USING GENETIC ALGORITHMS FOR CREDIT SCORING SYSTEM MAINTENANCE FUNCTIONS

David J. Fogarty

Practitioner Faculty Member at the University of Phoenix, Phoenix, AZ, USA
davidf1967@email.phoenix.edu

ABSTRACT

Due to recent advances in high-speed computing technology sophisticated tools for prediction have been introduced which include genetic algorithms. Genetic algorithms operate with the objective of replicating the manner in which genes are passed from one generation to the next. The surviving genes create models that form the most durable and effective offspring. Due to a high-level of understanding of the domain it is has proved to be very difficult for advanced techniques including genetic algorithms to outperform traditional techniques of credit scoring such as logistic and least squares regression. However, in large organizations it is difficult to continuously update complex scoring systems. This paper explores the use of genetic algorithms to automate and institutionalize the scoring maintenance function within an organization.

KEYWORDS

Classification; Consumer Loans; Marketing Models, Risk Models Credit Scoring; Logistic Regression; Genetic Algorithms

1. INTRODUCTION

This paper will attempt to show how useful advanced semi-automated modeling techniques specifically genetic algorithms can be to facilitate scoring system maintenance functions. Scoring is the term used to describe formal statistical methods used for classifying applicants and customers for credit into certain classes such as "good" and "bad" risk, propensity to buy, revolving behavior and all aspects of the consumer credit lifecycle. Such methods have become increasingly important with the dramatic growth in consumer credit in recent years. A wide range of statistical methods has been applied, though the literature available to the public is limited for reasons of commercial confidentiality. Standard statistical methods used in the industry for developing scoring models are discriminant analysis, linear regression, logistic regression and decision trees. In addition to the consumer credit industry scoring is also prevalent in other businesses such as the mail-order/internet catalogue, direct marketing and insurance businesses [1], [2], [3], [4].

Historically, discriminant analysis and linear regression have been the most widely used techniques for building scoring systems predicting a discrete outcome (e.g. good vs. bad). However, a seminal research paper written by Eisenbeis [5] on the pitfalls of using discriminant analysis for credit scoring paved the way for the use of logistic regression which is now the staple technique used in these applications. On theoretical grounds, logistic regression is a more appropriate statistical tool than linear regression for discrete classes given the fact that these classes have already been defined [1], [2], [3], [4]. Regardless of which statistical method is used scoring algorithms take a great deal of skill to develop especially given the regulatory requirements and the usual state and location of the data within many businesses. At a minimum most firms scoring professionals to have a Master's degree in Statistics and often PhD's with statistical programming skills to be able to join scoring teams. Having many of these resources could be extremely costly to firms from a labor perspective. For instance large financial services firms can often have over 500 analysts who specialize in this distributed across the entire enterprise. Given the economic reality most forms respond by capping hiring in this area and therefore do not have not enough resources hired to maintain all the systems. Scoring teams typically respond by focusing on new scoring development at the cost of updating existing scores.

One of the key challenges to maintaining scoring systems in firms include *population drift*. Population drift is a phenomenon responsible for the deterioration of scores over time. This describes the tendency of populations to evolve with time, so that the distributions change. This is to be expected since applicant populations are subject to economic pressures and a change competitive environment. Typically, population drift is detected through a model monitoring technique known as characteristic analysis. Population drift is fairly straightforward to detect with characteristic analyses and generally requires a rebuild of the scorecard [1], [2], [3], [4].

However, scoring professionals today are faced with data capture technology advancing faster than data technology. Moreover, in the field of predictive modeling, improvements are possible through being able to better explore growing data sources like the web and electronic payment systems. As these sources of data continue to grow companies are facing stiff competition in the global labor markets for highly-skilled employees who can analyze and extract value from the data. With this need to make improvements through exploring these new data sources most firms do not have adequate resources to address population drift and therefore do not devote enough resources to mitigating this situation.

One potential solution to deal with the key challenge of population drift are sophisticated semi-automated prediction tools which can work behind the scenes to insure that the models meet the minimal standards for efficacy or can rebuild models when necessary.

