Assessment of Mortgage Applications Using Fuzzy Logic
Swathi Sampath, V. Kalaichelvi

Abstract—The assessment of the risk posed by a borrower to a lender is one of the common problems that financial institutions have to deal with. Consumers vying for a mortgage are generally compared to each other by the use of a number called the Credit Score, which is generated by applying a mathematical algorithm to information in the applicant’s credit report. The higher the credit score, the lower the risk posed by the candidate, and the better he is to be taken on by the lender. The objective of the present work is to use fuzzy logic and linguistic rules to create a model that generates Credit Scores.

Keywords—Credit scoring, fuzzy logic, mortgage, risk assessment.

I. INTRODUCTION
Fuzzy logic has been widely used in Engineering and other aspects of technology which require modeling and control systems.

It is a form of approximate reasoning which is based on ‘degrees of truth’, as opposed to the usual Boolean or binary (0 or 1) logic, which the modern computer is based on. Most activities in the universe are not easily translated into the absolute terms of 0 or 1, hence making fuzzy logic a progressive attempt at better codifying and better explaining the reasoning processes and also providing an intuition-friendlier treatment of information.

II. FUZZY LOGIC IN FINANCE
The growing internationalization, the globalization of financial markets and the introduction of complex products have increased the volatility and the number of risks in the business environment [1]. One of the major thrusts of economic science is to describe the behavior of individual units such as consumers, firms, government agencies and their interactions. But a large number of economic or financial concepts are vague, or fuzzy in nature [2]-[4]. Fuzzy logic, if successful in supplanting mathematical methods, has the potential to be a very useful and powerful tool in financial analysis [5], [6].

There is no ideal method or framework for risk assessment [7]. Risk is about balancing strengths and weaknesses and weighting their interaction with each other. The failure process is influenced by many factors, internal and external, that cannot be precisely defined.

When dealing with risk, probability models are typically used. However, probability models built upon classical set theory may not be able to describe some risks in a meaningful way [8].

An incorrect understanding of cause-and-effect relationships also makes it difficult to assess the degree of exposure to certain risk types using only traditional probability models. Further, analytic dependencies among the variables of a process or a system are often unknown or difficult to construct [9].

The fuzzy logic rule base provides a framework in which experts’ input and experience data can jointly assess the uncertainty and identify major issues, thus making it easy to model risks that are not fully understood [10]. These models are in corporate information by describing them using linguistic terms, or ‘linguistic rules’ (If-Then) to explicitly consider the underlying cause-and-effect relationships and recognize the unknown complexity. The ability to utilize linguistic rules is an advantage of fuzzy rule based systems over other information processing systems [11].

It is found that such models are more adaptable to cases with insufficient and imprecise data [12]. Data reported in financial statements may not be exactly comparable due to differences in accounting practices and may include inaccuracies in reported numbers. The observed value may thus be better considered as a fuzzy phenomenon, which means employing the use of an interval instead of a single value, for financial variables.

Using a fuzzy model in a problem relating to finance has the advantage of being faster and more accurate [13], as there now exists a method to define customer attributes by quantifying the approximate values of these attributes using fuzzy variables and rules.

III. PROPOSED FUZZY MODEL
As is the case in modeling any Fuzzy Logic Controller, the steps followed are the usual. The first step is to formulate all the influencing factors and how they affect the output. Once the required set of inputs and outputs are established, modeling of the system can be accomplished using any kind of fuzzy software. For the purpose of this paper, the fuzzy logic toolbox in MATLAB was used to create the Fuzzy Inference Systems for generating each of the outputs. The model was structured as is seen in Fig. 1.
The inputs and outputs for the fuzzy inference system are as follows.

A. Input Variables

1. Consumer Evaluation [0, 10]
   - Demographics [0, 10]
     - Age [18, 65]
     - Education [0, 3]
     - Marital Status [0, 1]
     - No. of Children [0, 5]
   - Finance [0, 10]
     - Income [1000-90000]
     - Length of Employment [0-15]
     - Type of Employment [0-2]
   - Financial Security [0, 10]
     - Current Living Arrangement [8000, 100000]
     - Value of Car [10000, 100000]
     - Value of Assets [5000, 45000]

2. Market Value of House [90000, 15000]

3. Income [1000, 100000]

4. Interest on Loan [0.2 – 10]

B. Output Variable

1. Credit Score [0, 10]

Once there is a clear understanding of the input and output variables, the range for each variable (according to their respective units) is approximated, as is seen next to the inputs and the outputs. For example, the range for the variable Age, is 18 years to 65 years, which is a norm followed by most banks.

