Speech Recognition for Finance

ZHANG Chao@CSLT, RIIT, THU

History

- Prof. Fredrick Jelinek (1932-2010) created a team with highest reputation in the area of ASR at IBM from 1972 to 1993.
    - Seeing speech recognition as an acoustic signal passes a noisy channel.
    - Developed most of the state-of-the-art stochastic ASR techniques: phone based AM, n-gram LM, decision tree based state clustering ...
    - Used the above approaches to two other applications: machine translation and stock value prediction.
- In 1993, Prof. Jelinek retired from IBM and went to JHU as the head of CLSP until his death. He created the famous JHU Summer Workshop there.

History (Cont.)

- Dr. James Harris Simons (1938-) created one of the most successful hedge fund, the Renaissance Technologies, in 1982.
    - He is a famous mathematician on Geometry who contributed the famous Chern-Simon Theory co-worked with Prof. Shiing-Shen Chern (陈省身).
    - 12 researchers from Prof. Jelinek’s IBM team (include Peter Brown and Bob Mercer) joined the company in 1993. They used ASR techniques for finance and got legendary performance: 80% a profit lasting for 20 years!
- Peter Brown and Bob Mercer became the CEO of Renaissance Technologies after Dr. Simons retired in 2010. It is said they loved people with ASR background very much.

Basis

- Finance: “the management of money in which a business decides how the money will be used to generate more profit.” Economists generally regard financial markets an efficient mechanism function for the financial system.

- Stocks: “are the original capital invested in the business by its founders.”

    - Qualified companies can be listed at a stock market to raise money by selling stocks (represents shares of the company) to the public.

    - More money a company earns, higher market value it gets. Company shares part of its profits to the stockholders.

    - Investors deal stocks at a price they both agree with, so stock price is determined by confidence (that’s why it belongs to virtual economy).

Basis (Cont.)

- Stock Market Index Future: a cash-settled futures contract on gambling how a stock market index (a set of stocks on the market) changes.
  - If A believed the index will fall and B disagreed, A sells stocks to B: if A is right B has to sell them back to A at the lower price: A wins; vise versa, A has to buy the stocks back at the higher price: B wins.
- Earning by trading stocks and derivatives (e.g. Stock Market Index Future):
  - Long: buy some stocks and sell them when the price rises.
    Short: sell some stocks now and buy them back when the price falls.
  - A key factor in getting profits is to predict how the price changes.

Basis (Cont.)

- Hedge: “a technique used to reduce any substantial losses suffered.”
    - α Arbitrage: long stocks which can rise more and fall less than the index, and short the same amount of index. If the stocks rise, earn more from stocks; otherwise, earn more from the index.
- Hedge Fund: the investment risk of the fund get hedged by varies investment strategies and assets.
- Other assets for investment: currency swap, commodities futures, fonds, mortgage back securities, and *etc*.
- Quantitative Analysis: analyzing finance according to different mathematic features (e.g. predicting stock value from current price).

Why?

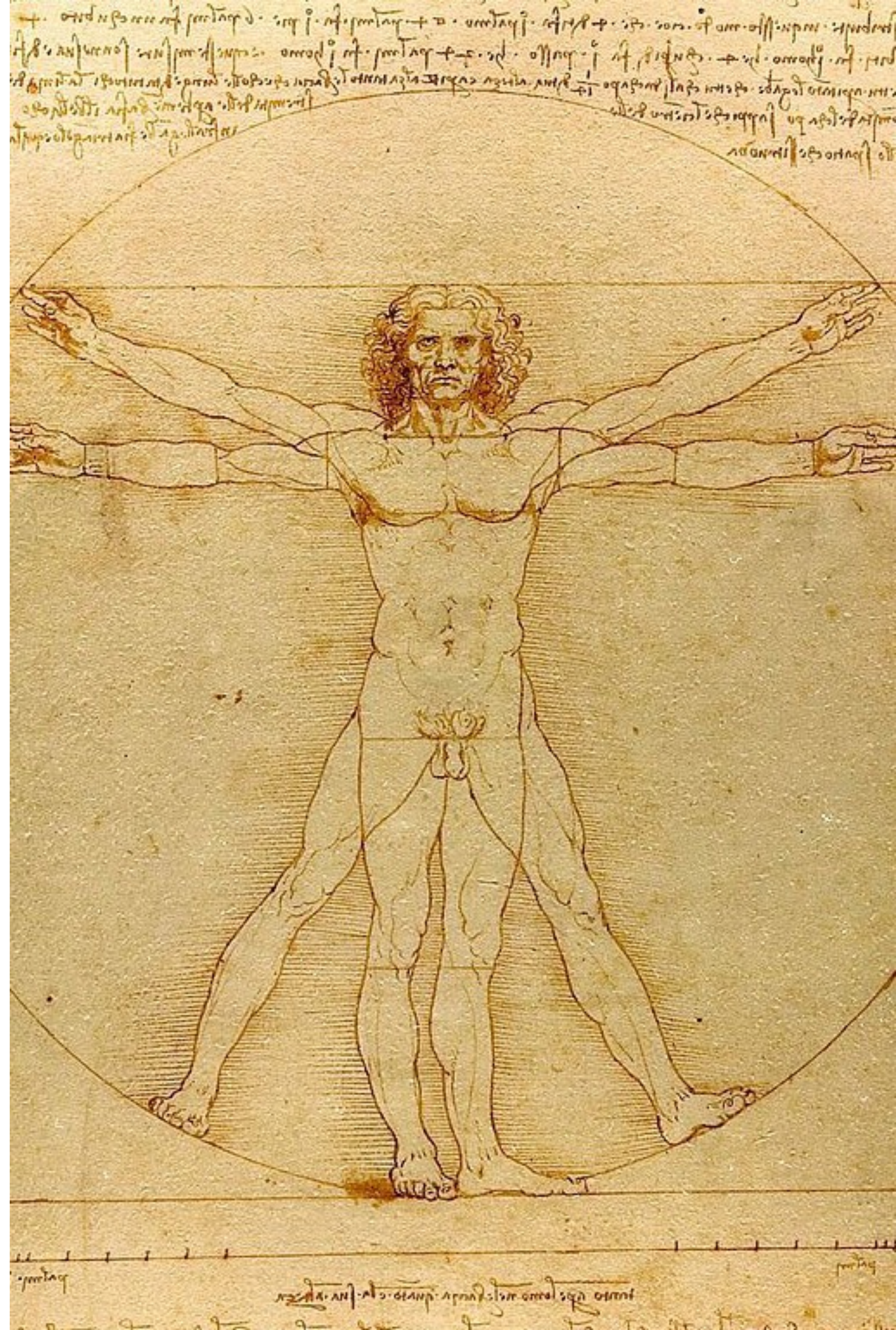
- Correctly predicting the price can generate large profits.
  - Financial markets are supported by confidence and human behaviors that are subject to human error and emotion.
- Behavioral Patterns: assume patterns are embedded in quantitative features indicating how will the price vary, and we can gain profit with these patterns.
  - Case 1: the price of a stock rises 1% 10mins within opening everyday. Buying it before closing and selling it ASAP the next day for a year will earn $1.01^{250} - 1 \approx 11$ times the initial capital (easy money indeed).
  - Case 2: find the evidence that someone is lifting the price, long the stock and sell it before his selloff will make a significant profit.

