Inducing Practice Guidelines from a Hospital Database

Karen Crowley Abston, M.S., Department of Medical Informatics, University of Utah,
T. Allan Pryor, Ph.D., Intermountain Health Care / University of Utah
Peter J. Haug, M.D., Jeffrey L. Anderson, M.D., LDS Hospital / University of Utah
Salt Lake City, Utah

Improving health care quality requires the elimination of unnecessary variation in the care process. Decision support applications already exist that can foster adherence to standards. The challenge resides in developing standards consistent with good medical practice. In this paper we present our efforts in determining where sufficient clinical data are captured electronically to automatically define a care process, and what analyses can be done to identify additional data that would allow a care process to be defined. Data routinely collected by a hospital information system have been examined. The analysis tools utilized include logistic regression, a neural network, a Bayesian network, and a rule induction program.

INTRODUCTION

Health care consumers and clinicians are changing the face of health care delivery across the United States. Consumers are demanding high-quality care at reasonable costs. Clinicians are placing increased emphasis on the practice of evidence-based medicine. Health care systems are seeking to meet these demands by employing the techniques of continuous quality improvement.

Improving health care quality while reducing costs requires the elimination of unintended and unnecessary variation in the care process. Decision support applications already exist to foster adherence to standards. These applications can help reduce variability. The challenge resides in developing standards based on scientific evidence and yet consistent with local norms of practice.

We postulate that by employing tools and techniques from the field of knowledge discovery in databases (KDD), we can induce models from a clinical data repository that reflect local physician ordering patterns. Information relating patient-specific parameters to treatment orders can then serve as a foundation for explicit, evidence-based practice guidelines.

BACKGROUND

Continuous Quality Improvement

Donald M. Berwick is a current health care champion of the quality improvement techniques pioneered by Shewhart and Deming.[1,2] Berwick discriminates between intended (based on reason) and unintended (not anticipated) variation.[3-5] Yet to judge variation as intended or unintended, it is necessary to measure it against a standard. Brook[6] and Eddy[7] have outlined the difficulties in establishing standards such as practice guidelines in health care.

To be most effective, practice guidelines should prompt the clinician to identify relevant signs, symptoms, and diagnostic findings, and then present specific care actions to be initiated. A simple example would be: "If patient is 65 years or older, give influenza shot."

Guideline Development

Guidelines vary in complexity, explicitness, and validity. Some are the product of committee consensus. Others are rooted in scientific data[8]. Guidelines should include rules about when to initiate and/or when to avoid health care interventions and should be based on hard data[9]. Nevertheless, they are frequently developed solely through expert consensus. This method is subject to the dangers of recall bias and the elevation of unsubstantiated opinion to the status of fact.

Most practice databases were primarily established to transmit and store data to facilitate clinical care. Though health care researchers make use of this record of clinical experience, data collected in the course of patient care represent an underutilized resource in guideline development.[10-12] By mining this rich data source, we can provide a framework for developing evidence-based care processes.

KDD Applications in Health Care

One can find an increasing number of KDD applications in health care. These studies and others have focused on the prediction of clinical diagnosis or outcome. Tu and Guerriere used a neural network

1091—8280/97/$5.00

to predict length of stay in the intensive care unit following cardiac surgery.[13] Lapuerta et al. used a neural network to predict the risk of coronary artery disease.[14] Clarke and Waclawiw applied the ITRULE generalized rule induction algorithm to analyze a database of a 5-year cohort study of the effects of obesity on major cardiovascular disease risk factors.[15] Hadzikadic et al. compared concept formation and logistic regression in predicting death in trauma patients. [16] Kukar et al. compared the performance of several algorithms including decision tree induction and Bayesian classification in the prognosis of femoral neck fracture surgery. [17]

This research focuses on applying KDD tools to explore the relationship of patient parameters and treatments. Following the knowledge discovery process set forth by Fayyad[18], we analyze data routinely collected by a hospital information system (HIS). The goal is to infer patient-specific orders using the clinical data available at the time when critical health care decisions are made.

The HELP System

The data for our experiments were extracted from the HELP Hospital Information System. This HIS has been in use from more than 20 years in the LDS Hospital, a 520 bed tertiary care center in Salt Lake City, Utah. It is a product of decades of research, development, and testing.[19] The system's Tandem mainframe is interfaced with a variety of departmental computer systems. These include systems for medical records, pharmacy, laboratory, electrocardiography, and a collection of local area networks that support clinical, research, and administrative tasks.

