

Comparison of Different Information Content Models by Using Two Strategies: Development of the Best Information Algorithm for Iliad*

Di Guo¹, Michael J. Lincoln^{4, 1, 2}, Peter J. Haug¹,
Charles W. Turner^{3, 1, 4}, Homer R. Warner¹

¹Department of Medical Informatics, ²Department of Internal Medicine,

³Department of Psychology

University of Utah, Salt Lake City, Utah

⁴Salt Lake City Veteran's Administration Medical Center
Salt Lake City, Utah

*Supported in part by NLM Grants 1R01-LM-04604 and 1R01-LM-052020

ABSTRACT

Iliad is a diagnostic expert system for internal medicine. Iliad's "best information" mode is used to determine the most cost-effective findings to pursue next at any stage of a work-up. The "best information" algorithm combines an information content calculation together with a cost factor. The calculations then provide a rank-ordering of the alternative patient findings according to cost-effectiveness. The authors evaluated five information content models under two different strategies. The first, the single-frame strategy, considers findings only within the context of each individual disease frame. The second, the across-frame strategy, considers the information that a single finding could provide across several diseases. The study found that (1) a version of Shannon's information model performed the best under both strategies --- this finding confirms the result of a previous independent study, (2) the across-frame strategy was preferred over the single-frame strategy.

INTRODUCTION

Iliad is a personal computer-based expert system which provides decision support and may be used as a teaching tool for medical students and practitioners. Iliad can run under both Macintosh systems and MS DOS-Windows. The program requires a 68030 Macintosh or 80386 SX DOS (or higher) processor with 2 megabytes of RAM memory. It currently recognizes over 6300 disease manifestations and covers 1350 diseases and intermediate diagnoses from internal medicine.

The Best Information Mode in Iliad

Iliad is based on a model of diagnosis that stresses the assignment of probabilities to patho-physiologic states. However, Iliad not only functions as a diagnostic engine, but also has a number of features that support other aspects of medical reasoning. For instance, Iliad can assist clinicians in choosing which clinical data to collect next. This function is called Iliad's best information

mode. This mode is primarily designed for two purposes: teaching medical students to pursue a cost-effective medical work-up, and evaluating students' performance when they try to solve simulated patient cases in Iliad [9, 10]. The performance scores for inquiry skills are measured by comparing the students' questions to the best question calculated by the best information mode.

Enhancing the performance of Iliad's best information mode has been a continuous effort during the development of Iliad. The objective is to ensure that students who use Iliad will receive accurate training. Iliad evaluates alternative work-up strategies by employing a "best information" algorithm. The algorithm evaluates the information content expected per dollar for uncollected data and selects the finding with the maximum information at the least cost. The cost for each procedure is stored in Iliad's knowledge base using the actual dollar charge at the University of Utah Medical Center. Other medical centers may modify the charges as needed. History findings are set to an arbitrarily low value of \$1 and physical exam items are \$2.

A user can select a subset of the diagnostic hypotheses in which to pursue the next most cost-effective work-up strategy. If no selection is made, Iliad automatically selects a work-up suggestion for the most likely diseases. The user can ask Iliad to restrict the best information analysis to specific categories: history findings, physical exam findings, or lab test procedures. However, the default is to produce the best information over all categories. The algorithm does not take into account other factors such as risk to the patient for certain lab procedures, or the time delay cost of waiting for results. We realize that these factors may play important roles in the selection of the best items to pursue. However, our first step has been to investigate different information content models combined with the direct cost of medical findings and implement the best information content model in a way preferred by human clinicians.

Five Information Content Models

Iliad's approach to the process of pursuing a group of diagnostic hypotheses is based on simple assumptions. The amount of diagnostic uncertainty in a case can be reduced by obtaining additional patient findings. The information provided by the patient findings can be measured quantitatively as the change in the level of uncertainty associated with a particular disease[2]. Several mathematical models are available to quantify information[4]. The model derived from standard information theory is Shannon's equation[1, 4, 6, 7, 8]. In the context of medicine, the Shannon model represents the average amount of uncertainty as to whether a patient does or does not have a disease. The basic mathematical equation is:

$$H(D) = -P(D)\log_2 P(D) - P(D-)\log_2 P(D-) \quad (1)$$

Here, $H(D)$ is the Shannon's uncertainty or entropy. The logarithm to the base two of the probability of a disease gives the $H(D)$ measured in unit of "bits". $P(D)$ is the probability that a patient has a disease D , $P(D-)$ is the probability a patient does not have the disease D . If a medical finding's result is known as F , the information content contributed by the result F , $I(D|F)$ can be calculated as the difference in entropy before and after the finding result F is known.

