Multi Data Source Stock Market Prediction

Yuhan Su, Ruichuang Cao, Wei Xu
Tsinghua University

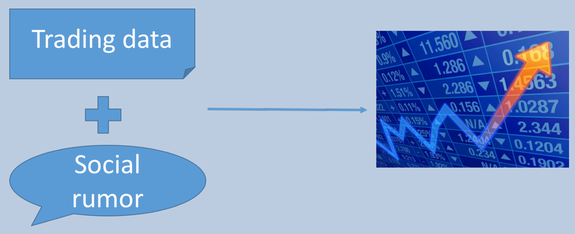
Introduction

- Everyone dreams to predict stock price changes. Some even succeeded – to some degree.
- China stock market is influenced by rumors on social media.

- We build a real-time data stream stock prediction system on IBM SuperVessel Cloud.
- We can use multiple data sources, including trading data and social media rumors.

Contribution

- A system to process real-time data stream from multiple sources.
- Use multi data source, including trading data and social rumors, to predict stock market in China.
- A scalable system using the state-of-the-art cloud technology.
- An intuitive web UI to let users edit and analyze code and data.



Data Source

- SSE50 index history. Daily closing price. Also our predicting target.
- Sina Stock Forum. Posts and comments from http://guba.sina.com.cn.
- Financial news. Collected from Tushare.
- NASQAF index. Daily closing price.
- RMB Exchange rate. Daily value.

Sina Stock Forum

SSE50 Index
(Prediction Target)

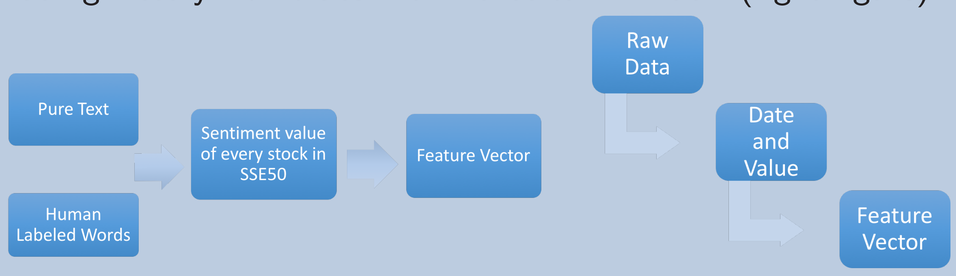
NASDAQ Index

RMB Exchange Rate

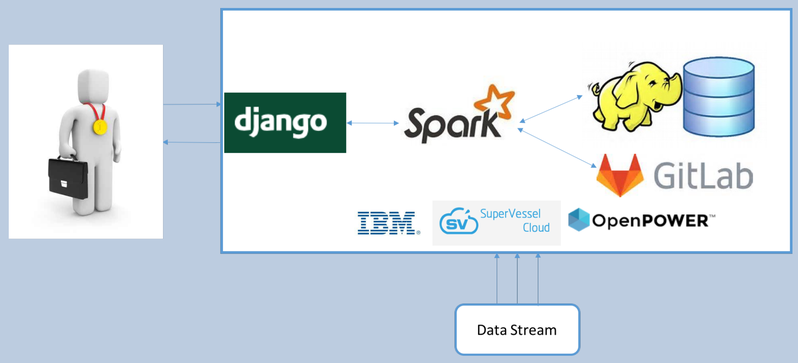
Financial News in China

Data Pre-Processing

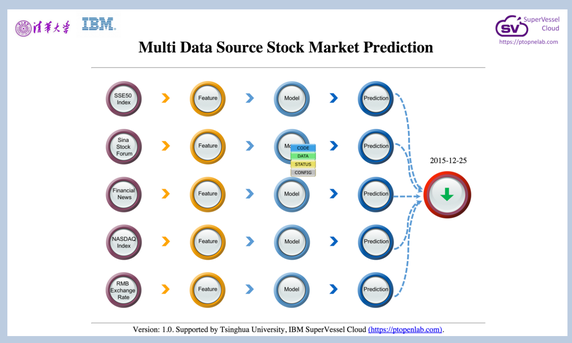
- Leveraging the cloud to do multi-step data pre-processing as streams.
- Rumors and news → sentiment (left figure)
- Trading history → values within a certain window (right figure)



System Architecture

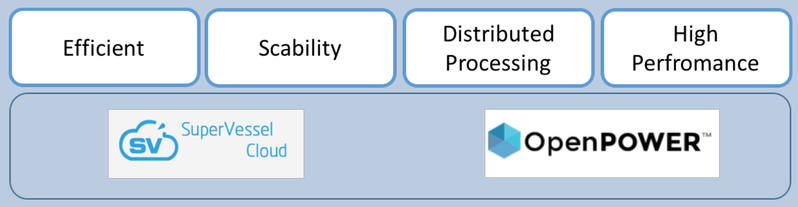


- Building blocks: Spark, HDFS, Gitlab, Django
- They all run efficiently on IBM SuperVessel Cloud.
- Integrates the acquisition, preprocessing and learning of data source streams, and shows the final prediction result.
- Everything under a single web UI. Users can change the model parameters, modify the codes, and check the running status online.



SuperVessel Cloud

- We use SuperVessel Cloud, based on OpenPOWER technology and provides the high efficiency cognitive computing infrastructure for frontier science with high performance heterogenous platform (GPU/FPGA).
- SuperVessel provides three handy services: basic cognitive cloud service, cognitive computing service platform, and application acceleration store for new technology sharing.
- SuperVessel provides us with a scalable and easy-to-maintain cloud infrastructure to build our systems on.



Conclusion

- Using multiple data sources improves stock prediction.
- We can provide the entire analysis under an intuitive web UI.
- Highly efficient cognitive computing infrastructure at SuperVessel greatly simplified our development.

Acknowledgments

- IBM China Research Lab provided the SuperVessel Cloud.
- IBM partly funded this project.
- Tushare website provided datasets.

Contact Information

- Web: http://iiis.tsinghua.edu.cn https://ptopenlab.com
- Email: Yuhan Su syhmartin@yeah.net
  Ruichuang Cao create0818@163.com
  Prof. Wei Xu wei.xu.0@gmail.com