

A News-Based Portfolio Management System

Development and Application of Data Mining and Learning Systems

Sattarov, Timur

Universität Bonn

May 21, 2014

Goal

Prediction

Based on the history of the stock and the corresponding news, predict its direction with a certain confidence.

Strategy

Using both prediction and confidence, a sound and efficient strategy is to buy and sell such amount of stocks, that will give a maximum profit at the end of a certain period.

Why SVM?

- It was selected as one of the “Top Ten Algorithms in Data Mining” [1]
- It was much superior to other machine learning algorithms in prediction of the stock market [2]
- Strong regularization properties
- Confidence level
- Handled very good with the tasks with high dimensionality

SVM prediction

Having the news and past stock prices as an examples, the task is to predict whether the stock price will increase significantly from the features derived from the news.

- Features
 - Daily stock news
- Labels
 - Based on the price movements (1, 0, -1)
- Classification target
 - Stock change class (1, 0, -1)
 - Confidence degree for each class [0:100]

Extracting features for SVM

Based on the daily stock news:

- 1 Mining the news text
 - remove stop words (conjunctions, particles, common function words)
 - stemming (chopping the ending of words)
 - lemmatization (morphological analysis to bring words to common lemma)
- 2 Ranking the words by importance for the stock
 - frequency
 - TD-IDF

Randomly generated prediction

Generates random values for the prediction and confidence.
In [3] and [4] the accuracy level was reached of around 70 - 80%.
Thus, we would like to see what will be in the opposite case.
Consider 20% of correct classification as a good classifier.

- 20% of the prediction is correct
 - confidence degree in the interval [0:100]
- 80% of the prediction is random (1,0,-1)
 - confidence degree in the interval [0:30]

Buying strategies

The stocks that are predicted as increasing are chosen to fill in the portfolio





- 1 Buy equally (buyEqually)
- 2 Buy relatively to the confidence (buyRelConf)
- 3 Buy top K relatively to the confidence (buyTopKRelConf)

Selling strategies

The stocks that are predicted as decreasing are chosen to be sold from the portfolio

- 1 Sell all stocks (sellAll)
- 2 Sell falling stocks (sellFalling)
- 3 Sell relatively to the confidence (sellRelConf)
- 4 Sell top K relatively to the confidence (sellTopKRelConf)
- 5 Sell all top K (sellAllTopK)
- 6 Sell relatively to the confidence and current portfolio (sellRelConfPortf)

Databases used

	Database Name	Period
	Bloomberg	03.01.2011 - 30.11.2012
	Reuters	01.01.2011 - 19.04.2013
	Standard & Poors	01.01.2004 - 28.12.2012
	Randomly Generated	01.01.2004 - 28.12.2012

NaN or zero values are replaced by the same price of the previous day. If the previous day does not have any price, then the current price is zero.

Precision

$$\frac{tp}{(tp + fp)} \quad (1)$$

tp - number of true positives

fp - number of false positives

- shows the hit rate of correctly classified stocks in the prediction
- impacts the profit / loss at every moment

Recall

$$\frac{tp}{(tp + fn)} \quad (2)$$

tp - number of true positives

fn - number of false negatives

- shows the proportion of correctly classified stocks, with respect to all the stocks, that should be predicted of a certain label

F1-score

$$2 * \frac{\textit{precision} * \textit{recall}}{(\textit{precision} + \textit{recall})} \quad (3)$$

- combination of precision and recall
- gives a broad outline about the correctness of the prediction

Questions ?

Prediction :

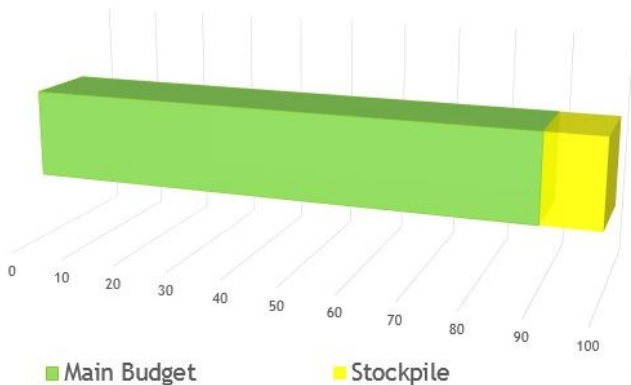
- Result of SVM prediction.
- Analysis of strategies with SVM prediction.
- Analysis of strategies with random prediction.
- The best combination of buying and selling strategy.

Initial parameters :

- **Initial Budget**
- Best **TopK** for strategies.
- Find **Price Difference** parameter.

Budget

Initial Budget = 100 000 dollars.



