Neural Network based Bankruptcy Prediction System

Anany Dwivedi

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 I 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, I 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 confidence.

Missions/Targets:

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 the model.
   - I intend to use a recently reported technique called Net2Net [1] to develop FCN model. I will compare the final model I create using Net2Net with a conventionally trained Neural Network.
   - Since the data sets is highly skewed, I will use methods such as undersampling, supersampling and balanced weight method to exclude the effects of skewed data.
   - I will also implement an Auto-Encoder based anomaly detection algorithm.

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

3. **Benchmark:** Benchmark for the problem is based on the work done by Zieba et. al [2]. 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. For their best network (Ensemble boosted trees) they get the AUC values of 0.959, 0.944, 0.940, 0.941 and 0.955 repectively.
I achieved AUC values of 0.985, 0.992, 0.989, 0.955, 0.966 respectively for the Auto-Encoder based anomaly detection algorithm. I am currently working on training a FCN.

Data set information:

The dataset can be found at[3]: [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. 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;

References:

[1] Chen, Tianqi, Ian Goodfellow, and Jonathon Shlens. "Net2net: Accelerating learning via knowledge transfer." *arXiv preprint arXiv:1511.05641* (2015).

[2] 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.

[3] 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