Clinical Trial and Evaluation of a Prototype Case-Based System for Planning Medical Imaging Work-up Strategies

Charles E. Kahn, Jr.

Section of Information and Decision Sciences
Department of Radiology, Medical College of Wisconsin
Milwaukee, Wisconsin
ckahn@mcw.edu

Abstract

This report describes the preliminary clinical trial and evaluation of ProtoISIS, a prototype case-based reasoning (CBR) system for medical decision support. ProtoISIS learned the domain of ultrasonography and body computed tomography from 200 consecutive cases of actual requests for imaging procedures. ProtoISIS was tested by presenting it with 100 new imaging-procedure requests. ProtoISIS correctly classified 72% of the imaging-procedure requests overall, and 84% of the cases in the last two test series. We evaluated ProtoISIS in terms of performance, utility, and acceptibility to physicians. CBR can be applied successfully to the selection of diagnostic imaging procedures and holds potential for use in clinical decision support aids. Further work is necessary to realize a clinically useful system.

Introduction

ISIS (Intelligent Selection of Imaging Studies) is a case-based decision support tool being developed to help physicians select appropriate radiological procedures (Kahn 1993; Kahn & Anderson 1994). Its goal is to provide comprehensive computer-based expertise in the domain of diagnostic medical imaging procedures such as computed tomography (CT), ultrasound (US), and magnetic resonance imaging (MRI). To better assess the applicability of CBR techniques to the problem of imaging procedure selection, we conducted a pilot study in a more limited domain, that of ultrasound and body CT. This report describes the construction, testing, and evaluation of ProtoISIS, a prototype version of ISIS.

System Development and Training

ProtoISIS was based on the Protos learning apprentice (Bareiss, Porter, & Wier 1989; Porter, Bareiss, & Holte 1990). We implemented CL-Protos, a version of Protos in the Common Lisp language (Dvorak 1989), in Macintosh Common Lisp 2.0. Cases consist of actual requests for imaging procedures, where the clinical indications and questions to be answered are the features of the case, and the imaging study performed is the classification. ProtoISIS elicits explanations from the user that establish relationships between the various terms, such as features and imaging-procedure categories. These explanations allow ProtoISIS to create a semantic network (Figure 1) that relates the cases, features, and imaging procedures and to establish "remindings" from a case's features to its imaging-procedure category.

ProtoISIS uses relations somewhat differently than Protos, and incorporates new relations. The PART OF relation indicates components of imaging-procedure categories; *e.g.*, CT-CHEST is part of CT-CHEST-ABDOMEN. A new verb VISUALIZES relates imaging procedures and the conditions they reveal.

ProtoISIS was trained with 200 consecutive cases of actual ultrasound and body CT procedure requests abstracted from one week of radiology department records; this number of cases was chosen to provide a small, but representative sample of cases. Cases were presented in order of accession; there was no attempt to present the more typical cases first. Each case included patient identification, procedure requested, procedure performed, clinical data (indications and questions to be answered), and results. As each ultrasound or body CT case was presented, synonyms and abbreviations for the terms used as features of the case were entered. Appropriate relationships were defined to explain the category assigned to the case.

After training, ProtoISIS incorporated a total vocabulary of 527 terms: 200 case names, 28 imaging procedures, 37 abbreviations, 40 synonyms, and 222 features. Of the nine CT procedures, CT-ABDOMEN-PELVIS, CT-CHEST-ABDOMEN-PELVIS, and CT-CHEST had the most exemplars (33, 22, and 16, respectively). Among the 19 ultrasound procedures, US-KIDNEY, US-ABDOMEN, DOPPLER-ABDOMEN and US-HEAD had the most exemplars (29, 26, 14, and 11, respectively). All other imaging-procedure categories had six or fewer exemplars.

Supported in part by the 1993 American Roentgen Ray Society Scholarship.

Each exemplar consisted of one or more features: 66 exemplars (33%) had only one feature, another 77 (38%) had two features, and none had more than seven features. All but 13 (5.9%) of the 222 features had remindings to one or more imaging procedures. The great majority of features (76.4%) had remindings to only one imaging procedure; none had remindings to more than three imaging procedures.

