IMIA Yearbook of Medical Informatics 2005: Ubiquitous Health Care Systems. Haux R, Kulikowski C, editors. Stuttgart: Schattauer; 2004. p. 451-453.

G. Hripcsak

Department of Biomedical Informatics
Columbia University
New York, USA

Synopsis

Decision Support, Knowledge Representation and Management

The four papers in this section cover decision support, knowledge representation, and knowledge management. One of the core goals of medical informatics is to support clinical decision making, and all four papers contribute to that goal. The first paper applies a hybrid learning approach to improve prognosis prediction, the second uncovers the relative importance of variables in a neural network predictive model, the third applies a deep modeling approach to influence decisions, and the fourth demonstrates the value of information and knowledge in clinical practice.

Abu-Hanna and de Keizer [1] describe a combined machine learning–logistic regression approach to develop a prediction model for mortality in intensive care. Prognostic models are essential to monitor the quality of care delivered by institutions, because institutions differ in the kinds patients they treat. By producing an expected mortality for each institution, the models permit fair comparisons.

In classification tasks, the ability of a model to discriminate is of paramount importance: are individual cases assigned to the right class? The authors point out that in prognosis, unlike classification, predicting the correct probability of belonging to a class (life or death, in this case) for a group of patients is of paramount importance. This probability serves as the expected mortality for that institution or a subgroup within the institution. The authors refer to the correctness of the probability as "precision," although readers may be more familiar with the term "calibration."

Most current prognostic models are based on logistic regression. The advantage of current models is their simplicity. A limited number of standard variables are collected, a simple set of criteria converts them to a weighted score, and a simple formula converts the score to a probability. One can envision a more complicated algorithm based on more variables that produces a better calibrated probability, but adoption may be limited due to difficulty implementing it in the field.

The authors took a hybrid approach, which maintained much of the simplicity of the current models. Health care workers collect the same variables and produce the same score, but based on the initial variables, they chose among several different formulae to convert the score to a probability. Each formula is tailored to the subgroup of patients with similar initial variables.

The hybrid "local logistic regression" approach used classification trees to divide patients into subgroups and logistic regression within each subgroup to generate accurate probability estimates from the scores. The classification tree used systolic blood pressure, admission type, urine output, and Glasgow coma score to create five subgroups. A data set was broken into five subsets, and 2-fold cross validation was applied to each subset to generate the logistic models and to test them. The authors compared a global logistic regression algorithm, which is analogous to standard models; the hybrid approach in which each fold was divided into subgroups according to the classification tree and a logistic model was run on each subgroup; and 5-nearest neighbors. A logarithmic score that was sensitive to calibration was used to quantify performance.

The authors found that the local logistic regression approach produced the best performance most of the time, producing slight improvements in the logarithmic score, and that 5-nearest neighbor always did much worse than either logistic model. Graphically, the local logistic approach appeared better calibrated. In effect, the local logistic approach exploited information contained in the underlying variables while maintaining a relatively simple and familiar scoring algorithm.

Furthermore, the classification tree divided the population into clinically relevant subgroups with very different average mortalities. No single logistic model needed to cover a wide range of severity of illness, so each could provide a better fit. In assessing institutional performance, one can then easily stratify by the subgroups to better tease out why an institution is failing. Thus, through a hybrid method, the authors were able to achieve better prediction of probabilities and a better explanation for the findings.

Future work might consider whether there exists a better set of variables or a better scoring method to feed into the local logistic regression algorithm, and whether other classification methods might produce better subgroups.

Heckerling and coauthors [2] describe an approach to determine which variables are most predictive of community-acquired pneumonia. Methods like stepwise logistic regression are designed to select variables. Neural networks are often employed in prediction because of their ability to handle dependencies, but teasing out which variables are most predictive is more challenging. The authors therefore employed a number of measures to estimate the relevance of variables directly from the neural network.

The authors used a feed-forward, back-propagation neural network to predict the presence or absence of pneumonia based on 35 demographic, symptom, sign, and comorbidity variables for which clinical relevance was at least plausible. The data set contained 1023 patients who had no missing data and who had an unambiguous diagnosis of pneumonia or no pneumonia.

The general approach was to eliminate each variable in turn and then measure the drop in predictive performance of the network. The drop was interpreted as relevance. The authors tested full retraining, in which the network was fully retrained after the variable was dropped; weight elimination, in which weights emanating from the variable were dropped without retraining; constant substitution, in which the variable was replaced with its mean value; linear substitution, in which the variable was replaced with a prediction of its value based on the other variables; and data permutation, in which the variable was negated or complemented.

The top ten most relevant variables were picked for each method. A model using only those variables was compared to a model using random variables. The methods' variable selection was compared to that of a weight analysis and logistic regression with backward elimination.

The authors found that the five methods produced similar lists of variables, with similar but not exactly the same ranking. Weight analysis and logistic regression also produced similar lists. All the methods selected variables that produced models that were significantly better able to predict pneumonia than random variable selection.

