Asian Journal of Managerial Science
ISSN: 2249-6300 (P) Vol.5 No.1, 2016, pp.30-40

DOI: https://doi.org/10.51983/ajms-2016.5.1.1198

An Improved Gradient Boosted Algorithms Based Solutions Predictive Model (Trade)

P.Senthil
Associate Professor in MCA Computer Science,
Kurinji College of Arts and Science, Tiruchirappalli, Tamil Nadu, India
Email: senthilkanapathy@gmail.com

(Received 17 February 2016; Revised 1 March 2016; Accepted 31 March 2016; Available online 8 April 2016)

Abstract - **In this paper, we describe a general process on how to integrate different types of predictive models within an organization to fully leverage the benefits of predictive modeling. The three major predictive modeling applications discussed in this paper are marketing, pricing, and GBA-Algorithm models. These applications have been well applied and published over the past several years for the Property and Casualty Manufacturing Industry, but this paper and discussions focused on their individual application. We believe that significant value can be realized if they are fully integrated, offering *manufacturing* companies the opportunity to take an enterprise wide view of managing their business through analytics. Therefore, the paper will discuss a general process on how they can be integrated and how the integrated result can assist insurance companies with managing the complex insurance business, such as minimizing the GBA(Gross Building Area) Algorithm cycle and achieving profitable growth and reacting to external market forces faster than their competition.**

I. INTRODUCTION

In recent years, predictive modeling has been widely used as a new strategic tool for manufacturing insurance companies to compete in the market place. Originally introduced in personal auto insurance to improve pricing precision [1], predictive modeling has been extended to homeowner's and small commercial lines as well [2]. Predictive modeling and the use of generalized linear models (GLM) have been individually applied widely in three key areas of insurance operations: GBA-Algorithm, Pricing, and Marketing. In this paper, we will discuss the value in integrating results from three traditionally distinct predictive modeling applications and the additional strategic and tactical benefits companies can achieve by taking an enterprise wide view of predictive analytics. Through the integration of predictive modeling results across multiple business operations, insurance companies can maximize their benefit and differentiate themselves in a competitive market environment where everyone seems to be using predictive modeling in some fashion. For instance, the integration of predictive modeling could enable existing GBA-Algorithm and marketing predictive model results to drive enhancements to pricing models and to align pricing with the GBA-Algorithm market cycle.

II. THREE TYPES OF MANUFACTURING INDUSTRY PREDICTIVE MODELING APPLICATIONS

In this section, we will discuss the similarities and differences as to how predictive models are built and applied to three different types of insurance business applications - GBA-Algorithm, Pricing, and Marketing. We will also discuss the data and modeling issues associated with each application.

III.IMPLEMENTATION PRICING MODELS

In predictive models for pricing, the main focus is on predicting loss cost, determining premium to charge, evaluating rating adequacy, or determining rating class plan factors. One typical result developed from a pricing model is a rating plan, which displays the rating variables, factors and loss cost relativities across the rating variables.

In developing the rating plans, actuaries often use the standard GLM frequency and severity approach, where the Poisson distribution is used to fit frequency data and the Gamma distribution is used to fit severity data. Recently, it has become more popular to combine the frequency and severity models into a pure premium model, where the Tweedie distribution, a Poisson – Gamma compound distribution, is used to fit the pure premium data directly.

A.Policy Level Variables

For pricing models, the source data files used to build the models need to be set up at a detailed exposure level. For example, for private personal auto (PPA), a pricing predictive model is generally set up at the vehicle and coverage level (i.e. – lowest form modeling data level). With regards to the rating variables, they are very different from one line of business to another, within the line of business, and can also differ from one coverage to another. Some complicated PPA rating plans may allow policy level variables across coverages and interaction between rating variables.

Perhaps, the most significant development for personal line rating plans in recent years is the usage of personal financial credit score [3]. Some states allow the usage of credit scores in class plans or tiering, others allow credit scores for GBA-Algorithm or target marketing activities only, while few states completely ban the use of credit scores. In addition to credit scores, other regulatory restrictions for pricing models include using not-at-fault accidents, capping the factors for youthful drivers or economic disadvantage territories, or enforcing forgiveness rules of prior years' loss and violation records, to name a few.

