

Unlocking value with predictive modelling

Mr Martin Houle of **RGA Reinsurance Company Middle East Limited**

looks at how life insurance companies can use predictive modelling to optimise internal data, and the various techniques that may be utilised.



Life insurance companies around the world have sizeable and growing portfolios of policyholder and experience data. For the past several years, this data has been leveraged to build sophisticated predictive models that have sharpened mortality and morbidity forecasting, strengthened pricing development, and enhanced underwriting and claims assessment capabilities.

Today, companies are taking predictive modelling a step further: using these models to leverage internal data assets, supplemented as needed by data from third-party providers and/ or their reinsurers, to enhance risk assessment and segmentation for marketing purposes.

Specifically, customers are being targeted by these models who would be most likely to qualify for, benefit from, and act upon cross-sell or upsell opportunities. Those customers would then be offered supplemental simplified- or guaranteed-issue policies.

Many uses

For life insurance companies, the successful use of predictive models depends upon several aspects: quality of data, access to model-building expertise, solidity of the implementation framework, and organisational support. Many companies are also currently upgrading their data collection, storage and management capabilities, moving hard-copy data assets to electronic formats for effective access and use in predictive modelling.

A myriad of opportunities exist for insurers to employ predictive modelling. A few examples include:

- **Cross-selling/ Upselling**
For a company with lower penetration rates across products lines, predictive models can be used to in two ways. “Propensity to buy” models can be developed specifically to target customers more likely to purchase particular products, resulting in lower marketing costs and less wasted effort. Predictive models can also be structured to identify the best risks and offer simplified underwriting, which can increase take-up.
- **Pricing**
Predictive models can also suggest pricing structures that can allow rate improvements for the best risks in order to attract these customers.
- **Experience analysis**
Predictive models can identify and analyse a company's most important experience factors in a true multivariate setting; i.e. considering all factors and their interdependence.
- **Claims fraud mitigation**
Identifying and fighting fraudulent claims is both

time-consuming and costly. Predictive models can help by identifying the claims most likely to be fraudulent. Claims fraud predictive models can also take a comprehensive view, examining and identifying patterns in the aggregate.

- **Persistency management**

A predictive model focused on persistency can be combined with a customer lifetime value model to improve retention among the most profitable customers.

A comprehensive process

For a life insurance company, the building of effective predictive models involves high levels of collaboration for a myriad of functions, from pricing and underwriting to IT, data scientists, business development, senior management, and from the company's reinsurance partner as well.

The model development process should incorporate the following steps:

- **Clearly define the model's purpose**
A company should have a clear understanding of why a model is being developed, what it seeks to accomplish, and its potential applicability. In practice, this can be more difficult than it sounds. Without a clear focus, the modelling team may try to accomplish too much at one time.
- **Identify and collect supporting data**
Data can come from many different sources, but it should be reliable and relevant to the predictions a model is attempting to make. The identified data should include an outcome variable and all potential predictor variables to be tested. A successful model requires a sufficient quantity as well as quality of data. High-quality data will lead to more accurate models and results, but a large quantity of data will help models to recognise and predict outcomes. Gathering and identifying the right data can be challenging, but is crucially important in the process.
- **Create a model**
When building a model, choosing the right structure for the problem and the variables to analyse are crucial. Adding variables to a model or incorporating additional dimensions to the data typically improves model accuracy, but eventually the model will start explaining ‘noise’ in the data rather than the actual underlying relationships.
This problem is referred to as ‘over-fitting’ the model. This is a subtle problem that can have serious consequences as an over-fitted model is likely to have poor predictive performance when used on future data. Careful validation techniques should be used throughout

the building process to understand the potential ‘real-world’ performance of the model and to limit over-fitting problems. One commonly used technique to test a model for over-fitting is to break the data identified and collected for study into two segments, run the model for the first segment and then test the predictive performance of the model of the second segment.

- **Interpret test results and deploy model**

The modelling team and business development experts can work together to understand and interpret the model's results, and then work towards applying it for business needs. It is also important to consider any possible current constraints to the model, which can range from regulatory pressures to internal buy-in, sales force limitations, customer acceptance and others.

- **Deploy the model**

Many companies start with a pilot programme in order to test the model and understand its impact on the organisation. Sharing the results of these pilot programmes and demonstrating the model's lift are great ways to gain acceptance for the model.

- **Monitor results and update model as needed**

Once a model has been deployed, its performance should be tracked to ensure if it is working correctly and meeting the original objectives. Models may be periodically refreshed to stay current.

Methods

An abundance of techniques are available and applicable to the insurance industry. However, due to the unique financial aspects of the insurance business and data structure, we find certain tools are more applicable than others. A few selected approaches are described below:

Generalised Linear Model (GLM)

Powerful, versatile, flexible and broadly applicable, GLM is currently the most common predictive modelling approach used in insurance today. GLM is a natural extension of ordinary linear models, and is a way to unify other statistical models. The expected mean value of the target variable is a function (“link function”) of a linear combination of independent variables. The introduction of this link function makes it possible to include a non-linear relationship between target and independent variables. This approach has been used successfully to develop several types of predictive models, including models for underwriting, fraud and health pricing.

Decision tree

Another popular approach to building a predictive model is the decision tree. There are several decision tree models, and the most popular currently is the Classification and Regression Tree (CART) (CART models have been used to refine group disability pricing structures).

The main advantage of CART is that it produces intuitive, simple results. Users can grasp the essence of the CART model without involving complicated math. CART's non-parametric nature also means distribution does not need to be specified as one of the assumptions. CART-based predictive models can quickly reveal important data relationships, making it excellent for data mining. Primary concerns when using a CART model are its low efficiency in dealing with linear relations and its sensitivity to random noise.

Data clustering

Clustering algorithms organise data points into distinct groups, where the data in each group has similar distributions. This modelling approach is ideal for applications in classification, especially when a target variable is unknown or not certain. Clustering can be applied to the identification of effective advertising, market segmentation exercises and recommendation systems. In insurance, clustering is useful for in-force cell compression and scenario reduction, when detailed seriatim studies are needed or a large number of scenarios have to be simulated.

Other models

Many methods in addition to those listed above have been utilised under names such as business analytics, big data or data mining. Neural networks, for example, have been widely applied and could have great potential in insurance, but their black-box approach may limit their acceptance. Support Vector Machine (SVM), also known as support vector networks, is

another possible model. It is generally more accurate than other models, very robust to noise and less likely to have problems of over-fitting.

Choosing the right approach

When selecting an approach for predictive modelling, data scientists have a wide variety of methods from which to choose. Choosing the right one requires a flexible, judicious approach and an understanding of the needs the model seeks to address. In many business cases, the most advanced, sophisticated model might not necessarily be the best choice. Frequently, simpler models such as GLM can meet accuracy requirements, effectively address business needs and produce desirable results.

Many good approaches are available for insurance industry predictive modelling needs. Building successful predictive models requires access to quality data, and it's never too late to start building the systems that will collect and store data in the most accessible and usable fashion.

Better data – and better data administration systems – will lead to better models, and a thoughtful, considered approach to predictive modelling can lead to innovative solutions and unlock significant value for your company. 

Mr Martin Houle is Chief Actuary with RGA Reinsurance Company Middle East Limited, regulated by the DFSA.

Many good approaches are available for insurance industry predictive modelling needs. Building successful predictive models requires access to quality data, and it's never too late to start building the systems that will collect and store data in the most accessible and usable fashion.