Mining Bayesian Networks to Forecast Adverse Outcomes Related to Acute Coronary Syndrome

Andrew J. Novobilski, Ph.D.*
Francis M. Fesmire, M.D, FACEP **
David Sonnemaker *

*The University of Tennessee at Chattanooga
615 McCallie Ave., Dept. 2303
Chattanooga, TN 37403
Andy-Novobilski@utc.edu

Erlanger Heart/Stroke Center
UT College of Medicine, Chattanooga Unit
960 East Third Street, Suite 100
Chattanooga, TN 37403

Abstract

One fascinating aspect of tool building for datamining is the application of a generalized datamining tool to a specific domain. Often times, this process results in a cross disciplinary analysis of both the datamining technique and the application of the results to the domain itself. This process of cross-disciplinary analysis often leads not only to improvements of the tool, but more importantly, to a better understanding of the underlying domain model for the domain experts involved. This paper presents the results of applying a datamining tool for identifying a Bayesian Network to represent a dataset of triage information taken from patients arriving at the emergency room with symptoms of Acute Coronary Syndrome. Specifically, a domain expert generated Bayesian Network and a mined Bayesian Network, both trained using the triage dataset, are compared for their accuracy in forecasting 30-day adverse outcomes for the patients represented in the dataset. The comparison, done using ROC curves, shows that the mined Bayesian Networked slightly outperformed the domain expert generated network. The results are discussed and direction for future work based on the complexity of the mined network versus the expert's network are presented..

Introduction

One of the fascinating aspects of tool building for datamining is the application of a generalized datamining tool to a specific domain. Often times, this process results in a cross disciplinary analysis of both the datamining technique and the application of the results to the domain itself. This process of cross-disciplinary analysis often leads to improvements of the tool, but more importantly, to a better understanding of the underlying domain model for the domain experts involved.

One area of cross disciplinary research that lends itself to this type of collaboration is found in medical informatics. Over the past 18 months, faculty from the University of Tennessee at Chattanooga and the University of Tennessee College of Medicine, Chattanooga Unit have begun an initiative to research the use of Bayesian Networks in forecasting outcomes related to Acute Coronary Syndrome in an emergency room setting (Fesmire and Novobilski 2003). This is viewed as a key research area as 11,000 patients with acute myocardial infarction (heart attack) and an even greater number of patients with unstable angina (chest pain) are inadvertently discharged from emergency departments nationwide (Htlatky 1997, Pope et al 2000). Adverse outcomes in these patients represent a significant cause of death as well as greater than 25% of malpractice awards.

The domain experts' acceptance of Bayesian Networks (Heckerman, Mamdani and Wellman 1995) as a the real-time forecast model was facilitated by the white-box nature of Bayesian Networks. The probability based nature of the predicted outcome that is inherent to working with Bayesian Networks is something readily understood by clinicians. Secondly, the ability of the network to handle missing data gracefully was important due to constraints sometimes present at the point in time the data to be input is collected (during triage). Third, the shifting of individual node probabilities as evidence was added to the network allowed the domain experts to compare the model to their own belief system for what "should" be happening given the available information.

A second reason for the acceptance of a network based model was the familiarity with related work in the use of forecast models for ACS that focused on the use of feed forward neural networks that have been trained using back propagation (Ebell 1993). More recently, Baxt et al reported on the ability of a neural network to identify heart related problems in patients arriving at the emergency department with chest pain (Baxt et al 2002). In 2,204 patients, the network had a true positive rate (sensitivity) of 88% and a true negative rate (specificity) of 86% for Acute Coronary Syndrome despite the fact that a mean of 5% of all required network input data and 41% of cardiac marker data were missing (Baxt et al 2002). The authors conclude that neural networks have the potential to be utilized as a real time aid to identify the presence of AMI and ACS.

Its important to note, however, the point in time at which the data snapshot was taken. In the work done by Baxt et al, forty variables were collected on patients that included an Electrocardiograph (ECG) and blood work necessary to check the first set of cardiac marker enzymes. In contrast, the work being described in this paper focuses on the real-time use of a forecast model where the data presented to the model is that which has been collected during triage. In fact, there is only one variable, the physician's classification of the chest pain being exhibited by the patient, that is subjective in nature. The remainder of the information collected consists of demographic and prior history information that can be collected as soon as the patient arrives at the emergency room, or even earlier if Emergency Medical Technicians (EMTs) are dispatched to the patient.

