Development of an Expert System for Credit Card Application Assessment

Tetsuo Tamai and Masayuki Fujita
Mitsubishi Research Institute, Inc.

Abstract

This paper discusses an application of practical expert system development in the finance industry. In particular, it describes the development of an expert system for credit card application assessment. In the problem domain of credit assessment, the method of 'credit scoring' has been widely used, but this system uses a 'profiling' method to simulate more closely the human process of decision making.The authors have adopted a decision tree as a form of knowledge representation for the profile design, and a decision tree generating algorithm is described. Finally, the implemented system, called ACAS-PRO, is outlined.

Keywords: expert system, decision tree, profiling system, scoring method, decision tree generating algorithm, credit card application assessment

1 Introduction

The manufacturing industry was the first to start exploring the possibilities of expert systems, developing typical systems like fault diagnosis expert systems for plant or machinery. Such problems have much in common with MYCIN, a famous expert system application for medical diagnosis.

Recently, much interest has been shown in the applications of expert systems to the finance industry. The MYCIN-type approach, however, is not necessarily suitable for this problem domain. This paper describes an application of practical expert system development in the finance industry.

There are various subjects in financial applications that are expected to be treated effectively by knowledge engineering approach: i.e. loan judgment, insurance underwriting, investment selection, assets liabilities management, bond/stock/foreign-exchange dealing support, and money market forecasting. Such application areas possess following properties.

1. Human factors play a certain role in their problem domains, thus, unlike the engineering domain, they do not necessarily follow rigid rules like physical or chemical laws.
2. The objectives of application systems development of such areas are to enhance the appropriateness and to ensure the uniformity of decision making rather than cost or labor saving, though the latter has been a principal goal of implementing online business systems or office automation applications, for which the financial industry has been spending such a huge energy.

Our problem, credit card application assessment, shares similar characteristics as mentioned above. Besides, it has some other features as follows.

1. From the beginning of this project, our expert system was planned and has been developed aiming at building a practical system, not just a prototype.
2. As a form of knowledge representation, the decision tree is adopted instead of other typical representation like the production system. Efforts are made in systematically generating an appropriate decision tree.
3. As the knowledge source, both human experts and the results obtained from analyzing the past credit applicants and their behavior data are exploited.

2 Problem Domain

2.1 Overview of Credit Card Application Assessment

For a credit card company, the task of screening out credit risky persons forms a crucial part of card application acceptance processes. Too strict judgment will result in the loss of expected profit opportunities; too loose acceptance will certainly bring large unredeemable credits.

The business process of credit card application acceptance usually proceeds as follows. An applicant fills in an application form and submits it to a credit card company. Items on the application form are about 20 to 30, including age, sex, address, affiliation, house status and others. Preliminary checking is done to find false information and examine past credit history, through calling to the phone number designated in the application sheet and accessing to a personal credit information center. Then, a credit officer investigates the application and decides acceptance or rejection. In this judgment, the officer matches the application to a certain pattern in his or her past experience that indicates the degree of credit risk. If the application is accepted, the related administrative process is taken and the card will be issued usually within a week.

2.2 Problems of Card Application Acceptance Procedure

Credit card business in Japan has relatively shorter history than in the United States. But the competitive entry into this market by the banking and the retailing industry has caused the ever increasing volume of card issues. This situation affected the application acceptance processes in the following ways.

1. Shift of stress from risk care to profit pursuit

   To be competitive in the "card war", more profitable customers must be acquired. It is often difficult to distinguish between the profitable customers and the risky ones, because both share the common characteristics of using their cards well. Therefore, the reliability of application judgment is required ever more strongly.

2. Requirement of speeding up card issue process

   There are many applicants who need a card for a specific immediate use such as travels abroad. In general, quick card issues are welcomed by customers. As the credit judgment process occupies a large part of the card issuing business, its speed-up is highly expected.

3. Increase in the number of applicants

   As the volume of card demands grows, the load on credit officers is increasing. An officer typically has to process a few hundred applications, sometimes as many as one thousand per day. As this work requires considerable experience, it takes time to supply new personnel. At the same time, over a half of cases expert officers handle are quite simple for them to judge so that they decide them almost mechanically. If such routine judgment is supported by computers, it will save much time of those officers.

4. Instability of judgment

   Most of the judgment are done by a single person. There is no way of completely getting rid of individual preferences and thus it is difficult to preserve stable and uniform judgment.

5. Difficulty in verifying acceptance criteria

   As t he current process depends on individual decisions, the results are not evaluated and verified systematically.

