MINING YOUR DATA FOR HEALTH CARE QUALITY IMPROVEMENT

Greg Rogers - SAS Institute, Inc., Cary, NC.
Ellen Joyner - SAS Institute, Inc., Cary, NC.

ABSTRACT

Quality improvement in the health care industry can best be defined by examining the driving forces that are effecting the industry. The evolution of the health care in this country is driven by a single purpose. How can health care organizations lower cost, raise quality, and still remain competitive? The implementation of patient-focused quality improvement programs are at or near the top of the list for both information technology (IT) and business unit managers alike in their efforts to lower cost.

For IT managers, the key benefit SAS software brings to the table an environment of seamless integration of tools, data, and solutions across the enterprise. Data is at the heart of any patient-focused quality improvement program. The SAS System's access methods and data warehousing capabilities give IT managers what they need to implement the infrastructure necessary for a successful quality improvement and data mining solution. For the business unit managers interested in quality improvement, the key benefit of the SAS System is its depth and breath of end-user analytical and decision support tools. Numerous health care organizations are using the SAS System's industry-standard quality improvement tools to raise the quality and efficacy of health-related products and services.

In this paper, we address how SAS software data mining technology can be utilized as a solution for improving the quality of care as well as help control spiraling costs in the health care industry. More specifically, we will review some of the driving forces in the market today and discuss the SAS software methodology and approach for data mining. Finally, you will see how this technology can be applied to help solve critical business solutions within the health care industry.

HEALTH CARE TRANSFORMATION

The once strong national legislation reform movement has created a firestorm within the industry. This has put health care into the hearts and minds of the public. Contrary to popular belief, the national health care reform effort didn't create this change in the industry; it has only served to speed up the process. The industry is reforming itself with or without national legislation. Reform efforts have grown out of the industry's need to control spiraling costs and inefficiencies. Today, real reform manifests itself in the form of market pressures. In other words, the industry is changing itself; no more of the status quo. The traditional, not-for-profit health care industry is now learning how to run health care like a business. Over the last two decades, most American industries have had to learn the value of cost control and higher quality through pressures created by foreign competition and investment. The health care industry is now being forced to learn the same lessons.

In addition to market pressures reform, many states are in the process of enacting their own health care legislation—involving everything from legislation that supports the growth of managed care, to the shifting of Medicaid. States have now gained federal approval to put their billion-dollar Medicaid programs into the hands of managed care organizations. Many feel this could be problematic. A lack of a national coherent regulatory framework could lead to what Dr. David Lawrence, CEO and chairman of the board for Kaiser Foundation Health Plan called a "rapid nonlinear discontinuous change" in the industry. Some agree that this lack of a coherent framework is a model for failure and potential collapse, while others in the industry would wholeheartedly disagree. Within this complex set of circumstances, the great challenge is how to lower the cost of health care while improving the quality and accessibility of that care.

THE EXPLOSIVE AVAILABILITY OF DATA AND TECHNOLOGY

Over the past decade or so, businesses have amassed exponentially growing amounts of data in large databases. These stockpiles mainly contain customer or patient data, the unique "gold" the competition lacks. But the data's hidden value, the potential to predict business trends and customer behavior, has largely gone untapped. To collect and store large amounts of data is a

waste of resources. Unless the data is used productively, the organization is sitting on an unexplainable overhead. It becomes an untapped resource costing the organization money. This is especially apparent within the health care industry. As the number of new alliances, consolidations, buy-outs, mergers, and acquisitions continues to grow, the industry will further find itself in a state of data confusion. With it comes a proliferation of non-integrated, stand-alone systems with useless collections of data.

To convert this potential value into strategic business information, many companies are turning to data mining, an emerging technology based on a new generation of hardware and software. Data mining combines techniques including statistical analysis, visualization, induction, and neural networks to explore large amounts of data and discover relationships and patterns that shed light on business problems. In turn, companies can use these findings for more profitable, proactive decision making and competitive advantage.

Data mining was designed for exploiting massive amounts of data. This process can be more efficient if you first define what the business problem is, and then determine the amount of data you will need to solve the problem. By taking this "bottom up" approach to data mining and involving upper management in the understanding of business problems and the potential return on investment (ROI), the process will be much more acceptable and the goals attainable.

