Passing Partial Information among Bayesean and Boolean Frames

Dean K. Sorenson, Ph.D., Robert M. Cundick, Ph.D.
Chinli Fan, M.S., and Homer R. Warner, M.D., Ph.D.

Department of Medical Informatics
University of Utah
Salt Lake City, Utah 84102

ABSTRACT

A method is described for passing information between frames in an expert system. These frames may be either Bayesean or Boolean. The system can accept the status of any frame as true, false, or any degree of uncertainty in between. An algorithm is described for determining information contributed by a finding which is linked to a given frame through one or more intermediate frames. This algorithm generates a dynamic weighted listing of diagoses (or decisions) and a means for determining the "best information" to acquire next at any time during a diagnostic workup. These features play a major role in making the Iliad expert system attractive to medical students as a teaching tool.

INTRODUCTION

A number of medical expert systems are currently being used for teaching diagnostic skills [1-6]. Iliad is a frame-based expert system implemented on a Macintosh microcomputer and designed primarily as a tool for teaching and consultation [1,2]. One of the most unique features of Iliad is its ability to pass partial information between decision frames. Iliad uses two types of frames to represent knowledge, probabilistic and deterministic. These frames are linked in a tangled hierarchy and must communicate their present state to one another in order to determine:

1. how close each frame (hypothesis) is to being true or false based on those features of the patient"s illness available at a given stage in the diagnostic workup
2. how useful the acquisition of any particular new finding might be at each stage of the diagnostic process.

How "close" is a frame to being true?

It is desirable to allow the knowledge engineer to choose whichever type of frame representation seems most appropriate to model a given disease or decision. We have consequently, developed a means for expressing the state (degree of certainty or closeness to being true or false) of a Bayesean or a Boolean frame in common units.

Boolean case

In the case of a Boolean frame we define the term "Close" as a measure of how near the Boolean logic is to being satisfied, either positively or negatively. The state of each frame will be expressed by two terms, "close_t" and "close_f", for true and false respectively. Each is a number between 0 and 1.

For example, if the logic in a frame was stated as "true if a and b", and if item a was unknown and item b was true, then close_t for this frame is 0.5 and close_f is 0.0. This is because we have 50% of the information necessary for the frame to become true. However, item a may be another frame and its value may be anything from 0.0 to 1.0.

As a second example, if the logic in a frame was stated as "true if a or b", and if finding a were true and finding b was unknown, then close_t = 1. If either close_t or close_f = 1, then the other = 0.

In a more complex example, if the logic in a frame was stated as "true if 2 of (a,b,c,d) and 3 of (e,f,g,h,i)", and if c, g, and h were true and e were false, and the other items were unknown, then close_t = (.5 + .66) / 2 = .58. This number represents the value of the frame since one of the two required items among (a,b,c,d) is present and, thus, adds a weight of .5 or half of what is needed to make this component true. Since g and h are present, the terms in the second group (e,f,g,h,i), are 2/3 of what is needed to make this component true. Since the two components must both be true, we divide by 2. The negative logic (derived from the "true" statement) for this frame would be "false if 3 of (a,b,c,d) are false or 3 of (e,f,g,h,i) are false", close_f = .33. The inference engine in Iliad derives the negative logic and performs these logical operations as each finding is entered during the patient workup.

Bayesean case

To represent the state of a Bayesean frame using a commensurate measure (to the Boolean case) of closeness to being true or false, we have defined an algorithm for expressing this from the difference between the probability (P) of the frame being true given the currently available findings and apriori probability (AP) of the frame. The rule is:

If P > AP
close_t = (P - AP) / (1 - AP)
close_f = 0, while
if P < AP
close_t = 0
close_f = (AP-P) / AP

[Equation 1]

In almost all clinical diagnostic cases, AP is close to 0 so 1-AP is approximately 1. Consequently, close_t is approximately equal to P, i.e., the posterior probability. Note also that, when P=0, close_f = 1.

0195-4210/89/0000/0050$01.00 © 1989 SCAMC, Inc.

Using Partial Information for Decision-making

"Working up a patient" or making a diagnosis is a problem-solving scenario whose success depends on the ability of the clinician to form appropriate hypotheses from partial information as to the nature of the patient"s problem at each stage of the process. Iliad must capture the essence of this algorithm if it is to serve as a source of consultation for the student or physician or as a standard against which to judge human decision-making.

In a Bayesean frame

For example, what effect does the finding of "rales" have on the likelihood that a patient has pneumonia? In the Iliad knowledge base; pneumonia is represented as a Bayesean frame.

Pneumonia

a priori = 0.025

	$P_{f/d}$	$P_{f/not\ d}$
a. @Lung consolidation	.99	.07
b. @Signs of systemic infection	.90	.20
c. @Hypoxemia	.40	.10
d. @Pleuritic chest pain	.25	.02
e. @Pleural effusion	.05	.01
f. @Acute productive cough	.95	.06
g. History of shortness of breath with exertion of recent onset	.40	.10
or		
History of shortness of breath at rest of recent onset	.40	.10

Figure 1. Pneumonia: an example of a Bayesean frame. The "@" signs at the beginning of an item indicates another frame or cluster of findings.

