Smart Phone Based Microcredit Risk Assessment using K-Means Algorithm

Girish Joshi (06329007)
Prathab K (06329902)
A. Radha (06329024)
KReSIT, IIT Bombay

November 28, 2007

Abstract

One of the key issues in microfinance is credit risk assessment. In this project, we propose to use K-Means algorithm to assist in making decisions regarding credit risk assessment. The choice of smart phones is due to its increased processing power and lesser cost.

1 Introduction

Finance is required for any business activity. With the proliferation of banking sector the financial needs of people are easy to satisfy. However, the banks require some form of assurance that their money is safe, when granting loans to people. Generally banks evaluate the risk in granting loan based on 3 C's, viz. Character, Credit and Collateral [1]. People from low income groups however, find it difficult to satisfy these requirements and hence most of them remain deprived of the benefits from banking. Microfinance is a term used to refer to the activity of provision of financial services to people who are excluded from the traditional financial system on account of their lower economic status [2].

However microfinance is not cure all. Microfinance is not a charity but it is indeed a business with different types of customers. Microfinance Institutions (MFIs) face similar or sometimes more serious problems when it comes to the risk associated with the money lended. For MFI to sustain its business and existence, the people taking loans and credit should be creditworthy. The 3 C's for assessing risk of loan are still applicable in Microfinance, but the attributes to be used for evaluation are different than those used by other financial institutes or banks [3]. Credit risk scoring involves evaluating a borrower with low income based on such attributes as

- Age of borrower
- Moral character of borrower
- Loan history
- Enterprising nature of borrower
- Social status of borrower

1.1 Scoring in Microfinance

Scoring assigns points to attributes such as the ones mentioned earlier which are associated with a loan, adds up the points and then links the points with a probability of going bad. A

simple scorecard can be constructed to find the risk that a loan, if approved, will at some point have 30 days of arrears. Here, more points corresponds to more risk. As an example consider a sector of business. The point scheme for this sector can be described as follows:

1. Retail or wholesale trade may receive 0 points, while services or manufacturing receive 3 points, indicating that they carry more risk.
2. Repeat applicants may get 0 points, and new applicants may get 2 points.
3. Repeat applicants also get one point (up to 7) for each day in the longest spell of arrears in their previous loan. (New applicants and repeat clients with no previous arrears get 0 points.)
4. Finally, applicants with a savings account get 0 points, while others get 4 points.

Going with this scheme, a new applicant in retail trade without a savings account would score $0 + 2 + 0 + 4 = 6$ points. A repeat applicant in manufacturing who had 4 days of arrears in the previous loan and who has a savings account would score $3 + 0 + 4 + 0 = 7$ points. For more information on other scoring schemes and scorecards refer to [4].

Experiments in some countries like Bolivia and Colombia suggest that scoring for microfinance based on such attributes can improve the judgment of risk and thus cut costs [4]. However, the banks and other financial institutes have rare presence in rural areas. The accessibility of the banks in such areas is thus a major problem, as the banks cannot provide their services in each and every rural area of the country. This inturn results into wide usage of unorganized MFIs which involve private money lenders or loan sharks. This sector being unorganized is not controlled and generally it is the poor who suffers as a result. A borrower in this sector generally provides some collaterals which are never returned by the lender. Moreover the interest rates charged are sometimes exorbitant.

2 Approach

Entry of banks into rural areas would thus help towards economic growth of this sector. To increase the access of the banks or MFIs to rural areas, smart phones can be used in the form of Decision Support System (DSS). The banking officer can not visit various rural areas of the country. However literate people with some knowledge of finance and smart phones can definitely become representatives of banks. Following are the steps involved in the usage of the application:

1. Some Representative of a Financial Institution like bank will visit rural areas on periodic basis. The person should be familiar with computer and mobile usage.
2. The Representative will carry with him a Smart Phone with The MicroCredit product loaded.
3. When there is some person who wants to take benefit of the Microcredit facility, the Representative will collect required information from him/her.
4. This information will then be fed to the application by the Representative.
5. Application will carry out a MicroCredit Assessment based on the information entered.
6. Finally, application will give the result based on the risk assessment carried out, as to whether the Person can get Microcredit facility from the Bank or not, and if possible how much amount can be given.