Top K

- Choose the strategies, where only **Top K** stocks should be chosen.
- Create all possible combinations with respectful strategies.
- Use prediction as a random prediction.
- Plot profits of these strategies with respect to K [1:100]

buyEqually
buyRelConf
buy**TopK**RelConf

sellAll
sellAll**TopK**
sellFalling
sell RelConf
sellRelConfPortf
sell**TopK**RelConf

Top K in buying strategy

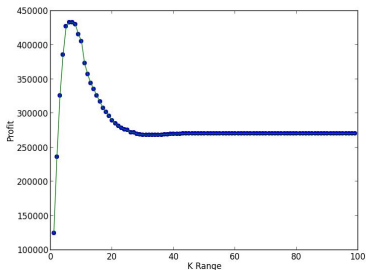


Figure: buyTopKRelConf and sellFalling

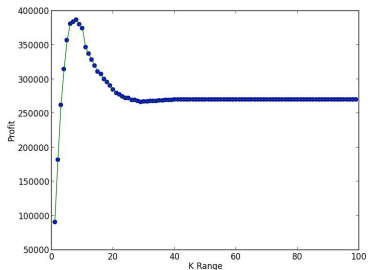


Figure: sellAllTopK and buyTopKRelConf

Price Difference

Definition

Price Difference parameter is the difference between the prices of the two consequent days of a certain stock.

Impacts

On labeling of the samples for the training of SVM.

Reason

Try to minimize the trading risk and make the profit more effective.

Price Difference

In [3] this parameter was used 0.0% and in [4] only 0.5% of the price of a stock. Thus, the initial parameter was chosen as 1% and was incremented on 1% more in each new experiment.

Experiments are done :

- With the real SVM prediction.
- Results are represented using “classification report”

Comparison with [3]

In [3] having the price difference parameter of 0.0%, Precision, Recall and F1-score were in the interval from 70% to 80%.

	Precision	Recall	f1-score	support
fall down (-1)	0.48	0.13	0.21	23990
remain the same (0)	0.00	0.00	0.00	34
rise up (1)	0.53	0.77	0.62	26576
avg / total	0.50	0.47	0.43	50600

Table: Classification report with 0.0% of price difference

Price Difference

	Precision	Recall	f1-score	support
fall down (-1)	0.33	0.02	0.04	10384
remain the same (0)	0.58	0.99	0.73	28963
rise up (1)	0.31	0.01	0.01	11253
avg / total	0.47	0.57	0.43	50600

Table: Classification report with 1% of price difference

	Precision	Recall	f1-score	support
fall down (-1)	0.00	0.00	0.00	4314
remain the same (0)	0.83	1.00	0.90	41762
rise up (1)	0.00	0.00	0.00	4524
avg / total	0.68	0.83	0.75	50600

Table: Classification report with 2% of price difference

Results

Having the price difference parameter greater than 2% SVM predicts none of increasing, or decreasing stocks during the whole period of the stock market !!!

Possible reason :

Feature space is very large : 5504 dimensional (number of words)
Number of examples are very small : 253 instances (number of days)

As a result, there is not enough information for classifier to make a correct prediction.

Dimensionality reduction (PCA)

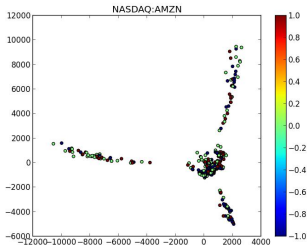


Figure: NASDAQ:AMAZON in 2D

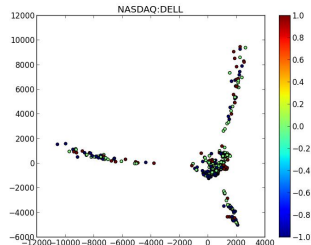


Figure: NASDAQ:DELL in 2D

SVM prediction

- Training period for SVM was chosen 1 year, from 01.01.2011 - 30.12.2011
- The prediction was during 01.01.2012 - 28.12.2012 (around 250 days)
- Initial Budget = 100 000 \$
- Top **K** = 5
- Price Difference = 1%

SVM prediction

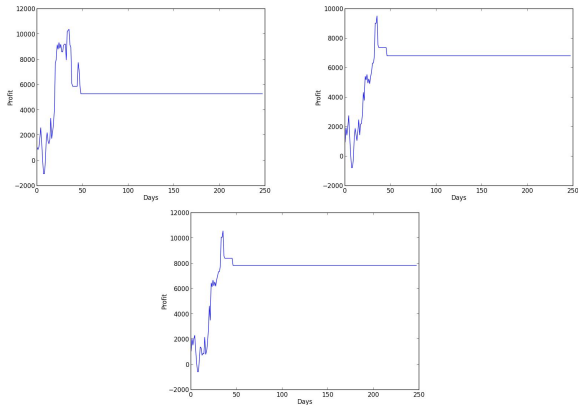


Figure: Profits of sellAll strategies with respect to the buying strategies



SVM prediction

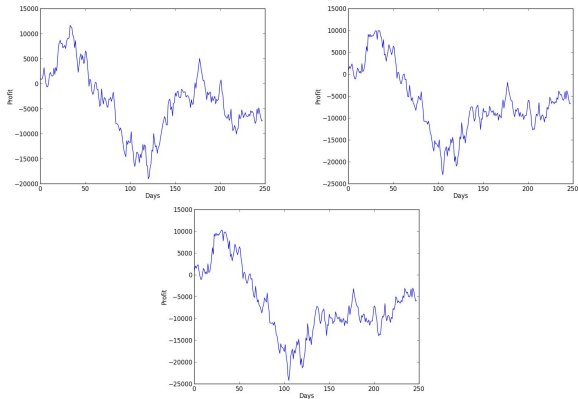


Figure: Profits of combinations of all the rest strategies with respect to

SVM prediction

Strategy	Profit (points)	Profit (%)
buyEqually	-7389.34	-7.3%
buyRelConf	-6584.22	-6.5 %
buyTopKRelConf	-5864.36	-5.8%