In the past several years, there has been a wealth of research, literature, seminars, and training classes in the Casualty Actuarial Society (CAS) community on using GLM to build pricing models [4,5]. Therefore, we will not repeat these theoretical discussions for GLM pricing models. Instead, we would like to discuss, based on our past experience, several typical data and modeling issues that arise when building the pricing models:

1. First, the commonly known data issues, such as missing data, miscoding information, information not captured in a insurance company data repositories, and unavailability of historical data due to purge, will hinder the development of predictive models.
2. Compared to personal lines data, commercial lines data posts an even greater challenge during the development of pricing models:
    a. Due to less regulation and scrutiny of commercial lines business operations, commercial lines data typically has much more commonly known data issues, as stated above, than personal lines data with regards to missing information, miscoding, and information availability.
    b. For personal lines, the exposure is well defined and fairly homogeneous: car-month for auto and home-year for homeowners. On the other hand, the exposure base for commercial lines is less defined and can even vary within the same line of business. For example, for General Liability (GL), some classes use sales and revenue for exposure, while other classes use payroll for exposure. Given the complexity associated with exposure, applying the pure premium approach for pricing within commercial lines is fairly difficult.
    c. For commercial lines, their data structure is heavily driven by rating bureau requirements. Therefore, the data is typically kept at the "industry class code" level, not at the exposure level. For example, for a commercial auto policy with multiple classes and multiple vehicles, the premium and loss information may be coded at the class level, but not at the vehicle coverage level.
    d. For commercial lines, more data credibility issues exist than they do with personal lines. Even for a mid-size regional personal carrier, it is fairly easy to collect millions of records for building up personal auto and homeowner's models. However, for commercial lines, there poses significant challenges regarding the availability of unique data points and it is very common that the data size is at least 10 times less than what is available with personal lines.

B.Business Reasons

In general, some major pricing variables are excluded in a company's analysis due to complex data structures, issues with data credibility, market competitiveness, or other business reasons. For example, "territory" and "vehicle symbol" are typically excluded from a modeling process of a PPA rating plan development. For these two variables, there exists many different values and therefore it is rare that a single company's data can provide fully credible data to evaluate these two rating variables. Another example for commercial lines is that most of the business, such as commercial Auto, GL, Property, Commercial Multi-Peril (CMP), and Workers Compensation (WC), will follow the industry class loss cost by ISO or National Council on Compensation Insurance, Inc. (NCCI). There exist hundreds of industry classes for each line of business. One way to appropriately consider their impacts on the model results is to adjust the exposure or pure premium by their indicated relativity. Another way is to use the GLM offset options, and this approach is discussed in a separate paper [6].

One data issue that needs to be considered for pricing model development is Catastrophic (CAT) losses for property lines, such as fire or hurricane loss, and extreme large losses for liability lines. Therefore, it is prudent to exclude CAT losses or cap large losses and then build the long term estimates for large loss loads or CAT loads back to the modeling data set.

For property coverages, the losses are net of the deductible. For liability coverages, the losses are capped by the liability limit. Therefore, we do not have the "complete" loss information to establish the entire severity distribution curve. This is a challenge in building up the severity models.

Another issue for building up the severity models is that for some of the segments in pricing, the severity data can be very thin and the modeling results can be extremely volatile with a great deal of "noise". The issue is significant for low frequency and high severity coverages, such as BI for PPA, and GL. This is why the pure premium models based on a Tweedie distribution have attracted more and more interest in recent years.

IV. GBA-ALGORITHM MODELS

The major business objective of an GBA-Algorithm model is to assess the risk quality for an insured on a prospective basis. One difference between GBA-Algorithm models and pricing models is that pricing models focus on determining the final class rates, while GBA-Algorithm models focus on evaluating risk quality beyond the class rating and the currently charged rate. The GBA-Algorithm models can assist Linear Model or product managers with their GBA-Algorithm decision making, such as company placement, crediting or debiting, limitation of coverage, payment plan selection, new business acceptance or rejection, renewal business referral and cancellation, and customer service and marketing activities. Regarding the modeling design, one difference is that pricing models use the pure premium approach at the exposure and coverage level, while GBA-Algorithm models use the loss ratio approach at a policy level.

Ideally, if a perfect rating plan exists, all risks are priced at their adequate rate level and there is no need for GBA-Algorithm models, or even GBA-Algorithm because generally speaking GBA-Algorithm models sit on top of pricing models and are designed to address pricing inadequacy through improved GBA-Algorithm precision. However, ideal rating plans do not exist due to various internal and external restrictions, including regulatory constraint, dynamic changes in the external economic environment, long delays for filing approvals, inability of using certain variables in rating plans, and limitation on rating structure (e.g non-linear pattern, interaction between rating variables, interaction between exposures at a policy level, etc.). Therefore, GBA-Algorithm models are used to evaluate the risk quality by identifying potential deficiencies in the rating plan.

The information used by Linear Model can vary widely and is sometimes highly subjective. Also, GBA-Algorithm actions are not always truly risk-based, but instead are influenced by the market, subjective decision making and external competition. This issue of a "market-driven" price is a more prevailing concern for commercial lines than for personal lines. Therefore, predictive modeling can be used to build up objective GBA-Algorithm models to assist Linear Model with making consistent and fact-based GBA-Algorithm actions each and every time and ensuring alignment with external market cycles.

