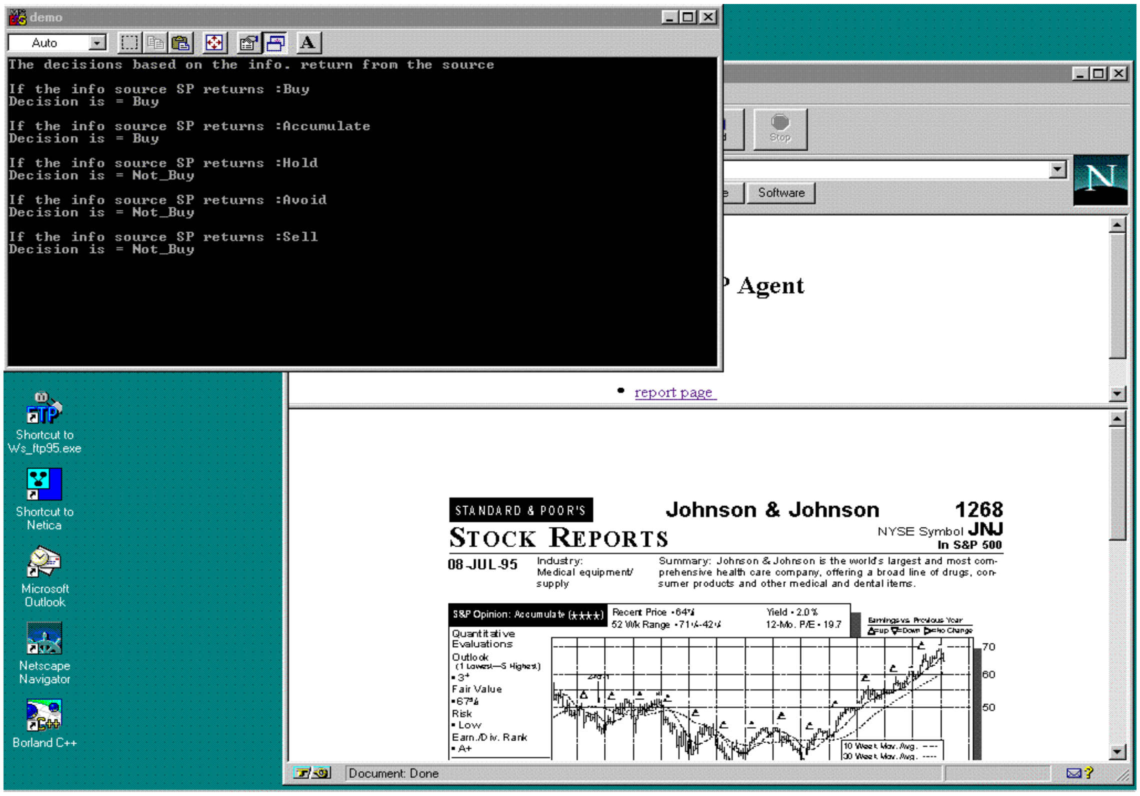

Figure 3. Under this condition: The stock broker wants to know about whether to invest in a company A when the Consumer Index is Good and Interest Rate is Low. The Information agent suggests the SP (S&P) information source to retrieve the information. In addition, the information agent triggers the internet agent, retrieves the information from the internet, and returns it to the user.

References

Etzioni, O. 1996. Moving Up the Information Food Chain: Deploying Softbots on the World Wide Web. In Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference,1322-1326. Menlo Park, Calif.: AAAI Press/The MIT Press.

Howard, R. A. 1966. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*,22-26.

Howard, R. A. 1990. From Influence to Relevance to Knowledge. *Influence Diagrams, Belief Nets and Decision Analysis*,3-23.

Horvitz, E., and Barry M. 1995. Display of Information for Time-Critical Decision Making. In Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence. San Mateo, Calif.: Morgan Kaufmann Publishers, Inc.

Gmytrasiewicz P. J., and Durfee E. H. 1993. Elements of a Utilitarian Theory of Knowledge and Action. In Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, 396-402. Menlo Park, Calif.: International Join Conferences on Artificial Intelligence, Inc.

Grass J., and Zilberstein S. 1997. Value-Driven Information Gathering. *AAAI 97 Building Resource-Bounded Reasoning Systems Workshop.*

Jensen F. V. 1996. An Introduction to Bayesian Networks. New York: Springer-Verlag.

Jensen F. V., and Liang J. 1992. drHugin A system for hypothesis driven myopic data request, Technical Report R-92-2021, Department of Mathematics and Computer Science, Aalborg University.

Jensen F. V., Liang J. 1994. drHugin A system for value of information in Bayesian networks. In Proceedings of the 1994 Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 178-183.

Matheson J. E. 1990. Using Influence Diagrams to Value Information and Control. *Influence Diagrams, Belief Nets and Decision Analysis*,25-48.

Pearl J. 1997. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference Revised Second Printing. San Francisco, Calif.: Morgan Kaufmann Publishers, Inc.

Russell S. J., and Norvig P. 1995. Artificial Intelligence, A Modern Approach. Englewood Cliffs, NJ.: Prentice Hall, Inc.

Zilberstein S., and Lesser V.. Intelligent information gathering using decision models. Technical Report 96-35, Computer Science Department, University of Massachusetts.