Ongoing Critiquing During Trauma Management*

Abigail Gertner
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104
(Internet: agertner@linc.cis.upenn.edu)

Introduction

A major challenge facing the developers of intelligent decision-support systems is how to present pertinent information to physicians in such a way that it will be effective in influencing their behavior. The successful integration of medical decision-support systems into clinical environments has been a widely reported problem ever since such systems began to appear. While recognizing these systems' potential for improving the quality of patient care and for controlling costs, physicians have tended to reject new technologies which they see as intrusive, time-consuming, or a challenge to their judgment or autonomy as clinical decision-makers [1]. At the same time, the information processing demands on physicians have been increasing dramatically. Utilizing available clinical data to make appropriate decisions about patient care can be challenging to physicians, who are susceptible to information overload which can lead to biases in data acquisition and processing [8].

To counteract the limitations of human information processing, computer-based decision-support systems have been developed to monitor clinical data as they become available and make physicians aware of pertinent events by means of *reminders* [4] or *alerts* [2]. Other systems follow the *critiquing* approach, providing a more extensive off-line discussion of the risks and benefits of alternative management plans [5, 9].

Critiquing Trauma Management

My research is concerned with the problem of providing effective decision support to physicians in time-critical situations, in which a great deal of information must be processed and responded to within an limited time frame. This work is being implemented within the framework of the TraumAID system at the University of Pennsylvania, whose purpose is to provide assistance during the initial definitive management phase of patients with acute trauma, specifically penetrating wounds to the chest and/or abdomen [10]. The TraumAID system is designed for use in the hospital trauma bay. An extensive graphical input interface has been designed to replace the Trauma Flow Sheet, the standard paper form on which a "scribe" nurse enters patient data as it is reported by members of the trauma team. Information entered by the scribe nurse includes bedside findings, vital signs, results of diagnostic tests, reported diagnoses, therapeutic actions that have been done, and orders for diagnostic and therapeutic procedures. For the purposes of output, a second monitor may be mounted in view of the physician in charge of the trauma team. Alternatively, output may be delivered using synthesized speech.

TraumAID's knowledge of plans and actions in its domain is represented in terms of *management goals*, *procedures*, and *actions*. Each goal is associated with a disjunctive list of procedures that may be used to address it. The procedures are listed in the order that they are preferred by the system in the absence of contraindications or interactions with other goals. Each procedure comprises a sequence of actions to be performed. Upon receiving new information about the patient, TraumAID derives a recommended management plan combining diagnostic and therapeutic actions, through a process of *reasoning* to derive a set of active goals, and then *planning* how best to achieve those goals, choosing the most appropriate combination of procedures given the current set of goals.

When the first version of TraumAID was experimentally introduced into an Emergency Center environment for a 12-month period, the physicians using the system expressed dissatisfaction with the fact that the system presented its management plan at every opportunity. It turned out that a great deal of the time, the system's recommendation was in agreement with the physicians' intended plans. Having the system present its entire recommended plan at any given point made it difficult for the physicians to identify just those items that would be helpful for them to attend to. To address these concerns, we began work on a *critiquing* interface, TraumaTIQ, which would take a proposed plan from the physician and produce a critique of that plan, focusing on those items where there was a potential disagreement [3]. In this way, important points could be emphasized, and the physicians would not have to be presented with

*This work has been supported by the National Library of Medicine under grant R01 LM05217-01.

a great deal of information that simply confirmed what they were planning to do already in order to find what was different. This approach to critiquing differs from previous critiquing systems in that the critique is delivered *during* the planning and delivery of patient care. Therefore the critique must be updated and made available rapidly, and must be as clear and succinct as possible, so as to be accessible to physicians whose attention is focused on the patient rather than on the computer system.

The critiquing process in TraumaTIQ is triggered whenever a new piece of relevant information is made available to the system. This information can be in the form of (1) bedside findings, (2) diagnostic test results, (3) therapeutic actions performed, or (4) actions ordered by the physician. TraumaTIQ interprets the physician's management orders in a goal-directed manner, using TraumAID's representation of goals, procedures and actions, so that the critique can address the likely reasons underlying any discrepancies and can suggest alternative means of addressing a particular goal.

The architecture of TraumaTIQ comprises *plan recognition*, *plan evaluation*, and *critique generation* processes. The function of the plan recognition component is to infer the underlying goal structure motivating the physician's orders. The plan recognition algorithm takes advantage of the fact that the physician using the system will have expert or near-expert knowledge of the domain, and will therefore usually develop plans that are similar to the plans generated by TraumAID. The algorithm works as follows:

1. When an action, α, is ordered by the physician, TraumaTIQ checks whether α is currently a part of TraumAID's recommended plan as a means of satisfying all or part of goal γ, or all or part of each member of a set of goals Γ.
2. If so, γ or Γ is ascribed to the physician and is incorporated into TraumaTIQ's model of the physician's plan.
3. If α is *not* currently in TraumAID's plan, TraumaTIQ determines whether there is a relevant goal that α might address.
   (a) If any of the goals that might lead to doing α are present in TraumAID's current set of active goals, TraumaTIQ will assume that α is being done to address that goal or goals.
   (b) In the case that there is no relevant goal to explain why the physician is ordering α, the goal is left unspecified and the intention to do α is added to the representation of the physician's plan with no goal attached. There is one exception to this rule:
   (c) If the system only knows of one possible goal that would lead to performing α, TraumaTIQ assumes that α is being done to address that goal, even though it does not consider the goal to be relevant.

