CONSUMER CREDIT LIMIT ASSIGNMENT USING BAYESIAN DECISION THEORY AND FUZZY LOGIC – A PRACTICAL APPROACH

Uttiya Paul, Angshuman Biswas
Analytics Centre of Excellence,
Genpact, Kolkata

ABSTRACT

The market for consumer credit is diversified and multilayered. This makes it a challenging task to come up with a generalized theory of consumer credit limit assignment. However it is natural to surmise that an accurate and dynamic credit limit assignment strategy will certainly reduce the magnitude of the random and non-random uncertainty embedded in future credit limit management activities. Traditionally credit risk scorecards are employed for such strategic decisions. In order to minimize the misclassification error, Bayesian decision theory can be employed instead of scorecards for a more accurate credit limit strategy. However, estimation of the probability density function is a challenging task under practical circumstances. In this paper, we have proposed a modelling framework algorithm to estimate the class conditional density functions using frequency probability stemming from statistically independent simulations. Also for the continuous variable components within the feature vector, the class membership of a new entrant is being assigned using fuzzy logic. This makes the model robust, easy to handle, comprehend, implement and control.

Key words: Credit limit assignment, Bayesian Decision theory, Fuzzy logic.

JEL Classification: C65, G20

http://www.iaeme.com/JOM/issues.asp?JType=JOM&VType=4&IType=2

1. INTRODUCTION

Consumer credit is defined as the loan amount extended to individuals as opposed to group level entities. The market for consumer is diversified and multilayered. This makes it a challenging task to come up with a generalized theory of consumer credit assignment. Traditionally in order to obtain credit, consumers have to go through a typical process: In the
1st stage of the process, potential customers submit an application in which they declare their demographic details and financial credentials heretofore. In the very same application, they also state the characteristics of the desired loan such as amount and term. In the 2nd stage of the process, the lender takes the accept-reject decision (usually known as the underwriting decision) regarding the desired loan. Such a decision is traditionally based on the probability of default estimate (PD) of the customer [1]. There are number of statistical models to estimate the probability of default and the most widely used one is known as the credit scoring model. The business objective of credit scoring models is to identify distinct groups within a population where each group has different risk measures across and same risk measure within [2]. This idea and technique of identifying homogeneous groups based on several characteristics within a population is not new. It was introduced by Fisher in 1936 in the form of a linear discriminant function. In his analysis Fisher considered several features (sepal length, sepal width, petal length and petal width) of the flowers for fifty plants corresponding to two species [3]. Interestingly in 1941, Durand proposed that Fisher’s technique can be successfully employed to distinguish between good loan and bad loan [4]. Since then credit scoring model had experience extraordinary improvement in terms of technical nuances as well as practical usage. However in recent times it was well understood that it is equally important to model the profitability aspects apart from risk. In fact there has been a significant pool of academic literature focusing on this aspect. Among many factors pertaining to profitability (such as likelihood of responding to a marketing campaign, likelihood of attrition , propensity to revolve account balance , flexible pricing), Though, Wu and Hand (2007) attributed importance on account management activities like credit limit increase and credit limit decrease. Unfortunately when particularly focused on the issue of customer reaction on account management activities, there is a dearth of research work [5]. The effects are ambiguous and views are sometimes conflicting. One school of thought is that an increase in credit limit would normally lead an increase in revolving credit balances which in turn increases revenue, customer loyalty and therefore customer lifetime value [5]. There are opposite views depicting the fact that an increase in the credit limit on a credit card, on average leads to increase in consumer debt [6].

Hence it is natural to assume that an accurate and dynamic credit limit assignment strategy will certainly reduce the magnitude of the random and non-random uncertainty embedded in future credit limit management activities. Usually a typical credit cycle can be divided into four phases namely acquisition, account management, collection/recovery and write-off [7]. Credit limit assignment strategy is designed in the acquisition stage of the cycle. As a general practice, in the acquisition stage using decisionbureau data, credit scoring models are developed using logistic regression technique. In present day, this is the technique followed by most of the financial institutions to assess credit risk. These credit scoring models are used for credit authorization [8]. However, one shortcoming of credit scoring models is that they are inadequate to address aspects like line utilization, profitability. The main reason is traditional credit scoring models do not treat credit line as an endogenous variable [10].

