
To Create and Compare the Predictive Accuracy of a Genetic Program and an Artificial Neural Network to Predict Company Failure: Final Report

CARL SAPTARSHI
STUDENT NUMBER: 640032165

APRIL 2017

Contents

1	Background and Introduction	1
1.1	Background into Bankruptcy	1
1.1.1	What is Bankruptcy?	1
1.1.2	Who does it affect?	1
1.2	Algorithms for CF prediction	2
1.3	Motivations For this Project	2
2	Summary of literature review and specification	3
2.1	Literature Review	3
2.2	Project Specification	4
3	Design	5
4	Development	6
5	Testing	7
6	Description of the final product	8
7	Evaluation of the final product	9
8	Critical assessment of the project	10
9	Conclusion	11
10	Yalda is great	12

1 Background and Introduction

Due to the dynamic and volatile economy that we live in, the number of companies filing for bankruptcy are on the rise, especially during times of economic uncertainty, for example, during a period of recession.

In turn, being able to predict the likelihood of a company failing and filing for bankruptcy is very important and has been a focal point of issue in accounting research and analysis over the past thirty years.

1.1 Background into Bankruptcy

1.1.1 What is Bankruptcy?

When a company (the *debtor*) takes out a loan or borrows money from somewhere else such as a financial institution like a bank (the *creditor*), it is up to the debtor to ensure that the creditor is repaid the full amount that was borrowed subject to the creditors terms and conditions.

If the debtor starts to fall behind on their payments to the point that they are unable to repay their debts or unable to keep up with the incremental loan repayments, the debtor may file for a Chapter 7 bankruptcy in which the court will appoint a trustee to shut down the company and liquidate their assets, for example, by selling machinery, land and company shares to recover some money which they the trustee can give back to the creditor to clear the company's debt. If the company is still unable to pay back the debt even after this, then they will file for bankruptcy and the company will be terminated. Since the 1960's, as economies have grown, especially in the western world, this problem has been recognised on grander scale in more economically developed countries.

1.1.2 Who does it affect?

Bankruptcy does not just affect those that are employed within that company; it also affects third party members such as shareholders, investors, suppliers, and company clients. CF has been a critical point of focus, especially in the field of financial analysis and for stakeholders who are interested in the performance of the company. Since the 1960's, empirical risk assessment models have been developed which have been used to predict CF. Using current and previous financial data of a company, the likelihood of CF can be predicted for 'n' number of years ahead. This in turn means that loan companies, such as banks, can use this information to determine whether loans should be granted to other firms, knowing the likelihood of a company defaulting or not. This has helped to give banks competitive advantages, as they become aware of how likely a company will be to default, and therefore is able to predict customer behaviour in times of difficulty.

Looking at reports from the American Bankruptcy Institute showed that in the year 2000, 35,742 companies filed for bankruptcy, 43,546 companies in 2008 and 60,837 by 2009, at the peak of the recession. By 2012 this fell to 40,075 and 24,114 by 2016. These statistics clearly indicate the volatility and uncertainty in the economy as it changes, which is part of what makes CF prediction incredibly important, especially for those involved with the company at hand.

1.2 Algorithms for CF prediction

Since the aim of the project to classify whether a company is likely to fail or not, this can be called a binary classification problem which will take in a series of inputs and return a classification which determines whether a company is financially distressed or not. On top of this, a company likely to suffer from financial distress will have certain characteristics associated with them, similarly this would occur for companies not facing this problem. This means that the data should be linearly separable when classifying the data, making this a linear binary classification problem.

There have been several techniques that have been used to predict CF, some of which will be introduced here.

Individual Ratio Selection (IRS) - In the 1960's, Beaver introduced IRS. This process involved selecting thirty financial variables, converting these to ratios. Based on a certain threshold for each variable, this would determine if a company is likely to fail or not.

Multivariate Discriminant Analysis (MDA) - Altman created this technique in the 1960's, which takes uses a discriminant function to score a company. This function uses five financially weighted ratios and based on the overall discriminant score, the company can be classified.

Genetic Programming (GP) - GP's are inspired by Darwin's Theory of Evolution, used for prediction and classification. A population of functions is created and fitter individuals in the population are more likely to survive and produce offspring that are even more suited to the environment. The aim is to create an optimal function which can give the most accurate prediction in terms of classification accuracy.