Due to recent advances in high-speed computing technology these sophisticated tools for prediction have been introduced which include genetic algorithms [6]. The genetic algorithm method which focuses largely on the nature of volatility stretches the limits of even the most high-powered computers. Genetic algorithms operate with the objective of replicating the manner in which genes are passed from one generation to the next. The surviving genes create models that form the most durable and effective offspring [7].

When considering a GA for a particular application Mitchell [8] points out that there is no prescribed answer. However, in general if the space to be searched is large, is known not to be perfectly smooth and unimodal (i.e. consists of a single smooth "hill"), or is not well understood,

or if the fitness function is noisy, and if the task a sufficiently good solution is enough – a GA will have a good chance of being competitive with or surpassing other more traditional methods. These rules of thumb, however, do not represent prescribed predictions of when a GA will be an effective search procedure when compared with other procedures. Instead, a GA's performance will depend very much on details such as the method for encoding candidate solutions, the operators, the parameter settings, and the particular criterion for success [8]. In consumer credit scoring the space to be searched is relatively well-known. However, if you factor in the all the combinations associated with variable transformations and interaction terms in addition to using every variable in lie of a variable reduction technique associated with traditional model building you end up with a very large search space that would task any analyst and force them to limit their manual search of all possibilities due to time and resource constraints.

Some of the theoretical limitations genetic algorithms as suggested by Bernstein [7] were that while these algorithms provide important insights into the complexity of reality, there is no real proof of cause and effect in the recognition of patterns that precede the arrival of other patterns ranging from financial activities to the spin of a roulette wheel. Furthermore, Bernstein [7] reported that likeness to truth is not the same as truth and that with the absence of any theoretical structure to explain why patterns seem to repeat themselves over time or across systems, tools such genetic algorithms provide little assurance that today's signals will trigger tomorrow's events. What we are left with is the subtle sequences of data that the enormous power of the computer can reveal. Thus, forecasting tools based on nonlinear models or on computer gymnastics including genetic algorithms are subject to many of the same hurdles that stand in the way of conventional probability theory: the raw material of the model is the data of the past.

However, while we need to be aware of the risks outline by Bernstein [7] above there are some definite advantages in using these tools. The first advantage includes include being able to mitigate some of the negative effects of Simpson's Paradox and corollary's which state that an association between two variables can be reversed upon introduction of a third variable. In addition, variables that by themselves may seem insignificant but when included with others can become important (multivariate and interaction). Finally, variables that by themselves may seem significant but when included with others can become irrelevant. Therefore, given Simpson's Paradox and corollary's suboptimal models are yielded when data reduction is performed early in the model-building process and analysis is restricted to 1st order correlation. However, the vast majority of credit scoring techniques require and recommend data reduction prior to model building and render primarily univariate consideration. Moreover, modeling in complex systems requires the simultaneous optimization of: variable selection, coefficient setting, missing data management, interaction detection and variable transformation. Searching for optimal solution through all possible variable combinations is impractical. For example, selecting just 10 in 25 variables equals 11 trillion combinations of variables. There are even more possibilities when transformations or interactions are included.

Therefore, the genetic algorithms enable some key improvements for model development. Further evidence of this was described by Ong et al., [9]. One benefit is significantly reduced model development time by 50 to 74% through advanced data preparation processing. Also, genetic algorithms should create models at least as powerful as skilled analysts using traditional techniques and given certain conditions can even allow the realization of an increased model "lift" typically ranging from 5-15% over traditional approaches such as logistic or linear regression.

Given the benefits of genetic algorithms described above one may ask why not switch to this approach versus traditional approaches. If we make the assumption that GA's will only produce a score that is just as good or slightly better than traditional techniques then one reason why this is not recommended is the widespread acceptance of traditional scoring techniques in the industry. Another reason is due to legislation and regulatory agencies requiring the transparency of scoring systems. Given this requirement there is little incentive to introduce techniques which require more detailed explanation or cannot be readily explained with traditional theoretical underpinnings.

While it is not recommended that GA's be used for primary model development they do have a place as an approach to maintain existing scoring systems developed by experienced modelers. These scoring systems often undergo population drift and are not updated on a consistent basis which can cause a great loss in a firm's ability to target the right customers for the right product or offer or will allow a degradation in the overall credit quality of the portfolio. The idea here is that once the professional modeler has developed an initial score using traditional techniques the GAs could then be used to refine the score under the domain created by the initial model on a go forward basis.

