Using Artificial Intelligence to Predict High Cost Workers' Compensation Cases

Neil Smithline, MD and Nora Blay, MBA
January 2, 2001

Abstract

Applying their years of experience in outcomes research, the authors and a team of researchers sought to develop a scientific predictive model to be used to reduce the cost of Workers' Compensation injuries. The purpose of the predictive model is to identify, within a few days of injury, those workers' compensation (WC) cases destined to become high cost claims. Early identification makes it possible for corrective actions to be initiated within a time frame that would reduce the escalation of costs.

The team applied an advanced form of neural net technology known as Artificial Life Networks (AL Nets) to a WC database of 1.5 million administratively closed claims. For purposes of this discussion, the model will be called "ALpha Predictor."

To test the validity of the ALpha Predictor, the analytic team ran a horse race, comparing it to the industry's current best business practices. ALpha Predictor excelled by 70 percent.

The final step was to translate this highly predictive model into a business tool that assists claims organizations in allocating resources to the 12 percent of claims that drive over 80 percent of the losses. The team determined that the Internet would be the best engine for this approach and incorporated ALpha Predictor into Internet-based software. With its rules-based object-oriented programming, the software automates best business practices, such as referral and transfer of information to appropriate individuals, so that case management, ergonomic, safety, and other necessary activities can begin immediately—to control medical and indemnity costs before they spiral out of control.

Neil Smithline, MD, FACP has served as HealthTies' President/CEO since founding it in early 1999. In the past, Dr. Smithline founded and was president of Current Health Concepts, Inc., was vice president and co-founder of The Peabody Group, maintained a private internal medicine and nephrology practice for 13 years, co-founded and was CEO of the Dialysis Foundation Southern Arizona, and published the *Physicians' DRG Newsletter*. Dr. Smithline received his BA and MD degrees from Tufts University. He is a fellow of the American College of Physicians and has published extensively in the health care field.

Nora Blay, MBA is a co-founder and vice president. Prior to joining HealthTies, Ms. Blay has worked directly with the state's major workers' compensation insurance companies as a consultant to the Pacific Business Group on Health (PBGH) and as an officer of William

M. Mercer, Inc. Ms. Blay was formerly director of Planned Parenthood in San Francisco. She received her BA in English and Spanish from the University of San Diego and her MBA in Health Services Management from Golden Gate University, San Francisco.

HealthTies.com
100 Woodward Avenue
Sausalito, CA 94965
415 331 7407

Using Artificial Intelligence To Predict High Cost Workers' Compensation Cases

Neil Smithline, MD and Nora Blay, MBA
December 27, 2000

Introduction

In 1999 the Workers' Compensation Insurance Rating Bureau issued a bulletin indicating that California WC insurers are $5.4 billion under reserved—an entire year's premium—and with loss ratios exceeding 140 percent, this is a clear call to action. A large portion of this loss is due to the industry's inability to effectively identify potential high loss cases—before those losses occur. To assist the industry, we applied the power of artificial intelligence to an extremely robust workers' compensation (WC) database to develop Internet-based software that is designed to reduce these losses and allow payors to aggressively manage claims in today's highly competitive climate.

Getting From Here to There

Developing A Predictive Model

Our goal was to develop a highly accurate predictive model that would identify, within a few days of injury, those cases destined to become high cost, so that corrective actions could be taken to *prevent* these cases from becoming high cost. Drawing on years of experience, and a WC database of 1.5 million administratively closed cases spanning over six years, we began by creating three random data samples: 30 percent of the database used to develop models, another 30 percent used to refine the models and 40 percent for validating models. All cases were classified into one of four class cutoff categories based on final medical and TD costs.

- Trivial– up to 75th percentile of cases
- Moderate– from 75th to 90th percentile of cases
- Serious– from 90th to 97th percentile of cases
- Extreme– at or above the 97th percentile

Building on its previous experience applying classical statistical approaches to WC data modeling, the company identified correlations between various predictive elements and outcomes. As is typical when working with large databases, there were many employee, employer and injury characteristics that correlated with final cost at very high levels of statistical significance, but had only weak correlations. As a next step, three different types of models were developed.

1. Least Squares Regression: prediction to the class cutoffs that defined the case types stated on a logarithmic scale.
2. Logistic Regression: Each case was specified as Trivial, Moderate, Serious, or Extreme by using the class cutoffs. Separate logistic regression models were then fit for each of these indicators.
3. Gamma Distribution: Generalized linear models were fit with an underlying gamma distribution to allow for the observed positive skew of cost distribution.