Clinical Trial

We presented ProtoISIS with 100 new, consecutive ultrasound and body CT cases from radiology department records. The cases were grouped into four sets of 25 cases each, presented sequentially; we divided the test cases into groups to assess the improvement in ProtoISIS' performance as it gained experience. After each case's identifier and clinical features were entered, the system attempted to assign the correct category to each case. If ProtoISIS was unable to assign a category or assigned an incorrect category to a case, we added that case and pertinent explanations into memory. ProtoISIS incorporated into its knowledge base all new terms – such as abbreviations, synonyms, and features – that were encountered in the test cases whether or not the case to which they belonged was itself added.

Because the underlying CBR system, Protos, is an incremental knowledge acquisition system, it is susceptible to order effects in the training it receives. The current study was designed to assess overall feasibility, and did not account for this phenomenon: training cases were presented in the same order in which they were received, rather than presenting the most exemplary cases first. Future evaluations will control for order effects by not including the test cases in the system's memory and by presenting the training cases in several randomized orders to determine average performance.

ProtoISIS demonstrated satisfactory performance in the four test series. Only three of the 100 test cases required new imaging-procedure categories: CT-ABDOMEN-DRAINAGE, CT-LIVER-BIOPSY, and US-AORTA. Overall, ProtoISIS correctly classified 72% of the imaging-procedure requests on the first attempt. Its performance improved as it gained experience: in the last two test series, it correctly classified 84% of the cases presented, compared with only 56% in the first series. In many of the incorrectly classified cases, the correct imaging procedure received the second highest matching score.