Another advantage of GBA-Algorithm models is that the models can help insurance companies improve their GBA-Algorithm efficiency. This is because the models can segment "good risks" versus "poor risks", and with such segmentation, Linear Model can spend their major time and effort on poor risks, while good risks can flow through the process with minimum GBA-Algorithm touch. In addition, GBA-Algorithm models can be used to segment good and bad risks within classes of business, which is a significant improvement over traditional pricing and GBA-Algorithm decisions which are made on a class basis.

V.PREDICTIVE MODEL

In general, the target variable of an GBA-Algorithm predictive model is the loss and allocated loss adjustment expense ratio. Since GBA-Algorithm is mostly performed on a policy basis, the predictive variables and the data files used for developing an GBA-Algorithm model are at the policy level. For predictive variables, there are many more candidate variables: rating versus. non-rating, internal versus external, credit and territorial, among others. There is less restriction for GBA-Algorithm models than pricing models. For example, there is a trend in the industry with using insured's premium payment records from historical billing data, such as late payments and bad checks, as GBA-Algorithm variables. The trend of using billing information makes logical sense, since an insured's premium billing records are essentially a proxy for personal financial credit data and an insured's ability to pay bills on time.

FOR GBA-ALGORITHM MODELS, THE POTENTIAL DATA AND MODELING ISSUES ARE AS FOLLOWS:

1. Several data issues stated before for pricing model development are equally applicable to GBA-Algorithm model development, such as data quality and data availability and data completeness issues.
2. Many candidate variables can be included in GBA-Algorithm models that generally cannot be

included in pricing models. Creating and selecting the candidate variables demands a look at the availability of the underlying information, internal or external, to insurance companies and the ease of implementing these variables and gaining GBA-Algorithm acceptance on their use. Here are several examples:

While there is a trend with using billing information for GBA-Algorithm models, some companies may purge their billing data on a frequent basis; therefore, such information is not available in the historical data. Over a long term, companies need to devise a master data quality initiative to maintain and update historical data in their corporate data repositories to support these GBA-Algorithm models and devise mechanisms to ensure that these data elements are available to be extracted. The role of data quality and data governance as a key strategy to successfully maintaining and gaining value from predictive modeling applications is taking on even greater significance in the manufacturing Industry as more companies seek new ways to differentiate themselves in today's market.

A.Section Analysis

This paper is organized as follows. Section **I** describes the materials and methods. In this section the proposed GBA algorithm is presented. The methods for Analysis as well for categorization **III** and also the online Recovery part **IV** of the system and significance response **V** will describe in this section **VI**. In section **VII**, experimental results are shown **VIII**. The results are discussed is section **IX**, while conclusion is mentioned in section **X**.

B.Dataset

Another example is that some GBA-Algorithm information is kept on paper instead of in electronic files or in back-end data repositories. For example, for new business GBA-Algorithm, while many insurance companies ask for prior loss experience or other external data, such as motor vehicle records (MVR) for commercial auto, rarely do they store this information in their back-end data repositories. Therefore, it is difficult to use such information during the development of GBA-Algorithm models, even though it is common for Linear Model to use prior loss information in GBA-Algorithm new business.

1. When loss ratio is used as the target variable for modeling, we need to apply due actuarial consideration to adjust the data, such as rate on-leveling, loss development, and trending. By applying the appropriate actuarial adjustments, the Linear Model can have a higher level of confidence so that when they use the GBA-Algorithm model, the indicated results on the quality of the risk as derived from the model are based on up-to-date information with the appropriate longitudinal adjustments made.
2. Since GBA-Algorithm models are constructed at the policy level, whether the results can be carried, or how the results can be carried, to the underlying pricing, is a difficult question. For example, driver age is commonly used as an GBA-Algorithm factor even though it is used for pricing already. If an GBA-Algorithm model indicates that youthful driver policies are worse than average, it may not suggest that the underlying youthful pricing factors are wrong, but rather it may indicate the inadequacy of the pricing structure, such as purely multiplicative structure, or potential interaction of youthful drivers with other variables, such as vehicle type. The answer can be difficult to find without in-depth research and analysis.
3. Sometimes, GBA-Algorithm is not only performed on a policy level, but also on an account level. For example, it is very common for personal line carriers to cross-sell auto and homeowner's policies, and for commercial line carriers to cross-sell all the major small commercial lines of business, including BOP, Commercial Package, Auto, and WC. Therefore, the full value of GBA-Algorithm models may not be realized until they are built for all lines of business for account-driven companies and GBA-Algorithm models take a holistic view of assessing the quality of a risk.

I. MARKETING MODELS

The earliest, classical business application for predictive modeling is for marketing and sale operations, such as mail solicitation and response models. In general, the purposes of marketing and sales predictive models include identifying prospective customers, increasing the hit rate for solicitation, and assisting with retaining existing customers [7]. This is not for the manufacturing industry alone but historically predictive modeling has been used for marketing and consumer business related applications across multiple industries.

