H.J. Suermondt

Hewlett-Packard Laboratories,
Palo Alto, CA, USA

Synopsis

Decision-Support Systems

The area of decision-support systems has traditionally been an active field of research in medical informatics. Medicine provides some of the most challenging decision-making problems. As a result, over the past decades medical informatics has been a fertile area for research into decision-support systems. The complexity of medical decision making has led to numerous new problem-solving methodologies that are applied in many other domains, and medical decision-support problems have served as testbeds for scores of problem-solving methods that were originally developed in other fields.

The papers in this section focus on a number of practical as well as theoretical aspects of medical decision-support systems: how to generate various types of decision-support systems automatically from data; how to provide diagnostic assistance for a large domain in a manner that is consistent with clinician reasoning; whether and how easily systems can be transferred from one setting to another; how decision support affects resource use; and what types of assumptions underlie our decision-support systems.

The knowledge-acquisition bottleneck remains an active problem area in medical decision-support systems. A number of papers in this section address this problem by focusing on methods of machine learning. Bohren et al. [2] present a classification system that builds a decision tree to make outcome predictions. They demonstrate their technique using databases from three medical domains. The papers by Doyle et al. [5], Livieratos and Chappell [6], and Su [7] present theoretical and practical contributions in the area of neural networks. Doyle et al. present a type of network that handles missing data well. Livieratos and Chappell develop a neuron that can show the level of abnormality of image sections. Su develops a class of neural networks in which the weights and links are more easily interpreted and explained.

The article by Boon-Falleur et al. [3] discusses some of the logistic issues involved in applying decision-support systems in clinical care. These researchers present a system that automatically applies practice guidelines to determine whether test orders are justified given the indications provided on the order form. They also discuss the problems involved in porting a system from one institution to another, especially if this involves settings in different countries. Do Amaral et al. [4] present a comprehensive system for psychiatric diagnosis. The system reflects a large knowledge-acquisition effort, resulting in a system that combines uncertain and deterministic reasoning to model the successive phases of diagnosis employed by an expert. Aliferis and Miller [1] address an interesting philosophical issue that has been the source of much debate throughout the history of medical decision-support systems. They provide an overview of the assumptions underlying decision-support systems, and propose definitions that clarify the heuristic nature of practically all problem-solving systems.

Bohren et al. [2] present a general classification system called INC2.5. The objective of their system is to build decision trees from data in order to predict the outcome of new patients. The system has the ability to build (and incrementally refine) a classification tree in an unsupervised manner. It can be used to predict patient outcomes, and also provides information about the predictability of the outcome and about the relevance of patient variables to the result. The authors present in detail an algorithm to build the decision tree from data. Once the tree is built, the classification method of INC2.5 uses two parameters, a certainty threshold and a variable threshold, each of which can be fine-tuned to optimize classification performance. The variable threshold determines the required level of consistency required among cases in a category for a variable, and the certainty threshold determines whether there is enough evidence to make a prediction. The authors discuss how irrelevant variables can be detected and pruned from the decision tree. Such pruning has minimal effects on classification accuracy, while potentially reducing the cost of data acquisition. Bohren et al. demonstrate their classification algorithm in three domains: breast cancer, general trauma, and low-back pain. The authors show that, given the data sets from these three domains, classification performance gradually improves with the size of the decision tree until performance reaches a plateau where additional cases do not improve the performance of the tree. They illustrate the effects of changing the certainty threshold and the variable threshold, as well as the effect of reducing the number of input variables used, on the system's ability to classify cases.

Doyle et al. [5] address the classification problem using neural networks. Their objective is to develop a classifier that can be used to identify liver-transplant candidates who have a hepatoma, since standard screening techniques for this problem are not very sensitive. An interesting practical problem is that the database they used to train their neural networks contains a substantial portion of missing data. They use a network architecture that contains a separate encoder layer to address the problem of missing data. This strategy introduces only the assumption that the fact that the data were missing was not in itself significant. The authors generated an ensemble of networks, rather than a single network, to improve the performance of the classifier. To train and evaluate the resulting classifier, they used a database of 853 patients, 528 of which had no missing data. They determined the sensitivity and specificity of the resulting classifier with respect to detecting hepatomas, and found a sensitivity of approximately 88% and a specificity of around 75%. They concluded that their multivariate model offers substantial improvement in the detection rate of hepatoma.

The authors did an interesting study on the effects of missing data in the training set. They compared classifiers that were generated in three different ways. The first was built using the full dataset of 853 patients, which included missing data for some patients. The second was built using only the 528 patients who had no missing data. The third was generated by taking the 528 patients who had no missing data, and randomly deleting data from this dataset in a manner proportional to the distribution of missing data in the original dataset. Interestingly, they found that the three resulting classifiers showed no substantial differences in terms of their ability to detect hepatomas. They concluded that their method for handling missing data, using the encoder layer in the network, provided an effective means of controlling the effects of missing data on the accuracy of the classifier.