The remainder of this paper presents the initial effort by the cross disciplinary team to establish the value of applying a datamining tool to initial triage data to identify a Bayesian Network capable of forecasting adverse outcomes in patients arriving at the emergency room with symptoms of Acute Coronary Syndrome. This effort resulted in the comparison of two Bayesian models; a domain expert generated Bayesian Network and a mined Bayesian Network. Both networks were trained using the triage dataset, and then evaluated for their accuracy in forecasting 30-day adverse outcomes for the patients represented in the dataset. The evaluation, done using ROC curves, allowed for a comparison that showed the mined Bayesian Networked outperformed the domain expert generated network. The results are discussed and direction for future work based on the complexity of the mined network versus the expert's network are presented.

The Domain of Interest

The domain of interest was represented by a prospectively acquired database of 2,148 consecutive chest pain patients with absence of injury on an initial ECG who underwent a standardized chest pain evaluation protocol for suspected acute coronary syndrome (ACS). All patients were followed for a 30-day Adverse Outcome that was defined as having a heart attack (acute myocardial infarction), angioplasty and/or placement of a stent (percutaneous coronary intervention), by-pass surgery (coronary artery bypass grafting), life-threatening complication, or cardiac death within 30 days of their initial visit to the emergency room. Each record consisted of the fourteen clinical variables shown in Table 1. This information was obtained on initial patient triage and was restricted to population demographics, classic risk factors, presence of pre-existing ischemic heart disease, and the nature and duration of chest pain.

Table 1 - Erlanger Chest Pain Protocol Definitions

Name	Description	Values
AGE	Age (In Years)	
RACE	Race	White, Black, Other
SEX	Sex	Male, Female
HXMI	History of Previous Myocardial Infarction	Yes, No
HXCABG_PCI	History of Previous CABG or PCI	Yes, No
HXHBP	History of Prior Essential Hypertension	Yes, No
HXDM	History of Prior Diabetes	No, NIDDM (), IDDM ()
HXCIG	History of Current Cigarette Use	Yes, No
HXLIPID	History of Prior Hyperlipidemia	Yes, No
ESTROGEN	Estrogen Status for Women Only	Positive, Negative
FHX	Positive Family History for Coronary Artery Disease	Yes, No
OBESE	Current Obesity	Yes, No
INITIALCPCAT	Initial Chest Pain Category	Typical, Atypical, Probable_Non cardiac
DURATION_HRS	Duration of Chest Pain Prior to Arrival if Definitely Known in Hours	

For the first part of the evaluation, a domain expert was given a blank Netica (Netica 1997) screen and asked to produce a "basic" model representing the interactions between the variables in Table 1, and the variable of interest, "ACS", which represents 30-day Adverse Outcome (heart attack, angioplasty, by-pass surgery or death). The network was then trained using Netica's internal learning mechanism to produce the Bayesian Network shown in Figure 1. Notice that the model indicates that the all variables with the exception of the two representing prior history of a cardiac event influence the probability of ACS. The expert chose to indicate that