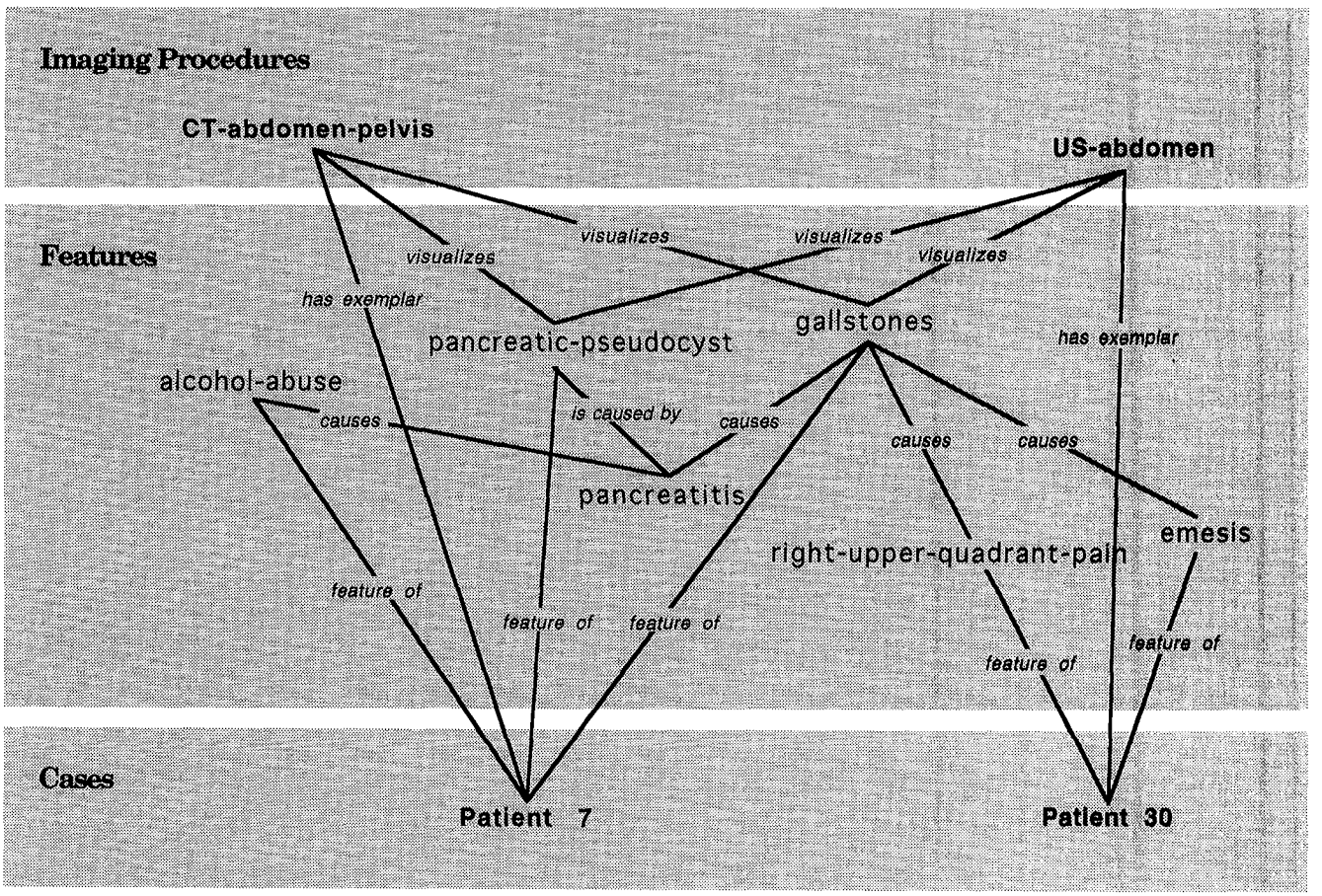

Figure 1. Semantic network of terms and relations in ProtoISIS.

On average, 40% of cases included terms that had not been encountered previously; given the small number of training cases, the large vocabulary of medicine, and the variety of ways that a single medical concept can be expressed, this finding is not surprising.

Evaluation

Performance

ProtoISIS performed reasonably well based on its small set of training cases, but not at a level suitable for clinical use. In order to be accepted into clinical practice, a decision support system must perform as well or better than physicians, and importantly, the physicians against whom one tests the system must be specialists or sub-specialists in the given domain. A very encouraging aspect of developing ProtoISIS was the relative ease of acquiring and entering the system's knowledge base.

More cases are needed to provide a suitable level of sophistication. Although one week of imaging requests — as used in ProtoISIS — provides a representative sample of cases, it does not provide sufficient depth of knowledge for ISIS to function at the level of an expert clinician. In addition, ISIS needs more information about each case, and more structure in that information. ISIS must distinguish between known features (patient history) and those being queried (clinical questions). In ProtoISIS, for example, the feature GALLSTONES might indicate either known gallstones or a question of gallstones.

In addition to the procedure requested and the clinical information provided, each case will include information about the procedure actually performed and the imaging technique or protocol. Each case will include clinical questions to be asked of the referring physician to determine the appropriateness of the requested imaging procedure and to assist the radiologist in formulating a diagnosis. ISIS' memory will incorporate "failures," such as non-diagnostic studies, inappropriately chosen studies, and cases with complications.

Interaction with Physicians

Although CL-Protos provides very efficient and easily understandable interfaces, they are not be suitable for use by physicians. We will need to develop user interfaces that limit the functionality of the system: several of the features available in CL-Protos are best used by programmers and "knowledge engineers." In addition, information being entered or presented must be clustered in ways that corresponds to typical clinical scenaria. We plan to "embed" ISIS into our radiology department's clinical information system; such integration offers the greatest potential impact on clinical care (Rossi-Mori, Pisanelli, & Ricci 1990). Most commercial radiology information systems are based on VT100-type 24-line, 80-column displays. Integration with these system will require special attention to the user interfaces to assure ease of use.

The ability of a medical decision support system to explain its reasoning is crucial to its acceptance by physicians. For each case, the expert user can provide comments or explanations which will be retrieved when viewing the case. We anticipate that these plain-text comments will correspond to the indexing features identified for the case. Future work will examine the ability to generate the explanations automatically from the case information. In addition, the comments for a case may include references to pertinent medical literature.

Some of the explanations generated by the Protos knowledge-based pattern matching algorithm did not represent valid reasoning. The most typical error was to present a chain of explanatory links that were too heavily dependent on the first exemplars seen by the system. In several instances, even though new explanations were offered, ProtoISIS continued to base its explanations on the earliest exemplars seen in each imaging-procedure category. Some of the errors may have been caused by incorrect use of Protos' relation verbs in several cases.