II.FOCUS

In general, the main focus of these marketing models is on the "success or failure" of converting or retaining a risk, so the target variable is typically a binary one. Whether the risk is profitable or not is not a consideration for these models but rather the probability that the risk will be acquired as a new policy or retained as a renewal policy.

III.USEAGE

Depending on the final usage of the marketing and sales models for insurance, there is wide variation in the types of models with regards to the predictive variables and the design of the target variable. For insurance applications, the marketing and sales models can be grouped into four main categories: new business qualification and targeting, new business conversion, renewal business retention, and renewal business conversion models. The details for these four types of models are as follows:

IV.MODEL ANALYSIS

For new business qualification and target models, the purpose is to identify a list of potential prospects for targeting. This list can be used for phone or mail solicitation campaigns. The data and variables used for the models are fairly limited, and are mostly from data sources external to insurance companies. There are numerous data vendors who sell consumer databases, and insurance companies can use the data for these models. Since there is a cost associated with the solicitation campaign, such as phone call cost or mailing postage fee, it is important to measure the cost versus return benefit, that is, the response rate, after the models are implemented.

V.CONVERSION MODELS

For new business conversion models, the key is to increase the new business hit rate when an insurance company has an opportunity to offer a quote to an insured. Insurance companies are very interested in knowing the overall hit rate, or conversion rate; for new business, how the hit rate varies by different segments of the book; and how to increase the hit rate. Many insurance companies do capture certain information in their insurance quote files, such as name, address, number of quotes, quoted prices, etc. For the conversion models, we can expect that one critical, if not the most important, factor that will influence the hit rate is how competitive the company's quoted price is compared to its competitors. The relationship between the hit rate and the quote price can be expressed through the "elasticity curve" commonly used for classical economic supply-demand theory. Without such price elasticity information, the value of new business conversion models will be significantly limited.

VI.RETENTION MODELS

For renewal business retention models, the main purpose is to understand the probability of an existing insured to stay for the next renewal term [8]. The reason that an existing insured does not stay for the next renewal term may be due to the insured's action, such as mid term cancellation, non-response to renewal request, or non-payment of premium, or insurer's action, such as non-renewal. Therefore, the renewal retention models will focus on understanding how an insured's characteristics correlate with the retention rate.

VI PREDICTIVE VARIABLES

For renewal conversion models, the model will measure the probability of the policy to be converted to the next term at the point of renewal for the existing policy. Therefore, these models exclude the mid-term cancelled policies. Similar to the new business conversion models, the renewal price offered and how it compares to the competitors will play an important role on the outcome. Obviously, for renewal models, much more information, especially information from the company's internal data sources, can be used. For new business models, the predictive variables are very limited, and sometimes the models may completely rely on external data sources. In the end, these marketing models may not be as accurate as GBA-Algorithm and pricing models but they do offer an opportunity to improve resource allocation and efficiency in the sales process by allowing insurance companies to focus their marketing and sales efforts on the risks that are most likely to be bound or retained.

A.NEW BUSINESS CONVERSION MODELS

In the remaining sections of the paper, we will focus on the new business conversion models because they are the most challenging ones to build, and they are very critical for insurance companies to sustain long term profitable growth. For the new business conversion models, predictive modeling techniques can be employed to find certain segments with a higher likelihood for responding to the quote, i.e. the response rate, and purchasing after taking quotes, i.e., the hit or conversion rate, as well as, segments with a higher or lower sensitivity with respect to the price. Similar to GBA-Algorithm models, the marketing models are often created on a policy level, and sometimes even on a household or account level.

B.ANALYSIS PROBLEM

As mentioned earlier, in order to analyze the response rate and hit rate, it is important to capture the price competitiveness for the quote, that is, the price differentiation between the company and its competitors. The competitors' pricing information can be obtained in published rating manuals, company's quote files, or industry competitive information vendors' data base. If the competitors' prices are well captured in the quote files, the core information of the price elasticity curve can then be established for the models.

C.SOLUTION PROBLEM

The typical data issues for building up marketing and sales models are:

Since quote files are not required for financial reporting or bureau reporting, the quality of the files are much worse than other files and data sources. In addition, insurance companies often purge their quote .files after one or two years, therefore little historical quote data is available for analysis. Once again this highlights the importance of corporate data quality and governance as a key strategy to maximize predictive modeling benefits.Typically, there is very limited information captured in the quote files and often only includes the following:

a. Name and address of an insured
b. Basic and key rating information
c. Agent information
d. Competitiveness information including prior carrier's name and price Insurance companies rarely capture information other than the above and therefore the number of variables that can be derived is very limited.