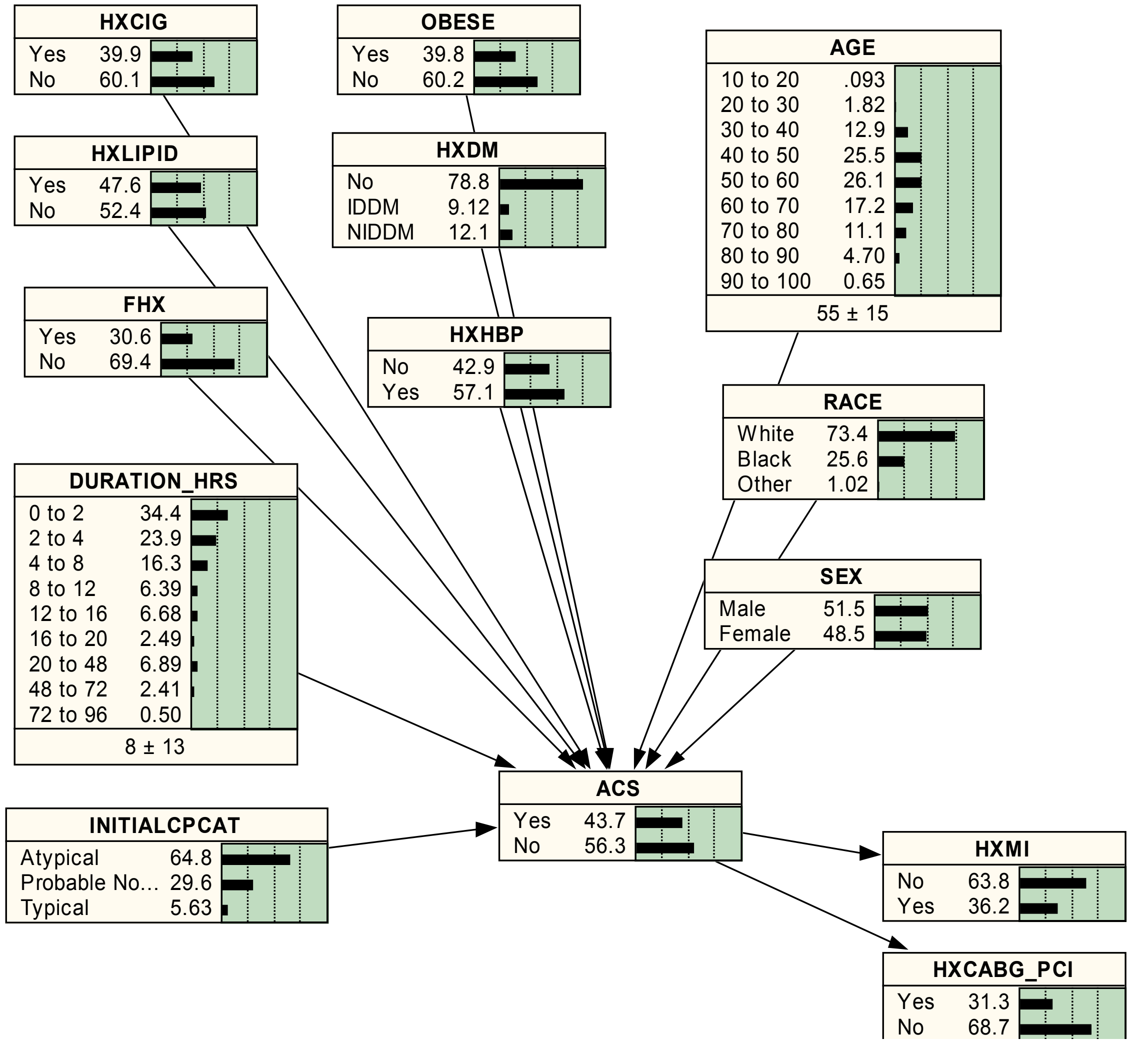

Figure 1 - The Expert Generated Bayesian Network

history of a prior cardiac event was a result of ACS at some time by making the prior history variables dependent on ACS.

Once the network was trained, it was used to evaluate the probability for each case that ACS would occur. The results were then used to generate the ROC curve shown in Figure 2. ROC (Receiver Operating Characteristic) curves were originally developed to measure the ability of radio receivers to discriminate signal from noise (Swets and Picket 1992). They are now used extensively within the medical field to measure the operating characteristics of a particular diagnostic tool or test (Zweig and Campbell 1993). The curve is a result of plotting True Positive Rate versus the False Positive Rate as the cutoff value indicating that ACS is true is varied from 0. to 1.

The key measure of accuracy of the network is obtained by computing the area under the curve with values of above .9 considered excellent, values between .8 and .9 good, values between .7 and .8 fair, and anything below .7 as poor/failing. Intuitively, this can be thought of as saying the perfect diagnostic test will have a key cutoff value that results in only true positives and no false positives being identified up to the key value.

Since the ROC curve and Area Under the Curve (AUC) are normalized, the AUC value can be used to rank diagnostic models against each other. Unfortunately, not all equal AUC values have the same importance. For example, a test good at discriminating true positives will have a greater AUC when plotted against the lower false positive rates. This leads to a second measure of accuracy that involves comparing the area under the ROC curve when the false positive rate is between 0 and .2. This value indicates the ability of the diagnostic tool to classify all true positives correctly and is viewed as more important in the clinical environment than being able to correctly classify true negatives.

