Financial Time Series Analysis and Prediction
金融時系列分析と予測

LIANG Zhenlin

Supervisor: Prof. LI Xiang

Lin Canguang Lab, The University of Aizu

1.Introduction

The prediction of **financial market**s is one of the most important concerns in the field of finance. **Usually financial data is time series data**, such as stocks, exchange rates, gold prices etc. all vary over time. **Time series is a sequence of values of a variable measured in time order over a certain period of time**. Time series analysis is the use of statistical tools to analyze the past of this series in order to model the changing characteristics of the variable and to forecast the future.

Any financial time series contains uncertainties and therefore **statistical theories** and methods are crucial in financial time series analysis. A time series of financial assets is often viewed as a realization of a sequence of unknown random variables over time. It is usually assumed that the sequence of random variables is a discrete stochastic process if it is defined only at discrete points on the time axis. The possible characteristics of a time series include the following:

- **Trend**: Trend is a time series that moves consistently in a certain direction. As shown in Figure 1, the USD/JPY exchange rate has been trending up or down at certain times.

Fig.1 USD/JPY exchange rate changes along time.

- **Seasonal Variations**: Many time series contain seasonal variations. In the financial sector, we often see seasonal variations in commodity prices, particularly those associated with growing seasons or temperature changes. As shown in Figure 2, “Snowboard” search trends follow the timing of the snow season.

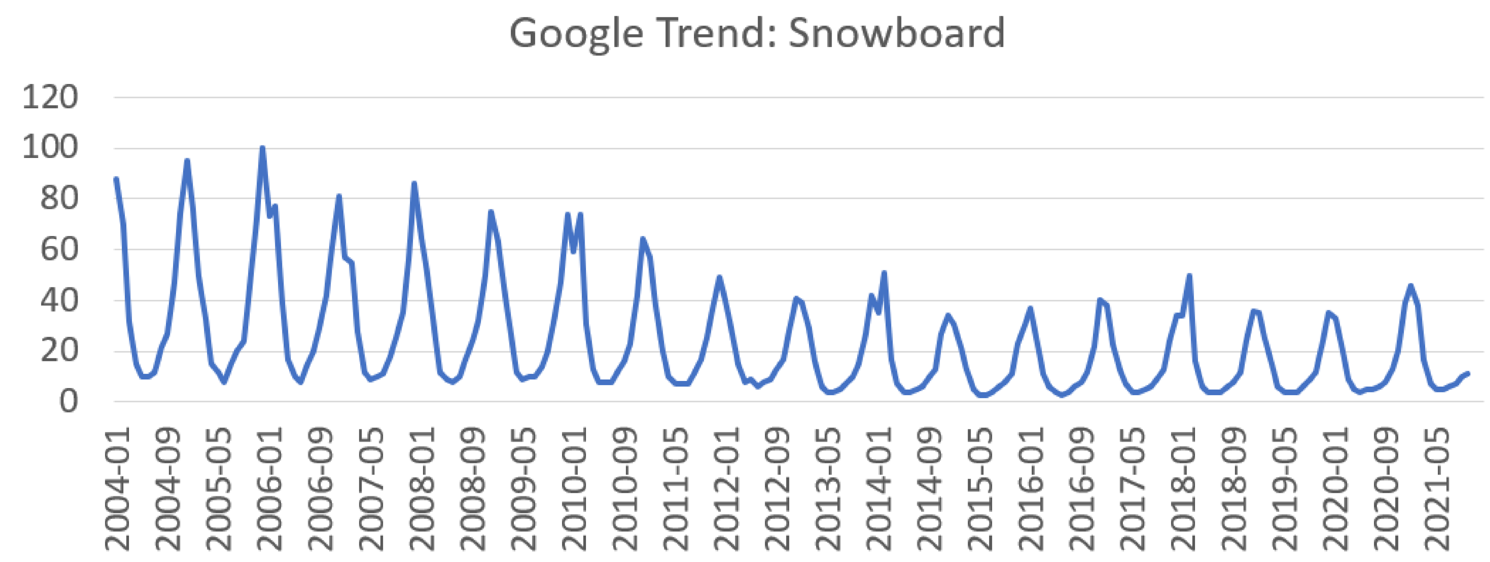

Fig.2 Changes in Google Trends for the keyword "Snowboard"

- **Serial Correlation**: One of the most important characteristics of financial time series is serial correlation, also known as autocorrelation, which is the degree to which a series is correlated at different points in time, for example, high volatility is often accompanied by high returns.

