From Tech to Finance: MLE perspectives

3/3/21 4:31 PM

Speaker: Ea-Ee Jan (Goldman Sachs)
Date: 03/03/2020

ML for the Front Office

Simplified Pair Trading algorithms

- Long and short on two highly correlation securities
    - Buy: the security is low in price
    - Short: the security is high in price

Pried trading using normalized spread

- If difference of price passes threshold, buy one short the other

Fixed income

- Generalized pair trading algorithms
- Return vs. volatility

Simplified Bollinger Band Trading Algorithm

- Bollinger Bands record bounds above and below the X day rolling mean (e.g. 20 days)
- If the current price goes above the upper band, should short

Sentiment and NLP for Trading

- How is market emotion, sentiment impact security return?
    - Document, paragraph or sentence-based sentiment
    - Entity or aspect-based sentiment analysis
    - Market impact on Sentiment may be short and instantaneously
- Topic detection and trending
    - Market focus

ML for Middle Office

Entity resolution via Fuzzy Match

- Find the best matching entities for a given query using search framework
- Only ranked documents from search results are needed instead of search scores

Contextual Search

- Use cases
- Heuristic type can be used for confusion pair disambiguation
    - Entity type
    - Industry, sector, subsector for corporate entities
    - Note: conditional independence assumption

Entity Resolution by Knowledge Graph

- When more contexture events are available
- General purpose of Knowledge graph is too expensive to build
- Application specific, tailer to business scenario are realizable

ML for the Bank Office

- IT Help desk is an essential services
- SLA, Ticket reduction, root cause finding are all critical for ITSM
- Cognitive technology, automation and chatbot are very actively involved in ITSM community

Finance Environment

Technology vs. business knowledge

- Work with stakeholders
- Focus on business requirement instead of algorithms

More closed environment

- Access public cloud?
    - Data stay in house
- Challenges in download open-source library and tools
- Upload download GitHub
- VM vs. laptop PC

Tips and Pitfalls

- Use every opportunity to test the system
- Understand the algorithms implementation
    - Feature normalization SVM vs. decision tree vs. regression
- Understand algorithm design
    - Feature selection vs. feature permutation
- 80% of time is for data process, 20% of time is for model development