

Financial Sentiment Analysis for Risk Prediction

Sarah Lu

February 11 2020

1 Introduction

The authors of this paper attempt to use the finance-specific sentiment lexicon to model the relations between sentiment information and financial risk. They formulate the problem as two different prediction tasks: regression and ranking.

2 Methodology

For the regression task, they aim to use sentiment information to predict a company's future risk, which is usually characterized by its real-value volatility. Instead of predicting the real-value volatility, in the ranking task, they try to employ sentiments to rank companies according to their relative risk levels. In regression task, they define the prediction by the following parameterized function:

$$\hat{v}_i = f(d_i; w).$$

Support Vector Regression is trained by solving the following optimization problem:

$$\min_w V(w) = \frac{1}{2} \langle w, w \rangle + \frac{C}{n} \sum_{i=1}^n \max(|v_i - f(d_i; w)| - \varepsilon, 0).$$

The authors use the following evaluation metrics:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\log(v_i^{+(12)}) - \log(\hat{v}_i^{+(12)}))^2.$$

3 Conclusion

From the two tasks, they observe that, trained on the finance-specific sentiment lexicon only, both the regression models and ranking models can obtain comparable performance to those trained on original texts, even though the word dimension is largely reduced from hundreds of thousands to only one and half thousand. In addition, the learned models suggest strong correlations between financial sentiment words and risk of companies. As a result, these findings provide readers more insight and understanding into the impact of financial sentiment words on companies' future risk.