3 AI Approach to Profiling Method

3.1 Profiling System and Scoring System

When human assessors judge credit card applications, they have certain images of applicant profile patterns and credit level of each profile, either consciously or unconsciously. They build a profile of a given applicant from the description given in an application form, match it to a certain pattern and judge whether to admit it or to reject it, according to the credit value of that pattern.

Based on this observation, we chose to represent the decision procedure simulating this human process, and named it “Profiling System.” In the problem domain of credit assessment, the method of “credit scoring” has been widely used. It applies statistics theories and determines discriminant functions over a set of applicants properties to divide a good class and a bad class. We think our profiling method has some advantages over the scoring method as follows.

1. To apply the scoring method, some kind of measure on linear scale is required for each applicants property. If a given property is of a continuous nature like amount of income or deposit, there is little problem. But if it has a combinatorial nature, like the status of home or industry type of the company the applicant is working with, then some way of quantification needs to be taken. It may be easy to give certain values, but not always easy to sort it on linear scale.
2. As the discriminant function for the scoring method is usually linear, the effect of combination of properties is treated in a limited way. In reality, there are such judgment as: “for a young person, it is not uncommon to live in a rented apartment, but for a middle aged person with family, it may be considered as a minus point.” Such case can be appropriately treated in the profiling method.
3. A human assessor does not make judgment according to some kind of scoring process, thus it is difficult to verify the method by comparing it with human decision making. The profiling method is natural for human experts to assess and to give constructive adjustment.

3.2 Knowledge Acquisition from the Past Data

Criteria for defining profile patterns A profile is specified by a set of values of applicant attributes. For example:

- sex: male,
- age: 25–35,
- house status: own,
- years in the current employment: 5–10,

may determine one profile pattern.

A desirable profiling system is required to have the following properties .

1. Each profile should be classified as “good” or “bad” as clearly as possible. Judgment will be based on the past performance of the applicants that have been sorted into the profiles. Thus, the assessment of a new applicant that belongs to a certain profile pattern should become more precise if the profiles are divided into the good and the bad more distinctively.
2. Any applicant should be classified into exactly one profile. If an applicant belongs to two or more profile patterns, then his or her evaluation cannot be determined uniquely. On the other hand, if an applicant does not fit into any profile, it is impossible to judge him or her by this method.
3. Statistically significant number of data should be collected for each profile. If very few applicants of the past belong to some profile, the judgment based on that profile will be uncertain. While it is important to construct a well classified profile set, the size of each profile should not become too small.

Available data A large amount of data are available from the past history of credit judgment. Samples were chosen so that a significant volume of data exist for each of the three customers groups: i.e. normal members, defaulted members, and rejected applicants. Various attributes can be collected from application forms, from which effective items for distinguishing credits are chosen, based on statistic analysis. In our case, 15 attributes have been selected. For the admitted applicants, either normal or defaulted, data on credit card usage activities are available: i.e. amount of payment by card, cases of payment failure, and amount of defaulted loans if any. A criterion for classifying profiles must be determined, which can be calculated from these data. In our case, we adopted the ratio of defaulted accounts vs. total accounts to evaluate goodness/badness of a group sharing a certain profile.

Representation of profiles by a decision tree We adopted a decision tree as a form of knowledge representation for the profile design. A decision tree is suitable for representing classification-type knowledge. An attribute is assigned to each node and the outgoing edges from that node correspond to possible values the attribute can take so that the selection of an edge at each node, starting from the root node, guides the classification process. The selection path ends at a leaf (a node without outgoing edges), which determines a profile.

If we construct a tree, where the outgoing edges from any node cover the whole domain of the assigned attribute and there is no overlapping between the value ranges of any two edges, then the second required property we described above is evidently satisfied. If the number of statistical data that belong to some leaf is inadequate, we can backtrack the tree upward to find a node with an appropriate size of data and reduce the subtree under that node into one node. This operation makes the tree meet the third requirement of assuring significant statistics for each profile.

To satisfy the requirement of profile patterns to distinguish the good and the bad as definitely as possible, we developed an algorithm, which is based on the work on learning in the research of artificial intelligence.

Decision tree generating algorithm The idea of our algorithm comes from an algorithm developed by R. Quinlan [3]. Our objective is to obtain a set of profiles that discern the customers who failed to pay their bills from the normal customers, based on the accumulated data. The algorithm can be roughly stated as follows.