D.COMPETITIVE POSITION

For the renewal retention process , insurance companies rarely follow up their non-renewal risks and find out the reasons for their non-renewal decision, the new company they took their business to, or the new price that they received from their new company. Without such competitive information, the value of the marketing and sales models will be significantly limited.

E.GAIN MARKET

It also minimizes a company's opportunity to gain market intelligence and assess its own competitive position because there is valuable business insight that can be gained from understanding why a company's customers are leaving.

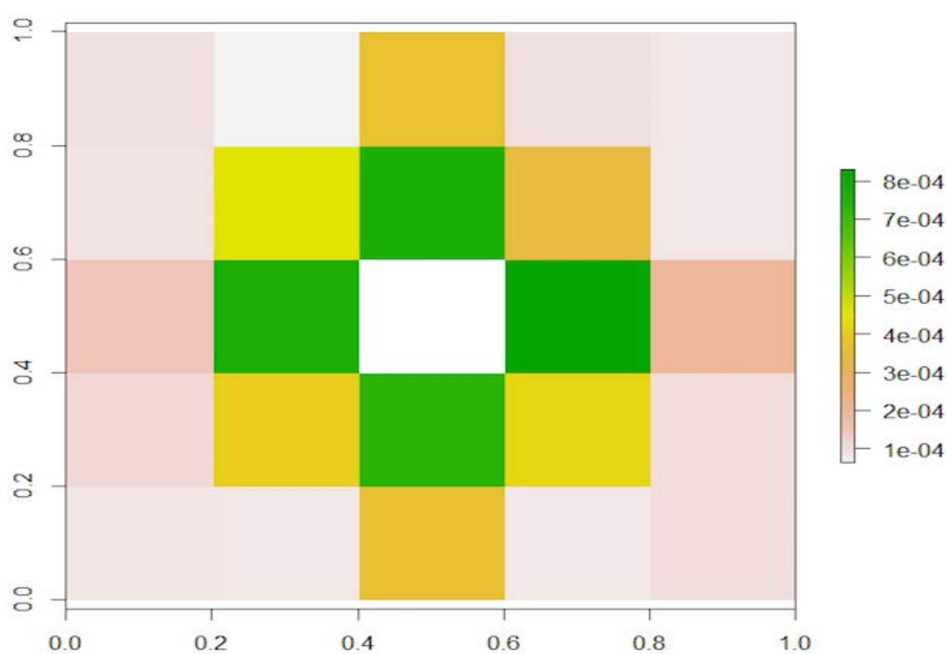

Fig. 1(a) Competitiveness information

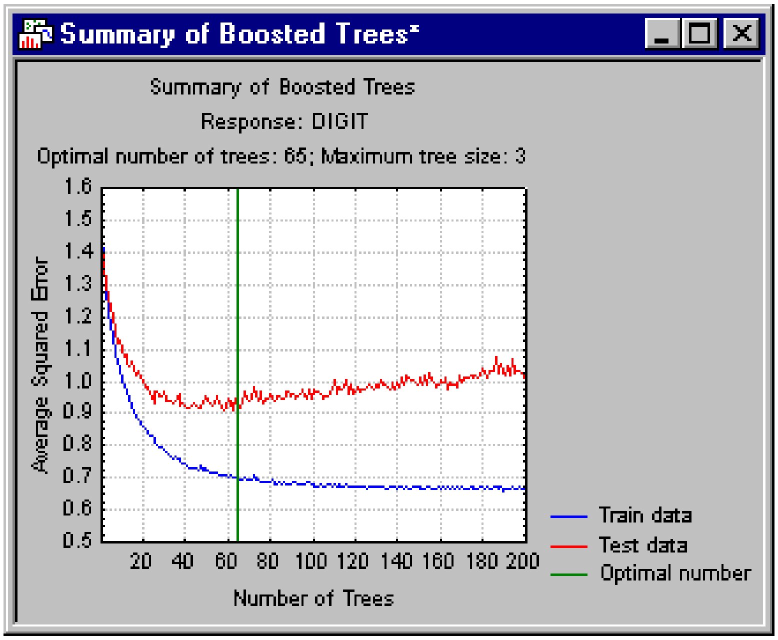

Fig.2 Analysis of Models

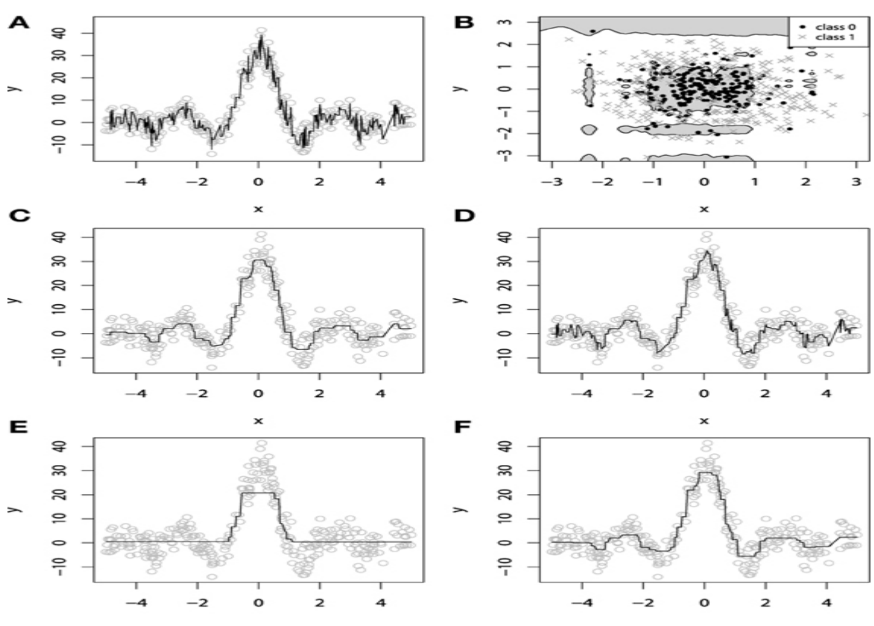

Fig.3 Methods Analysis

The third step, now that we have identified the key range for the rate adjustment-conversion rate relationship, is to use the results to adjust the GLM rating plan so that the parameters can be re-optimized with different adjustments. This step can be tedious and involves an iterative process but the benefits can be significant. At this step, the company's historical data is employed in the pricing model development. At the same time the marketing information is used along with the pricing information to improve the overall performance for the company's operation by striking a balance between profitability and growth.

The last step is to build the GBA-Algorithm models on top of the pricing and marketing models. There are several reasons that the GBA-Algorithm model is important to use along with the pricing and marketing models.

First, the GLM pricing model may still be far from addressing the overall rate adequacy because many significant variables are not used in the pricing models. Such information may include agent's performance data, credit score, and demographic and territorial information on a more refined level. A great deal of non-rating information can be used to enhance the segmentation of an insured's profitability.

F.GLM Model

Our experience indicates that for commercial lines, such GBA-Algorithm models are very important, since most of the commercial line carriers follow the bureau loss cost and rate structures for most of the major lines of business. They do not have their own GLM based pricing models.

G.RATE PLAN