Given this information, its interesting to note that the ROC curve in Figure 2 indicates that the forecast model performs very well initially, has difficulty, and then resumes in a "good" manner (AUC = 0.77776). Evaluation of the data showed a significant number of false negatives occurring when the cutoff value for ACS being true was set at .59. This deviation from the expected curve has been left for future evaluation.

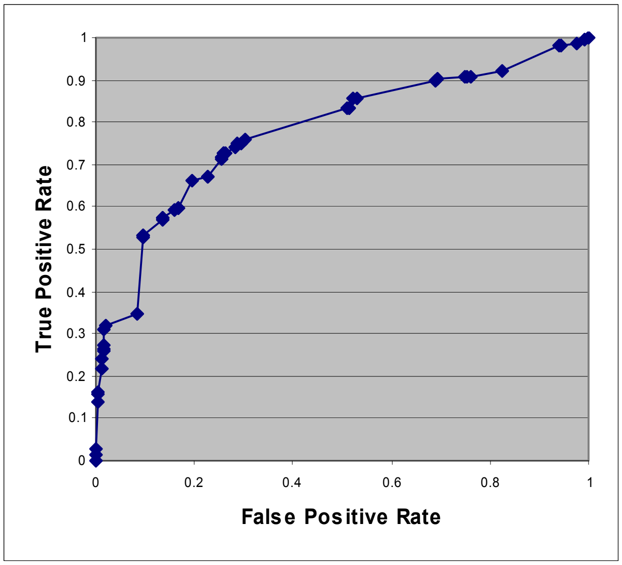

Figure 2 – ROC Curve for the Expert Network

The Mined Bayesian Network

Once the expert designed network was complete, the datamining tool described in (Novobilski 2003) was used to produce the Bayesian Network shown in Figure 3. The datamining process used a Genetic Algorithm approach by replacing the standard crossover operation with an alternative approach that defined three new operators for use in establishing and reproducing a population of legal fixed length encoded DAGs capable of describing Bayesian Networks without resorting to enforced node ordering or use of a repair operator. These operators support the Genetic Algorithm by providing both random selection of initial legal encodings and support for replacing the traditional crossover operator with two new operators, influences and joins, that serve the purpose of preserving and promoting "good" schemata while retaining legal encoding for the newly created members of the population.

As before, the network was used to evaluate the probability for each case that ACS would occur. The datamining process used the AUC value to rank candidate networks during the production of each generation for the genetic algorithm. The AUC value itself was computed by using a k-fold averaging process with k=4 for each network evaluated. The results were then used to generate