The key feature of the HELP System that facilitates research such as the proposed effort is the integrated clinical data repository. Clinicians throughout the hospital enter data pertinent to the health management of each inpatient. Large amounts of data flow in from departmental systems, e.g. pharmacy, lab, nursing, etc. Each patient's record is continuously available to personnel involved in direct care of the patient as well as researchers within the organization.

METHODS

Our research can be broken into four phases. Phase I consisted of preliminary data analysis to create a target data set for our experiments. In Phase II, we gained familiarity with the algorithms and function of the tools we would use to mine our data. In Phase III, we apply these tools in discovering data relationships relevant to the treatment of acute myocardial infarction (AMI). Phase IV will explore methods of refining our results by including information not available in the database.

Our target data set represents two periods of clinical practice at LDS Hospital. Cases extracted from 1990-1993 will be used in Phase II and III. Cases from 1994-1995 will be used in Phase IV.

The 1990-1993 population contains 854 cases. Each extracted case represents an inpatient with the following characteristics:

a) discharge diagnosis: acute myocardial infarction
b) source of admission: Emergency Department
c) reason for admission: AMI (confirmed as acute by estimating the time of infarct to be prior to the recorded admit time)
d) discharge disposition: live.

The data set is pre-reduced in the sense that not all available variables will be utilized. Case attributes include a unique patient identifier, age, gender, type of AMI, Killip's classification of AMI[20], discharge diagnoses (ICD9[21]), and medication orders placed during the first 24 hours from admission.

Cases exhibiting gross data abnormalities were excluded during extraction. It has been assumed that the absence of an electronic entry of a medication order indicates no order for that medication. This assumption may not always hold, yet there exists no electronic facility for their validation.

We have chosen to employ four prediction methods represented by four KDD tools: CN2, the Netica™ application, NevProp3®, and logistic regression. CN2 is an implementation of a production rule induction algorithm developed by Clark and Niblett.[22,23] Netica™ is a commercial software package for working with Bayesian networks.[24] NevProp3®, is a backpropagation artificial neural network simulator developed by Goodman.[25] Logistic regression was performed using SPSS®.[26]

Each KDD tool was trained with 80% of the study population cases. These cases were randomly sampled without replacement and presented in the same order to each tool. NevProp3® trained using its optional bootstrap function with the number of boots = 50. The logistic regression models were developed using both the backward stepwise and forced entry

methods. CN2 and Netica™ followed their default training parameters without modification. After training, the resulting models for each tool were tested with the remaining 20% of the cases. The decision threshold was set such that an output ≥ 0.50 constituted a positive "medication order present" decision.

For Phase II, we chose a relatively well-defined problem to familiarize ourselves with the functions and prediction capabilities of our KDD tools. The goal of these initial experiments was to predict correctly an order for a diabetic agent. The independent variables or attributes used for prediction were a discharge diagnosis of diabetes mellitus and the highest serum glucose value obtained in the first 24 hours from hospital admission.

The clinical focus of Phase III and IV is the management of AMI. Many of the medications used to treat AMI must be administered within a few hours of symptom onset to achieve their maximum effect. The American College of Cardiology has developed detailed guidelines concerning the administration of these medications.[27] One of our goals is to discover the magnitude of variation between these guidelines and documented practice. We will then be able to specifically target deficient areas for improvement. Another goal is to determine the electronic availability of patient data necessary to make treatment decisions.

Type of AMI, Killip class, gender, and age group were chosen as the initial variables of interest. Type, gender and age are captured electronically and are easily downloaded. Killip class, which represents the degree of heart failure, was inferred from discharge diagnoses. Variables such as electrocardiography results and time from infarct to Emergency Department presentation are not present in the electronic records for these cases. These variables, which are assumed to be more predictive of the administration of certain medications, will be the focus of Phase IV.

RESULTS

In Phase II, CN2, NevProp3®, and logistic regression performed comparably on the data set. Approximately 67% of the patients who actually received glucose lowering agents were correctly classified as receiving a medication order by these three tools. Netica™ performed poorly in relation to the other tools. Only 33% of the positive medication orders were correctly classified. Netica's™ model is constructed by the user. The models for the other three tools are created without user intervention during the course of training. Hence, Netica's™ poor performance may be attributed to an omitted critical dependency.

