Risk Analysis and Default Prediction for Taiwan Companies

Milestone II Report

Team Member: Yuan-Hsi Lai, Pei-Ling Tsai

Current Status

In milestone I (The green bubbles in ***Figure1.***), we finished retrieving data from news, Twitter and financial statement (Please refer to Appendix I for screenshots of source codes).

A close up of a logo

Description automatically generated

Figure

**Accounting Data**

we obtained several numbers and ratios that are relevant to the credit analysis: (TBD) total asset value, book-to-market value, long-term debts/ total invested capital, debt/equity, operating income/ received capitals, net income before tax/ received capitals, gross profit margin, earnings per share, EBIT interest coverage, quick ration and Merton’s distance to default.

**News and Twitter Posts**

We put the Web crawlers onto GCP (Google Cloud Platform) to run, obtaining such huge amount of data. For news, we currently collected 14 days of news from Google News, which aggregates comprehensive news sources from all the world. For Twitter posts, we collected data in one year and take the first 50 items.

**Relationship Graph of Companies**

We used the GraphDB to build the graph database of companies. In this graph, we obtained the “distance” between two companies. This information will be used to count the score of default prediction for one company, which doesn’t get much news coverage, by using the score of another company that is close to it and multiply a weight vector.

Challenges & Solutions

**Accuracy of Sentiment Analysis**

We tried some simple sentences to check the accuracy of sentiment; however, we found that the accuracy is only 50%. To improve the accuracy, we are going to handle some cases. For example, “not good” should be tagged as negative, but it will be counted as neutral since “not” and “good” would cancel the score for each other. Moreover, the grammar in Chinese is somehow more complicated. Thus, we may not cover all the special cases.

In addition, (TBD) we need to enlarge the neutral word set. There are many financial vocabularies that are neutral, such as, “earnings”, “investment”, “deal” and so on. However, they are tagged as “positive” or “negative” under our algorithm.

**Making Use of Relationship Graph**

We obtained the distances among companies… (TBD)……

**Combining Financial Ratios with Public Opinions**

In convention, banks use Altman Z-score\*1, to do the default risk analysis. The formula is static:

*Z* = 1.2*X*1 + 1.4*X*2 + 3.3*X*3 + 0.6*X*4 + 1.0*X*5.

Since we need to take public opinions into consideration, the formula with fixed dimensions is insufficient for us to build the NLP model. Therefore, we decided to adopt the method in, “Credit Rating Change Modeling Using News and Financial Ratios”\*2, to put financial data and public opinion variables into a vector, (), running the NLP model and obtain a weight vector, W(). If the value is large (e.g., total asset), then we will take “log” to prevent the value in the weight vector from being too small.

**Evaluating the Result**

The rating report of companies is hard to get (need to pay a certain amount of money). Thus, we are going to track the trend of stock prices instead. The mindset is: If the stock price is rising in this year, that means investors are confident with the company and are willing to invest. This will generate more equity in the company and makes it less likely to default.

Next Steps

Training model (TBD)

Other Enhancements (on-going)

* Adding timestamps to sentiment scores (TBD):
* Extend the collecting period of news: Since the financial statements are revealed each quarter. It’s impractical to collect news only in 14 days. Thus, we are going to collect news in each quarter and randomly choose 100 articles.

References

\*1. Altman Z-score: <https://en.wikipedia.org/wiki/Altman_Z-score>

\*2. “Credit Rating Change Modeling Using News and Financial Ratios”: <https://www.researchgate.net/publication/262323524_Credit_Rating_Change_Modeling_Using_News_and_Financial_Ratios>

Appendix I. Coding Part