Associative Classification for Quantitative Stock Trading

Giuseppe Attanasio

4th SmartData@PoliTO Workshop
27 & 28 February 2019

Giuseppe Attanasio

PhD student @ Department of Control and Computer Engineering

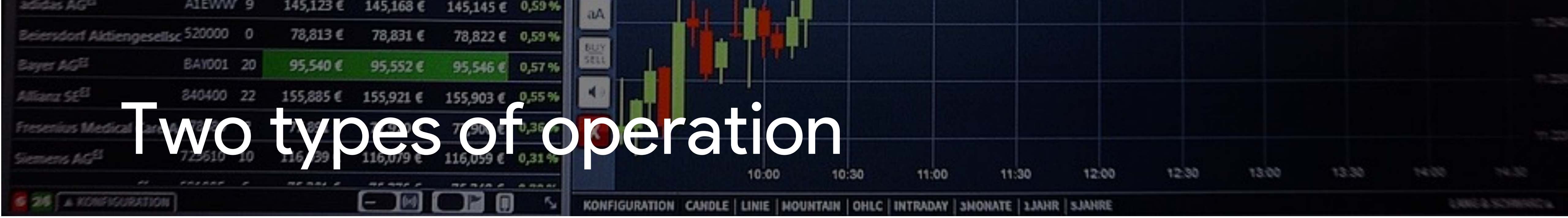
Enrolled: November 2018

Supervisor: Elena Baralis

Trading fundamentals

Trading fundamentals

- Stock exchange markets
- Price movements and trend forecasting
- Long- and Short- selling operations
- Technical and Quantitative analysis



Two types of operation

Long trades

1. Buy from the market
2. Sell back to the market

$$(gross)\ profit = \frac{C_f - O_i}{O_i}$$

Short trades

1. Sell to the market
2. Buy back from the market

$$(gross)\ profit = -\frac{C_f - O_i}{O_i}$$

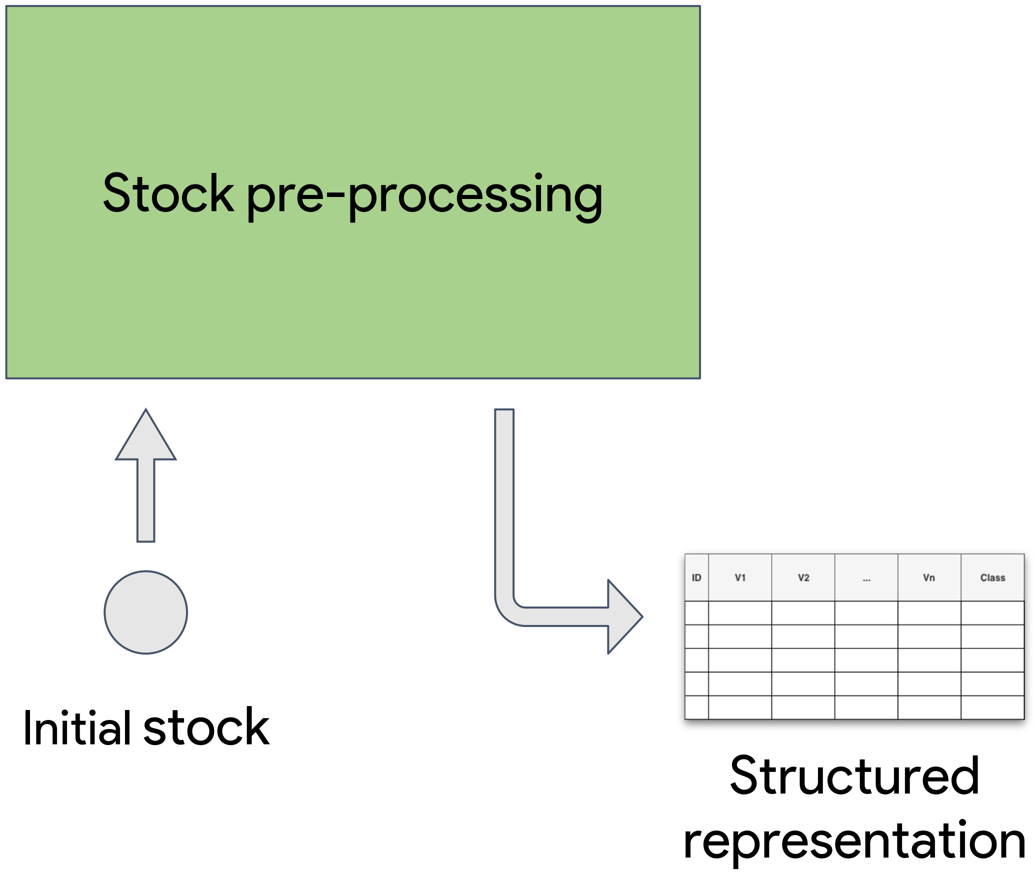
Trading fundamentals

Chart featured by Yahoo Finance, *https://it.finance.yahoo.com/*

The chart includes Moving Averages and Relative Strength Index indicators.

