Predictive Modeling: Financial Momentum

Abstract

How does a company’s prior performance affect its future share price? The aim of this project is to see whether a company’s previous financial success or failure carries over into the next performance period.

To discover whether such a correlation exists, this project analyzes companies based off of their previous financial statements, in addition to their pricing history.

Data is first scraped from NASDAQ and Yahoo Finance using a scraper written in Python which implements Beautiful Soup, a popular HTML processing package. This scraping process creates the database.

After the database has been populated, the preprocessing module, written in R, removes noisy data objects and performs dimensionality reduction on the data set to simplify both the algorithmic complexity and the difficulty of the predictive model growth.

After the data has been preprocessed and written to a new .csv file, which represents the refined database, the new database is fed into the predictive classification module, where the decision tree is grown and its performance is evaluated.

Results

There were two main classification sets used for the data. One set consisted of three possible classifications (Strong Buy, Buy, and Sell), while the other consisted of only two (Buy and Sell). On average, the decision tree produced from the ternary classification set, using a training set roughly one tenth the size of the entire dataset, produces a classification accuracy of 42%.

The binary classification set produces an average predictive accuracy of roughly 60%, exceeding the outcome of randomly guessing buy or sell by 10%, but also falling short of the expected accuracy: 70%.

The summary of the tree’s predictions is given in matrix form, such as in Figure 5 below.

<table>
<thead>
<tr>
<th></th>
<th>Buy-P</th>
<th>Sell-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy-A</td>
<td>0.383</td>
<td>0.233</td>
</tr>
<tr>
<td>Sell-A</td>
<td>0.187</td>
<td>0.197</td>
</tr>
</tbody>
</table>

Conclusions

From the results shown from the predictive model, it’s reasonable to say that, by only looking at a company’s financial statements and pricing history, we cannot improve our predictions much more than the evenly distributed binary probability of 50/50, or in the case of the ternary classification, 1/3 for each class. However, by performing attribute tests on the data (as done with the decision tree), we can increase the predictive accuracy to 1.2 times the original random choice accuracy.

For our ternary model, this means 33\(\times\)1.2 \(=\) 39.6%, whereas for our binary model, this yields 50\(\times\)1.2 \(\approx\) 60%.

Acknowledgements

- Trinity College Computer Science Department
- Lattanzi, R. (2012, September) Stock Selection using nonlinear classifier: a Decision Tree based approach, CCFEA
- Cont, R. (2010, April 08) Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance, 1/2, 223-236