For Phase III, 32 medication classes were identified as directly related to the care of a cardiac patient. Twenty-five of these medications were ordered for less than 10 or greater than 80 percent of the population. The remaining seven were chosen for further study: 1) antiarrhythmics, 2) loop diuretics, 3) combination alpha-beta stimulators (e.g. epinephrine), 4) calcium channel blockers, 5) anti-fibrin enzymes, 6) digitalis and related cardiac glycosides, and 7) beta blockers. The percentages of patients with orders for these medications and the uncertainty concerning the proper use of some of the drugs indicate instances in which the assistance of a treatment guideline may be warranted.

Sensitivity and Specificity

Tables 1 and 2 present the sensitivity and specificity of each tool in predicting individual medication orders. Sensitivity describes the cases which contain a medication order and for which the tool predicted an order. Specificity describes the cases without a medication order for which the tool predicted no order should be made.

Though logistic regression, either backward stepwise or forced entry, achieved the better sensitivities overall, none of the tools attained a sensitivity that would be considered clinically acceptable.

Table 1. Tool Sensitivity in Predicting Medication Orders

Medication	1	2	3	4	5	6	7
Netica™ (Bayesian Network)	0.26	0.74	0.42	0.22	0.36	0.15	0.59
NevProp3® (Backpropagation Neural Network)	0.52	0.66	0.00	0.00	0.00	0.00	0.00
CN2 (Rule Induction)	0.53	0.67	0.56	0.30	0.00	0.50	0.50
Logistic Regression: Backward Stepwise	0.53	0.68	0.67	0.54	0.00	0.60	0.58
Logistic Regression: Forced Entry	0.53	0.68	0.67	0.52	0.00	0.67	0.59

Table 2. Tool Specificity in Predicting Medication Orders

Medication	1	2	3	4	5	6	7
Netica™ (Bayesian Network)	0.70	0.83	0.80	0.81	0.97	0.97	0.57
NevProp3® (Backpropagation Neural Network)	0.63	0.85	0.86	0.66	0.84	0.85	0.57
CN2 (Rule Induction)	0.65	0.86	0.88	0.66	0.84	0.86	0.62
Logistic Regression: Backward Stepwise	0.63	0.88	0.89	0.69	0.84	0.86	0.62
Logistic Regression: Forced Entry	0.63	0.88	0.89	0.67	0.84	0.86	0.63

The failure of each tool to successfully model the order for anti-fibrin enzymes (medication 5) was expected. The two clinical factors that determine the use of thrombolytics, time from onset of chest pain and bleeding risk, are not present in the electronic record and, therefore, not in our data set. The addition of these factors to our models will take place in Phase IV.

Netica™ achieved the best sensitivity in predicting orders for anti-fibrin enzymes (0.36). However, Netica™ did not perform as well as the other tools in predicting orders for antiarrhythmics, eponephrines, and digitalis.

The apparent inability of NevProp3® to successfully model medications 3, 4, 6 and 7 may be attributed to the limited modifications we made to the default optimization parameters.

DISCUSSION

We were not surprised by the less than adequate performance of our first models. The Netica™ and NevProp3® software, in particular, allow for many more modifications of learning parameters than we have tried to date. Future examination of sensitivity and specificity through the use of receiver operating characteristic curves may also provide a clearer picture of each tool's overall accuracy and the relative strengths and weaknesses among tools.

We have already modified NevProp3® to produce better sensitivities in predicting orders for antiarrhythmics (from 0.00 to 0.52) and loop diuretics (from 0.00 to 0.66) by adjusting the default sigmoid prime offset optimization setting.

The current version of Netica™ does not automatically model dependencies among independent variables. This functionality is present in the other tools. We hypothesize that Netica™'s sensitivity scores will improve as we supplement the tool with this information.

CN2 performs comparably to the other tools in many instances, but at a high cost. The number of rules generated ranged from 28 to 51 with no predictable trend associating the number of rules with sensitivity. To be useful, the output should contain a small number of easily understandable, clinically relevant rules. We will attempt to tune the software to prune more aggressively, though this may not be feasible given the present functionality of the CN2 software.

Furthermore, we initially conjectured that data routinely included in the electronic medical record would be insufficient to accurately predict most admit medication orders. In Phase IV of this research, we will add to the models data extracted manually from the paper chart.

In addition to the challenge of model fitting, potential hazards we have encountered so far include recoding the data extracted from the HELP clinical repository into the forms required by the KDD tools. Though an apparently simple task, it is nevertheless tedious and subject to error. The data must be frequently audited to preserve their quality.

The data themselves are subject to an uncertain amount of noise. These data were originally collected in the course of clinical practice in a fast-paced setting. It is possible, and highly probable, that the electronic record does not reflect the completeness of the paper-based medical record.