Fig.3 USD/JPY return value, the red box is volatility clustering (a type of autocorrelation)

2.Prediction Methods

- **Recent Study**

In the field of finance, researchers always use linear models such as Autoregressive model (AR), Moving Average (MA) and Autoregressive moving average model (ARMA) to model financial data and study the autocorrelation of their return.

However, **financial time series** often **contain random noise** and are **influenced by external information**, such as policy or news that affects stock prices. The relationship between these characteristics and the target variable is often **non-linear** and requires more sophisticated models to model. With the development of machine learning and deep learning, more and more researchers have started to apply them to finance area, such as some of Kaggle's financial market prediction competitions, where the top solutions always see **LightGBM** and **LSTM** methods.

- **Model**

LightGBM is an open source **Gradient Boosting Decision Tree(GBDT)** algorithm developed by Microsoft. Compared with the neural network method, LightGBM has ***strong interpretability*** and can **rank features in order of importance.** Using this feature, we can evaluate and improve feature engineering. As shown in Figure 4, LightGBM consists of **several decision trees.**

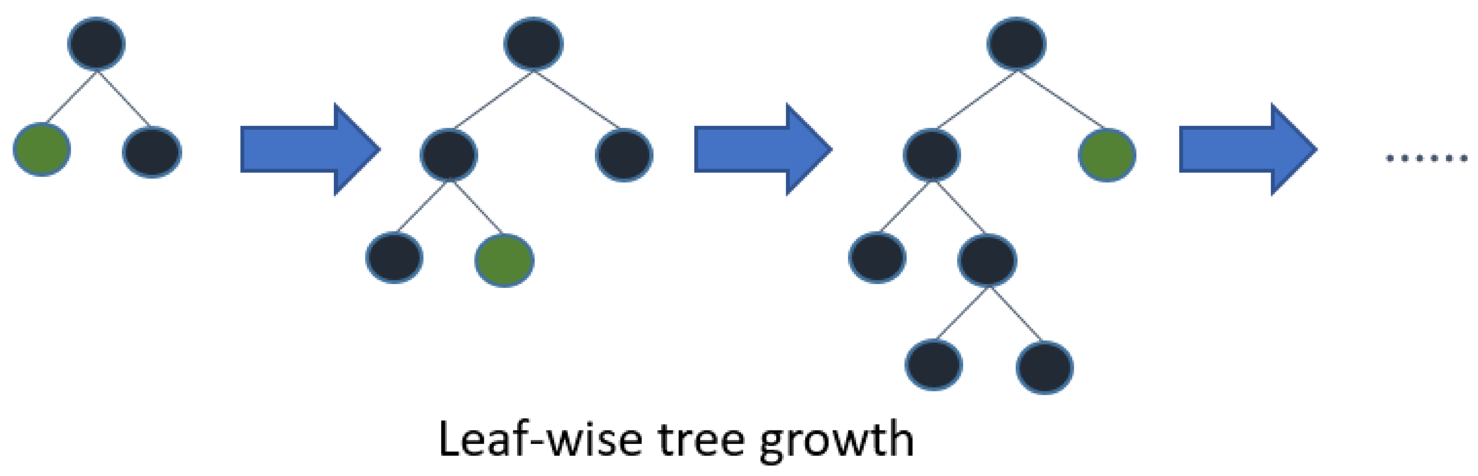

Fig.4 A basic structure of LightGBM.

The majority of financial data is time series, recurrent neural network (RNN) has the ability to process sequential data, so RNNs are widely used for time series problems. LSTM is a special RNN, which uses memory cells to store information and is better at modeling long-range dependencies. The basic unit of vanilla RNN and LSTM is shown in Figure 5.