The magnitude of measured relevance (drop in performance) was greatest in those methods that did not replace the lost information with information that might be available from correlated variables. For example, full retraining and linear substitution allow the network to exploit information from other variables. In effect, they can substitute one variable for another. Weight elimination and data permutation, on the other hand, do not support such substitution. This probably explains why logistic regression, which does allow substitution of information, produced a list most similar to that of linear substitution.

Future steps will depend on the clinical goal. If the clinical goal is to produce a neural network that provides accurate predictions with the fewest variables, then perhaps a neural network analog of stepwise logistic regression with both forward and backward steps will be helpful. If the clinical goal is to determine important factors in understanding pneumonia, then a method to unravel the dependencies among variables will be helpful. If the clinical goal is to improve the credibility of the neural network output, then further work understanding the behavior of the network based on combinations of variables will be helpful.

Plougmann and coauthors [3] describe a decision support system that models the effect of alcohol on diabetes patients. Their goal is to create a diabetes decision support tool for patients and providers that will help them better manage a patient's diet and insulin doses. Based on input about diet, past insulin doses, and patient-specific parameters like insulin sensitivity, the system simulates blood glucose response, allowing the patient to tailor their insulin dose and diet.

The authors carried out a clinical study of the effect of alcohol on glucose levels and found that glucose levels dropped after the alcohol was metabolized. They proposed a physiologic mechanism for the finding, based on alcohol metabolism, glycogen store depletion, and failure to drop insulin levels naturally as glycogen stores are being repleted. They cite earlier studies to support their proposal.

This physiological mechanism was incorporated into their carbohydrate metabolism time slice model. The original model tracks carbohydrate in two compartments and simulates the processes of organ systems. The new model adds an alcohol metabolism state, alcohol and glycogen compartments, and other process variables.

The model was evaluated using data from the clinical study. The original model displayed a large deviation (offset and mean error) from the observed glucose of patients who drank alcohol. The new model had a much lower offset and mean error for patients who drank alcohol. Both models matched the glucose of patients who drank water.

Future work includes refinement of the model to accommodate varying alcohol doses and further development of the tool.

Tamblyn and coauthors [4] describe a clinical trial of a decision support intervention. Their goal was to reduce adverse drug events by providing relevant prescription information to primary care physicians when they needed it. The study was conducted in Quebec, where a central prescription-claims database is maintained. The intervention consisted of information about dispensed prescriptions from the central database, which included information about their own and other physicians' prescriptions, and alerts for 159 predefined prescribing problems.

The study population comprised 107 physicians (of 440 approached) who cared for 12,560 study patients (of 20,109 approached) aged 66 years and older. Only one physician was selected from any practice. Randomization was by physician. The intervention and control groups both received a computer with software to document health problems and medications prescribed locally. Trained personnel entered the data into the computer based on reading the physicianís chart. The outcome measures were the rate of initiation of problematic prescriptions and the rate of discontinuation of problematic prescriptions.

A baseline study revealed similar rates of prescribing problems (32% intervention and 33% control). During the 13-month intervention period, intervention physicians initiated fewer problematic prescriptions (44 per 1000 visits versus 52 per 1000 visits). The most significant subtypes of problems were excessive duration of therapy and drug-age contraindication. Discontinuance of problematic prescriptions was not significantly different between the two groups overall or among the subtypes of prescription problems.

Subgroup analysis by subtype of prescription problem and by physician involved (study physician, outside physician, or both) revealed two improvements unadjusted for multiple hypotheses. The effect on prescription initiation was greater for physicians with more computer experience than those with less. Early technical difficulties and a change in copayments for prescription drugs may have affected the measured outcomes.

The study's results corroborated expectations. Physicians are more likely to heed advice when writing prescriptions than when reviewing existing prescriptions, especially if another physician wrote the prescription. The correlation with computer experience supports the assertion that the effect really was due to computer use and not some confounder.

Future work might include a more advanced and stable information system intervention, a study of outcomes, and a larger study population (or a longer duration).

Taken together, the four papers illustrate several trends. It is usually insufficient in health care to simply produce a prediction model; explanation and credibility are important factors for adoption by providers and patients. In some cases, detailed physiological modeling and a deep representation may be required. Hybrid methods may provide advantages over machine learning or statistical approaches taken alone. Automated decision support has strong potential in health care today.

References

1. Abu-Hanna A, de Keizer N. Integrating classification trees with local logistic regression in intensive care prognosis. Artif Intell Med 2003;29:-23.
2. Heckerling PS, Gerber BS, Tape TG, Wigton RS. Entering the black box of neural networks. Methods Inf Med 2003;42:287-96.
3. Plougmann S, Hejlesen O, Turner B, Kerr D, Cavan D. The effect of alcohol on blood glucose in Type 1 diabetes-metabolic modelling and integration in a decision support system. Int J Med Inform 2003;70:337-44.
4. Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, Hanley J, McLeod P, Laprise R. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. CMAJ 2003 Sep 16;169(6):549-56.

Address of the author:
George Hripcsak, MD, MS
Department of Biomedical Informatics
Columbia University
622 West 168th Street, VC5
New York, NY 10032
USA