2. USING GENETIC ALGORITHMS FOR CREDIT SCORING SYSTEM MAINTENANCE FUNCTIONS

In order to test the research hypothesis that genetic algorithms can be used to facilitate scoring system maintenance functions a proprietary genetic algorithm was created which facilitated the credit scoring function. Experts in the domain of credit scoring were consulted to model the system functions and insure that all of the features were retained.

The problem from a scoring perspective can be represented as:

$$f(x_1 + x_2 + ... + x_n) = y \in \{0,1\}$$

Where the model tries to learn the function *f(x)* from the given data

Genetic algorithms can be used to learn the function without making any assumption about the nature of the function. The genetic algorithm applied to scoring models basically works by applying Darwin's "Survival of the Fittest" principal in that better scoring models survive and poorer performing models perish. It accomplishes this by genetically encoding the problem into a series of bit strings according to the standard genetic algorithm framework. In this the genes represent the data attributes or independent variables. The chromosomes are then made up of genes and represent a model. Finally, the models represent the potential solution to the business problem. In terms of the evolutionary process it begins by breeding initial population of random models and evaluating the fitness of each of the models. Then you continue to breed new generations of models through the process of cloning, mating and mutating with the best evolved models becoming the final solution.

3. RESULTS

The proprietary genetic algorithm model was then applied to the datasets on 25 models in developed in a traditional manner using logistic regression in a large finance company. The models represented a cross sectional sample from all aspects of the consumer credit lifecycle including acquisitions, activation, cross-sell and retention and covered both the risk and marketing domains. The data for these models come from several sources depending on the business function being modeled. Internally, the data comes from application and receivable systems which house all the consumer information for the consumer lending and banking products. Externally, the data comes from consumer credit bureaus including Equifax, Experion and Trans Union in addition to external demographic and psychographic data sources. All of the models which have been developed using traditional techniques such as logistic are currently in production at this firm and have yet to be updated using the newest information. The idea here was to use the developmental data sets from these models and compare the performance of the traditional techniques with the genetic algorithm models. The implications for the genetic algorithms performing as good or better than the traditional techniques is that they would be fit for model redevelopment maintenance functions. A hold out or development sample was prepared and used to validate each score which is standard industry practice to prevent overfitting. The models were compared on the basis of a decile analysis. Deciles are created when probabilities output from the model are converted to scores and rank ordered from high to low and then split into ten equal groups of 10% based on the size of the overall population base. The theory being that if the scores derived from the models are performing better a random generated number then we should see a better performance in the higher deciles and worse performance in the lower deciles. A decile analysis allows us to pinpoint where the critical decisions are being made. For instance, in a direct marketing model budgetary constraints and the size of the mailing universe define which deciles can be mailed. If there is a population of 100,000 to be mailed and there is a budget to mail 30,000 then one would want the models to perform the best at the third decile. Therefore in this situation we compared the performance of the models at this decile. The results from this exercise are shown below in Table 1. What is evident at the outset it that the genetic algorithm propensity models are performing up to their theoretical expectations being better or no worse than their traditionally derived counterparts. In other words, in 72% of the trials the Genetic algorithms models performed significantly better than traditional techniques and in 16% of the trials the genetic algorithm models performed equal to the traditional techniques. In 12% of examples the models failed to perform satisfactorily in the out-of-sample validation due to overfitting the data. In no cases during this experiment did the models actually perform worse than the traditional techniques. The implications for these findings is that given that the initial scores are developed and the data has been prepared according to standardized scoring preparation processes the genetic algorithms do an excellent job at building the scores.