SOLVING BUSINESS PROBLEMS

The SAS data mining solution has been used in health care to overcome a wide range of business issues and problems. Some of these include:

- Segmenting customers/patients accurately into groups with similar health patterns.
- Rapidly identifying who are the most profitable customers and the underlying reasons.
- Understanding why customers leave for competitors (attrition, churn analysis).
- Planning for effective information systems management.
- Preparing for demand of resources.
- Anticipating customers'/patient's future actions, given their history and characteristics.
- Predicting medical diagnosis.
- Forecasting treatment costs.
- Predicting length of stay in a hospital.
- Identifying medical procedure expenditures and utilization by analyzing claims and point-of-care data.
- Predicting total cost of patient care.

Once the business problems have been defined and agreed upon, the next logical step is to determine the type and amount of data that will be necessary for making business decisions. As a precursor to data mining, a data warehouse strategy and implementation is suggested.

Integration with SAS software gives the SAS data mining solution several distinguishing characteristics which allow faster, easier and more accurate conversion of data into knowledge useful to decision makers.

Data diversity: The SAS data mining solution is designed to accept a wider range of data formats than any other data mining product currently on the market. It will accept data from relational and hierarchical databases, flat files, and other data formats, and it will accept this data from all major hardware platforms.

Distributed client/server: The SAS data mining solution supports both the "data server" model of client/server computing, in which data located on a remote machine can be accessed, and the "compute server" model, which allows data to be processed on a remote server and then forwarded to a client. This is particularly well suited to analytical tasks involving large volumes of data that require superior processing capabilities.

Consistent implementation on multiple platforms: The SAS data mining solutions are fully integrated with SAS Software and give users the flexibility to use their platforms of choice, ranging form desktop machines to powerful servers.

Integrated data management: SAS software's data management facilities guarantee data integrity without the need for re-keying or additional validation of data. No other data mining solution includes seamless integration with such a comprehensive range of data management functionality. Once the business objectives and data issues have been resolved, the methodology and approach to data mining can begin.

THE SEMMA METHODOLOGY

The methodology and approach that SAS Institute proposes is referred to as SEMMA, for *Sample*, *Explore*, *Modify*, *Model*, and *Assess*. Beginning with a statistically representative sample of data, users can

apply exploratory statistical and visualization techniques, select and transform the most significant predictive variables, model the variables to predict outcomes, and affirm the model's accuracy.

Sample: The first step is to extract a portion of a large data set big enough to contain the significant information yet small enough to manipulate quickly.

Explore: This phase involves searching speculatively for unanticipated trends and anomalies so as to gain understanding and ideas. This can reveal which subset of attributes will be the most productive to work with the modeling phase. Data visualization delivers intuitive tools for business professionals, while statistical techniques offer added detail for specialist.

Modify: The insights that are gained from the exploration phase enable knowledge workers to group the most productive subsets and clusters of data together for further analysis and exploration.

Model: This process involves searching automatically for a variable combination that reliably predicts a desired outcome. Data mining techniques such as neural networks, tree-based models, and traditional statistical techniques can help reveal patterns in the data and provide a best-fitting predictive model.

Assess: During this evaluation process, assessment of the results gained from modeling provides indications as to which results should be conveyed to senior management, how to model new questions that have been raised by the previous results and thus proceed back to the exploration phase.

SEMMA is a process that allows SAS Institute to distinguish ourselves by being the only vendor that can offer all of these components, as well as the ability to seamlessly integrate them with a company's existing hardware and software strategy.

SAS SOFTWARE: A MULTI-ANALYTICAL APPROACH

The Gartner Group, a leading technology analyst firm, states that "a number of underlying techniques can be applied to the various functions of the data mining effort. Visualization, statistics, induction, and neural networks are the most popular." Most of these techniques are mentioned above in the various phases of the SEMMA process. Following is a brief discussion on the usage and advantages of each.

Neural Networks

Often defined as a computer application that mimics the neurophysiology of the human brain, a neural network is capable of learning from examples to find patterns in data.

The neural network is first trained, which involves reading sample data and iteratively adjusting the network's weights to produce optimum predictions. Then new data can be applied to this model to quickly generate predictions. Neural networks are reputed to produce highly accurate results and, in practical applications, can contribute to profitable decisions.

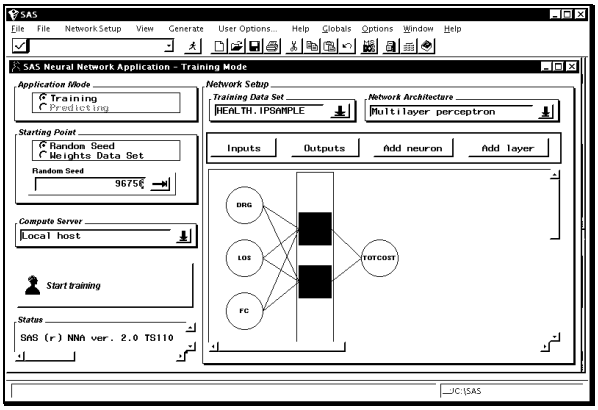

Figure 1. A neural network is just one of the many data mining analytical tools that can be utilized to make predictions on key health care indicators such as cost or facility utilization. In this example, the SAS neural network application is being used to determine key factors associated with predicting the total cost for an episode of care. An intuitive user interface is available to help build a multilayer perceptron network with three inputs, one hidden layer, and one predictor.

