Bankruptcy Prediction System with Fuzzy Logic

Low Q.W., Chin K.Y., Chong W.J., Nur Farahin Binti Abdul Halim.

***Abstract*—Studies have shown that qualitative data mining approaches are capable of eliciting and representing experts' domain knowledge. The paper presents 3 different approaches of building a bankruptcy prediction system. The result shows that Neural Network is better at predicting bankruptcy status than the fuzzy counterparts.**

Keywords—Bankruptcy prediction; Fuzzy Inference System; Fuzzy Clustering Method; Neural Network; Risks.

I. Introduction

Corporate bankruptcy triggers economic losses for management, stakeholders, employees, customers and others, together with great social and economic cost. Risk assessment is critical because the complicated network of mutual credit obligations can make the actual risk exposure of the entire system invisible at the level of individual institutions. Although sophisticated data mining algorithms have been developed to help assess risk in an unprecedented accuracy, in an actual risk assessment process, the discovery of bankruptcy prediction knowledge from experts is still regarded as an important task because experts' predictions depend on their subjectivity [1]. This paper demonstrates 3 different approaches of building a bankruptcy prediction system according to attributes that came together with the dataset, which we will discuss in the sections below.

II. Data Exploration

A. Understand the Attributes

The Qualitative_Bankruptcy dataset is being acquired from the UCI Machine Learning repository. The dataset contains 6 risk factors which are Industry Risk (IR), Management Risk (MR), Financial Flexibility (FF), Credibility (CR), Competitiveness (CO), and Operating Risk (OP). All 6 factors are made up by 3 attributes, which are Positive (P), Average (A), and Negative (N). The parameters from this dataset is referred from the paper *The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms* [2]. Also, the dataset contains 250 instances with 2 classes which is Bankruptcy (B) and Non-Bankruptcy(NB). Table I shows an example row of the dataset.

TABLE I. Example Data Row

IR	MR	FF	CR	CO	OP	Class
N	N	P	A	A	N	NB

B. Risk Factors

According to the paper [2], experts evaluate the qualitative risk factors through the risk estimation process and assign appropriate levels such as positive, average and negative to these factors using their subjective knowledge. This section explains each of the risk factors in detail. Industry Risk is measured by the stability and the growth of the industry, the degree of competition within the industry, and the overall conditions of the industry. Management Risk is concerned with the efficiency and stability of management and organization structure. It is measured by the ability of management, the stability of top management, the stability of organization structure, management performance, and the feasibilities of business plans. Financial Flexibility means the firm's financing ability from direct and indirect financial market and other sources such as affiliates and the third parties. Credibility is concerned with the reputation of a company associated with credit history, reliability of information provided by the company, and the relationship with financial institutions. Competitiveness means the degree of competitive advantage determined by market position and the capacity of core technology. Operating Risk is the volatility and stability of procurement, the efficiency of production, the stability of sales, and the efficiency of collection policy of accounts receivable.

III. System Building

There are several method when it comes to building a fuzzy logic system. One can use the *Fuzzy Logic Designer* in *MATLAB* where it is possible to tailor a custom membership function and design rules for the input [3]. One of the obvious advantages is the Fuzzy Logic Designer has a beginner friendly GUI interface and it is possible to create a fuzzy system without any prior knowledge of coding. Another fuzzy system building method is by using any of the automated methods. For instance, Batch Least Square algorithm constructs a fuzzy model from numerical data which can be used to predict outputs given any input [4]. In this paper, we adopted 2 different fuzzy model building approaches, 1 neural network model and compared their effectiveness.

A. MATLAB Fuzzy Toolbox

With Fuzzy Toolbox, we built a mamdani based fuzzy system. We defined membership functions for all 6 risk factors (variables) in the *Membership Function Editor*. A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. For each factor, we defined 3 different Gaussian membership functions that represents Negative (N), Average (A), and Positive (P). The input ranges from 0 to 10, as illustrated in figure 1. It is worth noting that all input variables are weighted equally and have the same input range. The output variable is also defined as a gaussian distribution function in the editor.

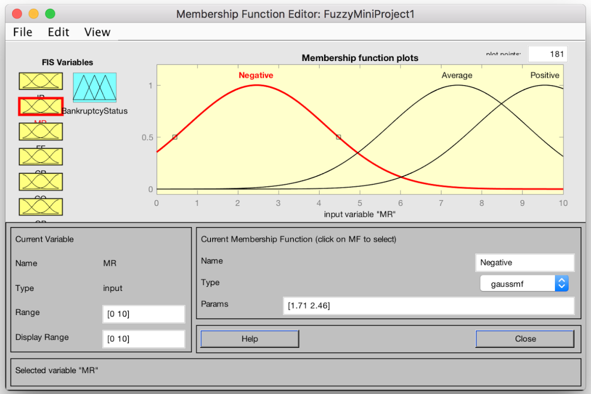

Fig. 1. Membership Functions.