Variables like Demographics and Consumer Evaluation have a range from 0-10 as they have normalized values following from the value that is evaluated from the fuzzy evaluation of (Age, Education, Marital Status, No of Children) and (Demographics, Finance, Financial Security), respectively. Following this, the Universe of Discourse of each fuzzy variable is partitioned into a number of fuzzy sets, assigning each a linguistic label. A set of ranges is also identified for each linguistic label. As can be seen in the code in Fig. 2, the variable Demographics was partitioned into 5 parts, and each was assigned a linguistic label, such as Weak, Medium, Average, Par and Strong.

![Fig. 2 The code for the Demographics ‘.fis’ file in MATLAB showing the ranges for each of the linguistic labels assigned](image)

The illustration of the ‘Triangular’ membership function as seen on the MATLAB GUI, for the variable Consumer Score, with five linguistic labels, High Risk, Average Risk, Medium Risk, Par Risk and Low Risk.

Different shapes can be used for forming membership functions, such as ‘Gaussian’, ‘Trapezoidal’, ‘Bell curve’, etc. The triangular membership function was used to make the model owing to its simple formula and computational efficiency [14].

The next step is creating a fuzzy rule base of If-Then rules, in order to determine how a variable affects the outcome. The following are some of the rules that were used to determine the Demographics score of an applicant.

If (Age is Young) and (Education is High) and (Marital Status is Single) and (Children is Few) then (Demographics is Average)

If (Age is Middle Aged) and (Education is Basic) and (Marital Status is Single) and (Children is Few) then (Demographics is Weak)

Each element in this credit scoring model has a detailed rule base, spanning from 52 rules (in the Demographics Fuzzy rule base) to 142 rules (For the Consumer Evaluation rule base).

It is possible to view a graphical illustration of the relation of each variable with the outcome in relation to the rule base, on MATLAB.

It is possible to observe the changing values of Income and Length of Employment on the Finance factor of an individual, as seen in Fig. 4.

As the Income increases, it is considered to be a favorable factor for the Finance score of the individual as it reflects on a better ability to repay the lender.
An increase in the Length of Employment is also a favorable factor as an increased duration of employment can be inferred as financial stability.

Fig. 4 A graph depicting the relations between the linguistic labels Income and Length of Employment and how they affect the output label Finance

Fig. 5 (a) Crisp values of Income and Length of Employment as entered to determine the Finance score

Fig. 5 (b) The previous 2 variables, along with Type of Employment are put through the fuzzy system to generate the Finance score

After the system has been designed in its entirety, crisp inputs can be entered to get the output of the credit score. The system will fuzzify the inputs, calculate the membership degree of each, evaluate them from the rule base, determine which rules fire, and finally obtain a fuzzy output, which will undergo defuzzification. There are different methods that can be used, such as ‘Bisector method’, ‘Weighted Average’, ‘Fuzzy Mean’, and ‘Mean of Maxima’, etc. For the purpose of this paper, defuzzification was carried out using the ‘Centroid’ method. Figs. 5 (a) and (b) illustrate the defuzzification GUI on MATLAB. When the inputs Income=45000, Length of Employment=14.7 years and Type of Employment =1 are entered, the fuzzy rule base generates a Finance score of 5.39 for the individual.

For the purpose defuzzification in this model, the centroid method was used, as it is generally better compared to the other methods in relation to consistency in results [15].

IV. TESTING THE SYSTEM

A list of 100 mock data was fed into the system to obtain the credit scores. The following representation of data is a continuous representation of the different inputs and how they influence the scores. In Table I, the Demographics scores are shown as an influence of the 4 Demographics inputs. Tables II and III depict the Finance and Financial Security scores respectively, and Table IV shows the Consumer Evaluation scores, which is a fuzzy function of the Demographics, Finance and Financial Security scores. In Table V the Consumer Evaluation score is weighed in with the Market Value of the House, Income and Interest of the loan to get the final Credit Score.