Not surprisingly, the statistical models showed that there is no single claim characteristic at the time of claim set up that by itself predicts final cost: there are many, many factors that drive the ultimate outcome of a claim, each contributing a small amount. Because of these findings, we moved to a second type of modeling that uses artificial intelligence and is known as Artificial Life Networks or AL Nets, a form of neural nets.

AL Net technology mimics the brain's own problem solving process. Just as humans apply knowledge gained from past experience to new problems or situations, an AL Net takes previously solved examples to build a system of "neurons" that make new decisions, classifications, and forecasts. The type of neurons that AL Nets use is much more powerful than standard neural networks. These nets are programmed using a process called training.

Both AL Nets and standard neural nets use neurons to make up the software brain. A single neuron is illustrated mathematically in Figure 1. Inputs into these neurons (such as age, length of employment, etc.) are then translated into a single number, the output.

Figure 1. An AL Net Neuron

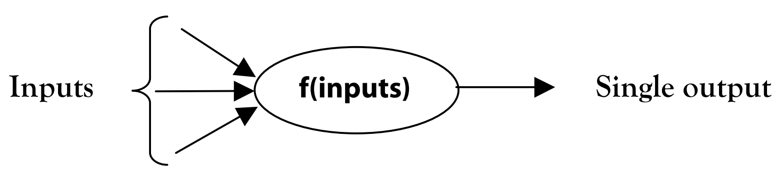
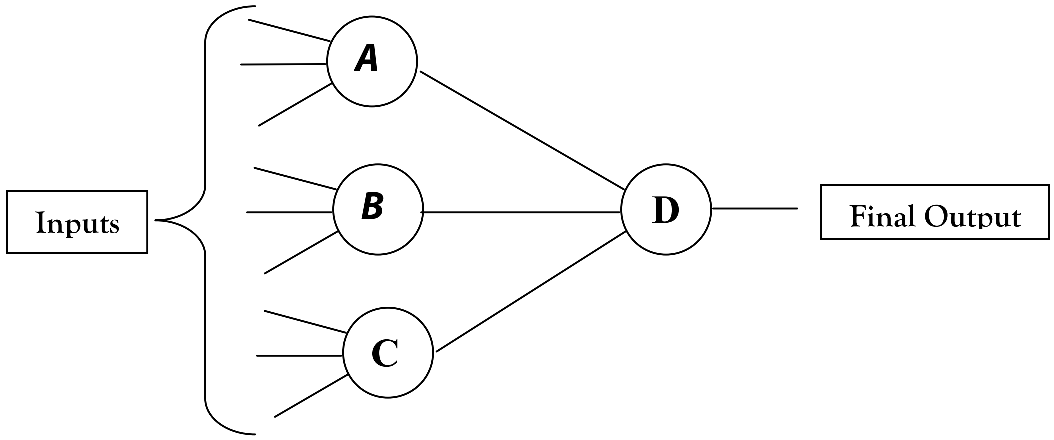


Figure 1 illustrates three inputs to the neuron. In practice there can be as many inputs as are required. Regardless of the number of inputs, there is just one output. The final model considered 30 such inputs, but only one output. To make a brain, many neurons are tied together as shown in Figure 2, which is an example of a four-neuron brain.

Figure 2. A Four-neuron Brain.



In the model shown in Figure 2, the inputs from the training data go to neurons A, B and C. The outputs from A, B and C are tied into neuron D. Finally, neuron D translates these numbers into the final output, which in the model is the predicted cost of the case at the time it is administratively closed. To train a brain, we take a set of input data where we know the outcome (final cost). The input data are the employee, employer and injury characteristics known at the time of claim set up. This input data is fed into the AL Nets' brain and the predictions are then recorded. Each prediction is then compared with the known outcome and a reward/punishment scheme is applied to train the brain. The top 10 percent of predictors are retained and through random mutation and breeding between brains we create 100 offspring. These offspring are then tested, and the top ten percent are retained and bred again. Using this genetic algorithm approach, the best predictors constantly beget better predictors. The process is continued until future generations do not substantially improve the prediction.

The Hybrid Predictive Model

The model integrated the best aspects of statistical and AL Net models into its predictive capabilities. This proprietary hybrid model, which we call ALpha Predictor was trained not only to identify the most expensive cases, but also to maximize potential savings from claim management interventions. A final estimate of precision of the model was validated using cases from 40 percent of the database that had not been used to build or test the model.