One of the findings in the "pneumonia" frame is "lung consolidation" which is a Boolean frame or cluster.

Lung consolidation

a. PE shows rales
b. PE shows bronchial breath sounds
c. PE shows egophony
d. PE shows increased vocal fremitus
e. PE shows dullness to percussion
f. PE shows whispered pectoriloquy
g. Chest x-ray shows alveolar infiltrate

True of g or ((a and b) and (c or d or e or f))

Figure 2. Lung consolidation: an example of a Boolean frame.

"Rales" is a finding in this "lung consolidation" cluster. The presence of "rales" by itself makes "lung consolidation" closer to being true (i.e., close_t=.33); this is because the frame becomes true if a and b and one other finding are present, i.e., finding a represents 1/3 of the necessary information.

When this partial or incomplete information about "lung consolidation" is passed to the Bayesean frame "pneumonia", the probability of "pneumonia" increases from 0.092 to 0.197. In making this calculation, Iliad uses the following relationship [7]:

$$P_{d/f} = \frac{P_d (P_{f/d})^a (1-P_{f/d})^b}{(P_d (P_{f/d})^a (1-P_{f/d})^b) + (1-P_d)(P_{f/not\ d})^a (1-P_{f/not\ d})^b},$$

[Equation 2]

where P_d (0.092) is the prior probability of "pneumonia" before considering "rales", $P_{d/f}$ (.197) is the posterior probability of "pneumonia" given the partial information about the finding "lung consolidation", $P_{f/d}$ (.99) is the sensitivity of the finding, and $P_{f/not\ d}$ (.07) is the complement of the finding's specificity (i.e., the false positive rate). The exponential terms a and b (not to be confused with findings a and b in a frame) are the close_t (.33) and close_f (.00) values for the "lung consolidation" frame being used as the finding. Note that in the limiting case where a=1 and b=0 (or vice versa), the equation becomes the familiar form of Bayes.

In a Boolean frame

Iliad allows a Boolean frame to deal with findings which are not just 0 or 1 (yes or no), as may be required in the case where a finding is another frame and its closeness to being true or false is somewhere between 0 and 1. To accomplish this, Iliad converts the logical expression to reverse Polish notation (RPN) as shown in the following examples:

An example of an "and" frame

Frame logic statement: "true if a and b"
RPN representation ---> [a],[b],+,2,/
where [a] is the "close_t" value of finding a and the result is the "close_t" value for the frame

The average of a and b is used to represent how close this frame is to being true since the values of a and b each range from 0 to 1 (i.e., each item, whether it be a single finding or another frame, represents half of the information) and the value of the frame is 1 when it is true.

An example of an "or" frame

Frame logic statement: "true if a or b or c"
RPN representation ---> [a],[b],[c],3,max
where max returns the maximum of the preceding 3 items in the stack

The max of a, b, and c is used here since the closeness of the frame to being true is only as close as whichever item has the highest close_t value (i.e., each item can make the frame come true individually).

A complex example

Frame logic statement: "true if 2 of (a,b,c,d) and 3 of (e,f,g,h,i)"
RPN representation --->
[a],[b],[c],[d],2,4,getval,[e],[f],[g],[h],[i],3,5,getval, +,2,/
where "getval" averages the 2 highest values among the 4 items a,b,c and d

The items in the frame can be subdivided into two groups separated by a logical "and". The close_t of the first group will be the average of the highest two items of a, b,c and d. The close_t of the second group will be the average of the highest three items of e, f, g, h, and i. The close_t for this frame will then be the average of the two subdivided statements which each contribute half of the information for the frame to come true.

Example 3 is taken from the Boolean frame for AIDS, where many of the findings in the definition of AIDS is a disease entity (such as "Pneumocystis Carinii pneumonia" and are represented as a Bayesean frame.

AIDS

a. @Constitutional manifestations of HIV infections (ARC)
b. @Kaposi"s sarcoma
c. @Primary lymphoma of central nervous sytem
d. @Pneumocystis carinii pneumonia
e. @Unusually extensive mucocutaneous herpes simplex of > 5 weeks duration
f. @Cryptosporidium enbterocolitis > 4 weeks duration
g. Esophagitis due to Candidia albicans, CMV, or HSV
h. Progressive multifocal leukoencephalopathy
i. Atypical mycobacterium species
j. @Pneumonia, meningitis or encephalitis due to opportunistic organisms
k. @Disseminated infection
l. Currently taking immunosuppressive drugs
m. Currently on immunosuppressive radiation therapy
n. History of hematologic malignancy
o. History of organ transplant
p. History of immunodeficiency disorder
q. Seropositive for HIV antibody

True if (a and ((b or c or d or e or f or g or h or i or j or k) and (q or not (l or m or n or o or p))

Figure 3. Example of a complex Boolean frame which incorporates the official definition of AIDS and calls multiple Bayesean frames

Selecting the "best information" to acquire next

The ability to decide which item of information to acquire next is an important feature of Iliad"s inference engine. For Iliad to behave like an expert in diagnosing a patient"s problem and for it to provide the standard against which student performance will be judged, the criteria for this decision must reflect what we accept as optimal performance. The definition of "best" used by Iliad is that item of information which will most influence (increase or decrease the probability) of the most likely diagnosis for the least cost. This can be expressed as:

$$\text{score for each finding} = \text{Inf} * \text{close} / \text{cost}^x$$

where "Inf" is the information content of the item for a given frame (defined below) and x is an empirically-determined exponent.

