Final Exam*

* Course: CPTS 515 (Advanced Algorithm)

1st Md Mahedi Hasan
School of Electrical Engineering and Computer Science (EECS)
Washington State University (WSU)
Pullman, USA
mdmahedi.hasan@wsu.edu

***Abstract*—This paper talks about the stock market prediction using an input of a stock chart that consists of 5 information, namely, highest price, opening price, closing price, lowest price, and the volume. Using only the stock chart, for a fairly long sequence, $k \geq 2000$, we have developed an algorithm for predicting the next day stock price. The Markov chain model has been imparted to predict the stock prices. The Markov model, calculates the transition probability and that multiplied with the initial state value to make the prediction for the next state. Besides, we also have compared to charts based on the conditional likelihood of two different Markov models.**

***Index Terms*—Markov models, transitional probabilities, stock market, prediction**

I. Introduction

It is commonly stipulated that the dynamic of the stock market is unpredictable. However,if we look at the fluctuations of the stock prices, we would see that the price variability is following a pattern, so, it is predictable. A majority portion (about 80%) of the stock prices are predicted by the algorithms. From a trader's perspective who wants to make profit from the market movement, would buy at a lower price and sell at a high price at some interval of time and make profit. So, the big idea is "buy low, sell high". But making profit by trading is not easy, in fact it is super difficult since it is hard to beat the market index (benchmark).The biggest problem (challenge) is that we don't know when to buy and when to sell. The key idea behind the stock trading is that we can only see the information at the current time (time when we decide to buy or sell), we don't know the future price or market movement. So, it is possible that the price will go up or go down, or may be a crash or a boom. Here, we need to have such an algorithm that can detect this future behavior of our stock price, and which is not easy!

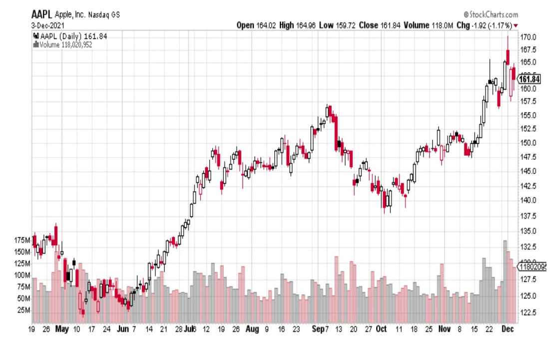
A. *Stock Chart and their intuitive behaviour*

In order to understand our problem well, we need to understand the stock chart / market very well. The following stock price chart of Apple that has daily price.

Identify applicable funding agency here. If none, delete this.

Fig. 1: Stock price chart

Source: Final Exam Question



In a bar (candle stick) of stock chart, we have information about the (i) Highest price, (ii) open price, (iii)closing price, and (iv) lowest price. And (v) the volume (number of shares traded). One bar just represent the data for a day (for daily chart) and the stock chart is the joining of all these bars for each of the days. These 5 distinct information about the stock, represented by a tupple of 5 numbers. Hence, the input is a long sequence of tupple of 5 numbers. And we use this array of tupple to predict the movement of the stock market for the next day. So, the stock chart can be represented in the following way:

$$\text{Chart} = \alpha[1, ..., n]$$

where, n represents the current day and reasonably large (e.g. $n \geq 2000$). Here, each $\alpha[i]$ is a tupple of 5 numbers.

II. Stock Market Trading Scenario

A. *Problem with high-frequency trading*

The "high-frequency trading" is one where the trading is done at a very short interval and of course very fast, such as, for an interval of second, few seconds (e.g. 5/10 seconds), minutes etc. In 1980's, mathematician, Jim Simons developed an algorithm for predicting short-term (high-frequency) stock price variability. This high-frequency trading is a mature technology following the mature algorithms. So,it is also difficult (almost impossible) to defeat this well established algorithms in trading.