Developers of medical decision support systems have abandoned the "Greek oracle" model that assumed that physicians would convey all necessary data to the expert system and await its decision (Miller 1990; Miller & Masarie 1990). The "critiquing" approach allows a physician to propose a solution; the decision support system then responds to the physician's proposal and identifies the evidence supporting it or against it (Miller 1983). The physician and decision support system can engage in a dialogue to refine the proposed plan. This approach allows more robust interaction between the physician and the decision support system. This approach was been applied to radiology procedure selection in a rule-based system (Swett et al. 1989).

ISIS must integrate case-based reasoning with a critiquing approach. The critiquing approach allows ISIS to propose a revised or alternative plan and allows the physician to override its suggestions. In a consultative specialty such as diagnostic radiology, this mode of interaction and knowledge sharing is essential to the relationship between the radiologist (who knows the imaging procedures) and the referring physician (who knows the patient). This mode of interaction resembles retrieve-and-propose systems such as REPRO (Simoudis & Miller 1993).

Open Issues

Noise. In the current domain, the problem of "noise" can manifest itself as two or more cases with identical clinical information but different proposed plans (*i.e.*, selected imaging procedures). Physicians may choose different plans based on identical clinical information due to overlap of the diagnostic abilities of the imaging procedures or difference of opinion among expert physicians.

To account for such "noise" within the system's memory, the system might count the number of

occurrences of each alternative plan; the one used most frequently thus becomes the "preferred" plan, with other, less common plans allowed on the basis of precedent. This approach permits diversity of opinion, yet can help discourage physicians from selecting suboptimal procedures. If the secondary approach comes to be chosen with increasing frequency, then its weighting factor will exceed the other's and it will be considered the "primary" modality for the given problem. This approach will provide the system with flexibility to accommodate changes in protocols and procedures due to evolving medical imaging technology.

Episodic and Prototypic Cases. ProtoISIS included only "episodic" cases: all were derived from actual clinical records. It is not yet clear whether or not we will need to incorporate prototypical cases to represent medical practice guidelines. In retrieving cases, ISIS might need to give precedence to prototypic cases, because these cases presumably represent the composite experience of several learned cases. Where a retrieved episodic case presents a serious conflict for the proposed plan, that case, too, would be retrieved to use in adaptation of the imaging plan.

Case Weighting. Another open question is the use of weights to indicate which cases merit greater attention as exemplars. Such information might be placed in a "statistics" slot of each case. If the case is a prototype, the number of actual cases that comprise the prototypic case could be encoded in the statistics field to indicate the "weighting factor" of the case's importance.

Future Directions

Case-based reasoning has been applied experimentally in medicine to clinical audiology (Bareiss, Porter, & Wier 1989; Bareiss 1989; Porter, Bareiss, & Holte 1990), diagnosis of heart failure (Koton 1989b; Koton 1989a), and planning of radiation therapy protocols (Berger 1992; Berger 1993). Although these systems have been validated, to this author's knowledge, they are not in routine clinical use.

Once completed, ISIS will be integrated with the radiology information system of United Regional Medical Services, which provides radiology services to two large teaching hospitals on the grounds of the Milwaukee Regional Medical Center. Our radiology department performs more than 200,000 procedures annually, including about 36,000 imaging procedures. The clinical implementation of ISIS will permit investigation of a CBR decision aid in day-to-day clinical practice. This setting will provide an excellent "production system" test of ISIS, and will serve as a pilot project for "scaling up" the system to include all radiology procedures.

Conclusion

Existing decision support systems for radiological procedure selection include rule-based systems (Swett et al. 1989; Kahn 1991b), hypertexts (Greenes et al. 1989; Kahn 1991a), and belief networks (Haddawy, Kahn, & Butarbutar 1994). Once ISIS has been completed and validated, it will integrated with our department's clinical information system to provide interactive, on-line expertise to physicians at all times of the day in work areas such as clinics, inpatient wards, intensive-care units and the emergency department. Such a system will have excellent potential to significantly improve the quality and cost-effectiveness of medical care, and will offer an opportunity to study the role of case-based reasoning in day-to-day medical decision making.