Table: Comparison of profits of buying strategies

SVM prediction

Stock	Price at 03.01.2012	Price at 27.12.2012	Increased/ Decreased (%)
NYSE:NOV	70.87	66.97	-5.5%
NYSE:BHI	51.02	40.07	-21.4%
NYSE:APC	78.65	73.77	-6.2%
NYSE:HAL	34.15	34.60	1.3%

Table: Comparison of the prices of the stocks in portfolio

Portfolio

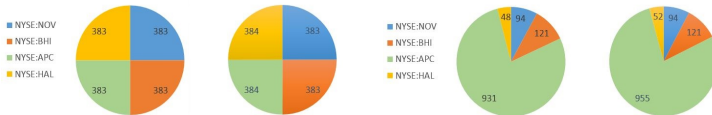


Figure: buyEqually

Figure: buyRelConf

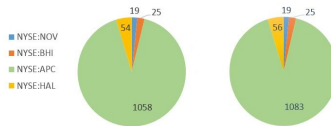
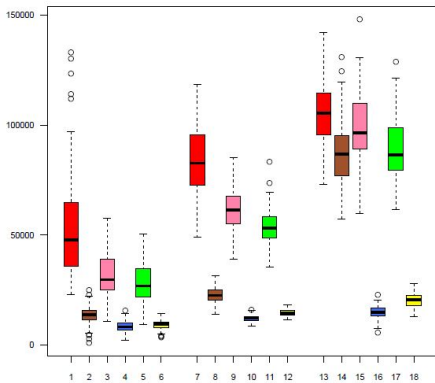


Figure: buyTopKRelConf

Figure: Portfolios in the beginning and at the end of the period

Randomly generated prediction



- 1 - buyEqually and sellAll
- 2 - buyEqually and sellAllTopK
- 3 - buyEqually and sellFalling
- 4 - buyEqually and sellRelConf
- 5 - buyEqually and sellRelConfPortf
- 6 - buyEqually and sellTopKRelConf
- 7 - buyRelConf and sellAll
- 8 - buyRelConf and sellAllTopK
- 9 - buyRelConf and sellFalling
- 10 - buyRelConf and sellRelConf
- 11 - buyRelConf and sellRelConfPortf
- 12 - buyRelConf and sellTopKRelConf
- 13 - buyTopKRelConf and sellAll
- 14 - buyTopKRelConf and sellAllTopK
- 15 - buyTopKRelConf and sellFalling
- 16 - buyTopKRelConf and sellRelConf
- 17 - buyTopKRelConf and sellRelConfPortf
- 18 - buyTopKRelConf and sellTopKRelConf

Figure: Profits of all strategies with description

Conclusions

Parameters :

- There is no single value for **TopK** for all strategies. However in most cases, portfolio reaches highest profit at the value between 4 and 10.
- Having the price difference parameter of 0% we have almost reached the same accuracy as [3] and [4] did, however starting from 1% the results are not so promising.

Conclusions

Predictions :

- Even though, SVM prediction did not give so effective results, among all buying strategies, **buyTopKRelConf** could be considered as the strategy that gives the highest profit.
- Concerning the selling strategies, in case of SVM prediction, it was impossible to find a superior strategy.
- However, in case of random prediction, **sellAll** was always superior in its group, in contrast with **sellRelConf** and **sellTopKRelConf**, which gave always the worst results.
- The results of random prediction prove, that even having a classifier with 20% of correct prediction, it is enough to have a positive profit at the end of a certain period.

Future research



- 1 Add more advanced strategies, which will provide higher profit.
- 2 Increase the training phase of SVM to obtain higher accuracy with higher price difference parameter.
- 3 Determine the best parameter for the *MainBudget* and *TheStockpile*.
- 4 ... and many others ...

Thank you for your attention!

References I

-  [1] Wu, Xindong and Kumar, Vipin and Ross Quinlan.
Top 10 Algorithms in Data Mining.
Knowl. Inf. Syst., December 2007.
-  [2] Phichhang Ou and Hengshan Wang.
Prediction of Stock Market Index Movement by Ten Data
Mining Techniques.
Modern Applied Science, 2009.

References II

-  [3] Ms. D. Preetha and Mrs K. Mythili.
KENERL BASED SVM CLASSIFICATION FOR FINANCIAL
NEWS.
*International Journal of Advances in Computer Science and
Technology*, November 2013.
-  [4] Rama Bharath Kumar and Bangari Shravan Kumar and
Ragiri Shiva Sai Prasad.
Financial News Classification using SVM.
International Journal of Scientific and Research Publications,
March 2012.