The clinical relevance of discharge diagnosis data is also questionable. These data are collected for the primary purpose of billing. However, we will attempt to estimate the extent of the data noise by performing a chart review of type of AMI and Killip classification. We will compare the electronically

abstracted categories with the written record of the first 24 hours of care.

CONCLUSIONS

We have presented the results of our preliminary study in the application of KDD tools in modeling routinely collected clinical data. This is a work in progress. Our next steps include refining the models by adjusting various learning parameters. In the final phase of this research, we will supplement model input to include factors that are more clinically predictive of a medication order. We will then measure the resulting improvement or deterioration in model performance.

It is our eventual goal to infer care process models based on specific patient profiles. These models can serve as the foundation for evidence-based guidelines to improve the quality of health care.

References

1. Shewhart WA. Economic Control of Quality of Manufacture Product. New York: Van Nostrand Reinhold Co., 1931.
2. Deming WE. Out of the Crisis. Cambridge: MIT Center for Engineering Study, 1986.
3. Berwick DM. Continuous improvement as an ideal in health care. N Eng J Med 1988; 320:53-56.
4. Berwick DM. Controlling variation in health care: a consultation from Walter Shewhart. Med Care 1991; 29:1212-25.
5. Berwick DM, Godfrey AB, Roessner J. Curing Health Care: New Strategies for Quality Improvement. San Francisco: Jossey-Bass, 1991.
6. Brook RH. Practice guidelines and practicing medicine: are they compatible? JAMA 1989; 262:3027-3030.
7. Eddy DM. Practice policies: where do they come from? JAMA 1990; 263:1265-1275.
8. Woolf SH. Practice guidelines, a new reality in medicine, II: methods of developing guidelines. Arch Intern Med. 1992;152:946-952.
9. McDonald CJ, Overhage JM. Guidelines you can follow and can trust: an ideal and an example. JAMA 1994; 271(11):872-873.
10. Safran, C. Using routinely collected data for clinical research. Stat Med 1991; 10:559-564.
11. Tierney WM, McDonald CJ. Practice databases and their uses in clinical research. Stat Med 1991; 10:541-557.
12. McDonald CJ, Hui SL. The analysis of humongous databases: problems and promises. Stat Med 1991; 10:511-518.
13. Tu JV, Guerriere MRJ. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Comput. Biomed. Res. 1993; 26:220-229.
14. Lapuerta P, Azen SP, LaBree L. Use of neural networks in predicting the risk of coronary artery disease. Comput. Biomed. Res. 1995; 28:38-52.
15. Clarke EJ, Waclawiw MA. Probabilistic rule induction from a medical research database. Comput. Biomed. Res. 1996; 29:271-283.
16. Hadzikadic M, Hakenewerth A, Bohren B, Norton J, Mehta B, Andrews C. Concept formation vs. Logistic regression: predicting death in trauma patients. Artif. Intell. Med. 1996;8(5):493-504.
17. Kukar M, Kononenko I, Silvester T. Machine learning in the prognosis of the femoral neck fracture injury. Artif. Intell. Med. 1996; 8(5):431-451.
18. Fayyad UM. Data mining and knowledge discovery: making sense out of data. IEEE Expert 1996; 11(5):20-25.
19. Kuperman GJ, Gardner RM, Pryor TA. HELP: A Dynamic Hospital Information System. New York: Springer-Verlag, 1991.
20. Roberts R, Morris D, Pratt CM, Alexander RW. Pathophysiology, recognition, and treatment of acute myocardial infarction and its complications. The Heart, vol.1, 8th ed. Schlant RC, Alexander RW, eds. New York: McGraw-Hill, Inc., 1994.
21. St. Anthony's Color-Coded ICD-9-CM Code Book. Alexandria, VA: St. Anthony Publishing, 1992.
22. Clark P, Niblett T. The CN2 algorithm. Machine Learning Journal. 1989; 3(4):261-283.
23. Clark P, Boswell R. Rule induction with CN2: some recent improvements. Machine Learning: Proceedings of the Fifth European Conference (EWSL-91), Kodratoff Y, ed. 1991: 151-163.
24. Netica™ Application User's Guide. Vancouver, BC, Canada: Norsys Software Corp., 1996.
25. Goodman PH. NevProp® software, version 3. Reno, NV: University of Nevada, 1996.
26. SPSS® Reference Guide. Chicage, IL: SPSS Inc., 1990.
27. American College of Cardiology. Guidelines for the early management of patients with acute myocardial infarction. JACC 1990; 16(2):249-292.