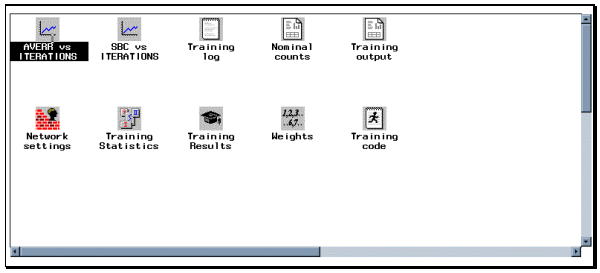

Figure 2. Once the neural network has been trained, there are various results windows available for exploring.

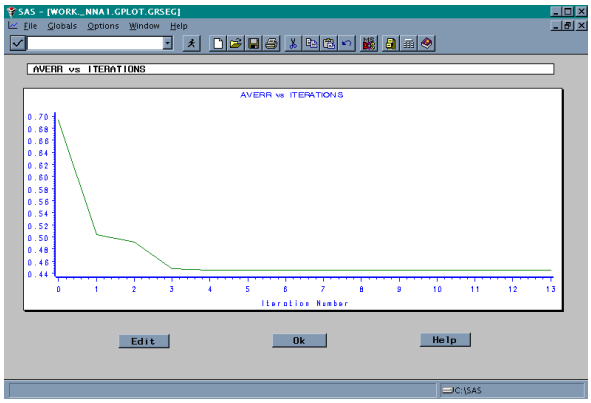

Figure 3. Above, we review the average error vs. iterations through the network. You can see that the error takes a significant drop after the third iteration. This allows us to graphically assess the model and error predictions.

Induction

Tree-based models, which include classification and regression trees, are the most common implementation of induction modeling. Tree-based models automatically contract decision trees from data, yielding a sequence of rules, such as "If income is greater then $60,000, assign the customer to this segment"(see figure 4). Many companies have been using the SAS System for years to perform classification tree modeling.

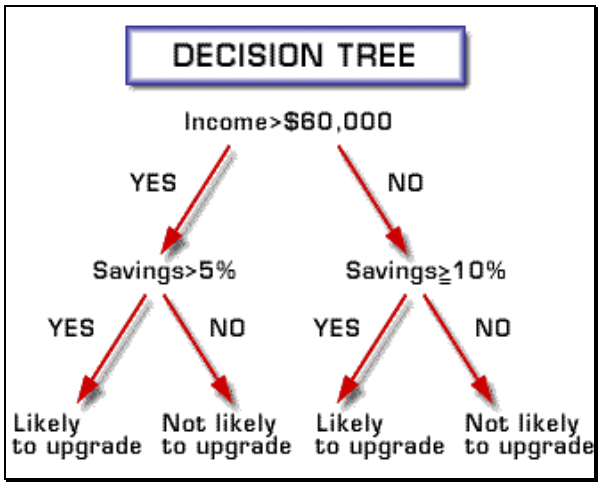

Figure 4. An example of a decision tree with decision points and rules associated with customer income and the likelihood to upgrade.

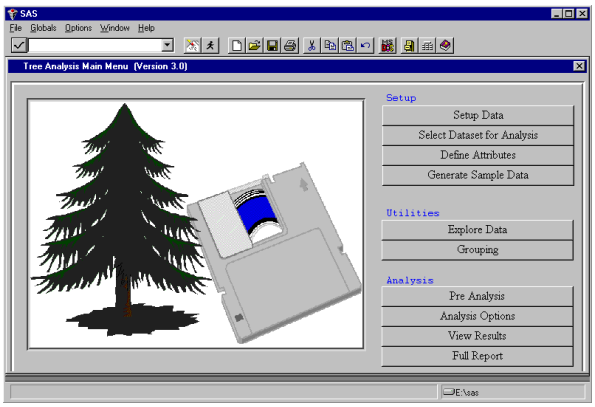

Figure 5. SAS Institute's data mining solution provides a point-and-click, easy-to-use GUI interface to guide a user through the SEMMA approach of tree-based modeling.