There can be several ways to address this problem. One approach is to create a number of clusters on historical accounts based on customer’s pattern of spending and paying bills in the 1st phase of the analysis. In the 2nd phase, CART based decision tree can be used to determine the degree of cluster membership. Finally in the 3rd stage of the analysis, by maximizing the overall expected net present value of customer over all values of the deal, the optimal deal structure can be determined. There are few drawbacks of this methodology. Firstly there is no guarantee that this overall expected net present value will have a unique maximum. Secondly it is not compatible with the change in the consumer behaviour [9].
2. NEED FOR THE STUDY

In this paper, we have prescribed a methodology to overcome these problems by designing a classification scheme using Bayesian decision theory based on feature vector (which may composed of variables like risk score, line utilization, thickness of file, historical trade counts, policy indicators), we can assign each customer a specific credit limit using Bayesian decision theory. Here we have adopted an application based approach from a professional data consultant’s perspective without comprising on the theoretical sanctity. One reason we have chosen Bayesian decision theory is because when applied, Bayesian decision theory leads to the minimization of probability of misclassification error (or maximizing the accuracy) [10]. However estimation of the probability density function is a challenging and time-consuming task in the real world. There are several methods and each one has its own pros and cons. Moreover, the decision surface constructed using Bayesian decision theory may not always be mathematically tractable. Such a problem may arise when in the feature vector we have mix of continuous as well as discrete variables. Such a feature vector is highly plausible in real world.

We have designed a modelling framework algorithm where the class conditional probability density functions (likelihood functions) can be estimated using frequency probability stemming from statistically independent simulations. Also for continuous variables, the class membership of a new entrant is being decided using fuzzy logic. This makes the model not only robust but also easy to handle, comprehend, implement and control.

3. AN OVERVIEW OF BAYESIAN DECISION THEORY

If $\omega_1, \omega_2$ are two states of nature then based on a supervised learning mode, we can generate two probabilities $p(\omega_1)$ and $p(\omega_2)$. These probabilities are known as a priori probabilities. However it is not appropriate to design a classification scheme based on $p(\omega_1)$ and $p(\omega_2)$ only. We need to combine the observation component in the form of a feature vector $x$ in the decision making process. Since we are considering supervised learning mode, it means we are taking into account objects only for which decision had already been taken in the history. So we are selecting objects from both $\omega_1$ and $\omega_2$. After this we need to find the probability of the feature vector $x$ in both class $\omega_1$ and $\omega_2$, given by $p(x|\omega_1)$ and $p(x|\omega_2)$. These probabilities are known as class conditional probabilities (also known as likelihood functions because these probabilities tell us how $x$ is likely to be distributed in $\omega_1$ and $\omega_2$). Now the decision problem is as follows: For a new entrant, we measure the value of the feature $x$ and it is from the measurement of $x$, we will have to decide whether we would put this new entrant in class $\omega_1$ or class $\omega_2$. So we are interested in computing $p(\omega_1|x)$ and $p(\omega_2|x)$. Thus the decision rule proposed in standard Bayesian decision theory is as follows: If we find that $p(\omega_1|x) > p(\omega_2|x)$, then the decision goes in favour of class $\omega_1$. On the other hand if $p(\omega_1|x) < p(\omega_2|x)$, we put the new entrant in class $\omega_2$. From the multiplicative rule of classical probability, we can write:

$$p(\omega_i|x)p(x) = p(\omega_i \cap x) = p(x|\omega_i)p(\omega_i) \quad \forall i = 1, 2$$

From this relation, we can write:
Also from the law of total probability, we can write:

\[ p(x) = \sum_{i=1}^{2} p(x | \omega_i) p(\omega_i) \quad \forall i = 1, 2 \]