Su [7] addresses the often-cited problem that neural networks are viewed as so-called black boxes. The numerical weights in the network, and even the links among nodes, are difficult or impossible to interpret, and the classifications of the resulting systems are impossible to explain. Su proposes a class of neural networks called hyperrectangular composite neural networks. These two-layer networks are architected in such a way that they can be interpreted in the form of production rules, with meaningful weights. Each node in the first layer checks the conjunction of a number of lower and upper bounds on a subset of the input variables. The second layer tests the disjunction of the nodes in the first layer. The author presents a new training algorithm that refines successively the bounds on links between input variables and intermediate nodes. The algorithm adds intermediate nodes as needed until all training examples are classified correctly. Once the network has been generated, it can be translated easily into a number of rules that represent the conjunctions of bounds in the first layer of the network. As an example of the classifier, the author used a data set of measurements to classify the phase of progression in the development of Type-II diabetes mellitus in monkeys. The study showed that the resulting network classified the test cases correctly in about 90% of cases. Moreover, the author indicated that the resulting network could be used to describe the phase classification in terms of a limited set of rules even though, according to the author, the relationships are complicated and experts have difficulty defining them clearly. This paper contributes a neural-network architecture that addresses the inability of neural-network classification systems to explain their results or contents. It would be interesting to see how the performance of networks with this architecture compares to that of classification systems based on other types of neural networks.

Livieratos and Chappell [6] present a neural-network system that can determine the level of abnormality of each part of an image. The goal of the system is to analyze medical images and to recognize abnormal regions of an organ, independent of the particular disease. The authors cite the importance of determining not only whether a particular region is abnormal, but also the degree of abnormality of the region. They develop a neural-network method to generate, given an input vector representing an image, an output vector that corresponds to the level of abnormality of each element in the input vector. The output can be rendered as an image itself. Thus, the system can be used as an information-extracting tool. Livieratos and Chappell illustrate their method with a system for analyzing gamma-camera images of the lung. The images are normalized for lung size and shape, and have normalized pixel values. The system is trained with eight images of normal lungs. The authors presented the system with unseen test images and compared the results of the system with experts' diagnoses for the images. The paper shows a number of examples of output images that highlight abnormal regions. The authors state that the system revealed hidden information and compared well with the experts, but they do not discuss the details of this comparison.

Boon-Falleur et al. [3] present the results of an international effort to provide decision-support in the area of ordering laboratory tests. They discuss progress on two separate systems in clinical use. The first system addresses the problem of inappropriate or inadequately justified test ordering. The authors developed a rule-based system that incorporates their institution's local guidelines and practice agreements concerning thyroid disease, infectious mononucleosis, and allergies, in order to assess whether the indications and clinical information provided with a test order provided sufficient reason for ordering the test. In a preliminary evaluation of the system, they studied TSH tests that were ordered to rule in or rule out thyroid disease. For each order, they used the test result to determine whether thyroid disease was actually present. They found that in 17% of the orders, the system detected that the order was unjustified and no thyroid disease was present. However, the system also marked a small number of orders as unnecessary in which the test result actually showed thyroid disease. Upon further review, the authors cited insufficient documentation of the orders as the reason the system ruled these unnecessary. This study shows considerable promise in terms of monitoring test-ordering behavior.

In the same paper, Boon-Falleur et al. report on the transfer of LUMPS, a test-ordering system used in the liver-transplant unit, from a hospital in Birmingham, England, to a university hospital in Brussels, Belgium. The system takes data from the laboratory information system and applies local guidelines to help generate patient-order sets. The original system in Birmingham had been shown to reduce the number of test orders and the time it took staff to order tests. The transfer of the system required the user interface to be made language-independent. In addition, it required implementation of local protocols in the system's rule base. The system was used by hospital staff to order tests for pre- and post-transplant patients on the liver unit. The authors showed that, initially, large numbers of orders were added to the order sets proposed by the system. Currently, about 83% of all orders are proposed by the system. The authors found that overall order volume decreased only very slightly after introduction of the system. Given the large number of additions to the system-generated order sets, the authors suggest that users may not have learned to trust the system to propose the appropriate tests, and that with time the impact on order volume will improve.

