

Predicting Return to Work with Data Mining

IN THE GOOD OLD DAYS (20 years ago or so) most long-term disability (LTD) claims were straightforward. The path to recovery or non-recovery was based on physical factors and seemed almost predictable. No longer.

Psychological factors play an increasingly larger role in long-term disability; from 1990 to 1998, the incidence rate of mental-nervous claims almost tripled. New diagnoses such as *fibromyalgia* (a loosely defined aching of the muscles, tendons, and ligaments) have appeared on the scene. Clinical depression has become an accepted, treatable disease. Staying home sick from work no longer carries the stigma it once did.

Early Intervention

With the increasing importance of mental health as a factor in long-term disability, and the increasing number and complexity of diagnoses, early intervention is becoming more and more important to claims management.

Early intervention can offer claimants the attention and assistance they need, at the time they need it, to make a successful return to work. It can also prevent the slide into the disability mind-set—the claimant’s feeling that staying away from work is inevitable, and return to work is impossible.

The Claims File

One of the major difficulties in practicing early intervention is the heavy demands it places on the claims management department. In a perfect world, every claims manager would have infinite time and resources and 100 percent objectivity. The reality, of course, is different.

LTD claims files are bulky. Some, in hard copy, are several inches thick and tend to thicken over time. Few, if any, are easy to peruse or review.

When a claims file is first presented to a claims manager, it will already contain a number of documents:

- Employer’s statement
- Employee’s statement

- Job description
- Income information
- Attending physician’s statement

- Insurer’s forms
- Over time, more will be added, such as:
- Supervisor’s comments
- Doctor’s reports
- Prognosis letters
- Test results



- Correspondence between insurer, claimant, employer, doctors, lawyers, rehab providers, and so on
- Telephone transcripts of discussions about the claim

Claims managers have to plow through a great deal of information before deciding which of their several hundreds of claims will receive priority and attention that day. Should they suggest a payout for Martha B.? Order an independent medical examination for Jerry T.? Call Jacob Z. again to monitor

his progress? Hire an investigator to check out Brenda B’s suspicious-sounding bad back?

Claims Scoring

We developed a data-mining technology to create a scoring tool to help claims managers make the best choices available and to optimize their use of available time and resources. Our system scores each claim with a number from one to 10, based on likelihood of return to work within a given time frame, usually six, 12, or 24 months.

This score gives claims managers a quick, initial assessment of each claim file they manage. It helps them to optimize resource allocation and to decide where to focus their time, energy, and available support. By offering a classification for each claim, it also helps the claims supervisor to allocate caseloads effectively among claims managers.

BARRY SENENSKY is president and **JONATHAN POLON** is vice president of Claim Analytics, a data-mining company in Thornhill, Ont.

Data Mining

Most statistical tools used in business predate the invention not only of computers but of the electric light and automobiles. These tools were designed with a now-antiquated constraint: They could only require calculations that could be done by humans in a reasonable amount of time, with the amounts of data available.

With the advent of modern computing, two changes occurred to stretch the boundaries of statistical analysis. First, computers could compute with incredible speed. Second, they began to gather and store vast amounts of data. The setting was ripe for the birth of data mining.

Data-mining tools are uniquely suited to dealing with sixth-sense problems, problems that defy the capabilities of traditional if-then-else programming. What data mining does is to rumble, wide-eyed and open-minded, through the data. By drawing empirical observations, based on historical behavior, data-mining tools succeed at accurately predicting human behavior.

Neural Networks

A neural network is an optimization tool that works to develop a weighted algorithm. The neural network starts by assigning random weights to a random series of inputs, and predicts a result. It then compares its predicted result to the actual result. If the predicted result differs from the actual result, the neural network will adjust its weights to better approximate the actual result.

In this way, neural networks are similar to humans; they “learn” by comparing known inputs to known outcomes, over and over again, until they can detect a pattern. Unlike humans, they can consider and weigh many factors simultaneously.

This process gives neural networks the ability to deal with an object as complex as a claims file. The neural network doesn't need to be programmed to “understand” the file, and go through a logical series of if-then-else steps to place it in a particular category. It simply compares groups of known inputs (age, gender, diagnosis, occupation) to the known out-