Artificial Neural Networks (ANN) - This technique is inspired by the interconnectivity of the brain and applies this to prediction and classification. ANN's are made of an input layer, hidden layers and the output layer which outputs the classification based on the input. The network uses the weights on the synapses of the nodes that connect one node to the next, which are tweaked to allow the network to learn and give a more accurate classification for unknown datum.

1.3 Motivations For this Project

2 Summary of literature review and specification

2.1 Literature Review

All the techniques mentioned in section 1.2 have been used extensively in the classification and prediction of problems. Altman and Ohlson, who are the pioneers of CF prediction since the 1960's used selected financial variables from multiple companies bank statements to predict CF. These variables (*key performance indicators* (KPI's)) were selected as they believed that these were important factors that indicated whether a company was financially distressed. To make companies more comparable, both techniques involved converting the KPI's into ratios as a method of standardising the data. Due to the success of both of their methods, newer techniques proposed are based around their financial ratios and use the prediction accuracy for both methods as benchmarks to compare their new proposed work against techniques already in use.

Overall, Altman's MDA technique was superior to Beaver's IRS method, but only if the KPI's were jointly distributed according to a multivariate normal distribution. Otherwise, the MDA technique was prone to errors. However, the MDA technique was more favourable as it could use multivariate data at once to get an overall prediction rather than taking each ratio and scoring that to give an overall prediction.

Wilson and Lensburg took modern approaches by implementing ANN's and GP's to this classification problem. Both techniques can handle noisy data that is not normally distributed better than Altman's MDA, showing that both ANN's and GP's have potential to be even more accurate than IRS and MDA.

ANN's tend to perform very well in terms of efficiency and accuracy. Wilson used an ANN approach with a 5 10 2 structure. To improve accuracy, the Monte-Carlo technique was used to give a better representation of predicative accuracy. Overall, they achieved a 97.5% accuracy on their testing dataset, making this much more accurate than MDA and IRS.

ANN's suffer from being unable to produce a readable function to indicate how the ANN came to the solution that it did other than using the synaptic weights. Another issue faced with ANN's is that since a function is unable to be produced, the ANN is unable to tell us what influence each KPI has when predicting CF.

GP's can overcome some of the problems of an ANN as they are able to produce a user readable function. It is therefore possible to see what kind of influence each ratio must predict CF and which KPI's are more significant than others. Lee used a decision tree (DT) method to predict CF. Unlike Wilson and Lensburg, Lee used eight different KPI ratios when approaching this problem. Using this GP method, the testing accuracy of 92.91%. Rostamy used a similar approach to Lee, using five different KPI's. After training the GP, it could correctly predict if a company would fail 90% of the time, which was like MDA, however with a lot more flexibility in terms of what type of data it could accept.

The problem when GP's is that they may work slower as they explore a large search space and may be restricted to certain limitations e.g. maximum tree depth. For each crossover and mutation, depth of tree may increase and this can increase the computation time rapidly. When designing and implementing the GP, these factors will typically be accounted for as seen in Etemadi's et al paper.

Through the research completed, it could be seen that many of these used Altman's ratios as

their inputs. However, as Altman suggested, these ratios may not necessarily be the most optimal, but these five input ratios still provided the best alternative discriminant function to work with at the time. Since then economies have changed significantly, and therefore these ratios may not necessarily be the best to use to predict CF, but may still be significant enough to give an accurate prediction. As seen by Back, Rostamy and Lee, they have used other ratios to predict CF, and achieved similar results to MDA, which could potentially be more significant now. Wilson used Altman's five ratios and still achieved the 97.5% accuracy, with far fewer ratios relative to Back and Lee, which must be taken into consideration.

2.2 Project Specification

After careful consideration of the techniques reviewed in this section and other, I will be focusing my attention in predicting CF using ANN's and GP's due to their strong accuracy rates and ability to handle noisy data. Even though they do have drawbacks, I have tried to minimise these through the project specification and design of both techniques in this paper. The title of the project has been developed through the literature that has read to predict CF .

3 Design

4 Development

5 Testing

6 Description of the final product

7 Evaluation of the final product

8 Critical assessment of the project

9 Conclusion

10 Yalda is great