$$I(D|F) = \text{abs}(H(D) - H(D|F)) \quad (2)$$

The use of eqn. 2 requires an appreciation of uncertainty as the function of the prior probability of disease. The standard Shannon model fails to capture reasonable intuitions about the quantity of information provided by a diagnostic finding[1]. For example, when the prior and the posterior probabilities are complementary (e.g. the prior is 0.1, the posterior is 0.9), the finding provides no change in uncertainty, and thus no information has been conveyed. To overcome the problem, we used the modified Shannon information content model[4, 6] whenever the disease probability passes through 50%.

$$I(D|F) = (H_{\max} - H(D)) + (H_{\max} - H(D|F)) \quad (3)$$

H_{\max} is the maximum value of Shannon's uncertainty, which is 1 bit. This value is obtained when there is a 50% chance of the disease being present.

The information content models tested are either based on different ways of expressing uncertainty or derived as "quasi-utilities". The five models were discussed in detail in our previous work[4]. A summary is shown in Table I.

One way of categorizing these five information content models is by whether the model depends on the prior probability of the disease. The first category, where prior probability has an effect, consists of three models: Shannon, $\log P_2 - \log P_1$ and $P_2 - P_1$. All these models measure a medical finding's information based not only on the effectiveness of the finding in changing the probability of disease but also on the patient status (prior probability) prior to acquiring the finding. The second

Table I. Summary of Five Information Content Models

Model Name	Uncertainty representation	Information content of a given finding F
Shannon	$H(D) = -P(D)\log_2 P(D) - P(D-)\log_2 P(D-)$	If $H(D F)$ does not pass through a maximum in moving from the prior state to the posterior state $\text{abs}(H(D) - H(D F))$, otherwise, $(H_{\max} - H(D)) + (H_{\max} - H(D F))$
$\log P_2 - \log P_1$	$-\log_2 P(D)$	$\text{abs}(\log_2 P(D F) - \log_2 P(D))$
$P_2 - P_1$	$P(D)$	$\text{abs}(P(D F) - P(D))$
$\log LR$	N/A	$\text{abs}(\log_2(\text{sen}/1\text{-spec}))$ for a positive finding, $\text{abs}(\log_2(1\text{-sen}/\text{spec}))$ for a negative finding
LR(Current model in Iliad 4.0)	N/A	$\text{sen}/(1\text{-spec})$ for a positive finding, $(1\text{-sen})/\text{spec}$ for a negative finding.

LR = likelihood ratio. N/A = not applicable. $P(D)$ = the prior probability of disease before the finding result F (either positive or negative). $P(D|F)$ = the posterior probability of disease after the finding result F . Sen = true positive rate, spec = true negative rate.

category, where the value of information is independent of prior probability, consists of models of $\log LR$ and LR (LR = likelihood ratio). These two models ignore the prior probability of the disease and depend only on the sensitivity and specificity of a medical finding. Whether one uses the $\log LR$ or the LR model should not make any difference in terms of ranking information value per se. However, the ranking in Iliad's best information algorithm is based on the information per dollar cost. Therefore, the $\log LR$ model is more sensitive to dollar cost because of the logarithm function. In addition, we note that $\log LR$ is additive and LR is not.

Iliad 4.0 currently utilizes the LR model. This model has the advantage that the parameters required are easily accessible from Iliad's knowledge base. In addition, a minimum of calculation is required. In the case of models like Shannon, $\log P_2 - \log P_1$ and $P_2 - P_1$, Iliad has to calculate the potential posterior probability of each hypothesis under consideration, given each possible medical finding, in order to find the best item. In the past, the computational burden of this approach discouraged use of more complex algorithms. However, with the rapid development of more powerful hardware configurations on both PC and Macintosh machines, we now wish to investigate more computationally intensive algorithms which may provide even better performance.