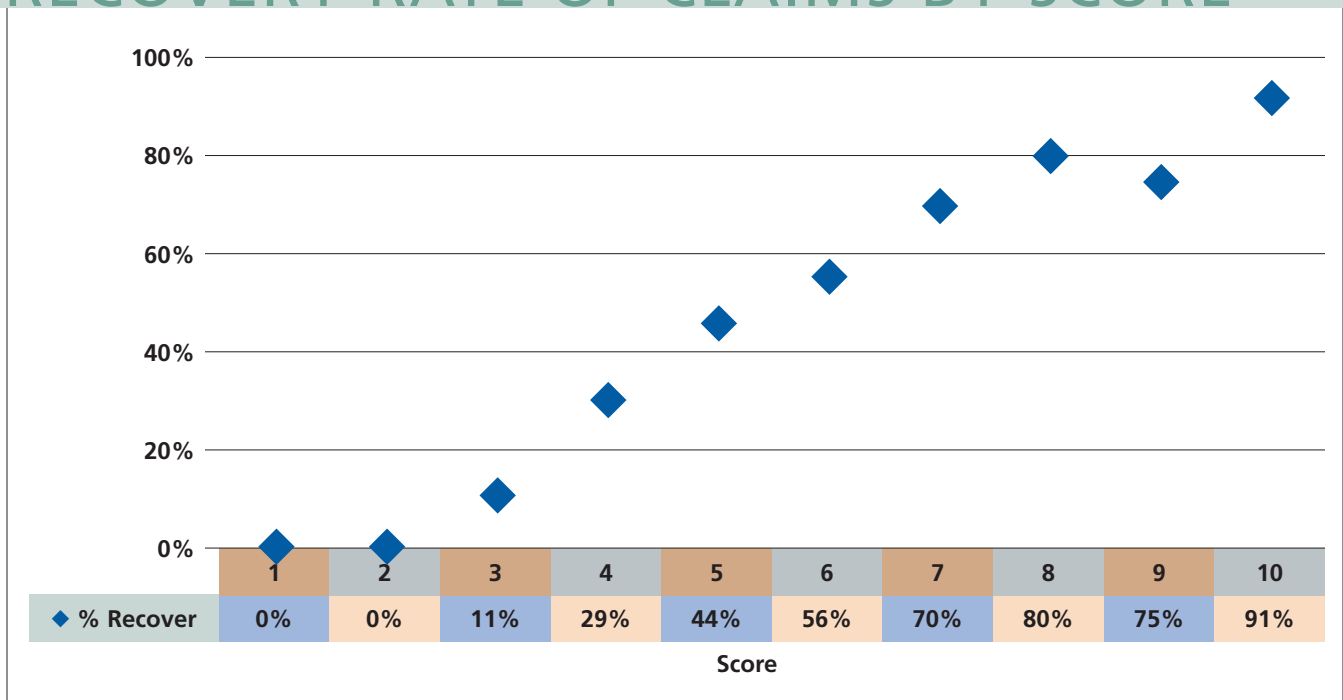
come, in order to detect patterns, some of which are very subtle indeed.

Does Data Mining Work?

As with any ground-breaking new technology, data mining has attracted a coterie of detractors who say that the answers it delivers are too complex, that the language of delivery is arcane, and that the solutions are impractical.

Many of these criticisms are representative of poor applications of the new techniques, rather than of the techniques themselves. Others are simply wrong. Certainly, in terms of economic payoff, data mining is posting some head-turning returns. A recent study reported in *ZDNet Australia* found that: “Among North American and European companies, the successful introduction and use of analytics has delivered returns on investment of anywhere from 17 percent to 2000 percent, with a median ROI of 112 percent. These kinds of numbers make a compelling argument for business analytics, and data mining is a significant part of those analytics.”

RECOVERY RATE OF CLAIMS BY SCORE



The rising line of blue diamonds in the chart shows the results from the blind test of the model. The model's predictions aligned very closely with real-world outcomes.

One article on data mining, published in a newsletter of the Association for Computing Machinery, shows just how much data-mining techniques have already penetrated the industrialized world. It describes a day in the life of John, an average American:

... John starts his day with the morning paper, which has improved over a few years ago; it uses a data mining process control method to maintain ink quality while maximizing print speed and minimizing cost. John pops his morning hair-loss pill as he reads. Although John's pill was first developed to treat prostate enlargement, it was recently approved by the FDA for treatment of hair loss. Major drug companies have been applying data mining techniques to patient information to discover beneficial side effects of drugs.

As he drives to work, John's car uses an onboard neural net to monitor information from his engine, exhaust, and fuel systems. Once he arrives, his neural net-based software helps him

find opportunities for cross-selling his products. Neural nets even help him find a birthday present for his mother, by suggesting books she may like. And when he uses his credit card to buy online, a neural net fraud detection program runs silently in the background, checking that he is, indeed, the owner of the card.

Building the Claims-Scoring Model

The one-page summary that forms a sidebar to this article gives an overview of how the claims-scoring model was initially built. A more detailed examination of the process can be found on the Claim Analytics website, in a case study about building the model sponsored by the Society of Actuaries.

One of the greatest challenges in creating such a model is to prepare the data in a way that allows the neural network to actually compare cases on an apples-to-apples basis. More than 400 different diagnoses were involved in the dataset used, some of which were very rare and offered few opportunities for comparison.

Creating one field for diagnosis, with 400 different possible inputs, would not have allowed the neural network adequate opportunities for comparison.

Accordingly, the "diagnosis" field was broken down into a number of subfields. This gave the neural network a far greater number of opportunities to compare. Similarly, repeated experimentation revealed that the "age" field was better classified as a categorical (non-numeric) variable and that there would be only three age categories (18-35, 36-50, 51-65). The neural network derived much greater meaning from the absolute differences found in the three age categories than it had from comparing the relative values of the ages in just one age category.