1) Pitfall of long term trading: Now, if we ignore the short term jig-jag (fluctuation) rather look a the long term behavior of the stock prices we would see a pattern or trend. If the candle sticks (bars) are for a week, a month, etc. that means we are looking for long term investment.We call this a low frequency approach. In this approach, for a long term investment, we don't need any computer algorithm. The big idea is, buy and hold.

B. Trading by following the daily chart (Midterm trading)

First of this an open problem. We solve this by indirect approach. One important note at this point is that we don't want to predict stock prices using any direct method (solution) because that will eventually incur loss in trading.

C. Efficient Market Hypothesis (EMH)

The principle of EMH is that "all market information available is already factored into the price of a stock." This implies that we just need the stock charts to predict the their behavior, no other information are needed.

III. Tasks to Perform

The assigned tasks are:

A. Task A

To design an algorithm for predicting the stock price for next day using the stock chart α with $k \geq 2000$.

B. Task B

To design an algorithm to measure the similarity between two stock charts (α) and (β)

IV. Task A: Markov Chain Approach

There are myriad of ways to solve this problem and each of the approaches has their own pros and cons. In our case, we are applying the **Markov chain model** for predicting the price for next day in the stock market. Before explaining how are we going to use the Markov chain model, we need to make few definition / related concepts clear.

A. Markov Chain and Markov Property

A Markov chain or Markov process is a stochastic model that describes a sequence of possible events where the probability of each event depends only on the state attained in the previous event [2]. And, the Markov property sates that the probability of moving from the current state to the next state depends only on the current state, and not on the previous state. Let, we have a sequence of random variable $X_1, X_2, ..., X_n$. This sequence is called a Markov chain if for all states, $s_0, s_1, ..., s_n$ in the state space the probability from the current state to the next state is

$$\begin{aligned} &P(X_{n+1} = x | X_1 = x_1, ..., X_n = x_n) \\ &= P(X_{n+1} = x | X_n = x_n) \end{aligned} \tag{1}$$

All possible values of X_j from the sequence form the state space of the Markov chain.

B. Transition and Transition Probability

The changes of states in a sequence is called the transition. In our case, we can think this as a day to day price changes of the stock market. Now, this changes comes with certain probabilities and these probabilities are called transition probability.

C. State Space

Once we look for the transition probabilities, we need to understand the concept of state space. State space is the process of obtaining the transition probabilities using a transition matrix. The matrix describes the transition probabilities and an initial state across the state space. By definition, we assume the process consists of all possible states and transitions, so, there is always a next state, and it guarantees that the process does not terminate.

D. Memory less Property of Markov Chain

The memory less property of Markov chain states that "tomorrow only depends on today, not yesterday or any of the past days". This is a big idea in our approach to solve the problem. This implies that using the Markov chain model, we can predict the stock prices for tomorrow by having the current day price.

E. Application of Markov Chain

First of all, our input α (stock chart) has 4 different types of price for the stock; highest price, opening price, closing price, and lowest price and the volume. And we want to predict all different types of prices and the volume for our stock of interest. We can see this in two ways, (i) developing a complex Markov chain model using all 5 information at a time and predict, and (ii) building a separate model for each of the attributes and predict.

For simplicity, we are considering developing a separate Markov chain model for each of the prices and volume and predict them. That's being said, we will have 5 different Markov chain models. Lets call them;

$$\begin{aligned} M_{HP} &\rightarrow \text{Markov model for highest price} \\ M_{OP} &\rightarrow \text{Markov model for opening price} \\ M_{CP} &\rightarrow \text{Markov model for closing price} \\ M_{LP} &\rightarrow \text{Markov model for lowest price} \\ M_{V} &\rightarrow \text{Markov model for volume} \end{aligned}$$