The Application of Different Information Content Models: Two Strategies

We implemented information models using two different strategies. The first strategy, the "single-frame" strategy, considers findings only within the context of individual disease frames. The second strategy, the "across-frame" strategy, considers the information that a single finding could provide across several diseases. Suppose that we have two hypotheses under consideration, pulmonary embolus and atypical pneumonia. Also suppose there are three unanswered questions exist for each disease frame, as shown in Table II,

Table II. A Scenario of Considering Two Hypotheses

Atypical Pneumonia	Pulmonary Embolus
present history: cough with purulent sputum	present history: cough with gross hemoptysis
vital signs: respiratory rate	vital signs: respiratory rate
chest x-ray shows alveolar infiltrate	chest x-ray shows alveolar infiltrate

The approach for this scenario by the single-frame strategy is as follows:

1. Calculate information content per dollar of each finding within each disease frame.
2. Rank the cost-effectiveness of each finding in the Atypical Pneumonia frame and the Pulmonary Embolus frame separately.
3. Select the finding which receives the highest of the six scores.

The approach by the across-frame strategy is as follows:

1. Calculate information content per dollar of each finding for each disease frame.
2. Sum across the information per dollar for the "common" findings, respiratory rate and the chest X-ray, across the two hypotheses.
3. Select the finding which receives the highest of four scores.

Single diagnostic procedures, such as chest X-ray examinations or batteries of laboratory tests (e.g. Chem-20), can produce multiple findings. We assume that these findings should be evaluated together to give a total value for the information from the procedure. Thus, in the implementation of the across-frame strategy, we sum across all information from the lab test or other procedures. Although history and physical exam findings are usually collected systematically in real life, we treated each history and physical exam finding individually in our current implementation of the across-frame strategy.

METHOD

To begin this study, we implemented the five information content models and two strategies in an experimental version of Iliad, so that any combination of an information content model and a strategy could be used

to pursue the most cost-effective work-up. Our objective was to compare the performance of the five information content models and two strategies to the judgments provided by expert clinicians.

Subjects Six academic internists certified by American Board of Internal Medicine served as the subject judges in the experiment. The physicians were all experienced in the use of computerized expert systems for medical decision making.

Procedure Six pulmonary cases were selected from real patient cases at the University of Utah Medical Center. Each case was divided into three stages. The first two stages denoted stages in the work-up when history and physical exam findings were acquired. The third stage denoted a later stage in the work-up when major competing diagnostic hypotheses were considered and finally accepted or eliminated through laboratory tests or other procedures. Thus there were eighteen medical decision points, or vignettes represented. Each information content model was applied under two strategies. Therefore, ten best work-up suggestions ($5 \text{ algorithms} \times 2 \text{ strategies}$) were generated for each vignette. Each expert was provided with a copy of each vignette containing a subset of patient findings and including the hypotheses Iliad considered, the ten sets of work-up suggestions, and a simple rating form. Based on Iliad's suggested work-up items, the experts were instructed to choose (1) the best strategy for each information content model and, (2) the best information content model for the single-frame and the across-frame strategies. "Ties" were allowed only when several information content models or two strategies produced the same work-up suggestion. Whenever the strategy or the information model was chosen as the best, the score was assigned to 1; otherwise the score was 0. All the experts completed their entire set of evaluation forms.

Experimental Design The independent variables are: Case (six cases), Stage (three stages in each case), Information Content Model (five models), and Strategy (two strategies). All raters had substantially the same level of training and experience. Hence, no effort was made to classify the experts by level or type of expertise.

The experiment was a $6 \times 3 \times 5 \times 2$ (Case \times Stage \times Information Content Model \times Strategy) factorial design. All independent variables are within subjects factors. There were three dependent variables. The first dependent variable was the frequency of being chosen as the best information content model under the single-frame strategy. The second dependent variable was the frequency of being chosen as the best information content model under the across-frame strategy. The third dependent variable was the frequency of being chosen as the best strategy overall. Each dependent variable represents the proportion of experts who chose the result of that outcome as the best.

RESULTS

Best Information Content Model Under the Single-frame Strategy

The judges' ratings of the best Information Content Model under the single-frame strategy were analyzed by using a $6 \times 5 \times 3$ (Case \times Algorithm \times Stage) factorial analysis of variance. Comparisons among cell means were based upon a Bonferonni adjusted confidence interval[5]. We divided the significance level (α) by the number of comparisons to be performed (k). In this study, four comparisons were made among the means so that the adjusted significance level was 0.0125 (0.05/4). The results indicated that the main effect for Information Content Models was statistically significant, $F(4,360) = 10.72$, $p < 0.0001$. The interaction between Stage and Information Content Model was also statistically significant, $F(8,360) = 4.83$, $p < 0.0001$. We used average scores (frequency of being chosen as the best) of each information content model at stage 1 and stage 2 to represent the effectiveness of the model in suggesting history and physical exam findings, and selected the score at stage 3 to reflect the effectiveness of the model in suggesting lab test procedures. The comparisons among the means for the algorithms indicated that the Shannon model was significantly better than the other models ($\alpha = 0.0125$) in terms of suggesting history and physical exam findings. However, the results revealed no significant differences between the Shannon's model and the current Iliad model in terms of suggesting lab test procedures during the late stage of work-up. No other models performed better than the current Iliad model at the late stage. The results also indicated that the Shannon's model was the best in terms of overall scores across all stages ($\alpha = 0.01$). The overall performance of five information models is shown in Figure 1.