So we finally have:

\[ p(\omega_i | x) = \frac{p(x | \omega_i) p(\omega_i)}{\sum_{i=1}^{2} p(x | \omega_i) p(\omega_i)} \quad \forall i = 1, 2 \]

Here \( p(\omega_i | x) \) and \( p(\omega_2 | x) \) are called posterior probability. The result demonstrated above is Bayes theorem. Using Bayes theorem, we can compute the posterior probabilities from the prior probabilities and the class conditional probabilities. After that, we can implement the decision rule proposed earlier. For example, to decide in favour of \( \omega_1 \), we need to show that following condition prevails:

\[ p(\omega_1 | x) > p(\omega_2 | x) \]

This condition further implies:

\[ \frac{p(x | \omega_1) p(\omega_1)}{\sum_{i=1}^{2} p(x | \omega_i) p(\omega_i)} > \frac{p(x | \omega_2) p(\omega_2)}{\sum_{i=1}^{2} p(x | \omega_i) p(\omega_i)} \]

Cancelling the common denominator, we can write the above condition as follows:

\[ p(x | \omega_1) p(\omega_1) > p(x | \omega_2) p(\omega_2) \]

So under a general framework, the decision rule takes into account both the observation and the history. If prior probabilities are equal, history does not add value in the decision rule. Similarly if the class conditional densities are equal, we have to frame the decision solely based on history. Following this idea, in generalized Bayesian decision theory, we construct a function \( g(x) \), known as discriminant function, which computes the functional value \( g_i(x) \) \( \forall i = 1, 2, \ldots, c \) where \( c \) is the number of possible states of nature. The class which gives the maximum value of \( g(x) \), we take decision in favour of that class. Thus for a two-class problem we have two functional values of the discriminant function \( g(x) \), namely \( g_1(x) \) and \( g_2(x) \). The decision rule we are proposing is as follows: if \( g_1(x) > g_2(x) \) we take decision in favour of \( \omega_1 \) and \( g_1(x) < g_2(x) \) we take decision in favour of \( \omega_2 \). The equation of the decision boundary is \( g_1(x) = g_2(x) \). Since discriminant functions are not unique, we can choose a mathematically tractable form as per our convenience. One such discriminant function \( g(x) \) we propose here is as follows:

\[ g(x) = p(x | \omega_1) p(\omega_1) - p(x | \omega_2) p(\omega_2) \]
The biggest advantage of Bayesian decision theory apart from its intuitive appeal is that there is no other decision rule (classifier) with lower misclassification error rate when compared the same with Bayesian decision theory rule. However there are some practical challenges of applying the Bayesian decision theory in the real world.

Firstly, even though the computation of prior probabilities is straightforward, the same for the class conditional probabilities is not. There are a number of methods to determine the functional form of the probability density function in statistical literature. Unfortunately when compared, there is no unambiguous winner. Curve-fitting may not work under real life situations. So under practical circumstances, the estimation of the density function can be a time consuming task without certainty for absolute precision. Secondly, if the feature vector consists a mix of continuous and discrete variables, the decision surface may not be mathematically tractable. After all it is not straightforward to estimate the density form of the multivariate distribution when the individual random variables (which may exhibit collinearity) follow normal and binomial distribution (an example). Of course with rigour and advanced mathematical tools, such challenges can be conquered. But in real world situations, professional data consultant may not have the sufficient opportunity window to explore such labyrinths of technical nuances.

So the question is how we can overcome these difficulties in real life from an application perspective. In this paper, we have prescribed a data modelling framework algorithm where the probability density functions can be estimated using frequency probability stemming from statistically independent simulation exercise.

4. THE MODELLING ALGORITHM

For simplicity we assume that we have two different levels of consumer credit available. They are denoted by $\omega_1$ and $\omega_2$. So this is a two class classification problem. Here we are considering a two dimensional feature vector having the following components:

- **Fico Score** ($x_1$): Fico is a credit score created by the Fair Isaac Corporation to assess credit risk. Fico score is widely used in loan industry. It is perceived as a robust indicator of risk and thus has wide acceptability in the industry. Being a scorecard, Fico scorecard is a continuous variable.