Information content of an item in a Bayesean frame

To determine the information content (Inf) of an item in a Bayesean frame, first calculate the posterior probability (P_p) is calculated for each possible value (or range of values) for the finding using the current probability of the frame (P_c) as the apriori. The highest and lowest values of P_p are used to calculate positive (inf_t) and negative (Inf_f) as follows:

If $P_p > P_c$
$$\text{Inf\_t} = (P_p - P_c)/(1-P_c)$$

or if $P_p < P_c$
$$\text{Inf\_f} = (P_c - P_p)/P_c$$

If $P_p = P_c$, then Inf_t = Inf_f = 0

[Equation 3]

The information content (Inf) of the item is taken as the maximum of Inf_t and Inf_f.

Information content of an item in a Boolean frame

The information content of an item in a Boolean frame is determined by how much the value of the item contributes toward making the frame closer to being either true or false.
This is determined as follows:

Inf_t = (close_t after item - close_t before item) / (close_t before item)
or
Inf_f = (close_f after item - close_f before item) / (close_f before item)

[Equation 4]

The information content (Inf) of the item is taken as the maximum of Inf_t and Inf_f.

Hierarchical propagation of "information content"

Iliad acquires all information about a patient in the form of findings (history, physical examination, or other tests performed on the patient or his/her body products). Where possible, any deductions or interpretations of these direct observations are made by the knowledge built into the frames themselves. Thus, estimation of the information content of a finding may involve its contribution to an interpretation frame (cluster) which in turn makes a contribution to another frames whose status depends upon that interpretation.

For example, "rales" are a characteristic of "pulmonary venous congestion" which may be a manifestation of "left heart failure" which can be the result of "hypertensive heart disease". Iliad evaluates the information contribution of finding or not finding "rales" on auscultation of the chest toward ruling in or ruling out "hypertensive heart disease" by propagation of the Inf_t and Inf_f values for "rales" thru each of these frames as follows:

Hypertensive heart disease (HHD)
  Congestive heart failure (CHF)
    Left heart failure (LHF)
      Pulmonary venous congestion (PVC)
        rales

Inf_t (rales-->PVC) = .33
* Inf_t (PVC-->LHF) = .16
* Inf_t (LHF-->CHF) = .08

The Inf_t term to be used as the exponent ("a") in equation 2 to calculate the probability for "Hypertensive heart disease", given that the probability of the presence of "rales", is 0.08. Thus, the presence of rales increases the probability of HHD from 0.003 to 0.006.

DISCUSSION

A set of algorithms have been described for allowing information about the state of one frame (be it Bayesean or Boolean) to be passed to another frame, and have the receiving frame (be it Bayesean or Boolean) draw appropriate inferences, even though the sending frame may be neither completely true nor completely false. The inference engine in Iliad currently employs these algorithms. They provide a level of sophistication which mimics expert behavior not only in the sequencing of questioning and test ordering, but it also provides a mechanism for explanation of the systems reasoning. Iliad acceptance by medical students is in large part attributable to these algorithms.

REFERENCES

1. Warner, H.R., P. Haug, O. Bouhaddou, M. Lincoln, H. Warner, Jr., D. Sorenson, J.W. Williamson, and C. Fan. Iliad as an expert consultant to teach differential diagnosis. Proceed. SCAMC, 1988, pp371-376.

2. Warner, HR, Jr. Iliad -Moving Medical Decision-Making into New Frontiers. Internatl. Sympos. Med. Inform. Edu., U Victoria, May, 1989, 267-270.

3. RA Miller & FE Masarie, Jr. Use of the Quick Medical Reference (QMR) program as a tool for medical education. Internatl. Sympos. Med. Inform. Edu., U Victoria, May, 1989, p247.

4. AH Levy. Factors Affecting computer-mediated instruction in medical education. Internatl. Sympos. Med. Inform. Edu., U Victoria, May, 1989, p312.

5. S. Merigliano, R. Rossoni, A. Segalin. Use of computers in medical educations learning clinical diagnosis. Internatl. Sympos. Med. Inform. Edu., U Victoria, May, 1989, p244.

6. RS Wigton. The use of computer analysis of diagnostic reasoning to teach medical diagnosis. Internatl. Sympos. Med. Inform. Edu., U Victoria, May, 1989, p271.

7. Warner, HR. Computer-Assisted Medical Decision-Making. Academic Press, NY, 1979, pp13, 138-143.

This work was supported in part by a grant from the National Library of Medicine (5 R01 LM04604-03).