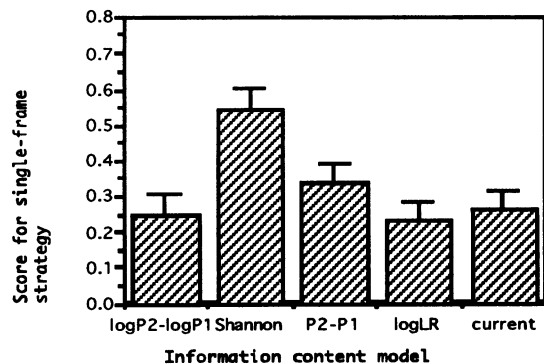


Figure 1. Overall (all stages) frequency each information content model being chosen as the best by experts under the single-frame strategy.

Best Information Content Model Under the Across-frame Strategy

The results showed that the main effect for Information Content Models was statistically significant,

$F(4,360) = 3.20$, $p < 0.015$. The interaction between Stage and Information Content Model was also statistically significant, $F(8,360) = 2.30$, $p < 0.02$. The Shannon's model and the P2-P1 model were not significantly different in terms of suggesting history and physical exam questions, but they were all significantly better than the current Iliad model ($\alpha = 0.0125$). During the late stage of work-up, no models performed better than the current Iliad model in terms of suggesting lab test procedures. The Shannon's model and the P2-P1 model were the best overall among the five models across all stages ($\alpha = 0.01$). The overall performance of the five information models under the across-frame strategy is shown in Figure 2.

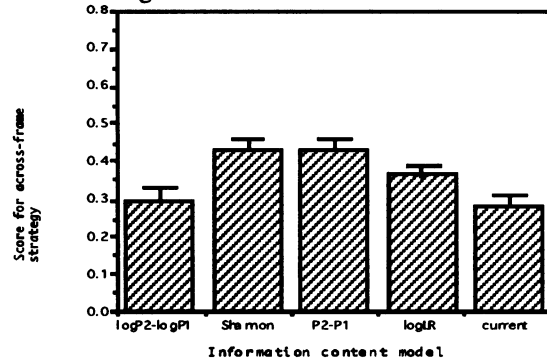


Figure 2. Overall (all stages) frequency each information content model being chosen as the best by experts under the across-frame strategy.

Best Strategy

We calculated the frequency of each strategy being chosen as the best based on the grand average scores for all the information models. Each strategy, single-frame and across-frame, was evaluated five times in each of the eighteen vignettes. The five times represented implementation of the five information models under each strategy for each vignette. The best strategy scores were analyzed by ANOVA using repeated measures. The results indicated that experts preferred the across-frame strategy to the single-frame strategy, as shown in Figure 3.

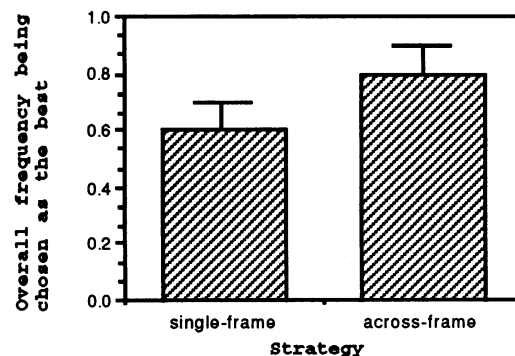


Figure 3. Overall (all information content models) frequency of each strategy being chosen as the best

DISCUSSION

The modified Shannon information model was the best model overall, regardless of strategy. The Shannon's model was significantly better than the current Iliad model during the initial encounter of the patient when history and physical exam findings were the major items to acquire. During the late stage of work-up when the patient's major history and physical exam features were known, the current model was just as good as the Shannon's model. This finding may reflect the fact that fewer choices were available at this period. Iliad's best information mode could be improved by employing the Shannon's model, especially during the initial phase of a patient case. These results also confirmed our previous findings which suggested that Shannon's model was preferable to other models[4]. Given each possible medical finding, Shannon's model requires the prior and the potential posterior probability of each hypothesis under consideration. As microprocessors improve, it may prove practical to adopt Shannon's model in Iliad.