Neural networks and tree-based models are both effective data mining tools, however they each have their own unique strengths. Both have advantages over linear models by being able to detect nonlinear relationships automatically. But they have different virtues when it comes to making predictions from a large number of predictor variables. Tree-based models are good at selecting important variables, and therefore work well when many of the predictors are irrelevant. Neural networks are good at combining information from many predictors without over-fitting, and therefore work well when many of the predictors are partially redundant.

Traditional Statistics

Data mining includes a variety of traditional statistical methods, such as cluster analysis, discriminant analysis, logistic regression, and time-series forecasting. Also included are the more general methods for model fitting and validation, which conduct automated searches for complicated relationships and apply fresh data to tentative relationships. Noteworthy here is the unique multiplicity of tools available in the SAS System for performing these tasks.

Visualization

Data visualization software is one of the most versatile tools for data mining exploration. Appropriate for both technical and business end users, this technology offers immediate, graphical identification of patterns and trends. The SAS System includes two powerful data visualization software products—SAS/INSIGHT® software for graphical statistical visualization and SAS/SPECTRAVIEW® software for volume visualization of large amounts of data.

Under pressure of intensifying competition and shorter product life-cycles, many companies are implementing data mining and database marketing applications with

the SAS System. They are discovering that a commitment to learning about customers' needs can add value to their products and services.

The SAS data mining solution builds on the proven strengths of SAS software for interrogating large volumes of complex data. In fact, the SAS data mining solution becomes all the more suitable for data mining as data complexity increases.

Integration with SAS software gives the SAS data mining solution several distinguishing characteristics which allow faster, easier, and more accurate conversion of data into knowledge useful to decision makers.

SAS DATA MINING IN ACTION: A CLOSER LOOK

Oxford Health Plans

Background and Business Problem

Oxford Health Plans is one of the largest health care providers in the country. It provides care services and coverage for about 1.5 million members. Oxford Health Plans serves five northeastern states and is one of the fastest growing companies in the industry, adding about one-half million members per year. In addition to its large business group segment, Oxford has one of the largest medicare plans in the country with over 98,000 seniors enrolled.

The highly volatile environment of today's health care market drives companies like Oxford Health Plans to put a premium on delivering high-quality care and services to its members. The competitive nature of the industry creates a real moving target in terms of with whom you work, how to measure effectiveness, quality, and how to demonstrate quality to members, employer groups, and regulators.

Data in the health care industry are often very massive and complex. Oxford Health Plans is no different, with data in excess of several hundred gigabytes existing across the enterprise in a wide range of sources and systems. Two years ago, Oxford developed a data warehousing strategy to handle the growing amount of data and information. In such an environment of "hyper-growth," problems often exist in analyzing such large amounts of data. Data mining tools and techniques offer ways of creating complex data models to make sense of information and best reach out to members. These technology-based tools help Oxford better understand and raise the efficacy of care and services, allowing for the delivery of the best possible care.

Data Mining Goals

Data mining goes well beyond queries and is more than a series of reports. It includes tools for visualizing relationships in data. At Oxford, data mining is used to help answer complex issues revolving around patient utilization. They want a better understanding of member utilization of care and services, as well as patient and member preferences of how, who, and where they like to deal with care. What are the key variables in predicting utilization, improving quality outcomes and improving compliance issues? To stay competitive, Oxford must empower hospitals, physicians, and members with information so they can interact with each other effectively to deliver high-quality care. Data mining tools are useful for understanding and improving physician and member satisfaction goals. To make this happen, information must be delivered and presented in a manner that's easy to understand and use.

The SAS Solution

At Oxford, the typical user is a physician or executive with a non-IT background. Oxford's goal is to create a system that is dynamic and interactive, allowing the user to ask "what-if" questions in a real-time, easy-to-use environment.

The SAS data mining tools and methodology create a framework and range of capabilities that Oxford needs to create an interactive data mining system. The system allows them to integrate all areas to solve problems rather than just to create applications. The SAS System's strengths for data mining at Oxford include:

- Ease of use for generating complex samples and analysis.
- Complete support for all types of statistical and analytical techniques including visualization tools, traditional statistics, neural networks, induction, advanced modeling, etc.
- Ability to pull different streams of analysis together to complete the feedback loop.
- Data management capabilities, including the ease of accessing differing data sources and operating systems.
- Scalability.

Oxford Health Plans uses SAS data mining tools in many ways. For example, they may want to evaluate a new communications campaign to improve asthma compliance. By analyzing interactions or "touches" between physicians, hospitals, employees, and members,

Oxford can better understand the most effective means of communication to improve the efficacy of these touches.