Running A Horse Race

The ability to identify potentially expensive cases, before they become expensive does not necessarily translate into savings for the WC industry. To test the ability of ALpha Predictor to reduce losses, a business model was created to identify savings generated by triaging cases to case management or senior claims personnel. The model was a "horse race" that compared savings generated by ALpha Predictor, to that generated by triaging cases using the industry standard: a series of red flags that rely heavily on the employee being absent from work at the time of claim set up (the TD Rule). To build this business model, experts were polled at WC insurers, case managers, and TPAs throughout the industry. Using an iterative approach, the assumptions about potential savings were identified and refined (See Table 1) and then applied equally to the ALpha Predictor predictive model and to the TD Rule.[1] The actual horse race consisted of testing the two models against each other using data from the database that had not been previously used to create either model.

Table 1. Triaging Assumptions

1. Referral of potentially expensive claims to highly experienced personnel (claims examiner or case manager) saves money
2. As cases get more expensive, there are proportionately more medical and TD dollars to be saved.
   - Zero savings on cases below $3,000
   - 15% medical and TD savings at $3,000, ramping up to 30% at $20,000
3. 10% savings on TD on cases with net medical and TD savings
4. The cost of referral of a case to case management is $500.

How Much Savings Can The ALpha Predictor Model Generate

In an ideal world, it would be best to identify all cases with potential savings. Knowing at the outset that no predictive model would be perfect, we constructed a series of business scenarios to determine where the greatest leverage would exist in developing a triage system. Analysis revealed that 90% of potential savings are obtainable by identifying all cases less then $1,300 and greater than $13,000 (medical plus TD plus PD). In identifying these cases, ALpha Predictor was far superior to the industry standard TD Rule as measured by specificity and sensitivity (See Table 2).

Table 2. Predicting Sensitivity and Specificity

	TD Rule	ALpha Predictor
Sensitivity	53	83
Specificity	93	91

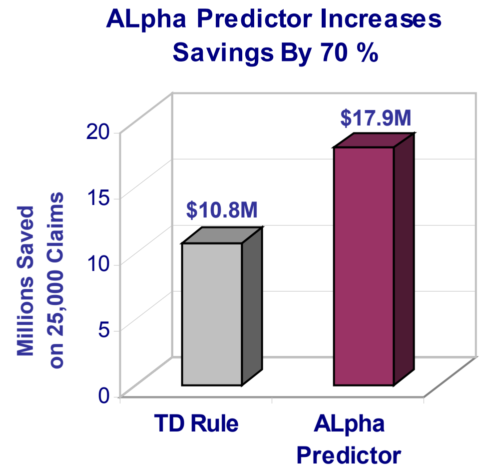
To dust off the cobwebs, sensitivity is the ability to identify these expensive cases so they can be referred to case management/senior examiner: ALpha Predictor clearly excels at this. Specificity is the ability to not mistakenly identify, and therefore triage to case management, the inexpensive cases, where the cost of referral outweighs any potential savings. Here the two systems perform at the same level.

---

[1] Because the identical assumptions were used for both ALpha Predictor and the industry standard, the relative differences between the two predictive systems will remain the same even if the assumptions vary.

From a business perspective, however, it is important to address not just the ability to identify cases within these broad ranges, but to identify those cases providing the greatest return. For example, a system that consistently identifies cases above $40,000 will provide far greater savings than one identifying cases predominantly in the $10,000 to $39,999 range, yet both might have identical specificity and sensitivity. In developing ALpha Predictor, the Al Nets allowed us to run various business simulations designed to maximize savings for the user, and the Al Net brain was trained to preferentially identify a very high cost case at the expense of missing one in the moderate range. This fine- tuning was extremely worthwhile, as shown in Figure 3, which illustrates the results of the horse race: applying the assumptions in Table 1, ALpha Predictor identifies 70 percent more savings than does the TD Rule. When applied to 25,000 cases, this amounts to $7 million.

Figure 3. Results Of The Horse Race



From Concept to Execution

The team knew there was substantial savings to be gained through modern data modeling. They believed that by unleashing the power of AL Nets on a robust WC database they could assist the industry by developing a highly accurate predictive model. But they also know that a scientific model that stood in isolation would not help the industry.

To be of true value, the model needed to be not only highly accurate, but also incorporated into technology that would be easy to use, fit transparently into the claims culture, and be compatible with existing software and systems. Therefore, we determined that using the Internet as the engine could best satisfy these conditions. The company has incorporated the ALpha Predictor model into software that will be available via the Internet to any licensed user with an Internet browser. With its rules-based object-oriented programming, this product is designed to automate best business practices, such as referral and transfer of information to appropriate individuals, so that case management, ergonomic, safety, and other necessary activities can begin immediately—to control medical and indemnity costs before they escalate.

###

For More Information…

The company, HealthTies™, has trademarked their Internet software, Compensa™, and is currently marketing it to insurance companies, TPAs, and claim systems. For further information regarding HealthTies and Compensa, please contact, Neil Smithline, MD or Nora Blay at (415) 331-7407.