Physicians typically generate a differential diagnosis early in the work-up of a patient case. They then pursue findings which allow them to separate these potential diagnostic competitors[3]. We have attempted to model this process using the single-frame and the across-frame strategies described above. The single-frame strategy, which is present in the current version of Iliad, evaluates the relative cost-effectiveness that each diagnostic finding has in relation to each hypothesis on the differential. This strategy allows us to rank-order each possible diagnostic finding and select the best one. However, this strategy treats each finding and disease link independently. In some cases, obtaining one finding may provide positive information for one hypothesis and negative information for another. For instance, a chest X-ray may be ordered to work-up a patient with shortness of breath when the physician is considering pneumonia versus pneumothorax. If the chest X-ray shows a pneumothorax, and not an infiltrate, information accrues (positive and negative) for both diagnostic hypotheses. The single-frame strategy does not reflect the combined information available in a group of diseases to which a particular finding may be relevant. The result is an underestimation of the total information value of findings that contribute to multiple diagnoses. This may explain why Iliad sometimes delays obtaining tests such as chest X-rays even when experts feel they are indicated. It appears that the across-frame strategy may improve this performance. If this result can be replicated, it may prove appropriate to adopt this strategy in Iliad. The relative performance of different information models was closer to each other when we used the across-frame strategy as compared to the single-frame strategy. This finding indicates that the performance of best information mode in Iliad depends on not only the information content model, but also the strategy with which the model is implemented.

During our experiment, we observed that experts sometimes selected the across-frame strategy as a better

one because it simply suggested more findings than the single-frame strategy, whereas the additional findings might be collected later by the single-frame strategy. This could have produced some bias in our results. We are designing a future study to control this potential source of bias.

Further study is needed to analyze factors such as the risks of certain lab procedures and costs associated with time delays while waiting for results. These potential additions to the best information algorithm should improve the ability of Iliad to simulate the multifaceted environment in which real data collection decisions are made.

References

- [1] Asch, D.A., Patton, J.P., Hershey, J.C., Knowing for the Sake of Knowing: The Value of Prognostic Information. *Medical Decision Making*, 47-57, Vol. 10, No 1, Jan-Mar 1990.
- [2] Bharath, R., Information Theory, *BYTE*, 291-298, December 1987.
- [3] Elstein, A.S., Shulman, L.S., and Sprafka, S.A., Medical problem solving--A ten year retrospective. *Eval. Health Prof.* 13:5-36, 1990.
- [4] Guo, D., Lincoln, M.J., Haug, P.J., Turner, C.W., Warner, H.R., Exploring a New Best Information Algorithm for Iliad, *Proceedings of the Symposium on Computer Applications in Medical Care*, 15, Washington, D.C.:MCGRAW- HILL, INC, 624-628, 1991.
- [5] Hays, William L., *Statistics*, Fourth Edition, Holt, Rinehart and Winston, Inc., 1988.
- [6] Pitkeathly, D.A., Evans, A.L., Hames, W.B., The Use of Information Theory in Evaluating the Contribution of Radiological and Laboratory Investigations to Diagnosis and Management. *Clinical Radiology*, 643-647, Vol. 30, 1979.
- [7] Shannon, C.E., Weaver, W., *The Mathematical Theory of Communication* Urbana IL: Univ. of Illinois Press, Chicago (1949).
- [8] Lee, C.Y., Carmony, L., Evens, M., Naeymi-Rad, F., and Trace, D., A Test Selection Module for MEDAS. *Proceedings of the Symposium on Computer Applications in Medical Care*, 15, Washington, D.C.:MCGRAW- HILL, INC, 706-710, 1991.
- [9] Lincoln, M.J., Turner, C.W., Haug, P.J., Warner, H.R., Williamson, J.W., et al., Iliad Training Enhances Medical Students' Diagnostic Skills. *Journal of Medical Systems*, 93-110, Vol. 15, No. 1, 1991.
- [10] Warner, H.R., Haug, P., Bouhaddou, O., Lincoln, M.J., Warner, H.R. Jr., et al., Iliad As An Expert Consultant to Teach Differential Diagnosis. *Proceedings of the Symposium on Computer Applications in Medical Care*, 12, Washington, D.C.: IEEE, Computer Society Press, 371-376, 1988.