The biggest payoffs of data mining for Oxford Health Plans are in leveraging vast data resources and information to maximize quality of care and improve satisfaction levels. An effective approach can help in determining the proper medical techniques to facilitate "right actions." Oxford wants to empower their physicians to improve quality of care through the use of technology. The system allows Oxford to present information in a manner and format that most effectively compels action from the physician. This system is used to rapidly extract information for slicing and dicing data interactively from different perspectives, and then to present information effectively to motivate and entice changes in behavior. Oxford wants to feed the physician information in a way that they feel comfortable and confident about its use to improve quality.

At this time, SAS data mining efforts are being supported by two groups within Oxford, the artificial intelligence group (AI), which works with neural networks and expert systems, and the statistical group. Oxford Health Plans takes advantage of many other SAS System capabilities as well such as the interactive applications development environment, the SAS multi-dimensional data base (SAS/MDDB™) and Scalable Performance Data Server™ (SPDS). SAS/MDDB software allows data to be viewed in various dimensions efficiently. The SPDS enhances response times when accessing, sorting and querying large data views.

Benefits to Oxford

SAS software was chosen at Oxford Health Plans because, quite simply, it is viewed as "strategic technology." In the long run, higher quality of care will result in lower cost. Oxford has a philosophy of doing it right the first time. They want to give physicians better, more useful information, so that they can become better doctors.

Future of Data Mining at Oxford

SAS software is the data mining technology tool of choice that helps Oxford Health Plans gain a competitive advantage in their market place. It is used by the highest levels of management for strategic decisions. As their market share and associated business needs continue to grow, so will the future for data mining applications. Areas of growth include:

- Continuing to use data mining in sales and marketing.
- Exploiting and moving data mining to the clinical side to apply the tools for identifying higher risk patients.
- Empowering more people across the enterprise who are not trained and/or educated in information technology or statistical techniques.

SUMMARY

The effective use of information and technology is crucial for health care organizations to stay competitive in today's complex, evolving environment. The challenges faced when trying to make sense of large, diverse, and often complex data source are considerable. In an effort to turn information into knowledge, health care organizations are implementing data mining technologies to help control costs and improve the efficacy of patient care. Data mining can be used to help predict future patient behavior and to improve treatment programs. By identifying high-risk patients, clinicians can better manage the care of patients today so they do not become the problems of tomorrow.

The SAS data mining solution offers tools and techniques for extracting, visualizing, modeling, and delivering information to make informed business decisions for gaining a competitive advantage. Leading health care organizations such as Oxford Health Plans have successfully implemented a SAS data mining solution to better understand key indicators involving member utilization, quality outcomes, and encounters of care. Those health care organizations that are successful at leveraging data mining technologies will improve quality, control cost, and produce measurable results of increased return on investment.

REFERENCES

Joyner, Ellen, Rogers, Gregory, (1996), "Health Care Transformation: The Role of Technology," *Proceedings of the Twenty-first Annual SAS Users Group International Conference*, pp. 1372-1380.

Lawrence, D. MD (1994), Plenary Session, Institute for Healthcare Improvement's 6th Annual National Forum on Quality Improvement in Health Care, December 7, San Diego, CA.

Rogers, Gregory, S. Luker, G. McBride, G. Buckner, C. Boldenow (1995), "Managing Information for the Delivery of Quality Health Care," *Proceedings of the Twentieth Annual SAS Users Group International Conference*, pp. 792-799

Sarle, W.S. (1994), "Neural Networks and Statistical Models," SAS Users Group International: *Proceedings of the Nineteenth Annual SAS Users Group International Conference, pp. 1538-1550.*

SAS Institute Inc. (1995), European Inform, The SAS System Journal, Issue 16.

SAS Institute Inc. (1995), SAS Communications, "*Data Mining Reveals Diamonds In Your Databases*" Third Quarter.

SAS Institute Inc., SAS Data Mining, Finding the Diamonds In Your Data, *http://websrv03.unx.sas.com/www-prod/feature/4qdm/intro.html.*

SAS Institute Inc. (1990), SAS/STAT User's Guide, Version 6, Fourth Edition, Volumes 1 and 2 Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990), SAS/GRAPH Software: Reference, Version 6, First Edition, Volumes 1 and 2 Cary, NC: SAS Institute Inc.

Stratton, Scott, Director of Medical Analysis, Oxford Health Plans, personal interview (November 1996).

Thearling, Kurt (1995), "From Data Mining to Database Marketing", Data Intelligence Group.

Ultragem, Web Examples Of Data Mining Projects To Stimulate Your Imagination, *HTTP://www.ultragem.com/projxmpl.htm.*