Not only that, it is required to design our own logic (rules) before executing the model. For instance, if an institute is identified with positive Financial Flexibility and positive Competitiveness, then we can construct a rule to represent the expert knowledge in the fuzzy system. In an actual risk assessment process, the discovery of bankruptcy prediction knowledge from experts is still regarded as an important task because experts' predictions depend on their subjectivity [5]. To effectively and reasonably represents expert knowledge as rule in the fuzzy system, we referred and adopted a Genetic Algorithm based data mining method to extract decision rules from experts' qualitative bankruptcy decisions [2]. A total of 12 summarised rules are generated from the aforementioned method. The GA-based rules are tabulated in table II.

TABLE II. GA-BASED DECISION RULES

RULE	DESCRIPTION
1	IF FF=(A || P) && CR=(A || P) && CO=(A || P) THEN NB
2	IF FF=(N) && CR=(N) && CO=(N) && OP=(N) THEN B
3	IF FF=(P) && CO=(P) THEN NB
4	IF IR=(A || P) && CR=(A || P) && CO=(P) THEN NB
5	IF IR=(A || P) && MR=(A || P) && FF=(A | P) CO=(A || P) && OP=(A || P) THEN NB
6	IF MR=(A || P) && CR=(A || P) && CO=(A || P) THEN NB
7	IF MR=(N || A) && FF=(N) && CR=(N) && CO=(N) THEN B
8	IF IR=(P) && MR=(A || P) && CO=(P) THEN NB
9	IF IR=(A || P) && CO=(P) && OP=(A || P) THEN NB
10	IF MR=(P) && FF=(N) && CR=(N) && CO=(N || A) && OP=(N && A) THEN B
11	IF IR=(P) && MR=(N) && FF=(N) && CO=(N) THEN B

Each rule generated from Genetic Algorithm is composed by multiple rules, which can be decomposed into 67 different rules. Each decomposed rule is then inserted into the *Rule Editor* manually, as shown in figure 2.

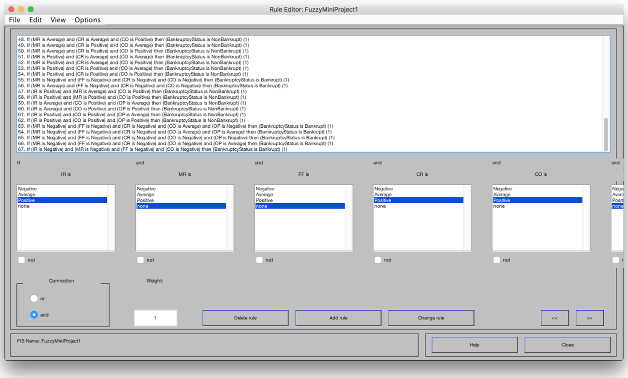

Fig. 2 .Fuzzy Rule Editor.

Similarly, the membership functions can be visualised with the Surface Viewer, as shown in figure 3. For Mamdani systems, the range of output depends upon both the input and the output MF ranges, since both inputs and outputs are fuzzy sets.

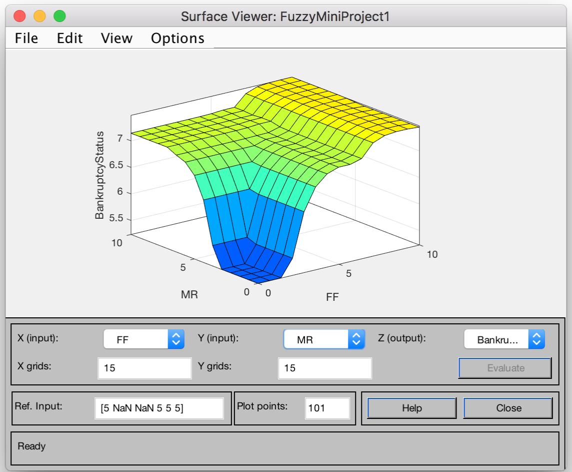

Fig. 3. Membership Function Surface Viewer.