1. Let T be a tree consisting only of a root node, to which all the samples to be classified are assigned.
2. Repeat until a certain terminating condition holds.
   (a) Choose an appropriate leaf (a node not split yet, i.e. with no outgoing edges), that can be split. (If the size of assigned samples is too small, the node cannot be a candidate for splitting. Non-existence of candidate nodes constitutes one of the terminating conditions.)
   (b) Determine an attribute that most clearly distinguish the good and the bad of the samples assigned to the node. We adopt the decrease of entropy after the splitting by the attribute as a measure for selection.
   (c) Partition the samples of the node by the attribute values and generate a node assigned for each partitioned sample group, creating an edge between the split node to each new node.

Here, entropy in the sense of the information theory is used for evaluating the degree of separability by the properties. For a leaf i or a profile indicated by the leaf, its entropy is defined as:

$$E_i = \sum_j -p_{ij} \log p_{ij},$$

where j runs over the classes to be discerned: in our case, the "good" group and the "bad" group. p is a probability that a member who belongs to the leaf i (or profile i) is in the class j . For an entire tree, its entropy is defined as a probabilistic mean of entropy at leaves:

$$E = \sum_i p_i E_i,$$

where p_i is a probability that an arbitrary member falls into the leaf (profile) i. As the probabilities p_i's and p_{ij}'s are usually unknown, we use N_i/N for p_i and N_{ij}/N_i for p_{ij}, where N is the size of total sample, N_i is the size of sample belonging to the leaf i, and N_{ij} is the number of samples within i belonging to the class j.

When a node i is split into a set of nodes $i_1, \ldots, i_l$, the entropy is changed by

$$p_i E_i - \sum_{i_k} p_{i_k} E_{i_k},$$

which is always positive and means the increase of information gained by the attribute used for the splitting. Thus it can be used for measuring the effectiveness of partitioning choice. The entropy of the leaves of the final tree also implies the degree of classification. If it is near zero, the corresponding profile gives definite judgment either good or bad. On the other hand, if the entropy is close to l, its classification is fuzzy, which we call gray profiles.

Example Assume following unreal applicant samples and attributes of the application form.

- Samples 100 (50 normal, 50 failed)
- Attributes

1	sex:	male,	female
2	house status:	own,	rental
3	telephone possession:	yes,	no
4	age:	under 30,	30 or more

Distribution of samples among the above attributes are as shown in Table 1.

attribute	sex		house status		tel. possession		age	
value	male	female	own	rental	yes	no	$30 >$	$30 \leq$
normal	23	27	30	20	41	9	26	24
failed	27	23	20	30	9	41	24	26

Table 1: Sample Data Distribution

16 profile patterns are obtained from all combination of attribute values, whereas 8 profiles are created from the decision tree generation as illustrated in Figure 1.

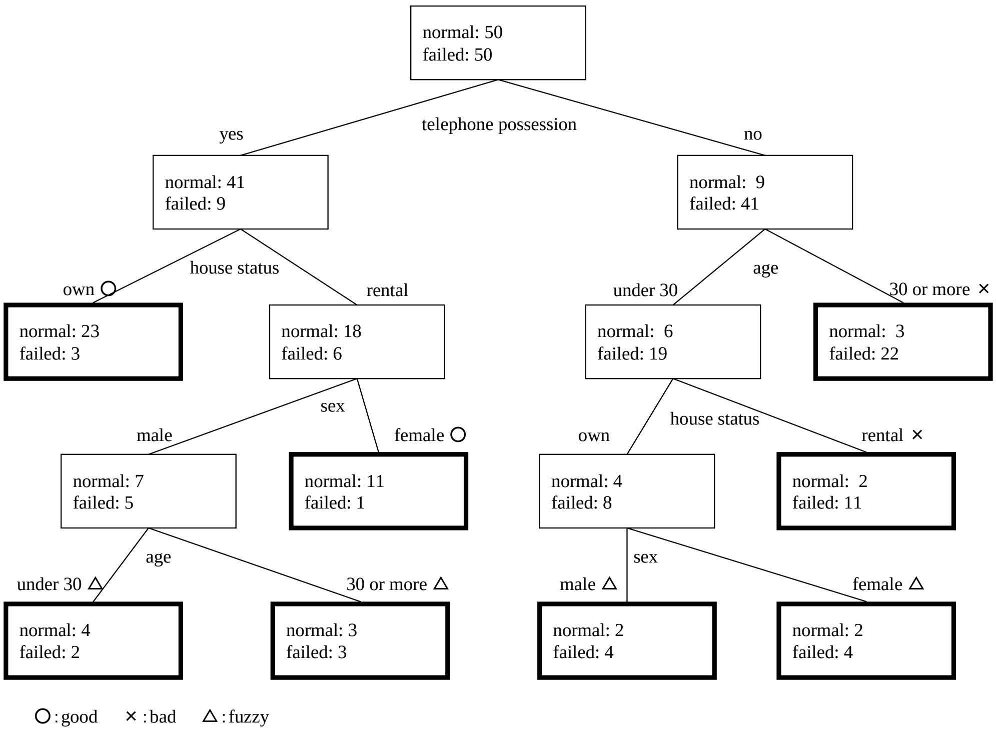

Figure 1: Example Decision Tree

Results are:

- number of profile patterns: 8
- well classified profile patterns: 4 (good 2, bad 2)
- number of samples occupying well classified profile patterns: 76
- well classified ratio of samples: 76%

Specifying attribute values For each attribute, its possible values should be either discrete or grouped into finite subsets; otherwise the edges generated from a node corresponding to selected attribute values cannot be confined to a finite number. However, the way of discretization is not always unique. Continuous attributes like the amount of income evidently have an infinite ways of partitions. Even for originally discrete attributes, it is sometimes desirable to reduce the number of values, because the node splitting by an attribute with many possible values will produce unnecessarily small sized nodes that harm the overall performance of distinction. In such cases, grouping of certain values are required but again there is no unique way of grouping.

Some procedures have been invented to make this type of decision based on reasonable criteria (e.g. [1]), but we adopted somewhat more naive method of asking to and discussing with credit officers, because such experts have certain intuitive ways of partitioning attribute spaces based on their experiences. We were afraid that if we determine the partition by some mechanical means, the results may not fit to the understanding of human credit officers.

Adjustment to the generated tree We chose a sample data set consisting of 2500 normal members and 1800 failed members. By applying the above described procedure, we obtained a set of reasonable profiles consisting of about 340 profile patterns. A problem with the result was that among 340 profiles, more than 70% were small-sized patterns with less than 30 samples sorted into each of them. Such small-sized profiles are undesirable, because their lack of statistical significance may cause inaccuracy in judgment.

Major factors of producing small-sized profiles are:

1. If attributes that have relatively many possible values are chosen in the earlier stage of tree generation (i.e. close to the root), sample data are divided into small groups too early, which does not only create many small-sized profiles but also inhibit the tree to grow deep enough to bring a variety of profiles combining many attributes.
2. Those classes which are difficult to judge credit tend to be partitioned many times to produce small-sized patterns.

As a measure to deal with this problem, we added the following processes.

1. Hierarchical divisions of attribute value groups
   For attributes with relatively many value classes, we specified two (or possibly more) layers, the higher one corresponding to rough classes merging some value classes and the lower corresponding to finer classes (usually the original values classes). For example, from an attribute A with 6 value classes, we make a new attribute A' with 3 value classes by properly grouping the original 6 classes and for the tree generation, we make a rule of preferring A' to A. A should be a candidate of partitioning only after A' is already used.
2. Subtree pruning
   When we get too small leaves, we can backtrack the tree to find an ancestor node with an appropriate size and merge the subtree under it to that node. This operation is often called "pruning".

Through these operations, we succeeded in reducing the tree to obtain 184 profiles, which were actually used as an initial profile set in our credit assessing system operation.

3.3 Knowledge Acquisition from the Experts

In addition to the profiles obtained from the past data analysis, we collected some profiles from human experts. We distributed questionnaire to credit officers and collecting personnel to specify their patterns of good or bad profiles. About 120 profile patterns were collected, though there were much overlapping between them and differences in specification levels. An experienced credit officer sorted the results with us and finally identified 3O profiles. We call them specific profiles, which do not cover the whole types of applicants but indicate important patterns to be used in the credit assessment process.

In the actual operation, these specific profiles are used for screening conspicuous patterns, and then profiles obtained from the data analysis are applied to give systematic information.

4 Implementation

The implemented system is largely divided into two parts. One is a knowledge acquisition subsystem, by which decision trees are constructed, profiles are determined, and statistics are gathered for those profiles. This subsystem runs on a mainframe, because it is supposed to deal with large volume of data up to 100,000 application samples for constructing an inductive decision tree. The other is an operation system for daily application assessment business. The latter is operated on a small business computer installed in the card company office environment. The profile specifications and related statistical data are transported from the mainframe system to this operation system.