Acknowledgments

CL-Protos, written by Daniel L. Dvorak, is based on the Protos system originally developed by Ray Bareiss and Bruce W. Porter at the University of Texas at Austin.

References

Bareiss, E. R.; Porter, B. W.; and Wier, C. C. 1989. PROTOS: an exemplar-based learning apprentice. *International Journal of Man-Machine Studies* 29:549-561.

Bareiss, R. 1989. The experimental evaluation of a case-based learning apprentice. In: Hammond, K., ed. Case-Based Reasoning: Proceedings of a Workshop on Case-Based Reasoning, 162-167. San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Berger, J. 1992. ROENTGEN: case-based reasoning and radiation therapy planning. Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care, 210-214. New York: McGraw-Hill.

Berger, J. 1993. Knowledge acquisition and design support in a medical domain. In: Leake, D. B., ed. Case-Based Reasoning: Papers from the 1993 Workshop (Technical Report WS-93-01), 141-146. Menlo Park, CA: AAAI Press.

Dvorak, D. 1989. *CL-Protos User's Guide.* Austin, TX: Artificial Intelligence Laboratory, Department of Computer Sciences, The University of Texas at Austin.

Greenes, R. A., et al. 1989. Knowledge management as a decision support method: a diagnostic workup strategy application. *Computers and Biomedical Research* 22:113-135.

Haddawy, P.; Kahn, C. E., Jr; and Butarbutar, M. 1994. A Bayesian network model for radiological diagnosis and procedure selection: work-up of suspected gallbladder disease. *Medical Physics* (in press).

Kahn, C. E., Jr. 1991a. A radiology hypertext system for education and clinical decision making. *Journal of Digital Imaging* 4:207-212.

Kahn, C. E., Jr. 1991b. Validation, clinical trial, and evaluation of a radiology expert system. *Methods of Information in Medicine* 30:268-274.

Kahn, C. E., Jr. 1993. Case-based selection of diagnostic imaging procedures. In: Leake, D. B., ed. Case-Based Reasoning: Papers from the 1993 Workshop, 164. Menlo Park, CA: AAAI Press.

Kahn, C. E., Jr, and Anderson, G. M. 1994. Case-based reasoning and imaging procedure selection: a pilot study. *Investigative Radiology* (in press).

Koton, P. 1989a. Evaluating case-based problem solving. In: Hammond, K., ed. Case-Based Reasoning: Proceedings of a Workshop on Case-Based Reasoning, 173-175. San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Koton, P. 1989b. A medical reasoning program that improves with experience. *Computer Methods and Programs in Biomedicine* 30:177-184.

Miller, P. L. 1983. Critiquing anesthetic management: the "ATTENDING" computer system. *Anesthesiology* 58:362-369.

Miller, R. A. 1990. Why the standard view is standard: people, not machines, understand patients' problems. *Journal of Medicine and Philosophy* 15:581-591.

Miller, R. A., and Masarie, F. E., Jr. 1990. The demise of the "Greek Oracle" model for medical diagnosis systems. *Methods of Information in Medicine* 29:1-2.

Porter, B. W.; Bareiss, R.; and Holte, R. C. 1990. Concept learning and heuristic classification in weak-theory domains. *Artificial Intelligence* 45:229-263.

Rossi-Mori, A.; Pisanelli, D. M.; and Ricci, F. L. 1990. Evaluation stages and design steps for knowledge-based systems in medicine. *Medical Informatics* 15:191-204.

Simoudis, E., and Miller, P. 1993. Automated support of developing retrieve-and-propose systems. In: Leake, D. B., ed. Case-Based Reasoning: Papers from the 1993 Workshop (Technical Report WS-93-01), 147-152. Menlo Park, CA: AAAI Press.

Swett, H. A., et al. 1989. Optimizing radiologic workup: an artificial intelligence approach. *Journal of Digital Imaging* 2:15-20.