B. *FCM with MATLAB GENFIS3*

Apart from the Fuzzy Toolbox, MATLAB also provide a command-based fuzzy system generation tool, *genfis3*. *genfis3* generates an Fuzzy Inference System (FIS) using fuzzy c-means (FCM) clustering by extracting a set of rules that models the data behavior. The function requires separate sets of input and output data as input arguments [6]. All data are converted to a numerical matrix, as *genfis3* can only accept numerical data for clustering. The number of clusters determines the number of rules and membership functions in the generated FIS. As a result, *genfis3* generates 6 membership functions and 2 rules, as illustrated in figure 4. The effectiveness of this method will be discussed in the discussion section.

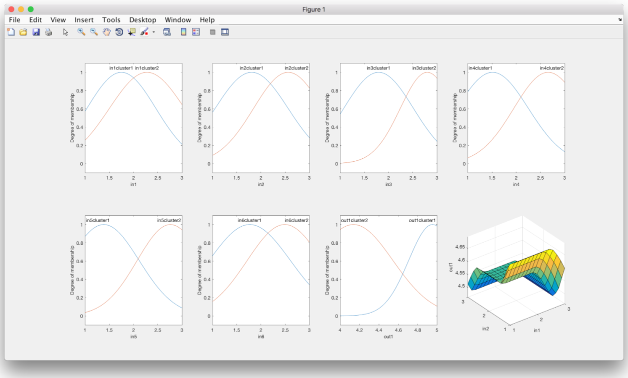

Fig. 4. Visualization of genfis3 results.

C. *RapidMiner*

In Rapidminer, we built a neural network model and tested the model's accuracy. In data preprocess phase, neural network model required numerical training data to build, thus Nominal to Numerical operator with parameter coding type of unique integers is applied to transform data type of non-numerical attributes to a numerical value. In model building and testing phase, nested operator Cross Validation is used. Cross Validation training and validate the neural network model. The trained model is then applied in the testing subprocess. The performance of the model is measured during the testing phase by using Apply Model and Performance operator, as illustrated in figure 5.

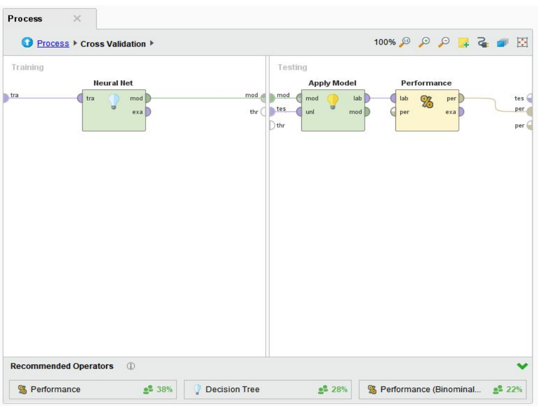

Fig. 5. Operators in nested operator Cross Validation.

IV. Results and Discussions

All three of the approaches we worked on were able to achieve a significant but rather debatable result. This section serves to discuss the results of all 3 aforementioned approaches. With the Fuzzy Toolbox, we defined 3 membership functions for each risk factors. The membership functions are built on the Gaussian distribution curve and we can expect a smooth fuzzy inference system. The fuzzy sets vaguely describe risks concepts like industry risk, competitiveness and all other risk factors. As for the GA generated rules, the antecedent of a rule can have multiple parts:

if IR is P and MR is A or P and CO is P then NB

in which case all parts of the antecedent are calculated simultaneously and resolved to a single number using fuzzy operator that are described in the Fuzzy Toolbox. For instance, applying OR (max) operator onto the antecedents will result in a selection of maximum membership degree between the antecedents. Next, use the degree of support for the entire rule to shape the output fuzzy set by applying implication method. By default the Fuzzy Toolbox provides *min* function or scaling using the *prod* function. The result is shown in figure below.

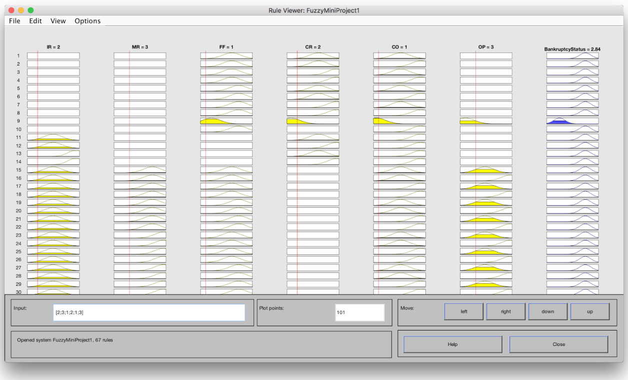

Fig. 6. Fuzzy Rule Viewer.

The testing input of [2;3;1;2;1;3] yields a BankruptcyStatus of 2.84, which indicates Bankrupt.