A state-based approach

By means of associative classification



The logical pipeline that composes the trading system

Stock pre-processing

- Move from time domain to state-based representation
- Relax temporal constraints among samples
- Describe time series data by means of a combination of state variables

State variables: technical indicators and oscillators - e.g. Moving Averages, MACD, RSI - to build a structured dataset

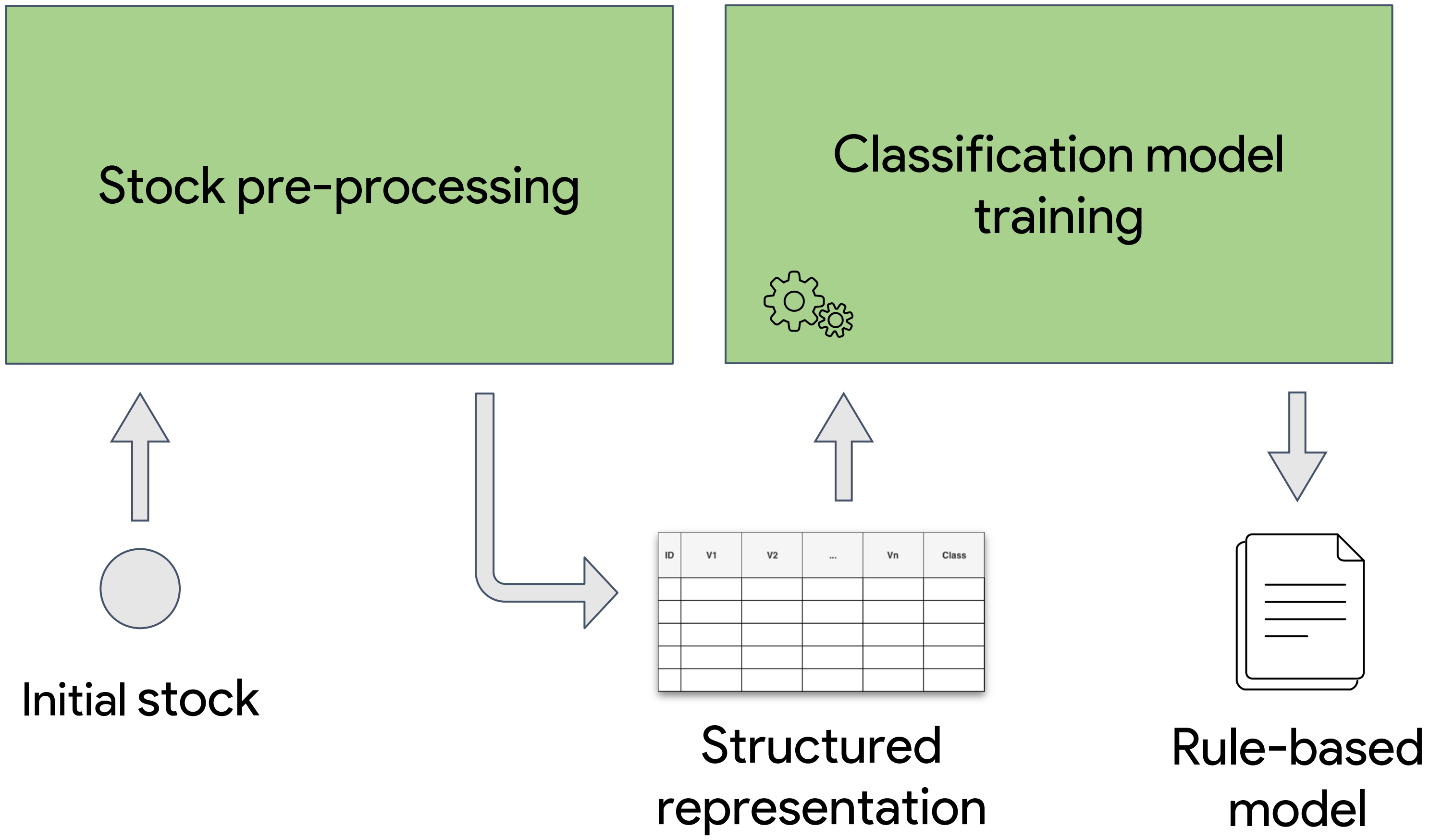
Stock pre-processing

- The state summarize market conditions on that day
- Class labels identify price variation with respect to the following day – e.g.:
  - $ROC > \mathrm{tr} \Rightarrow \mathrm{BUY};$
  - $ROC < -tr \Rightarrow \mathrm{SELL}$
  - $-tr \leq ROC \leq \mathrm{tr} \Rightarrow \mathrm{HOLD};$

