



Building Predictive Models that Estimate Differentials for Auto Insurance Pricing

Building predictive models to estimate claim risk is gaining traction among Canadian Property and Casualty (P&C) insurers. Multivariate analysis (MVA) and regression techniques evolved from military applications (bombing missions) in World War II and have been used to predict a given outcome or behaviour ever since. They are widely used among lenders in the banking industry worldwide and have played a key role in predicting claim risk for U.S. and European insurers for the past 20 years, yet acceptance of the techniques among Canadian Actuaries has been slow.

While Actuaries understand the mechanics of these techniques, challenging data environments have limited their ability to apply them. Through the implementation of new policy management systems and the growth of data mining applications and experience within P&C companies, Canadian Actuaries have the opportunity to leverage advanced statistical techniques to improve the accuracy of existing rating structures.

In most cases, predictive models are built to create a measure representing the likelihood that a prospect or customer (in this case policyholder) will behave in a certain way. The model's "score" assigned to the individual represents the degree to which the behavior will occur. In the case of Auto insurance, the score may be a dollar amount representing the expected (claim) loss amount for a vehicle in its policy year. To bridge the gap between traditional Actuarial rating methods and predictive modeling applications, data mining practitioners within P&C companies are building models that predict *differentials* as opposed to an actual score.

In the rate-making process, there are a limited number of different characteristics that can be used to assess the claim risk for a vehicle. Some of these characteristics include geographic location, vehicle type, distance driven annually, previous claim history, etc. A portfolio of vehicles is segmented into groups based on loss performance for a specified characteristic or "rating variable" (e.g. # of years claim free driving – 0, 1, 2, 3, 4, 5, 6+ years) and are analyzed to determine how much each group differs from the average of the portfolio for that characteristic. For example, a given portfolio of vehicles may have an overall claim frequency of 2%. Within that group however, vehicles that have had a claim in the past year are found to have a claim frequency of 4%, or a *differential* of 200 ($4\%/2\% * 100$). On the other hand, vehicles that have not had a claim in the past 6+ years have a claim frequency of 1%, or a differential of 50 ($1\%/2\% * 100$). In this example, the premium rate for vehicles with a claim in the past year would be 4 times more than the premium for a vehicle with no claims in the past 6 or more years. This observation represents one of a number of different characteristics that Actuaries assess when trying to determine the most important variables and most accurate differentials that can be used to form a rating structure.

Boire Filler Group

from scores to differentials, the models yield output that meets regulatory expectations, and rate filing submissions are more likely to be met with approval.

The business benefit of using predictive models over the more traditional approaches is becoming more clearly accepted throughout the P&C industry. Organizations that refuse to adopt these techniques will become uncompetitive relative to other organizations that are taking advantage of these techniques. The continued evolution of the data mining discipline will simply enhance the adoption and use of these techniques. Using these techniques to produce output such as differentials will further accelerate the use of predictive models as a business standard throughout the P&C industry.