- **Line Utilization** ($x_2$): As a general practice, line utilization is expressed a percentage of the consumer’s available credit that he or she had used. So traditionally it is also treated as a continuous variable. However, in our model, we assume that line utilization is an indicator variable having $k$ distinct levels. Each level specifies a range of line utilization. This specification makes the line utilization a discrete variable. So we can specify the space of line utilization as $\Omega = \{L_1, L_2, L_3, ..., L_k\}$.

In real world, both of these variables are available in the decision bureau database with certainty because of their ultra-frequent usage in credit strategy. In fact, line utilization has many variants in the bureau database. Without loss of generality, we are choosing one suitable representative as our $x_2$. So now we have a two dimensional feature vector $x = (x_1, x_2)$ based on which we need to design the classification problem. Since our objective is to prescribe a mathematically simple and easy to implement approach, we first create a partition $H$ of $x_1$ into $n$ number of intervals each having length $h_j$. The set of all the partitions is given as follows:
Here we have \( h_j = a_j - a_{j-1} \) \( \forall j = 1, 2, \ldots, n \). Let us denote the partition \( \{a_{j-1}, a_j\} \) by \( A_j \) \( \forall j = 1, 2, \ldots, n \). Here \( a_0 \) and \( a_i \) are globally accepted floor and ceiling value of \( x_i \). Now we introduce the set function \( f : \Psi \rightarrow N \) such that \( f(A_i) = i \) \( \forall i = 1, 2, \ldots, n \). Similarly we introduce the set function \( g : \Omega \rightarrow N \) such that \( g(L_j) = j \) \( \forall j = 1, 2, \ldots, k \). The Cartesian product between \( f(A_i) \) and \( g(L_j) \) can be defined as follows:

\[
\Theta = f(A_i) \times g(L_j) = \{(i, j)\} \quad \forall i = 1, 2, \ldots, n \quad \text{and} \quad \forall j = 1, 2, \ldots, k
\]

These functions are defined on the population data. Now using random sampling with replacement, we create \( p \) number of training samples each of size \( N \) from the population. So it is a randomized simulation exercise. Basically we are following the bootstrapping methodology to estimate the probability density function. Let us consider the \( m^{th} \) training sample. We introduce the real valued function \( h : \Theta \rightarrow R \) such that we have:

\[
h(ij)^m = \left( \frac{n_{ij}^m}{N} \right)
\]

Here \( n_{ij}^m \) represents the frequency count for the \( (i, j)^{th} \) group in the \( m^{th} \) training sample. So \( h(ij)^m \) represents the relative frequency of the \( (i, j)^{th} \) group in the \( m^{th} \) training sample. Given \( p \) is a large number, we can get an estimate of the probability of the \( (i, j)^{th} \) group in the population by averaging \( h(ij)^m \) over \( m \) \( \forall m = 1, 2, \ldots, p \). So we can write:

\[
P(ij) = \frac{1}{p} \sum_{m=1}^{p} h(ij)^m = \frac{1}{p} \sum_{m=1}^{p} \left( \frac{n_{ij}^m}{N} \right)
\]

Computing \( P(ij) \) \( \forall i = 1, 2, \ldots, n \) \( \text{and} \) \( \forall j = 1, 2, \ldots, k \), we get the probability distribution of the feature vector \( x = (x_1, x_2) \). Since for bureau information on the feature vector \( x = (x_1, x_2) \) is available for any new entrant, we can easily determine its crisp (binary) membership in the \( \Omega \) space. For \( \Psi \), we will go for a fuzzy membership. Formally a fuzzy set \( \tilde{S} \) is defined as the collection of ordered pairs \( \{(z, \mu_S(z)) : z \in Z\} \) given that \( Z \) is the universe of discourse and \( \mu_S(z) \) is the membership function. Bezdek (1993) stated that some plausible properties of the membership function may be (1) normality (2) monotonicity and (3) symmetry [11]. Suppose we consider a fuzzy set \( \tilde{B}_i \) defined by the imprecise linguistics “should belong to the group \( A_i \)” . The membership function we are proposing here is as follows:

\[
\mu_{\tilde{B}_i}(a_i) = \frac{1}{1 + (a_i - \overline{S}_{A_i})^2} \quad \forall a_i \in x_i
\]

