Predicting Financial Time Series using Deep Learning

Module1. Machine Learning for Financial Time Series Prediction

Jongho Kim

NICE Pricing & Information Inc.

Fall, 2018

Note. This content mainly refers the summer session of KAIST organized by Jiyong Park(2018)

Machine Learning for Financial Time Series Prediction

- Obviously, forecasting financial time series is a very difficult problem
- In this module, we will discuss two paradigms of applied analytics (Causation vs. Prediction)
- After then, we will answer on the question “Why is machine learning harmful or beneficial for financial time series prediction?”

Two Paradigms of Analytics
: Causal Inference and Prediction

“Correlation is Enough in the Era of Big Data”

“인간은 원인을 찾도록 길들여져 있었다. 하지만 분명한 것은 이제 사회가 인과성에 대한 그동안의 집착을 일부 포기하고 단순한 상관성에 만족해야 할 것이라는 점이다. 즉 이유는 모른 채 결론만 알게 되는 것이다. 이것은 수백 년간 이어져온 관행을 뒤집는 일이며, 우리는 의사결정 방식이나 현실에 대한 이해 방식을 기초적인 부분부터 다시 생각해야 할지도 모른다.

많은 경우에 우리는 그 정도면 충분하다. ***빅데이터에서 중요한 것은 결론이지 이유가 아니다.*** 어떤 현상의 원인을 알아야 할 필요는 없다. 우리는 데이터 스스로 진실을 드러내게 하면 된다.”

<Big Data (빅데이터가 만드는 세상)>, Mayer-Schonberger and Cukier

“Causality Lies at the Heart of Our Understanding”

“예측 알고리즘은 상관관계를 찾아내는 데 초자연적인 능력을 발휘할지 모르지만, 그 특성과 현상이 생기는 근본적인 원인에는 무관심하다.

하지만, 인간이 가진 이해력의 범위를 확대하면서 궁극적으로 ***우리는 지식 추구 활동에 의미를 부여하는 것은 인과관계의 해독이다.*** 이는 세상이 돌아가는 원리를 세심하게 풀어헤치는 것이다.”

<The Glass Cage (유리감옥)>, Nicholas Carr

To Explain or To Predict?

Statistical Science
2010, Vol. 25, No. 3, 289–310
DOI: 10.1214/10-STS330

To Explain or to Predict?

Galit Shmueli

Department of Decision, Operations and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742, USA (e-mail: gshmueli@umd.edu).

- Four major disparities between explanation and prediction (Shmueli 2010)
  - ~~Theory – Data~~
  - **Causation – Association (Correlation)**
  - **Bias – Variance**
  - Retrospective (*in-sample*) – Prospective (*out-of-sample*)

Shmueli, G., 2010. To Explain or to Predict?. *Statistical Science*, 25(3), pp.289-310.

Bias versus Variance

- Bias – Variance decomposition

$$True\ Relationship: Y = f(x) + \varepsilon$$

$$Empirical\ Model:\ \ Y = \hat{f}(x)$$

$$Expected\ Estimation\ Error = E\left[\left(Y - \hat{f}(x)\right)^2\right]$$

$$= Var[\varepsilon] + \left\{E\left[\hat{f}(x)\right] - f(x)\right\}^2 + E\left[\left(\hat{f}(x) - E\left[\hat{f}(x)\right]\right)^2\right]$$

$$= Var[\varepsilon] + Bias^2 + Var\left[\hat{f}(x)\right]$$

- **Explanatory modeling** focuses on **minimizing bias** to obtain the most accurate representation of the underlying true relationship.
- **Predictive modeling** seeks to **minimize the combination of bias and estimation variance**, occasionally **sacrificing theoretical accuracy** for improved empirical precision.

Shmueli, G., 2010. To Explain or to Predict?. *Statistical Science*, 25(3), pp.289-310.

What is Causal Inference?

- Causal inference is to connect one process (*cause*) with another process or state (*effect*) where the first is partly responsible for the second, and the second is partly dependent on the first.

- Goal of causal inference: (In-sample) **Unbiased estimates**
  - But, it is quite challenging in observational studies.

- Potential outcome framework
  - Ideal causality

    = *(Outcome for treated if treated) – (Outcome for treated if not treated)*
  - But, in reality…

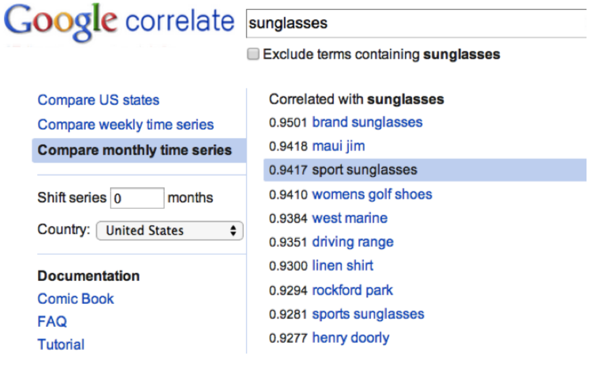
    = *(Outcome for treated if treated) – (Outcome for untreated if not treated)*

What is Prediction?