Stock pre-processing

- Technical indicators domain is quantized with semantical meaning - e.g.:
  - $0 < RSI \leq 30$: stock is oversold condition;
  - $30 < RSI \leq 70$: stock is in normal;
  - $70 < RSI \leq 100$: overbought condition.

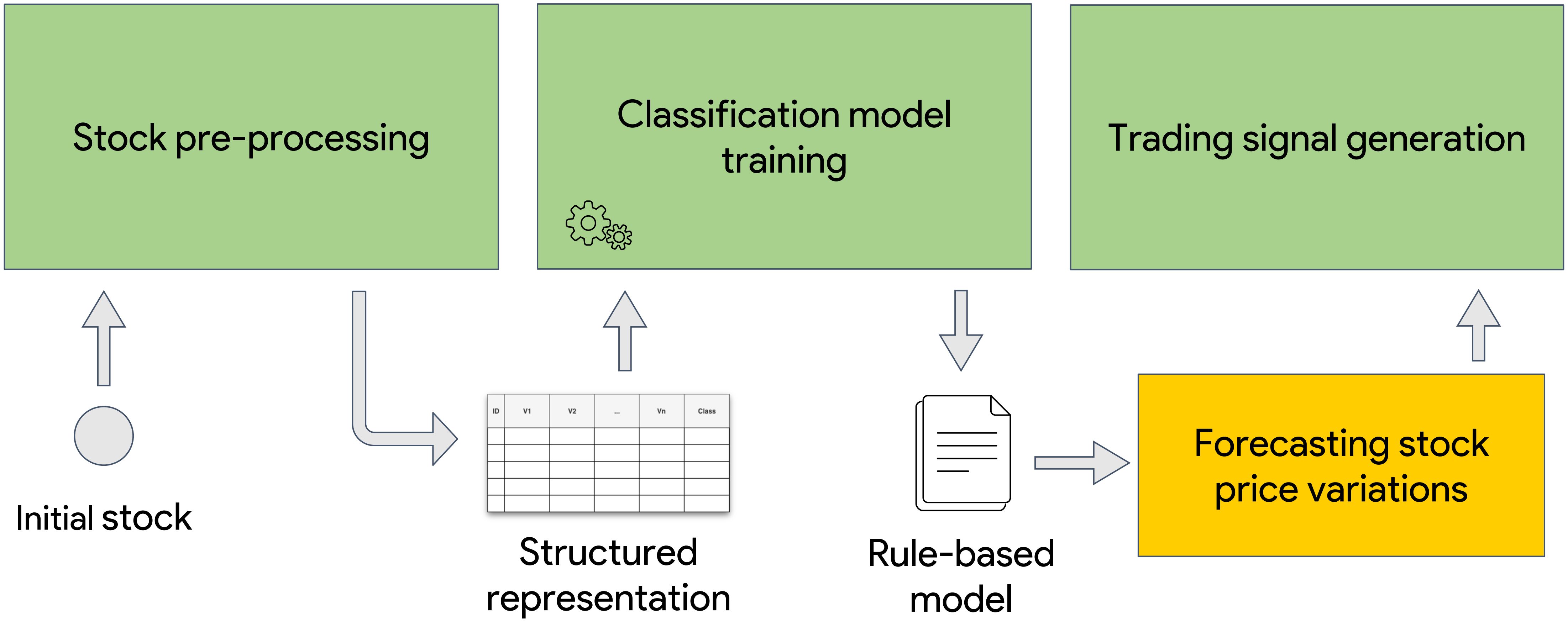
ID	RSMA	REMA	MACD	AO	ADX	RSI	PPO	...	Class
1	< 0	< 0	> 0	< 0	> 20	< 30	> 0	...	BUY
2	> 0	< 0	< 0	> 0	> 20	> 70	> 0	...	HOLD
3	< 0	< 0	> 0	< 0	< 20	30 < i < 70	> 0	...	HOLD
4	> 0	> 0	> 0	< 0	> 20	< 30	< 0	...	BUY
...	...	...	...	...	...	...	...	...	...