Here \( \overline{S}_{A_i} \) is the arithmetic mean of the group \( A_i \) in the \( \Psi \) space. Here the membership function serves a twofold purpose:
1. We can assign \( a_i \) to group \( A_j \) if \( \mu_{\pi}(a_i) = \text{Max}\{\mu_{\pi}(a_i)\} \) \( \forall i, j = 1, 2, \ldots, n \)

2. We can use the chosen the \( \mu_{\pi}(a_i) \) as a weight on \( P(j) \) to scale the probabilities.

Thus once we assign the belongingness of a new entrant characterized by the feature vector \( x = (x_1, x_2) \) in both \( \Psi \) and \( \Omega \) space, we state its chance of occurrence as \( \mu_{\pi}(a_i) \cdot P(ij) \) in the \( \Theta \) space. Now we can concentrate on the computation of the class conditional probabilities. For that we partition the data for \( \omega_1 \) and \( \omega_2 \) individually. Suppose we create \( p_1 \) number of training samples each of size \( N_1 \) from the population corresponding to class \( \omega_1 \). Similarly we create \( p_2 \) number of training samples each of size \( N_2 \) from the population corresponding to class \( \omega_2 \). Following the methodology described above, we can obtain the scaled likelihood functions as follows:

\[
P^*(ij | \omega_a) = \left[ \mu_{\pi}(a) \right]_{\omega_a} \cdot \frac{1}{p_a} \sum_{m=1}^{p_a} [h(ij)^m] = \left[ \mu_{\pi}(a) \right]_{\omega_a} \cdot \frac{1}{p_a} \sum_{m=1}^{p_a} \left( \frac{n_{ijm}}{N_a} \right) \quad \forall a \in x_i
\]

So the discriminant function can be constructed as follows:

\[
g(x) = P^*(ij | \omega_1) \cdot P(\omega_1) - P^*(ij | \omega_2) \cdot P(\omega_2)
\]

Substituting the values of \( P^*(ij | \omega_1) \) and \( P^*(ij | \omega_2) \) in \( g(x) \), we get:

\[
g(x) = \left[ \mu_{\pi}(a) \right]_{\omega_1} \cdot \frac{1}{p_1} \sum_{m=1}^{p_1} \left( \frac{n_{ijm}}{N_1} \right) \cdot P(\omega_1) - \left[ \mu_{\pi}(a) \right]_{\omega_2} \cdot \frac{1}{p_2} \sum_{m=1}^{p_2} \left( \frac{n_{ijm}}{N_2} \right) \cdot P(\omega_2) \quad \forall a \in x_i
\]

The equation of the classifier is \( g(x) = 0 \). So in explicit form, we can write the classifier as follows:

\[
\left[ \mu_{\pi}(a) \right]_{\omega_a} \cdot \frac{1}{p_a} \sum_{m=1}^{p_a} \left( \frac{n_{ijm}}{N_a} \right) = \frac{P(\omega_a)}{P(\omega_1)} \quad \forall a \in x_i
\]

### 5. CONCLUSIONS

Thus in this paper, we have shown a practical approach to assign consumer credit limit to new customers using Bayesian decision theory and Fuzzy logic. The simplicity of the data modelling structure proposed may benefit the professional data consultants to come up with a precise solution without delving much into mathematical details. Another advantage of this algorithm is whenever there is any change in the environment; the classifier adjusts itself.
automatically so that misclassification error does not increase abruptly. This is certainly a big plus point over traditional risk scorecards and objective segmentation based on CHAID and CART.

REFERENCES