The second reason is that, as the result of adjustments for the rating plan due to conversion consideration, it is likely that some segments can turn unprofitable due to the trade off for growth and retention. The decrease in profitability can be minimized with additional GBA-Algorithm information. For example, if it is determined that youthful policyholder factors need to be tempered to increase the conversion rate, the potential profitability impact can be minimized through the application of the GBA-Algorithm models by allowing profitable agents to write more youthful risks than unprofitable agents write (i.e. – offsetting the risk of youthful risks by focusing on youthful risks with favorable credit scores).

Finally, it is very important to note that, when developing the GBA-Algorithm models, the underlying premium should be based on the final pricing structures and rating factors. All historical premium data should be adjusted to the final selected pricing level.

H. OUR APPROACH

From a tactical perspective, our approach to integrating pricing, GBA-Algorithm and marketing predictive models is a four step integration process as outlined below:

Step 1: Develop the GLM based rating plan and pricing model.

$$P = O1 + (x / y)(O1 - On)$$

Step 2: Develop retention or conversion models to study the price elasticity behavior of insurance buyers.

Step 3: Adjust the rating plan and class plan factors based on the retention and conversion models to strike a balance between rate adequacy and conversion rate.

Step 4: Build up a series of GBA-Algorithm rules based on GBA-Algorithm models in conjunction with the pricing and market models to maintain the overall competitiveness.

$\mathcal{X}(1,2)$ *->Plan data*
$\mathcal{X}_n$ *->Analysis data*
$\mathcal{Y}(1,2)$ *->Base data*
$\mathcal{Y}_n$ *->End Data*
W-> World data

PROCESSING

VII. GRADIENT BOOSTED ALGORITHMS

Outlook:

Define sets of model parameter values to evaluate;

for each parameter set **do**

  for each resampling iteration **do**

    Hold–out specific samples ;

    Fit the model on the remainder;

    Predict the hold–out samples;

  end Calculate the average performance across hold–out predictions

end

Determine the optimal parameter set;

Main algorithms:

Initialize equal weights per sample;

for j = 1 ... M iterations **do**

Fit a classification tree using sample weights (denote the model equation as fj(x));

forall the misclassified samples do increase ample weight

end Save a “stage–weight” ("j) based on the performance of the current model;

end

Input: training set $T = (x_1, y_1),(x_2, y_2),...,(x_m, y_m); T \sim D^m$ a differentiable loss function Training error $\hat{\varepsilon}(h) \doteq \frac{1}{m}\sum_{(x,y)\in T} 1[h(x) \neq y] \doteq P_{(x,y)\sim T}[h(x) \neq y]$ number of iterations m