Now, we will develop one such Markov model and rest of them will follow the same approach.Intuitively, these models are independent to each others. Let us develop the Markov chain model for opening price. Now, for applying the model, let us first decide about the state space for the Markov model. Here, we are looking for the prediction of stock price, so, the states will represents the prices. We can use the range of prices instead of the exact value of price in the Markov models and that are represented in the states.Lets say, we have 9 different states, of which, one indicates that there is no change of prices, and 4 indicates a percentage of increases of price and other 4 indicates the decreases of prices. The states are $s_0, s_{+1}, s_{+2}, s_{+1}, s_{+4}, s_{-1}, s_{-2}, s_{-3}, s_{-4}$. Here, s_0 indicates no changes of price, and others indicates the changes of prices.

F. Calculating the Probabilities for States

In the following table, we have our hypothesized states with their counting of days for a certain changes in 2000 days.

TABLE I: Hypothesized values for 2000 days with 9 states

States	State Description	# Days
s_0	No increase or decrease of price	150
s_{+1}	(1-25)% increase of price	450
s_{+2}	(26-50)% increase of price	300
s_{+3}	(51-75)% increase of price	221
s_{+4}	(76-100)% increase of price	7
s_{-1}	(1-25)% decrease of price	422
s_{-2}	(26-50)% decrease of price	281
s_{-3}	(51-75)% decrease of price	100
s_{-4}	(76-100)% decrease of price	70

Therefore, we can calculate the initial state vector with corresponding probabilities.

$$\eta_0 = \{s_0, s_{+1}, s_{+2}, s_{+3}, s_{+4} s_{-1}, s_{-2}, s_{-3}, s_{-4}\}$$
$$= \{0.075, 0.225, 0.1505, 0.1105, 0.0035, 0.111, 0.1405, 0.05, 0.035\}$$

These probabilities adds up to 1. This also gives a kind of indication about the move, but this is very naive. Let us now calculate the transition probability matrix which incorporates more information from the data.

G. Calculating State Transition Probability matrix

Now, we select the state of the last day (2000-th day) and we want to predict the next day. Let us assume, the state of the last day is s_2, so, the price could vary from $(26-50)\%$. So, from our current state s_2, we can go any of the other states including the state itself. From the chart, we can count the number of transition from state s_2 to other states. For an instance, the number of transition from s_2 to $s_0 is 50. And since we are currently at state s_2$ on 2000-th day and we don't know our next move, so, we consider the number of occurring as 301-1 = 300, thus the transition probability from $s_2 \to s_0$ is $\frac{50}{300} = 0.1667$ and this way we can count the number of transition from s_2 to all other states and their probabilities. Hence, for a 8×8 matrix of transition probability, we will get the transition probabilities for every state to every other states. Let say our transition matrix is P, and the matrix could be like bellow;

$$P = \begin{bmatrix} 0.167 & .... & ... & 0.02 \\ 0.07 & & & . \\ . & & & . \\ . & & & . \\ & & & \\ . & & & . \\ 0.029 & .... & .... & 0.42 \end{bmatrix}$$

Now, our target is to get the prediction for 2001-th day. So, our initial state vector would be

$$\eta_0 = (0,0,1,0,0,0,0,0,0)$$

Now the probability of the opening price for 2001-th day would be

$$\eta_{2001} = \eta_0 \times P \tag{2}$$

This will give a matrix of order 1×8, thus we get the probability of each of our 8 states for the next day. This is how we can find the prediction of next day using the Markov model. And the similar process can be applied to other prices and get the prediction.

H. Algorithm

The following algorithm (steps) will give us the prediction for stock price using the Markov model.

- Step 1: Determine the states
- Step 2: Count the number of transition from each states to every other states
- Step 3: Calculate the transition probability matrix for the state space
- Step 4: Multiply the transition probability matrix with the initial state

V. Task B: Compare two charts

Here, we have two stock charts α and β. Our target is to find the similarities between these two charts. There are many ways to compare them, however, one effective way is to use our algorithm from part A. In part A, we have developed the Markov chain model, M_{OP} for the opening price. Now, similarly we also can develop a Markov chain model for our other stock chart β. Lets call this model as $M_{OP(\beta)}$ and our previous model as $M_{OP(\alpha)}$. We can compare this two model and that should give us a good indication about the similarity between two charts.