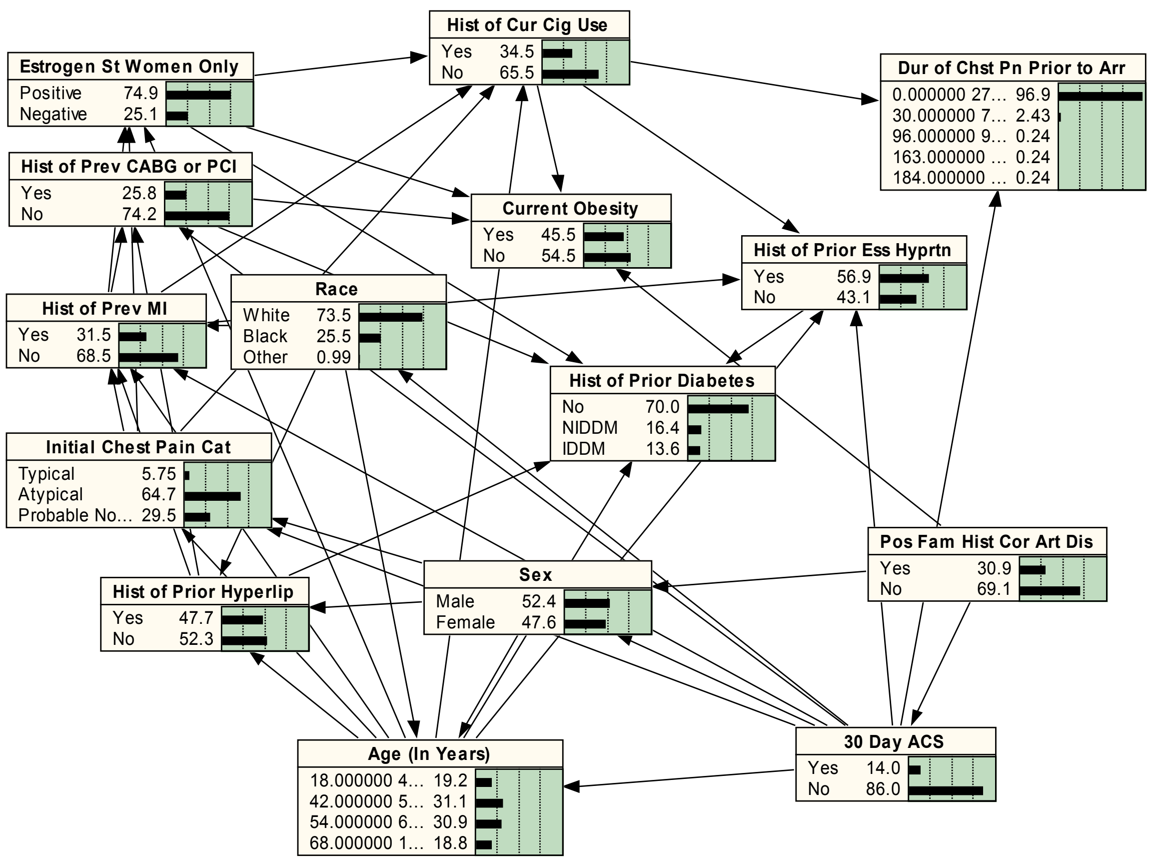

Figure 3 – The Mined Bayesian Network

the ROC curve shown in Figure 4. Once again, the ROC curve indicates that the mined Bayesian network initially performs very well and then tapers off. Unlike the ROC Curve in Figure 2, however, the second ROC curve conforms to the expected shape for a "good" curve, as indicated by its AUC value of 0.82216.

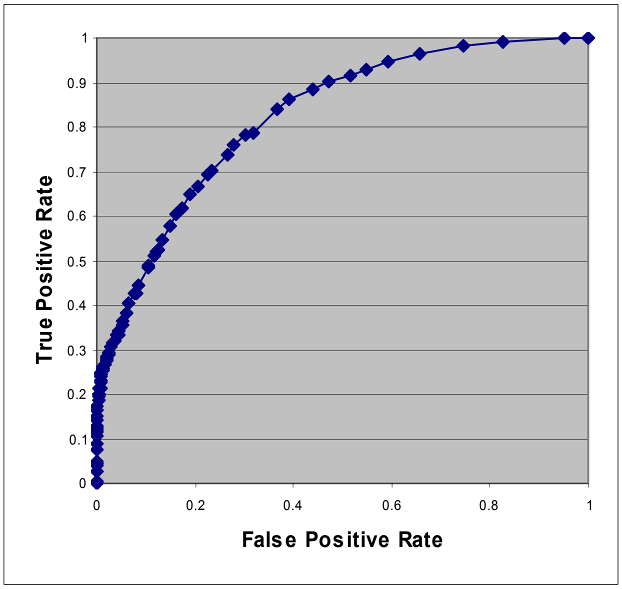

Figure 4 – The Mined Network ROC Curves

Conclusions

In conclusion, the data mining tool was able to produce a Bayesian network capable of forecasting a 30-day adverse outcome from triage data with a slightly better level of overall accuracy than the expert generated Bayesian network. Figure 5 shows the combined ROC curves for both the expert produced and the mined Bayesian networks. Looking at the values of the True Positive Rate for False Positive Rate between 0 and .2 indicates that the two curves are basically equivalent, except for the unexpected shape of the "expert" ROC curve at the previously discussed point. The reason that the mined net has a higher AUC value is due to its better performance at greater values of the false positive rate. Something that is typically not deemed useful in the domain (forecasting of adverse outcomes related to acute coronary syndrome) the model is being used for.