In the knowledge acquisition subsystem, there is another important component: a simulation facility that supports building card application acceptance strategies. One typical strategic problem is "What is the proper value of application reject ratio?" Using this component, one can simulate the policy of increasing the reject ratio or conversely decreasing the ratio. The simulation is performed by changing the criteria for judging each profile class.

The system under operation is called ACAS-PRO, whose implementation was conducted mainly by Nissho Electronics, Corporation. The system uses profiles obtained both from an induction tree and from expert knowledge. It handles not only the decision process but covers almost all the procedures of application handling: from accepting application forms, checking the credit data supplied by the third party, determining their profile patterns, to issuing the cards. The final decision is still given by human assessors but the decision process is expected to improve greatly in speed, consistency, and accuracy.

5 Discussions

As mentioned in Section 3, the most widely-used method for systematic credit assessment, so far, has been the Scoring Method. It was first introduced by Bill Fair and Earl Issac, who spun off from Operations Research Division of Stanford Research Institute and established a venture company, Fair, Issac & Co. in 1956. It took some time before their systems were widely accepted by the financial industry, but now about one thousand scoring systems are installed by consumer banking and credit card companies over 15 countries.

In the scoring system, each attribute of an applicant is given an appropriate quantitative value, and those values are summed with certain weights multiplied to give a score. To determine the weights, discriminant analysis, a theoretically well developed method of statistics, is usually applied. A few thousands of past applicant data, both normal applicants and defaulters are used to determine discriminant functions. This method has a reasonable ground, as it is based on a clear statistical formulation. However, it has some drawbacks as stated in Section 3.

The decision tree generating method adopted in our profiling system is based on an algorithm developed by J.R. Quinlan as one of inductive machine learning approaches [3]. But its basic idea has been known for quite a while, especially in the area of statistical analysis (see e.g. [1]).

Carter & Catlett also used a similar algorithm to generate a decision procedure for credit card application assessment [2]. They built a prototype system based on 600 samples and showed the method performs satisfactorily on this problem. Though their system is not for real use yet, their work is quite similar to ours both in the target and the approach. However, these two works have been conducted independently .

6 Conclusions

We developed a method of evaluating personal credit based on the combination of statistical analysis of the past data and accumulated experience of experts. It is implemented in a credit card application process system, ACAS-PRO, and entered into operation at the beginning of December, 1987. This system is expected to bring the following benefits.

1. Provide structural views of customers through the segmentation by appropriate measures that fit purposes: i.e. risk reduction, profit increase, and identifying customers taste.
2. Provide means for accumulating, analyzing, and exploiting past data for strategic use. As the volume of data gets larger, the statistical significance will increase. Also, profiles can be

reconstructed upon new data, if considerable change of applicants characteristics is recognized or card admission policy is altered.

3. Support handling assessment clerical works and enable effective administration of the process.

The number of profiles is 214, at the moment the operation of the system started. In six months or a year, the application data will have the volume of 50,000 to 100,000. At that time, it will be appropriate to reconstruct the decision tree based on statistically more significant samples. The resulting profiles will increase in number and will provide more elaborate classification.

This methodology developed for credit card application assessment may be applicable to other business fields as well. Probable subjects in credit card business include marketing promotion, authorization, loan judgment, renewal judgment, and collection administration. Encouraged by the success in supporting the credit issuing process, we plan to expand the scope of this methodology, first dealing with the domain of marketing such as promoting the use of cards and direct marketing of specific goods or services.

Acknowledgments

This project was sponsored by and conducted through the cooperation with Diamond Credit Co., Ltd. We deeply acknowledge the expertise and suggestions given by the members in Diamond Credit, especially Director Kaiji Koike, Yukio Ooishi, Toshiki Kadoi, and Taminori Miyakawa.

References

[1] Breiman, L. et al.: Classification and Regression Trees, Wadsworth, Belmont, 1984.

[2] Carter, C. and Catlett, J.: Assessing Credit Card Application Using Machine Learning, IEEE Expert, Vol. 2, No. 3 (1987) , pp. 71–79.

[3] Quinlan, R.: Learning Efficient Classification Procedures and their Application to the Chess End Games, in Machine Learning: An Artificial Intelligence Approach, Tioga Publishing Co., 1983, pp. 436–482.

[4] Fujita, M. and Tamai, T.: Credit Card Application Assessment with a Profiling System, Proc. 5th Knowledge Engineering Symposium, The Society of Instrument and Control Engineers, 1987, pp. 135–140, in Japanese.

[5] Miyakawa, T.: The Objectives of the Credit and Customer Information System Developed at DC, Consumer Credit, November Issue, 1987, pp. 70-73, in Japanese.