Do Amaral et al. [4] present a high-performance, comprehensive diagnostic system for clinical psychiatry called DSP (Diagnostic Decision-Support System for Psychiatry). This rule-based system helps classify psychiatric patients into diagnostic categories defined by DSM-III-R. The authors are motivated by their observation that non-specialists are usually able to classify psychiatric patients into broad categories, but are unable to refine the diagnosis. They develop a system that combines uncertain with categorical reasoning. DSP initiates the diagnostic process with a set of findings about the patient. It uses uncertain reasoning to delineate the problem, applying rules with certainty factors to compute an ordered list of possible hypotheses. Using a predefined hierarchy of diseases in the domain, the system groups hypotheses and selects the rules that will help differentiate among groups. At this point, the system switches to a deterministic strategy to focus on the diseases in the most likely diagnostic category; categorical criteria are applied to rule out diagnoses and confirm the final diagnosis. The system is helped here by the clearly defined diagnostic criteria in psychiatry. In an evaluation of the system on 53 cases from the literature, Do Amaral and colleagues show that the system's rules alone can rank the correct diagnosis as the leading hypothesis in only 52.8% of cases. By applying the second phase of categorical diagnostic criteria after the initial ranking, they can improve this performance substantially, correctly identifying the diagnosis in 73.6% of cases (where the diagnosis stated in the case description forms the gold standard). The authors hypothesize that an exclusive rule-based strategy cannot cover all the necessary steps required in psychiatric reasoning; they argue that the combination of uncertain and categorical reasoning is modeled after the diagnostic strategy used by general psychiatrists. In practical terms, one of their main contributions is a system that can form broad diagnostic categories and suggest questions that will help the non-specialist refine the diagnosis. The authors suggest that the system be used by medical students, residents, or non-specialists (as performance is better than those, but worse than that of an expert psychiatrist).

Aliferis and Miller [1] address the issue that the term "heuristic," in the context of decision-support systems, is often interpreted as meaning informal or possibly incorrect. They analyze carefully the types of assumptions underlying medical decision-support systems, and propose a unifying definition of heuristics that encompasses formal and ad-hoc systems. The authors explore the various definitions of heuristics that have been given by dictionaries and by researchers in cognitive science and in AI. They point out the need for guaranteed, verifiable solutions as the essence of non-heuristic problem-solving systems. They claim that due to the complex nature of the problems addressed by medical decision-support systems, such problems generally cannot be solved in a guaranteed, verifiable manner. Heuristic systems, as Aliferis and Miller define them, are systems that introduce provisional assumptions, simplifications, or less than ideal data or methods. Interestingly, they distinguish between two types of heuristic systems in practice: those that are explicit about their assumptions (but for which the assumptions have not been verified, cannot be verified, or for which the violation of the assumptions in practice is considered of small significance), and those that offer neither proof of correctness nor prerequisites for correctness. The latter systems, often based on expert knowledge or modeled after expert problem-solving methods, represent "initial efforts to give satisfactory but temporary solutions to otherwise intractable problems." The authors propose means through which the latter type of heuristic systems can be transformed into the former. They argue for system developers to "condition the validity of their systems upon well-defined and, eventually, testable conditions." They offer a number of reasons for the polarization in the medical decision-support community around this issue, and discuss the necessity of collaboration and communication among researchers to address the tremendously complex area of medical decision support.

An interesting caveat with regard to the success of medical decision-support systems is illustrated by the fact that only one of the papers in this section describes systems that are in actual clinical use. Other systems discussed in this section are illustrated with medical examples, or have been evaluated on clinical cases, but have not found their way into clinical care. Numerous reasons have been proposed for the lack of transfer from the research laboratory to the wards. However, given the promising results of many of the evaluation studies and the continuous need of health-care institutions to improve the quality and efficiency of care, the time appears right for researchers to make a strong effort to get their decision-support technology adopted in clinical care.

References

1. Aliferis CF and Miller RA. On the heuristic nature of medical decision-support systems. Meth Inform Med 1995;34:5-14.
2. Bohren BF, Hadzikadic M, and Hanley EN Jr. Extracting knowledge from large medical databases: An automated approach. Comput Biomed Res 1995;28:191-210.
3. Boon-Falleur L, Sokal E, Nightingale PG, Peters M, Otte JB, Ketelslegers JM, De Wilde E, and Hasman A. Utilization of laboratory resources: Developments in knowledge-based ordering systems. Int J Biomed Comput 1995;40:17-30.
4. Do Amaral MB, Satomura Y, Honda M, and Sato T. A psychiatric diagnostic system integrating probabilistic and categorical reasoning. Meth Inform Med 1995;34:232-43.
5. Doyle HR, Parmanto B, Munro PW, Marino IR, Aldrighetti L, Doria C, McMichael J, and Fung JJ. Building clinical classifiers using incomplete observations - A neural network ensemble for hepatoma detection in patients with cirrhosis. Meth Inform Med 1995;34:253-8.
6. Livieratos GN and Chappell PH. A new approach in artificial neuron theory for use in computer-aided diagnosis. Med Eng Phys 1995;17:366-71.
7. Su M-C. Use of neural networks as medical diagnosis expert systems. Comput Biol Med 1994;24:419-29.

Address of the author:
Jaap Suermondt, PhD,
Hewlett-Packard Laboratories,
3500 Deer Creek Road, Bldg. 26U-16,
Palo Alto, California 94304, USA.
E-mail: suermondt@hplb.hpl.hp.com