This algorithm will not infer a goal for every action, in particular when the action cannot be explained by any of TraumAID's currently active goals. However, given that the purpose of the system is to alert the physician to any potential problems, rather than to provide a tutorial analysis of the proposed management plan and its alternatives, this is an acceptable feature. Furthermore, since there are critical time constraints on the system to be able to produce a response before a potentially harmful action is carried out, it is necessary to constrain the amount of complex computation performed during the plan inference stage.

The plan evaluation component compares the plan structure inferred by the plan recognizer with the plan produced by TraumAID, and identifies four potential types of discrepancies:

- *Omission*: A goal that TraumAID considers relevant is not being addressed by the physician in a timely manner. This can be further analyzed as to whether:
  - The goal is not being addressed at all.
  - The goal is only being partially addressed – some but not all of the actions making up the procedure addressing the goal have been ordered.
- *Commission*: An action is present in the physician's plan that *does not* address a relevant goal. If a unique goal can be inferred to explain this discrepancy, it can be further categorized:
  - The goal has been proven incorrect by the failure of all the rules that lead to concluding that goal.
  - The goal is not fully proven. Some of the findings leading to concluding that goal are known, but not all of them.
  - The goal has already been addressed.
- *Procedure choice*: A relevant goal is being addressed, but not using the procedure preferred by TraumAID.
- *Scheduling*: Actions are not being done in the order recommended by Traumaid, e.g., satsifying urgent goals before non-urgent ones.

These discrepancies are then evaluated in terms of their *significance*. It is our understanding that physicians will not want to be notified each time there is a slight disagreement between TraumAID's plan and their orders, but only when the discrepancy could be understood to be important. To this end, we are currently in the process of developing criteria for classifying errors according to their potential impact on the outcome of the case. For example, an unnecessary chest x-ray probably should not be critiqued, while an unnecessary laparotomy should be. Each discrepancy will be classified as either:

1. Tolerable, probably harmless.
2. Non-critical, but potentially harmful.
3. Critical, potentially fatal.

Anything in the second or third category will be considered significant enough to be reported in the critique. The output of the plan evaluation component is a set of *communicative goals* to be conveyed to the physician. Each goal has a *propositional content* indicating the type of discrepancy and the particular TraumAID concepts involved, and an *illocutionary force*, such as URGE, INFORM, REMIND, or SUGGEST, indicating *how* the information should be realized linguistically.

Finally, the critique generation component serves to organize the system's comments according to the management goals they address, and to translate them into natural language sentences. Each (ILLOCUTIONARY FORCE, PROPOSITIONAL CONTENT) combination corresponds to a sentential template with syntactically marked slots to be filled in with the appropriate phrasal translation of a TraumAID concept. For example, consider the communicative goal:

(SUGGEST
(PROCEDURE_CHOICE GET_X_RAY_LAT_ABD
GET_CT_SCAN_ABD
RO_COMPOUND_FRACTURE_LUMBAR_VERTEBRA))

which would be produced in a situation in which the physician has ordered a CT-scan of the abdomen, and TraumaTIQ has inferred that this action is intended to address the goal of diagnosing a compound fracture of the lumbar vertebra, which TraumAID has instead chosen to address using a lateral abdominal X-ray. The template corresponding to the pair (SUGGEST, PROCEDURE CHOICE) is:

"Consider (ARG1 GERUNDIVE) rather than (ARG2 GERUNDIVE), to (ARG3 BareVP)."

The first and second slots are filled in with gerundive phrases corresponding to GET_X_RAY_LAT_ABD and GET_CT_SCAN_ABD, while the third slot is filled in with an untensed verb phrase corresponding to RO_COMPOUND_FRACTURE_LUMBAR_VERTEBRA, resulting in the sentence:

"Consider getting a lateral X-Ray of the abdomen rather than getting a CT-scan of the abdomen, to check for fracture of the lumbar vertebrae."

Another important aspect of critique generation is the *mechanism* of output delivery. Currently, the system is designed to display its comments as written text on a monitor placed within view of the managing physician. We are exploring the possibility of using *synthesized speech* to deliver critiques. This would both eliminate the need for the physician to divert his gaze in order to read the comments, and serve to attract his attention more easily when a critique is presented. Our collaborators in this area have developed an approach to speech generation that takes account of both the informational content and the discourse context of the proposition to be expressed in order to produce appropriate stress and intonational contours [6, 7]. This technique can drastically increase the hearer's ability to grasp the meaning of an utterance, particularly in a situation where a contrast is being made. For example, using the default lexical stress pattern for the word "thoracotomy," with the primary lexical stress on the third syllable, would produce the following spoken output:

"Consider doing a left thoraCOTomy rather than doing a right thoraCOTomy, to treat the left hemothorax."