How?

- Algorithm Trading:
  “use computer to enter trading orders with an algorithm deciding on aspects of orders.”

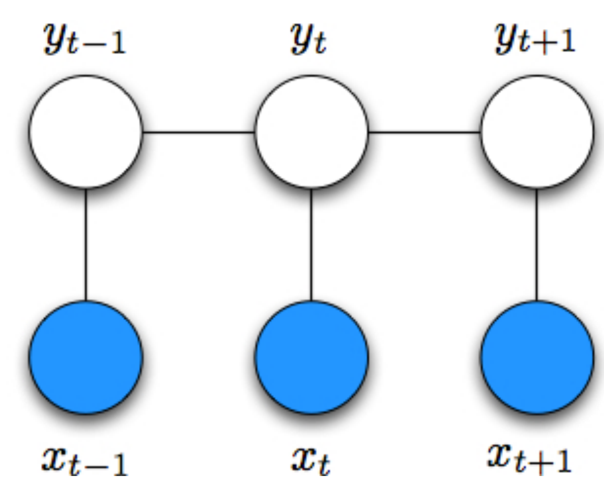
  - High Frequency Trading:
    trade rapidly (e.g. seconds) for even tens of thousands of times a day.

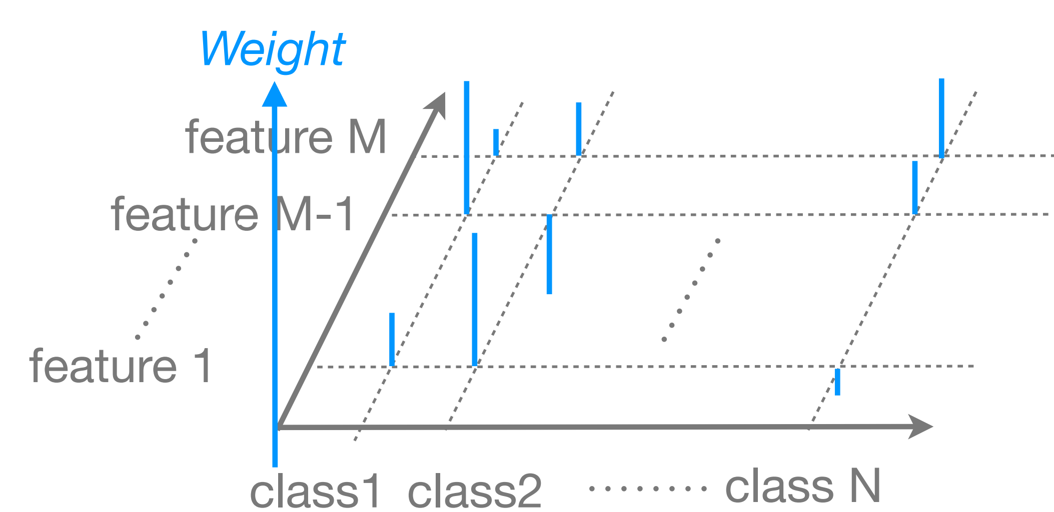
  - Slower Day Trading:
    e.g. daily.



A Trial

- Task - Use some variation of Conditional Random Fields (CRF) for stock market index futures (TianXiang 280) daily prediction.
    - CRF: an extension of HMM for some perspectives.
        - A discriminative model in nature: $p(\vec{y}|\vec{x})$.
        - Transitions rely on any necessary observations.
        - Fit for arbitrary and redundant features.
        - Features are automatically weighted.





A Trial (Cont.)

- Feature (daily): the price, volume, equal-string, and *etc*. for today.
- Output (daily): the closing price for the next trading day (rise/fall).
- Data Set: 1998~2008 for training; 2009~2010 for testing. Corr% = 58.6%.
- Actual Performance:
  - Accurate for price crash, suitable for being short.
  - 22% return in 2011.

Thank for your listening~!