In addition to the mined network producing an ROC curve with a conforming shape, its important to note that the mined network also has a greater complexity than the expert generated network. Also note that several of the arrows are "backwards" in the sense of causality between variables. Although it is possible to "mathematically" turn the arrows around, it still presents difficulties to a domain expert trying to evaluate the way in which the network "intuitively" models the domain being looked at. One possible way around this is to place the network behind a user interface that presents the clinician with a series of questions. As each question is answered, a probability of ACS score is updated in the background. This score is then related to an index that is adjusted based on the optimal cutoff point as defined by the ROC curve, with the index value being displayed to the clinician.

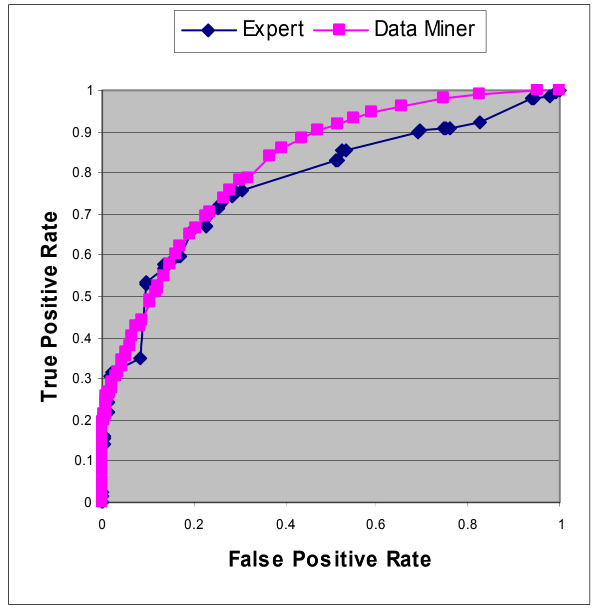

Figure 5 – The Combined ROC Curves

Future work planned by the cross disciplinary team includes applying the datamining tool to other datasets related to Acute Coronary Syndrome. These datasets are expected to be of different sizes, collected from different demographics, and containing different sets of collected variables. The team is also exploring the combined use of Neural Networks and Bayesian Networks in a two tiered architecture that would allow integration of data from sources such as continuous ECG and other diagnostic devices. Finally, the datamining tool itself continues to be expanded upon in order to take better advantage of available High Performance Computing environments.

References

Baxt WG, Shofer FS, Sites FD, Hollander JE: 2002 A neural computational aid to the diagnosis of acute myocardial infarction. Ann Emerg Med;39:366-373.

Baxt WG, Shofer FS, Sites FD, Hollander JE: 2002. A neural network aid for the early diagnosis of acute cardiac ischemia in patients presenting to the emergency department with chest pain. Ann Emerg Med;40:595-597.

Ebell, Mark H. 1993. Artificial Neural Networks for Predicting Failure to Survive following In-Hospital Cardiopulmonary Resuscitation. Journal of Family Practice. 36(3): p297-304

Fesmire FM, Novobilski A. 2003 First step in the Erlanger Artificial Intelligence Initiative: development of a Bayesian network utilizing initial triage history to risk stratify chest pain patients for thirty-day adverse outcome [Abstract]. Ann Emerg Med 42 (in press)

Heckerman, D., A. Mamdani, and M. Wellman. 1995. Real World Applications of Bayesian Networks. Communications of the ACM 38, no. 5: 25-30

Htlatky MA: 1997. Evaluation of chest pain in the emergency department. *N Engl Med*;337:1687-1689.

Netica Application API, DLL, and Users Guide. 1997. Norsys Software Corporation, Vancouver, BC, Canada.

Novobilski, Andrew. 2003. The Random Selection and Manipulation of Legally Encoded Bayesian Networks in Genetic Algorithms. The 2003 International Conference on Artificial Intelligence (ICAI)

Pope JH, Aufderheide TP, Ruthazer R, et al: 2000. Missed diagnosis of acute cardiac ischemia in the emergency department. N Engl J Med; 342:1163-1170.

Swets, J.A. R.M. Picket. 1992. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York.

Zweig, M.H., G. Campbell. 1993. Receiver Operating Chatracteristic (ROC) Plots: A fundamental Evaluation Tool in Clinical Medicine", Clinical Chemistry, 39, 561-577.