The logical pipeline that composes the trading system

Classifier training

- Applied algorithm: *Live-and-Let-Live* (L^3) associative classifier
  - Extract association rules between state variables and class labels
  - Lazy pruning of harmful rules
  - Rule set split in Level 1 and Level 2 rules
- Extracted rules suggest relationships between market state variables and price variations



The logical pipeline that composes the trading system

Automated trading system

Forecasting

Predict whether the percentage price variation is above the threshold *tr* (or less than -*tr*%) on the next trading day

Type of operations

Long- or short-selling operations

Operation length

Operations can last multiple days: close with a signal in opposite direction

Stop loss

Trading strategy to limit losses early closing positions.

Experimental results

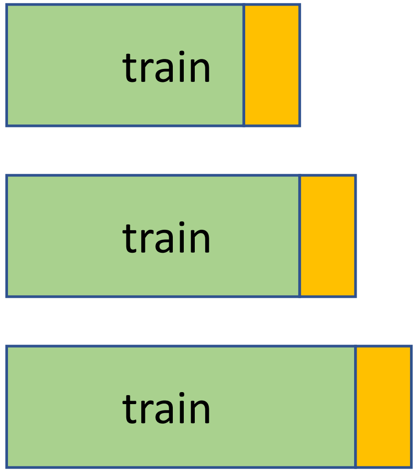
A comparative analysis

Comparative studies

- Trading sessions on *Financial Times Stock Exchange Milano Indice di Borsa* – or FTSE MIB. About 40 stocks
- Years 2011, 2013, 2015
  - Different market conditions
- Two validation strategies
- Performances comparison with:
  - Time series models
  - Other Machine Learning classifiers

Two validation strategies

Expanding window



models = |D| - | initial train size |

Hold-out

models = 1

Top performing configurations

Year 2011, Expanding window

Classifier	Total profit	Operations	Avg PPO
L3	43.34%	54	0.80%
ARIMA	79.87%	204	0.39%
MLP	122.24%	523	0.23%
EXPSMOOTH	317.92%	1600	0.20%
RFC	179.43%	1027	0.17%

Year 2011, Hold-out

Classifier	Total profit	Operations	Avg PPO
L3	185.04%	249	0.74%
MLP	102.36%	269	0.38%
RFC	129.72%	401	0.32%
SVC	83.11%	370	0.22%
EXPSMOOTH	49.10%	785	0.06%

Top performing configurations

Year 2013, Expanding window

Classifier	Total profit	Operations	Avg PPO
ARIMA	221.97%	104	2.13%
L3	129.02%	74	1.74%
VAR	149.33%	110	1.36%
EXPSMOOTH	252.04%	351	0.72%
MNB	95.23%	190	0.50%

Year 2013, Hold-out

Classifier	Total profit	Operations	Avg PPO
L3	79.95%	73	1.10%
MLP	90.84%	83	1.09%
MNB	32.30%	50	0.65%
SVC	36.71%	100	0.37%
EXPSMOOTH	17.19%	68	0.25%

Top performing configurations

Year 2015, Expanding window

Classifier	Total profit	Operations	Avg PPO
L3	115.29%	83	1.39%
MLP	129.82%	395	0.33%
EXPSMOOTH	474.79%	1528	0.31%
SVC	32.32%	689	0.05%
LINREG	40.41%	962	0.04%

Year 2015, Hold-out

Classifier	Total profit	Operations	Avg PPO
L3	39.71%	80	0.50%
SVC	27.41%	58	0.47%
MLP	26.24%	61	0.43%
RFC	134.46%	405	0.33%
EXPSMOOTH	38.70%	533	0.07%

Conclusions

And ongoing works

Conclusions

- Results are promising
  - L^3-based systems are comparable to ones that use other Machine Learning classifiers
  - Especially with mid-term forecasting horizon L^3 outperforms Time Series models

- Many configurations in play:
  - Statistical test on results are required
  - Different comparisons other than ranking by average profit

Which one should I choose?

- Choice between black-box and white-box
- L^3-based models are simpler to adopt in real trading systems:
  - Rules are interpretable and tunable
  - Refreshing the model each day is not required

Ongoing steps

- Extend the analysis to Standard & Poor 500 index
  - Same years
  - USA market

- Test different financial securities
  - e.g. Cryptocurrencies

- Address the problem with sequence modeling algorithms
  - Deep Recurrent Neural Networks
    - Long-Short Term Memory Networks


NYSE
NYSE
Oh, sell, sell, sell, sell,
sell, sell, sell !

Thank you!

Any question?