Table 1

Sl .No	Region	Product	**** Improvement (GA vs Benchmark)
MARKETING MODELS			
1	Europe	Credit Card	6.00%
2	Europe	Credit Card	8.26%
3	Pacific Rim	Credit Card	6.00%
4	Pacific Rim	Credit Card	6.12%
5	Pacific Rim	Closed-End Loan	0.02%
6	Pacific Rim	Revolving Loan	6.00%
7	Pacific Rim	Revolving Loan	7.60%
8	Americas	Credit Card	6.00%
9	Americas	Credit Card	8.16%
10	Pacifica Rim	Insurance	0.00%
11	Europe	Credit Card	6.20%
12	Pacific Rim	Auto Loan	10.83%
13	Pacific Rim	Auto Loan	10.26%
14	Americas	Closed-End Loan	0.46%
15	Europe	Credit Card	0.00%
16	Americas	Credit Card	1.40%
17	Pacific Rim	Closed-End Loan	2.13%
18	Pacific Rim	Closed-End Loan	6.00%
19	Europe	Credit Card	6.00%
20	Pacific Rim	Credit Card	8.49%
RISK MODELS			
21	Pacific Rim	Credit Card	6.00%
22	Europe	Closed-End Loan	0.08%
Out Sample Valid Failed			
23	Americas	Credit Card	6.00%
24	Pacific Rim	Credit Card	2.30%
25	Europe	Auto Loan	4.60%

The next step was to let experienced scoring professionals in the business evaluate the models from a qualitative perspective. A team of scoring professionals from the organization examined the models in detail and concluded that while some of the interaction terms needed more explanation in a consumer credit environment especially in the risk management applications the models in general tracked those developed by traditional modeling methods such as linear and logistic regression. Moreover, due to having a good comfort with the underlying data these experienced analysts expressed their approval in using these scores developed by the genetic algorithms in a live environment

4. CONCLUSIONS

This paper has made an original contribution to knowledge by demonstrating that given an initial scorecard developed using traditional methods the genetic algorithms can perform at least as good or better than traditional techniques. It is good to reiterate again that while it is not recommended that GA's be used for primary model development for the various reasons discussed earlier they have shown in this paper to have a place as an approach to maintain existing scoring systems initially developed by experienced modelers. Some of the limitations of using genetic algorithms for scoring in general include the fact that it is difficult to come out with a generic model for all class of problems and data. Also, the GA models are dependent on the data and the objective of the business. Another limitation is that these models require good processing power. Finally, getting to the optimal combination of parameter values for a given problem is difficult. One of the limitations of this specific study is that it utilized data from a single firm. Further research extending this to several firms will significantly improve the ability to generalize the findings. As this paper has demonstrated the efficacy of genetic algorithms for scoring system maintenance functions there is also the possibility to extend this to scoring development auditing functions and this therefore is a suggestion for further research. The question of quantifying the merit of GAs over a random search for a generic problem is also presents an interesting topic for further research, though challenging because one needs to have a good definition of a generic problem. Finally, genetic algorithms are but just one of the numerous semi-automated advanced techniques available. Further research on applying other advanced techniques to scoring maintenance functions such as neural networks and support vectors/kernel machines is a suggestion for further research.

REFERENCES

[1]   Hand, D., J. (2001) Modeling Consumer Credit Risk. IMA Journal of Management Mathematics, v.12, no. 1, pp. 139-155.

[2]   Hand, D J and Henley W E (1997), "Statistical Classification Methods In Consumer Credit" Journal Royal Statistical Society Series A, 160, 523-541.

[3]   Hand, D., J., Sohn, S., Y., Kim Y. (2005) Optimal Bipartite Scorecards. Expert Systems with Applications, v.29, no. 3, pp. 684-690.

[4]   Henley, W., E. (1995) Statistical Aspects of Credit Scoring. PhD. Thesis. The Open University, Milton Keynes, UK.

[5]   Eisenbeis, R., A. (1977) Pitfalls in the Application of Discriminant Analysis In Business, Finance and Economics. Journal of Finance, v.32, no. 4, pp. 875-900.

[6]   Goldberg, D., E. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, 1989.

[7]   Bernstein, P., L. (1998). Against the Gods: The Remarkable Story of Risk. John Wiley and Sons, Inc., New York.

[8]   Mitchell, M. (1998). An Introduction to Genetic Algorithms. MIT Press, Cambridge, Massachusetts.

[9]   Ong, C, S., Huang J., J. Tzeng G., H. (2005) Building Credit Scoring Models Using Genetic Programming, Expert Systems with Applications, v. 29, no. 1, pp. 1-7.

Author

David J. Fogarty, MBA, Ph.D. is currently a practitioner faculty member at the University of Phoenix in the USA. His research interests include: data mining, advanced statistical methods and quantitative analysis applied to business, international business management and hypothesis testing in business.