I.GBA-ALGORITHM:

1. Initialize model with a constant value:

   $$(x_1, y_1, 1),(x_2, y_2, 1),\ldots,(x_n, y_n, 1)$$

2. For *n* = 1 to :

   1. Compute so-called *pseudo-residuals*:

      $$(x_1, y_1, w_1^1),(x_2, y_2, w_2^1),\ldots,(x_n, y_n, w_n^1)$$

   2. Fit a base learner h(x) to pseudo-residuals, i.e. train it using the training set .

      $$T = (x_1, y_1),(x_2, y_2),...,(x_m, y_m); T \sim D^m$$

   3. Compute multiplier by solving the following one-dimensional optimization problem:

      $$(x_1, y_1, w_1^2),(x_2, y_2, w_2^2),\ldots,(x_n, y_n, w_n^2)$$

      $$(x_1, y_1, w_1^{T-1}),(x_2, y_2, w_2^{T-1}),\ldots,(x_n, y_n, w_n^{T-1})$$

   4. Update the model:

      $$F_T(x) = \alpha_1 h_1(x) + \alpha_2 h_2(x) + \cdots + \alpha_T h_T(x)$$

3. Output $f_T(x) = \mathrm{sign}(F_T(x))$

J.ALGORITHMS ANALYSIS

By integrating the three types of predictive models seamlessly, insurance companies can gain two major benefits. First, instead of adjusting their rates across the broad for growth, insurance companies can “target” the segments to gain a high return on growth with minimum price changes. Second, the potential profitability issue associated with rate cutting for growth can be minimized with GBA-Algorithm models. We believe that with such integration, the full value of predictive modeling can be realized. It can provide insurance companies with an effective way to deal with the key business challenges of achieving profitable growth and minimizing the impact of the GBA-Algorithm cycle. History tells us that companies that are successful and regarded as market leaders are the ones that can process information and make sound business decisions faster than their competition can. The manufacturing companies Insurance Industry should be no different and an integrated approach to predictive modeling gives manufacturing companies an opportunity to realize the full value of their predictive modeling investment and stay a step ahead of the competition.

VIII.TEST

When these three applications are integrated, modelers should be conscientious about the data and modeling issues and problems described in previous sections for each application. In addition, there exist unique, challenging data and modeling issues during the integration process.

TABLE I OVER TEST DATA ANALYSIS

Decile (based on model score)	Number of Leads (Hold-out Sample)	Sales	Conversion Rate (%)	Lift (Above random sample)
1	8,000	252	3.1%	4.0
2	8,000	115	1.4%	1.8
3	8,000	70	0.9%	1.1
4	8,000	59	0.7%	0.9
5	8,000	40	0.5%	0.6
6	8,000	29	0.4%	0.5
7	8,000	31	0.4%	0.5
8	8,000	15	0.2%	0.2
9	8,000	14	0.2%	0.2
10	8,000	8	0.1%	0.1
TOTAL	80,000	632	0.8%	

The first unique challenge is due to the fact that the data level is different between the pricing model and the GBA-Algorithm and marketing models. Therefore, how to "accurately" profile the policies identified by the conversion model and link the model results to the subsequent pricing model is a challenge. For example, a youthful policy may have all of or partial of its drivers as youthful drivers. When the marketing model profiles youthful driver policies to be targeted or not targeted, it needs to be very specific in defining whether the profile is partial (if partial, the percentage of youthful drivers on the policy) or all youthful driver policies. In other words, how to "roll up" exposure based pricing information from the pricing model to the policy level information for the GBA-Algorithm and marketing models needs to be prudently considered.

A.RESULT

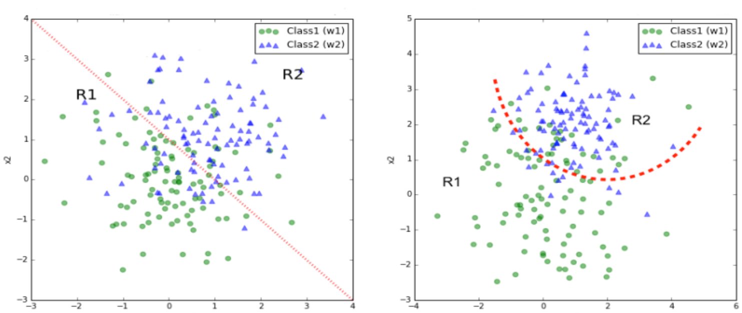

Fig.5 Model Analysis

Another challenge for integration is that the marketing application is "forward-looking" based, while the pricing and GBA-Algorithm application are based on "historical" information. Due to constant changes associated with the internal and external environments for insurance operation, the historical data distribution and composition may not serve well for the "forward-looking" integration application. For example, if a national insurance company would like to expand its business in certain a geographic region, such as in the northeast, it is possible that the northeast risks behave differently from the risks in other regions. Therefore, modelers need to pay extra effort as to how to prepare the data for the integration analysis, and, for this example, may want to use data in the northeast region only. Other considerations include the distribution change in industry class, affinity programs, or premium size.