Results

The proof is in the pudding. The most interesting discovery the model made was that it could accurately predict the likelihood of recovery. Upon being declared ready for final testing, the model was fed the input fields for a data set of historical cases whose outcome was known. The

HOW THE MODEL WAS BUILT

DEFINE THE GOAL

We decided that our model would predict likeliness of recovery within a given time span rather than time to recovery. As well as making this decision, we decided how we would evaluate the completed model.

DATA REQUIREMENTS

We started by trying to get a general idea of which factors influence recovery. Then we had to precisely define recovery, a surprisingly tricky task. We had to decide how many records were needed to create the model, and ascertain that the necessary data fields and records were available.

DATA PREPARATION

We first split the data into three parts. The largest, 80 percent of the data, was used for training the model. We also set aside 10 percent for testing and 10 percent for final validation.

We then checked data quality, and reviewed the data fields with a

knowledgeable user. Finally, we had to devise a way of transforming raw medical data into quantifiable terms the model could understand and evaluate.

CART

We used Salford Systems CART®—a decision-tree tool that automatically sifts large, complex databases, searching for and isolating significant patterns and relationships—as an initial filter to key in on which data factors affect recovery most.

BUILDING THE NEURAL NETWORK

Much more data preparation goes into preparing the data for a neural network than for CART®. For example, we modeled "age" in three ranges, and used fuzzy logic for the edges of the ranges. That is, we put claimants in three age buckets—18-35, 36-50 and 51-65—but used fuzzy logic on the bucket boundaries so the model would recognize the similarity between a 35-year-old and a 36-year-old claimant.

NEURAL NETWORK TRAINING SETTINGS

Choosing neural network training settings is akin to making design decisions for the model. Choosing the settings is an iterative process; it takes time and a great deal of trial-and-error testing.

DETERMINING THE BEST NETWORK

The experimentation involved in choosing training settings required training a neural network for each combination of settings selected. We then determined the best settings by comparing the various networks we trained, using measures such as percentage of records correctly scored, R-squared, etc.

GENETIC ALGORITHMS

When we had finished training our network, we decided the results were not adequately precise; too many scores were lumped in the middle. We decided to try the "Genetic Training Option" of our neural network software.

rising line of blue diamonds in the chart on page 69 shows the results of the blind test. A higher score (score shown on x axis, rate of recovery on y) was clearly linked to a higher rate of recovery. The model's predictions aligned very closely with real-life outcomes.

Different Actions for Different Scores

The score can be used as a factor in creating an overall policy for managing claims. For example, claims scored in the 4-to-7 range might be afforded the highest level of time, attention, and resources. Claims scored between 8 and 10 would require a certain level of monitoring to ensure that nothing stood in the way of the claimant's recovery and return to work. Claims scored in the 1-to-3 range could receive a similar level of attention to those scored between 8 and 10.

One of the most powerful features of the scoring model is its ability to distinguish between gray-area claims, those claims that are neither particularly promising nor particularly unpromising. This is often a very difficult task for a claims

manager. Yet it can be seen in the chart that the model differentiates, clearly and accurately, in the 4-to-7 range. "Fives" were more likely to recover than "fours," "sixes" more likely than "fives," and so on. This differentiation may offer the claims manager assistance in deciding which claims in the 4-to-7 range would most benefit from extra time and attention.

One route that can be (and is being) taken by claims departments is to focus most attention on claims scored between four and seven and, within this group, to start with the claims scored with a six. These claims have a good potential for return to work, but may not do so without the proper assistance, resources, and encouragement. This is where an intensity of intervention could offer the greatest rewards.

Conclusion

Claims management is neither a simple

nor a routine task. Yet it forms an appreciable portion of insurance activity for most large insurance organizations, and many smaller ones. A means of assisting claims managers to improve return to work could be highly beneficial. The predictive powers of data mining can now offer assistance in this complex, critical area.

References

- <http://www.claimanalytics.com/papers.htm>
- <http://web.zdnet.com.au/techcentre/storage/features/story/0,2000054002,20275647,00.htm>
- <http://www.acm.org/sigkdd/explorations/issue1-1/john.pdf>
- <http://www.claimanalytics.com/papers.htm>

Genetic algorithms, while similar in function to neural networks, use different methods: neural networks "learn" from data, while genetic algorithms "evolve" to a solution. They can be used together, combining the strengths of both. (Genetic algorithms are capable of global search and are not easily fooled by local minima; neural networks can apply gradient descent to find the minimum point on the error curve.) Using genetic algorithms, we produced a model we felt was ready for final validation.

VALIDATING THE MODEL

We validated the completed model by comparing the model's predictions for the validation data with the real outcomes. The results were excellent. The model's predictions of return-to-work behavior aligned closely with actual outcomes.

R² The Reinsurance Resolution
A life & disability income reinsurance administration system

**Reinsurance Administration
Never Looked So Good!**

Meets or exceeds Guidelines of the SOA Reinsurance Section's Administration Committee for printed reports and magnetic file transfer.

QUASAR SYSTEMS

8989 North Port Washington Road • Suite 227
Milwaukee, WI 53217-1662
(414) 228-8622
www.qsi-r2.com