Neural Network based Bankruptcy Prediction System

Anany Dwivedi (adwivedi), Zili Ma (zma2) , Jiexuan Sun (jsun2),
Worcester Polytechnic Institute, Worcester, MA

Abstract:

The bankruptcy prediction is always one of the hottest topics in economics. A lot of statistical analysis has been made in the solving this problem. However, as far as we know, a very less amount of research is available in solving this problem using Artificial Intelligence (eg. Neural Nets, SVMs, Decision Trees etc). In this paper, we aim to build a Neural Network based prediction model to that is able to predict if a company can undergo bankruptcy in following years with high accuracy.

Missions/Targets:

Our target goals and intended metric for performance evaluation are as follows:

1. **Goals**: Build the network model using Keras and try different techniques to improve the performance of our model.
   - We intend to use a recently reported technique called Net2Net to develop our model. We will compare the final model we create using Net2Net with a conventionally trained Neural Network.
   - Since the data sets is highly skewed, we might use methods such as undersampling, supersampling and balanced weight method to exclude the effects of skewed data. And then discuss the network performance for each method (If time permits, we will try them all).
   - The financial features here is sometimes highly correlated, we will try techniques like whitening matrix to whiten the data first.

2. **Metric of Performance:** Since our problem is binary classification problem, we will the method of area under curve (AUC) along with test accuracy to measure the performance of our model.

3. **Benchmark:** Our benchmark for the problem is based on the work done by Zieba et. al [1]. They compare different methods to solve the proposed problem on the proposed dataset. They report the NN they used had an AUC value of 0.543, 0.514, 0.548, 0.596 and 0.699 for years 1 to 5 respectively.

Data set information:

The dataset can be found at[2]: Polish companies bankruptcy data Data Set

The dataset is hosted on UCI's Machine Learning Repository, The dataset is about bankruptcy classification of Polish companies. The whole dataset is divided into five separate subsets. And the data in each subset might comes from different companies. Here are the descriptions of the subsets:

Description	Statics
The 1st subset contains all the attributes from 1st year of the forecasting period and companies' bankruptcy	271-bankrupted in the next 5th year; 6756-healthy in the next 5th year.
The 2nd subset contains all the attributes from 2nd year of the forecasting period and their companies' status	400-bankrupted in the next 4th year; 9773-healthy in the next 4th year;
...	...
The subset contains all the attributes from 5th year of the forecasting period and their companies' status	410-bankrupted in the next year; 5500-healthy in the next year;

In each dataset subset, there are 64 attributes, which are all about financial rate in its corresponding year. Some of them may include missing value. Here are some samples of those attributes:

X1- net profit / total assets
X52- (short-term liabilities * 365) / cost of products sold)
X30 - (total liabilities - cash) / sales

References:

[1] Zieba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction. Expert Systems with Applications.

[2] Department of Operations Research, WrocÅ‚aw University of Science and Technology, wybrzeÅ¼e WyspiaÅ„skiego 27, 50-370, WrocÅ‚aw, Poland
https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data