As for MATLAB *genfis3* command-based tool, the output are limited to between 4 and 5, where values near 4 indicates NonBankrupt while values near 5 indicates otherwise. The testing input of [2;3;1;2;1;3] yields a BankruptcyStatus of 4.7107, which predicts high degree of Bankruptcy. A 80/20 cross validation was ran and we concluded with 96% of accuracy. However, upon further inspection of the data, we reckoned that our data (which is categorical data) tends to cluster really bad beyond counting duplicates. It is simply irrelevant to do fuzzy clustering on our dataset.As a result, our FCM produced a rather rigid calculation due to the nature of our data. We believe our work can be improved in the future by using an Adaptive Neuro-Fuzzy Inference System (ANFIS) as a classifier.

Results in Rapidminer are pretty straightforward, a neural network is built with the data as in Figure 7. Input layer consist of 7 input nodes which are the 6 risk factors in dataset and 1 threshold node, one hidden layer and output layer with two nodes, only two results would be output by the model which are Bankrupt and NonBankrupt.

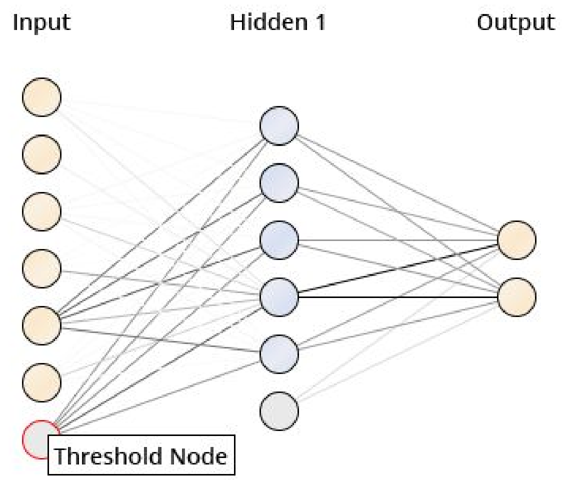

Fig. 7. Neural network built in Rapidminer.

The model is tested using cross validation of 10 folds, we get accuracy of 96.8% with slightly higher class precision and class recall of NonBankrupt, as confusion matrix shown in Figure 8. We can conclude that the neural network model trained performs better in predicting class NonBankrupt.

accuracy: 96.80% +/- 3.49% (mikro: 96.80%)

	true NB	true B	class precision
pred. NB	139	4	97.20%
pred. B	4	103	96.26%
class recall	97.20%	96.26%	

Fig. 8. Confusion Matrix of the Neural Network model.

V. Conclusion

3 different approaches were adopted to predict bankruptcy status given 6 different risk factors. It is apparent that both fuzzy system created can be used to predict bankruptcy status with high accuracy. Genetic Algorithm generated rules are used in building the first fuzzy inference system, where they represent the qualitative expert knowledge in the domain of risks and banking. We also came to acknowledged that the Qualitative_Bankruptcy is not suitable for fuzzy clustering tasks due to the nature of data. Lastly, we built a Neural Network, as a non-fuzzy approach to compare the effectiveness against fuzzy method. It is evident that the neural model performs consistently better than the fuzzy counterparts.

VI. Acknowledgement

We would like to acknowledge Universiti of Teknikal Malaysia Melaka for providing us the opportunity and resources to conduct this analysis. Also, we would like to thank our lecturer Associate Professor Dr Choo Yun Huoy for her willingness to guide us and her strong data mining knowledge. We are truly grateful to have her as our Machine Learning lecturer. It is a true honor to conduct this analysis under her supervision. Additionally, we would also like to express our gratitude to our parents for their unconditional spiritual supports. We would never achieve this far without all their supports.

References

[1] Risk Assessment for Banking Systems. Retrieved from https://www.fdic.gov/news/conferences/finance_banking/elsinger.pdf. Accessed 26 Dec 2017.

[2] Myoung-Jong Kim, Ingoo Han. The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms. Accessed 26 Dec 2017.

[3] MathWorks. Build Fuzzy Systems Using Custom Functions. https://www.mathworks.com/help/fuzzy/building-fuzzy-inference-systems-using-custom-functions.html. Accessed 9 Jan 2018.

[4] Timothy J. Ross. Fuzzy Logic with Engineering Applications. Page 205. Accessed 9 Dec 2018.

[5] Helmut Elsinger, Alfred Lehar, Martin Summer. "Risk Assessment for Banking Systems". Accessed 9 Jan 2018.

[6] MathWorks. Generate Fuzzy Inference System structure from data using FCM clustering. Retrieved from https://www.mathworks.com/help/fuzzy/genfis3.html. Accessed 9 Jan 2018