3 Credit Risk Assessment

To achieve the fast turn around time for the credit risk assessment, the microcredit scoring system is executed in the smart phone itself. The credit risk assessment decision system is implemented using K-Means algorithm.

3.1 K-Means Algorithm

K-means algorithm is an algorithm to cluster objects based on attributes into k partitions. It is a variant of the expectation-maximization algorithm in which the goal is to determine the k means of data generated from Gaussian distributions. It assumes that the object attributes form a vector space. The objective is to minimize the total intra-cluster variance, or, the function

$$V = \sum_{i=1}^{k} \sum_{j \in S_i} |x_j - \mu_i|^2 \qquad (1)$$

where there are k clusters S_i, i = 1,2,...,k and μ_i is the centroid or mean point of all the points $x_j \in S_i$ The algorithm starts by partitioning the input points into k initial sets, either at random or using some heuristic data. It then calculates the mean point, or centroid, of each set. It constructs a new partition by associating each point with the closest centroid. Then the centroids are recalculated for the new clusters, and algorithm repeated by alternate application of these two steps until convergence, which is obtained when the points no longer switch clusters (or alternatively centroids are no longer changed).

This algorithm can be formally be defined as given below: [5]
Input: The number of clusters k and a database containing n objects.
Output: A set of k clusters that minimizes intra-cluster variance.
Method:

1: Arbitrarily choose k objects as the initial cluster centers;
2: **repeat**
3: (Re)assign each object to the cluster to which the object is most similar, based on the mean value of the objects in the cluster. The similarity of objects to a cluster can be found out using equation 1. The object is most similar to a cluster for which value of V is minimum.
4: Update the cluster means, i.e. calculate the mean value of the objects for each cluster;
5: **until** no change in any of the cluster means

3.2 Using K-Means algorithm for Risk assessment

When applied to the particular problem in question, n objects correspond to n persons, whose data is accumulated in database. The k clusters correspond to the k levels of *Risk* assigned to borrower. Thus using an initial data about borrowers and K-Means algorithm, k clusters with their means can be obtained. Now for a new case in hand, it is required to find the closest cluster again using equation 1. The output indicates the % of closeness of new case to the known clusters. It is then upto the representative of MFI to decide upon whether to grant the loan or not to grant the loan. This program when deployed and run on a smart phone, the output looks similar to what is shown in the figure 1. The representative can use the output like the one shown in figure to further make some decision as to whether the loan can be approved or not.

Figure 1: Credit risk score

4 Further Enhancements

Many of the attributes used for assessing the risk of borrower in case of microfinance are subjective. K-means algorithm does not take into account the subjectiveness of the attributes. Thus a different approach such as Fuzzy clustering can be more useful. The reference [3] describes a variant of such method called Fuzzy C-means clustering.

5 Conclusion

We have used K-means clustering for delveoping cell phone based micro credit risk assessment. So it can helpful for low income people to easily access financial services from banks.

References

[1] BusinessTown.com,(2006),Getting a Bank Loan,[Online],Available:*http://www.businesstown.com/planning/creating-bank.asp*,

[2] Wikipedia,(2006),Microfinance, [Online], Available:*http://en.wikipedia.org/wiki/Microfinance*,

[3] Saket K.Sathe and Uday B.Desai, *Cell Phone Based Microcredit Risk Assessment using Fuzzy Clustering*,Electrical Engineering Department,IIT Bombay,2005

[4] Mark Schreiner,*Credit scoring for microfinance: Can it work?*, Journal of Microfinance, vol.2,no.2, pp.105-118,2000.

[5] Jiawei Han, Micheline Kamber *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.