The FinIR Grand Challenge Overview

Chang Jiangeng
NUS NExT++ Lab

Financial Deep Learning Method Time-series

Data Set	Period	Feature Set	Method
SPY ETF S&P500	2014-2016	Price data	FFNN
Dow30 stocks	2012-2016	Close data and Technical indicators	LSTM
Borsa Istanbul 100 Stocks	2011-2015	Technical indicators and OCHLV	CNN
Stocks in Dow30	1997-2017	RSI	DMLP
High-frequency record of all orders	2014-2017	Price data, record of all orders, transactions	LSTM
17 ETFs	2000-2016	Price data, Technical indicators	CNN with feature imaging
10 stocks from S&P500	_	Stock Price	TDNN, RNN, PNN

Financial Deep Learning Method textual data

Data Set	Period	Feature Set	Method
S&P500 Index, 15 stocks in S&P500	2006-2013	Financial News titles, technical indicators	SI-RCNN(LSTM + CNN)
10 stocks in Nikkei 225 and news	2001-2008	Textual information and stock prices	Paragraph Vector + LSTM
HS300	2015-2017	Social media news, price data	RNN-boost with LDA
News and Chinese stock data	2014-2017	Selected words in a news	HAN

The challenge target

- Deal with the information retrieve problem on the

  financial text data such as only some of data are useful.

- Explore the efficient method to aggregate the time-
  series data and text data during 2001-2018 together

- The targets are the one-day, twenty-day and sixty-day

  spot price movement

one-day: input--O~t   output--t+1_{spot}-t_{spot}
twenty-day: input--O~t   output--t+20_{spot}-t_{spot}
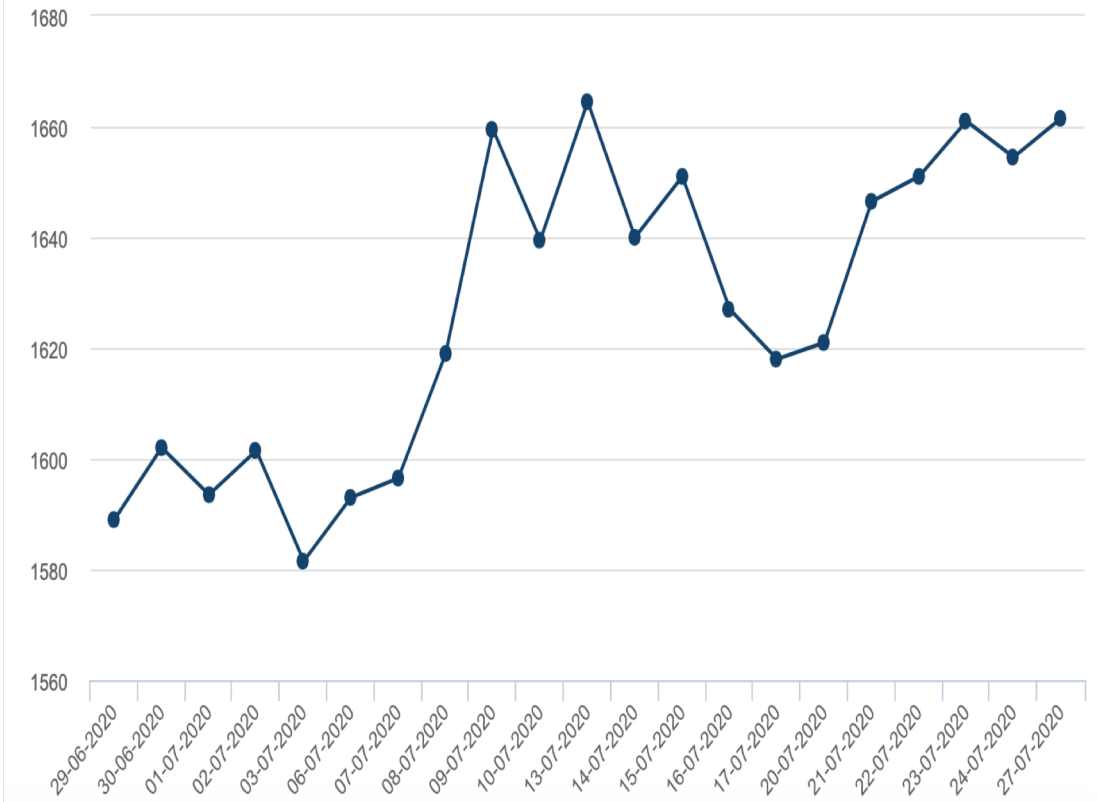sixty-day: input--O~t   output--t+60_{spot}-t_{spot}