B.DISCUSSION

As discussed in the previous sections, different applications may have different data available. In general, data is more sparsely available for the marketing application than for the GBA-Algorithm or pricing applications. For example, driver and vehicle details are fairly populated in the pricing and GBA-Algorithm data sources, but not for the marketing data sources. When the details are available in the marketing data sources, it is possible that they are more available for certain regions, branch offices, agents, or programs than for others. The inconsistency in data availability may lead to "bias" in the analysis results.

IX.COMPARATIVE STUDY

By combining a comprehensive GBA-Algorithm model with a pricing model, a company can more accurately estimate loss cost and profitability than by using the pricing model alone. Previously, we illustrate how to use the pricing model and the marketing model together first, and then develop an GBA-Algorithm model second. In theory, there is no limitation for the sequence of integration, and the GBA-Algorithm model can be used alone with the pricing model to fine-tune the marketing model. Of course, the challenge for this approach is that the GBA-Algorithm model is on the policy level, while the pricing model is on the exposure level.

X.CONCLUSION

Several years ago, merely using predictive models in some fashion to support GBA-Algorithm, pricing and marketing gave insurance companies a competitive edge. However, in today's competitive market, predictive modeling is not limited to just personal lines but is used widely in commercial lines as well. Therefore the first mover advantage no longer exists and insurance companies must find new ways to maximize the benefits of their predictive modeling investment and stay ahead of their competition.

Our paper illustrates the strategic and tactical approach of taking an enterprise wide view of predictive modeling and integrating the results from pricing, GBA-Algorithm and marketing models to support business decisions across multiple business operations. In today's market, companies that will succeed are the ones that incorporate analytics as a core business strategy and align multiple business operations with a single unified view of analytics.

ACKNOWLEDGMENT

This paper is made possible through the help and support from everyone, including: My wife M.suganya and Daughter S.S.Inakshi and My sir S.Syed Nazimuddeenand S.Irshath Ahamed , and in essence, all sentient beings. I sincerely thank to my parents, family, and friends, who provide the advice and financial support. The product of this paper would not be possible without all of them.

REFERENCES

[1] Predictive modeling methodology for obtaining optimally predicted results with reduced uncertainties: Illustrative application to a simulated solar collector facilityOriginal Research Article Solar Energy, Volume 119, September 2015, Pages 486-506

[2] Dan Gabriel Cacuci, Aurelian Florin Badea,Predictive model for the growth kinetics of Staphylococcus aureus in raw pork developed using Integrated Pathogen Modeling Program (IPMP) 2013 Meat Science, Volume 107, September 2015, Pages 20-25

[3] Evaluation and predictive modeling the effects of spice extracts on raw chicken meat stored at different temperatures Journal of Food Engineering, Volume 166, December 2015, Pages 29-37

[4] K. Radha krishnan, S. Babuskin, P. Azhagu Saravana Babu, M. Sivarajan, M. Sukumar,A supervised machine-learning approach towards geochemical predictive modelling in archaeology,journal of Archaeological Science, Volume 59, July 2015, Pages 80-88.

[5] S.V. Nanukuttan, P.A.M. Basheer, W.J. McCarter, L. Tang, N. Holmes, T.M. Chrisp, G. Starrs, B. Magee,A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings Energy and Buildings, Volume 97, 15 June 2015, Pages 86-97

[6] Hao Huang, Lei Chen, Eric Hu,Network based approach for predictive accident modelling Safety Science, Volume 80, December 2015, Pages 274-287

[7] Al-Amin Baksh, Faisal Khan, Veeresh Gadag, Refaul Ferdous,Predictive modeling and optimization of a multi-zone HVAC system with data mining and firefly algorithms Energy, Volume 86, 15 June 2015, Pages 393-402

[8] Yaohui Zeng, Zijun Zhang, Andrew Kusiak,Fuzzy modeling and predictive control of superheater steam temperature for power plant Article ISA Transactions, Volume 56, May 2015, Pages 241-251

[9] Xiao Wu, Jiong Shen, Yiguo Li, Kwang Y. Lee Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis Journal of Clinical Epidemiology, In Press,Corrected Proof, Available online 28 March 2015

[10] Berend Terluin, Iris Eekhout, Caroline B. Terwee, Henrica C.W. de Vet.,A summary of fault modelling and predictive health monitoring of rolling element bearings,Mechanical Systems and Signal Processing, Volumes 60–61, August 2015, Pages 252-272 .