- Prediction is to state about the future event or the unknowns, based upon experience or data.

- Goal of prediction: (Out-of-sample) **High predictive power**
  - ➢ Prediction does not require causation. Rather, it exploits a range of correlations as much as researchers can (*reducing bias*), while avoiding overfitting (*minimizing variance*).

Google's research on prediction



Google's follow-up service

Different Goals Need Different Methodologies

- Causal inference and predictive analytics do aim different goals
  - ➢ Causal inference aims to yield unbiased estimates.
  - ➢ Predictive analytics aims to make the most of correlations, while avoiding overfitting (ensuring out-of-sample prediction)

- Think first what question you want to solve

Machine Learning for Financial Time Series Prediction

Jongho Kim (quantic.jh@gmail.com)

Why Deep Learning Could be Harmful

- Lack of causality / Learn only correlation but not causation
- Model interpretability
- Lack of theory

머신러닝과 AI에 집착하지 않아야 하는 이유

리: 머신러닝, AI는 어떻게 보세요?

권용진: 이건 미국도 16년까지 생각보다 많이 안 썼어요. 왜냐면 결국 돈이 들어가니까요. 머신러닝과 AI의 치명적 단점이 패턴은 잘 찾는데, 이게 어떤 이론에 기반하는지 사람이 알 길이 없어요. 그냥 냅둬서 거래시킬 수도 있지만 정말 치명적인 데이터 편향이 발생했을 때 돈 크게 날리는 걸 방지하기 힘들어요. 실제로도 이 때문에 purely AI로 데이터 분석한 텍사스 헤지펀드 하나가 망하기도 했고요. 기간 설정에 따라서도 크게 달라지잖아요. 2008~2010년 데이터만 보면 폭락한 주식 무조건 매수하라고 나올 거고…

기계는 좀 더 빨리 망하게 할 수도 있다(…)

이승환 (2018), 전직 트레이더, 초단타 퀀트 매매법을 말하다: 권용진 대표 인터뷰,
Retrieved from https://ppss.kr/archives/177286

Why Deep Learning Could be Beneficial

- Although predictive analytics have different roles from casual inference, deep learning performs much better than casual models for prediction

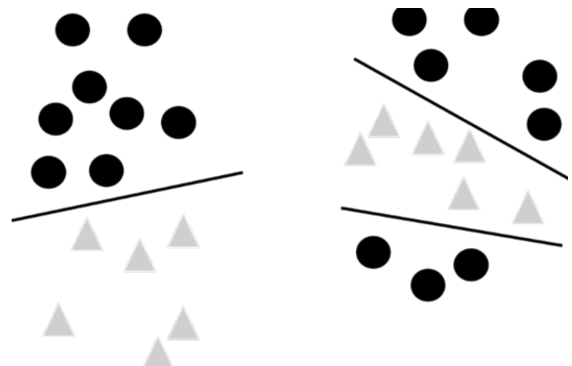
- Also, the deep learning predictor has a number of advantages over traditional predictors
  - input data can be expanded to include all items of possible relevance
  - non-linearities and complex interactions among input data
  - over-fitting is more easily avoided

Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). Deep learning in finance. *arXiv preprint arXiv:1602.06561.*

Why Deep Learning Could be Beneficial

- Machine learning (e.g., neural networks, random forests) can represent non-linearity, which is not possible in linear regressions.

Linearly separable problems



Non-linearly separable problems

- Automatic feature selection
  - ➢ Machine learning searches for model specification automatically to achieve the best model fit. (e.g., which variables or interactions should be included?) (Mullainathan and Spiess 2017)

Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). Deep learning in finance. *arXiv preprint arXiv:1602.06561*.

Why Deep Learning Could be Beneficial

- Over-fitting is more easily avoided
  - Cross validation: aiming to reduce over-fitting and increase out-of-sample performance
  - Dropout for model selection: Dropout is a model selection technique. It is designed to avoid over-fitting in the training process, and does so by removing input dimensions in X randomly with a given probability p.
  - Regularization techniques: It is common to add a regularization penalty to avoid over-fitting and to stabilize predictive rule

Thank you ☺

Contact Info: quantic.jh@gmail.com

References

- Jiyong Park (2018), KAIST Summer Session, Retrieved from https://sites.google.com/view/kaist-mis-session2018/overview?authuser=0
- 이승환 (2018), 전직 트레이더, 초단타 퀀트 매매법을 말하다: 권용진 대표 인터뷰, Retrieved from https://ppss.kr/archives/177286
- Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). Deep learning in finance. *arXiv preprint arXiv:1602.06561.*