The text information retrieve challenge

- Example 1: Copper price managed to form strong bullish rally yesterday to reach the initial resistance at 2.7300 to record the initial target in the previous report,…. Expected trend for today: bullish.

- https://www.economies.com/commodities/copper-analysis/copper-attacks-the-initial-resistance-–-analysis-–-25-1-2017-35557

Provided Data

- Time series data and Textual data:

	Time Period	Trade days	Time-series data	Analyst report	News
Training Phase	2001-2017	3790	22740	37827	27388
Phase 1 testing	2017-2018	253	1518	63825	4033
Phase 2 testing	2018-2019	253	1518	14074	2513

Text data is unbalance

Competition Summary

NUS NExT++ Research Center & Four Elements Capital Pte Ltd & MegaTech.AI & Hessian Matrix · 346 Single-Player Team · 33 Multi-Player Team · 437 participants

SIGIR 2020: FinIR

2020-06-01 - Team Merger Deadline

2020-03-31 - Launch   2020-07-01 - Close

65 teams made successful submission

Top Team

Team	Feature	method	One-day	Twenty-day	Sixty-day
maimang	Price, indicator	Lightgbm	62.6%	71.5%	75.8%
USTC_ 火锅真好吃	Price	LSTNet	53.7%	65.9%	72.3%
youdu	Price	LSTM, Lightgbm	52.4%	63%	57.4%
USTC_rainbow	Price	Adaboost	52.6%	59.3%	59.4%
USTC_wjsty	Price	Lightgbm, XGBoost	52.1%	60.1%	69.9%

Tree models are the most popular

Deep learning model can learn more information

Team	Target	problem
maimang	3M contract close price	Regression
USTC_ 火锅真好吃	3M contract close price	Regression
youdu	3M contract close price	Classification
USTC_rainbow	Spot Price	Classification
USTC_wjsty	Spot Price	Classification

Top team chose to replace the target as 3M contract close price

Top team changed the classification problem with a regression problem

The difficulties of using the text data

- The 404 error

	url	title	date	source
0	https://www.economies.com/commodities/copper-a...	cooper-achieve-some-gains-23-06-2014-11197	2014-06-23	https://www.economies.com
1	https://www.economies.com/commodities/copper-a...	cooper-achieves-gains-8-82013-4566	2013-08-08	https://www.economies.com
2	https://www.economies.com/commodities/copper-a...	cooper-and-some-losses-10-6-2013-3566	2013-06-10	https://www.economies.com
3	https://www.economies.com/commodities/copper-a...	cooper-around-the-initial-support-2-10-2013-5468	2013-10-02	https://www.economies.com
4	https://www.economies.com/commodities/copper-a...	cooper-attempts-to-settle-positively-13-2-2014...	2014-02-13	https://www.economies.com

- How to find useful information

useless

Example 1: Copper price managed to form strong bullish rally yesterday to reach the initial resistance at 2.7300 to record the initial target in the previous report,.... Expected trend for today: bullish.   useful

The problems during holding the competition

1. Data Leakage: We need to build a testing platform

- twenty-day: input--0~t $+10$ output--$t+20\_\{spot\}$-$t\_\{spot\}$
- sixty-day: input--0~t $+10$ output--$t+60\_\{spot\}$-$t\_\{spot\}$

2. Inaccessibility of the text data: We need to large the scale of the financial data

Congratulation to top teams

Winner list:
- First team : maimang
- Second team: USTC_火锅真好吃
- Third team: youdu

Specific time reward:
- One-day: maimang
- Twenty-day: maiming
- Sixty-day: maimang

Thank you