Here our objective is to compare two Markov models. One canonical way to compare two Markov model is by calculating the likelihood for the sequence observation for one Markov model $M_{OP(\beta)}$ with respect to the other model, $(M_{OP(\alpha)})$. We know, the likelihood for Markov chain model is

$$L = s_x s_y \alpha(s_x s_y)^{N(s_x s_y)} \tag{3}$$

Here, $N(s_x s_y)$ represents the observed transition from s_x to s_y. In case of the Markov model with chart β the state transition probability from s_x to s_y is $\beta(_x, s_y) = \frac{N(s_x s_y)}{N(s_x)}$, where, $N(s_x)$ represents the of times we observed a visit to s_x. Similarly, we can find the probability for a state transition of the Markov model for chart α and that will be $\alpha(s_x, s_y) = \frac{N(s_x) s_y}{N(s_x)}$, where $N(s_x)$ comes from the model of chart of α. Now, the observed sequence by the Markov model of chart α would give $N(s_x s_y) = nm(s_x \alpha(s_x s_y) + O(n))$ for each of s_x and s_y. Here, m comes from the stationary distribution of the Markov model.Therefore, we can write the likelihood function as

$$log\ L_A = n\ X_A + O(n) \tag{4}$$

Where, $X_A = \sum_{s_x s_y} m(x)\alpha(s_x s_y)\ log\alpha(s_x s_y)$. Now, with the help of central limit theorem, we find find the difference between α and β. Hence, we get the likelihood as

$$L = L_A^n \times L_{B|A} \tag{5}$$

Where, $L_{B|A} = s_x s_y \alpha(s_x s_y)^{N(s_x s_y) - nm(x)\alpha(s_x s_y)}$

Here, $L_{A|B}$ measures the difference between two models and with the estimate of this we can easily compare the difference between two Markov models and hence, two stock charts.

Similarly, we also can do the comparison for other prices and volume of two stock charts α and β

VI. Limitation and Difficulties

- In developing our Markov chain model, we have considered 8 states, however; more states would give more specific prediction of the prices. But we need to be also careful about the accuracy level of the prediction. So, there will be a trade-off between these two and we need to choose that carefully.
- The states are considering the price changes up to 100% but the price changes could be even more,so, in that case we may have more states or may be our last category with an open interval. And that might require the model to adjust accordingly and introduce difficulty.

VII. Conclusion

There are many ways to make stock market prediction. This is one of the application where we just have information of a stock chart.So, with a proper application of Markov model, we can find a good prediction of stock prices. This is applicable for next day prices prediction, as well, we can use this for predicting stock price of any next day. And the comparison of two stock prices would also give us good understanding our the performances of the stocks.

Acknowledgment

First of all, I would like to express my deep sense of gratitude to Dr. Dang for taking our class and making the learning of this course interesting. Throughout the semester we have just not learned algorithms rather we have been through a rigorous training on how to solve difficult algorithm problems. I am also thankful to the researchers whose works I have cited and learned from.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] Gagniuc, P. A. (2017). Markov chains: from theory to implementation and experimentation. John Wiley Sons.

[3] Bhusal, M. K. (2017). Application of Markov chain model in the stock market trend analysis of Nepal. International Journal of Scientific Engineering Research, 8(10), 1733-1745.

[4] Zhang, D., Zhang, X. (2009). Study on forecasting the stock market trend based on stochastic analysis method. International Journal of Business and Management, 4(6), 163-170.

[5] Yang, L., Dang, Z., Fischer, T. R. (2011). Information gain of black-box testing. Formal aspects of computing, 23(4), 513-539.