<table>
<thead>
<tr>
<th>Table I</th>
<th>Demographics Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Education</td>
</tr>
<tr>
<td>24</td>
<td>1.2</td>
</tr>
<tr>
<td>65</td>
<td>1.1</td>
</tr>
<tr>
<td>19</td>
<td>2.1</td>
</tr>
<tr>
<td>44</td>
<td>2.8</td>
</tr>
<tr>
<td>42</td>
<td>1.1</td>
</tr>
<tr>
<td>22</td>
<td>1.6</td>
</tr>
<tr>
<td>64</td>
<td>1.7</td>
</tr>
<tr>
<td>59</td>
<td>2.6</td>
</tr>
<tr>
<td>45</td>
<td>2.4</td>
</tr>
<tr>
<td>43</td>
<td>2.4</td>
</tr>
<tr>
<td>47</td>
<td>2.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II</th>
<th>Finance Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>Employment Length</td>
</tr>
<tr>
<td>31600</td>
<td>9</td>
</tr>
<tr>
<td>2400</td>
<td>8</td>
</tr>
<tr>
<td>18000</td>
<td>15</td>
</tr>
<tr>
<td>12000</td>
<td>15</td>
</tr>
<tr>
<td>1200</td>
<td>8</td>
</tr>
<tr>
<td>4200</td>
<td>2</td>
</tr>
<tr>
<td>3900</td>
<td>2</td>
</tr>
<tr>
<td>5200</td>
<td>4</td>
</tr>
<tr>
<td>8500</td>
<td>13</td>
</tr>
<tr>
<td>9000</td>
<td>1</td>
</tr>
<tr>
<td>14000</td>
<td>12</td>
</tr>
</tbody>
</table>
## Results and Discussion

Once the credit score has been determined, the lender can decide which applicant will best serve his/her purpose, by sorting the credit scores. The customers with the higher credit scores will be those that would have a better ability to repay the lender, or be beneficial to the lender in terms of lower risk scores when sorting the credit scores. The customers with the higher credit scores will be those that would have a better ability to repay the loan, and a shorter repayment term.

An important point to consider after formulating the input variables is to see how they will affect the outcome. For example, a healthy *Income*, longer *Employment Length* and a stable *Employment Type* would give a positive Finance score.

A healthy *Income* reflects on good chances of repaying the loan, a longer *Employment Length* and a stable *Employment Type* reflects on job stability.

Some variables such as *Employment Type* were given different values which cannot usually be quantified, such as around 0.5 would mean an internship, around 1 would mean a part-time job, and around 2 would mean a full time job.

As we have understood, Finance and Accounting are influenced by many aspects that are a direct influence of human behavior. The *Employment Type* instance brings in Fuzzy Logic, with its advantage of using linguistic rules, to simplify the manipulation of those which cannot usually be quantified or solved by use of an analytic method/mathematical equation.

Finally, we see that a favorable consumer score, a reasonable market value of the house in question, a proper income and a realistic interest on the loan, each count towards a favorable credit score.

One of the main advantages of this method can be seen as the ability of *Fuzzy Logic* in being able to break down a complex problem into simpler sub-problems.

## Conclusion

The field of Fuzzy Logic has come far in proving its usefulness as an aid to researchers and engineers alike, in its pursuit to help its user to gain an in-depth understanding of real world occurrences that are often affected by a host of different and complex factors that cannot be given a tag of a crisp number or process.

Fuzzy logic is now a lot easier to use due to the development of tools such as MATLAB, which was used to create the Fuzzy Inference Systems for this project.

Thinking about which factors are the most important and which should be used for the modeling is the most important step. One major drawback is the time and skill needed to form the fuzzy rule base. It takes a considerable amount of energy to skim through data to determine the relations between the different variables and formulate the rules. The formation of the rule base for this particular project spanned across 3 months.

Regardless, the amount of time taken to get the initial data and rule base ready is not without the obvious advantage of the Fuzzy model in being able to solve complex problems in a fast and efficient manner, which is uncharacteristic of traditional probability/mathematical models.

## References


