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

CARL SAPTARSHI

STUDENT NUMBER: 640032165

APRIL 2017

Contents

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, specially for those involved with the company at hand.

1.2   Algorithms for CF prediction

Since the aim of the project to classify whether or not 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 Darwins 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 of the techniques that were mentioned in section 1.2 have been used in classification and prediction, and have also been used in this particular area. Many of the current techniques already in industry are based around the works of Altman and Ohlson who are considered to be the leaders of CF prediction. Both of these use *key performance indicators* (KPI's) which are legitimate variables that can be found on a company's financial statement. Both of the techniques convert their feature inputs i.e. the variables into ratios as a form of data standardising, which can be used to determine the likelihood of a company failing or not. Due to the success of both of their methods, newer algorithms and techniques that have been 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.

Individual Ratio Selection (IRS) - Beaver selected thirty financial ratios from companies over a ten year period. Each of these ratios had a threshold to meet. If the ratio met the threshold, then the company is not at risk for this particular case. This same process is repeated for all the ratio that Beaver used to predict CF. Based on all the outputs from the variables, if the majority of the ratios met the threshold value, then his model would classify the company as "not at risk".

Multivariate Discriminant Analysis (MDA) -Altman took a set of empirical data from bank statements and identified five key ratios that he believed to relate to CF the most. The aim of this work was to create a discriminant function (Z) that would be able to discriminate between companies that are likely to fail and those that are not within a one year period. MDA has the ability to utilise multiple ratios at once and give a classification based on multiple characteristics, all at once. 66 medium tier firms were used and the discriminant function that was eventually found was this: $Z = .012x1 + .014 x2 + .033 x3 + .006 x4 + .999 x5$, producing a classification accuracy of 95%.

Genetic Programming (GP) - various GP's have been used to predict CF. For example, Lee used decision trees to predict CF. For this , each individual in the population is a decision tree. Here, eight ratios were used rather than Altman's five, on a dataset of 165 companies. Since this was a GP, the population size was 500, functional set included +, -, *, /, sin, cos, log, power, terminal set of 0,1, and a generation limit of 200 to predict the correct answer. This gave a predictive accuracy of 92% on the hold out sample, proving to be a very good technique to use for this area. Rostamy et al used a rounding threshold for the outputs and declared that if the output value was greater than 0.499 they are not likely to fail, otherwise they would be likely to fail within one year. As part of their design choices, they used a Number Of Hits to get the solution fitness. Overall they were able to correctly predict 90% of their hold out sample.

Artificial Neural Networks (ANN) -ANN's have been implemented in the field of prediction and classification since the 1990's. Odom et al used a simple 5 5 1 structure, using a sigmoid transfer function to send signals between each layer in the network. o classify the output, a threshold value of 0.5 was placed on this node. A value above 0.5 indicated "not likely to fail" and vice versa for below 0.5. This network also had a learning rate of 0.6 and momentum of 0.9 to search the space thorouhly. Using 74 companies, they achieved 81.84% accuracy on hold out sample and completed this faster than Altman and Ohlson. Wilson et al collected data from 129 firms and used the Monte-Carlo sub-resampling techniques to give a better representation of predicative accuracy.

When training the algorithm, back propagation was used in this network to ensure convergence of their 5 10 2 network. and used a training tolerance of 0.1 to prevent network memorisation when the error rates were smaller than 0.1. They were able to achieve a 97.5% accuracy for their hold out sample, making this more accurate than any of the other models described in the literature review summary.