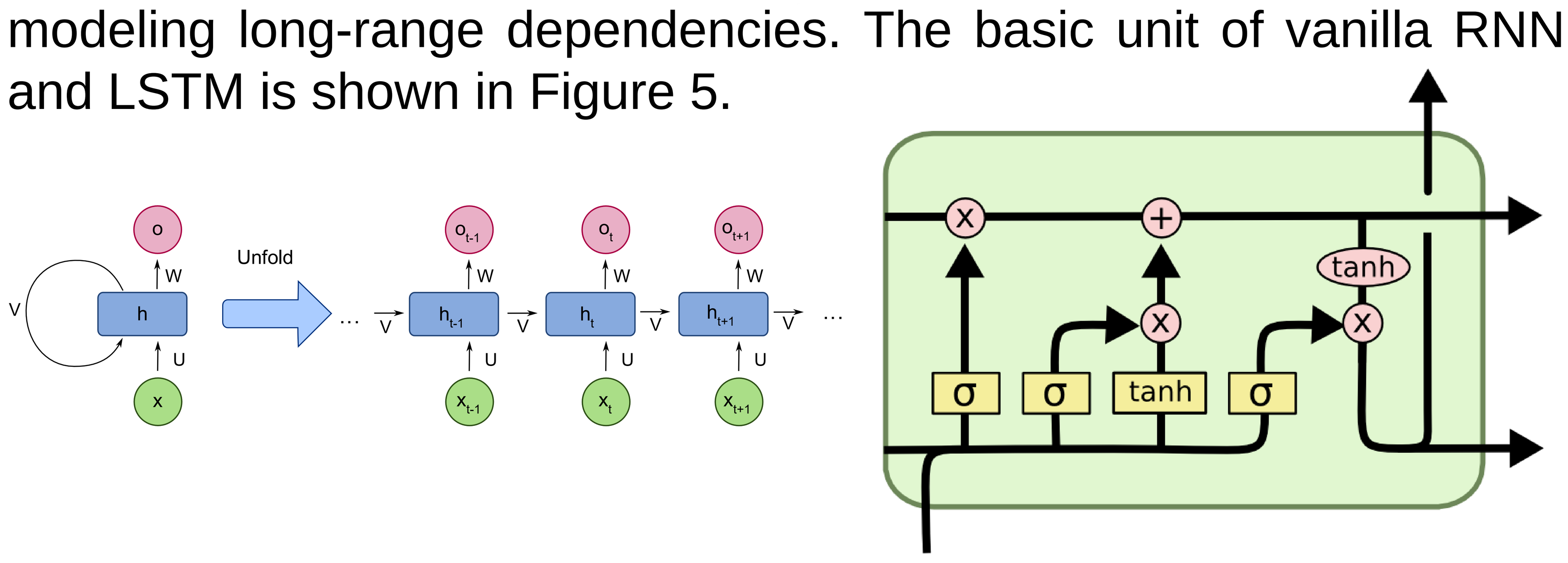

Fig.5 Vanilla RNN block and A memory cell of LSTM

3. Overview of Prediction Process

As shown in Figure 6, We need to collect historical data on the financial market and do **feature engineering** on it using **statistical indicators** such as **mean, standard deviation, return, moving average,** etc. to **extract** its autocorrelation features, and also collect **external information** that affects the market and **quantify** it into features. In addition, we can also use methods such as random forests to filter features. Finally, we feed the features into the model for prediction, and generally speaking, a reasonable **combination of different models** will give better performance.

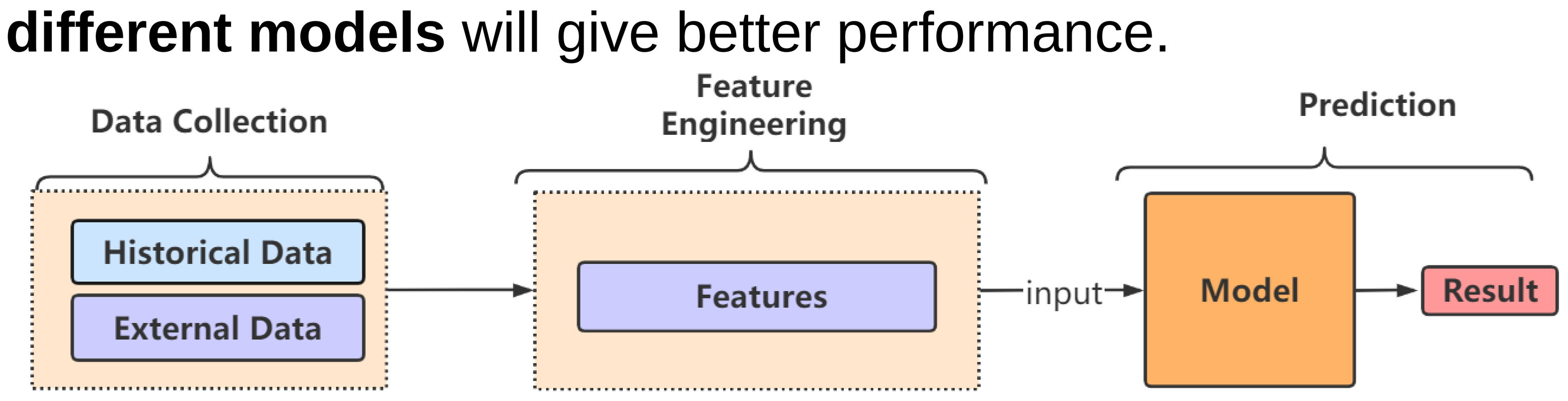

Fig.6 Overview of forecasting process