Where the contextually correct intonation would be:

"Consider doing a LEFT thoracotomy rather than doing a RIGHT thoracotomy, to treat the left hemothorax."

The latter would be much easier for a listener to interpret and ascribe the correct meaning to because it emphasizes the contrast between the two elements being compared.

Examples

As an example of the performance of the full TraumaTIQ system, suppose we have a patient with a gunshot wound to the abdomen and loss of sensation in both legs. These findings lead TraumAID to activate the goal of diagnosing a fracture of the lumbar vertebrae. TraumAID knows two procedures that can address this goal, a lateral abdominal X-ray or an abdominal CT-scan. The former is preferred as it takes less time.

If the physician orders a lateral abdominal X-ray in this situation, TraumaTIQ's plan inference module will ascribe to the physician the goal of diagnosing a compound fracture of the lumbar vertebrae (together with any other goals that it believes are relevant and can be addressed by the same procedure). It will also ascribe the intention to perform a lateral abdominal X-ray in order to address that goal. Since both TraumAID and the physician have chosen to get a lateral abdominal X-ray, no discrepancy will be found in this part of the plan during plan evaluation, and so no critique will be produced.

On the other hand, if the physician has ordered an abdominal CT-scan instead of the recommended X-ray, and the system can see no better reason for doing the CT-scan. TraumaTIQ will infer that the physician intends to do a CT-scan in order to address the goal of diagnosing a compound fracture of the lumbar vertebrae. This will be interpreted by the plan evaluation component as a procedure choice discrepancy with non-critical but potentially harmful consequences. The critique produced in this situation will be the comment seen earlier:

"Consider getting a lateral X-Ray of the abdomen rather than getting a CT-scan of the abdomen, to check for fracture of the lumbar vertebrae."

Now suppose that the patient shows hematuria as well as the previous findings of abdominal gunshot

wound and loss of sensation in both legs. Hematuria leads to a goal of diagnosing renal injury. The only procedure TraumAID knows for diagnosing renal injury is an abdominal CT-scan. Rather than including both a CT-scan *and* an X-ray in its management plan, TraumAID optimizes the plan so that both goals (diagnosing renal injury and diagnosing fractured vertebrae) are covered by the CT-scan procedure. If the physician were now to order a CT-scan, TraumaTIQ would recognize this action as being motivated by two currently active goals, and ascribe *both* of them to the physician. Once again, as there is no discrepancy between the physician's orders and TraumAID's plan, no critique is produced.

Conclusion

This approach to decision support in time-critical domains has the advantage that it can develop a model of the situation based on the partial information available to the system, and quickly determine what parts of that model are relevant in the context of the physician's current reasoning processes. Unlike previously developed reminder and alert systems, this approach evaluates the physician's proposed plan and attempts to intervene *before* problems occur. And unlike previous critiquing systems, it is able to provide decision support in real-time, *during* the planning and delivery of care. In the context of time-critical patient management it is, therefore, a more natural form of interaction, and we hope that it will prove to be a more effective one.

References

[1] J.G. Anderson, S.J. Jay, H.M. Schweer, and M.M. Anderson. Why doctors don't use computers: some empirical findings. *Journal of the Royal Society of Medicine*, 79:142–144, March 1986.

[2] Karen E. Bradshaw, Reed M. Gardner, and T. Allan Pryor. Development of a computerized laboratory alerting system. *Computers and Biomedical Research*, 22:575–587, 1989.

[3] Abigail Gertner. Real-time critiquing of integrated diagnosis/therapy plans. In *Proceedings AAAI Workshop on Expert Critiquing Systems*, 1993.

[4] Clement J. McDonald, Siu L. Hui, David M. Smith, William M. Tierney, Stuart J. Cohen, Morris Weinberger, and George P. McCabe. Reminders to physicians from an introspective computer medical record. *Annals of Internal Medicine*, 100:130–138, 1984.

[5] P. L. Miller. *Expert Critiquing Systems: Practice-Based Medical Consultation by Computer.* New York: Springer-Verlag, 1986.

[6] Scott Prevost and Mark Steedman. Generating contextually appropriate intonation. *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics*, 1993.

[7] Scott Prevost and Mark Steedman. Using context to specify intonation in speech synthesis. In *Proc. EuroSpeech '93*, Berlin, Germany, 1993.

[8] Barry G. Silverman. Expert bias research: Issues confronting knowlege engineers. In *AAAI-93 Workshop on Expert Critiquing Systems*, 1993.

[9] J. van der Lei. *Critiquing Based on Computer-Stored Medical Records.* PhD thesis, Erasmus University, 1991.

[10] B. L. Webber, R. Rymon, and J. R. Clarke. Flexible support for trauma management through goal-directed reasoning and planning. *Artificial Intelligence in Medicine*, 4, 1992.