Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2002 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-10-2002

Using Hidden Markov Model for Stock Day Trade Forecasting

Wen-Chih Tsai

An-Pin Chen

Using Hidden Markov Model for Stock Day Trade Forecasting

Wen-Chih Tsai An-Pin Chen
Institute of Information Management
National Chiao Tung University
Hsinchu, Taiwan
ibm@iim.nctu.edu.tw

Abstract

Around the world, the Hidden Markov Models (HMM) are the most popular methods in the machine learning and statistics for modeling sequences, especially in speech recognition domain. According to the number of patent applications for speech recognition technology form 1988 to 1998, the trend shows that this method has become very mature. In this thesis, we will make a new use of the HMM and apply it on day trading stock forecast.

However, the HMM is based on probability and statistics theory. In a statistics framework, the HMM is a composition of two stochastic processes, a Hidden Markov chain, which accounts for temporal variability, and an observable process, which accounts for spectral variability. The combination contains uncertainly status just likes the stock walk trace. Therefore, the HMM and the stock walk trace have the same idea by coincidence. In this thesis, we will try to learn the stock syntax; just like how the HMM model was used in speech recognition in different languages, and the take the next step ahead in price prediction.

Additionally, the stock market is the reflection of the economy. The stock trace is impacted by many factors such as policy, psychology, microeconomics, economics, and capital, etc. There, in this thesis, the TAIFEX Taiwan index futures (TX) and day trade are used to avoid all the uncertainty factors. After the all experiments, it is proven that the HMM is better than the benchmark method- Random Walk method and the Investment Trust & Consulting Association method- Modified Trading method. Moreover, the result is very conspicuous by the statistics testing of significance.

Key words：Day trade、Hidden Markov Model、Modified Trading method 、Time series forecasting

1. Introduction

A typical problem faced by stock managers is to take invest that various stocks, or stocks classes, in an optimal way. To do this the stock manager must first develop a model for the evolution, or prediction, of the rates of return on investment in each of these stock classes. A common procedure is to stock of return are drive by rates of return on some observable factors. In this research we propose a model for the rates of return in terms of observable and non-observable factors; we present algorithms for the identification for this model, (using filtering and prediction techniques) and indicate how to apply this methodology to the (strategic) stock prediction problem using a mean-variance type utility criterion.

Earlier work in filtering is presented in the classical work of Liptser and hiryayev. Previous work on parameter identification for linear, Gaussian models includes the paper by Anderson et al. [1]. However, this reference discusses only continuous state systems for which the dimension of the observations is the same as the dimension of the unobserved state. Parameter estimation for noisily observed discrete state Markov chains is fully treated in [2]. The present paper includes an extension of techniques from [2] to hybrid systems, that is, those with both continuous and discrete state spaces.

In recently, the Hidden Markov Model (HMM) is a popular technology in speech recognition domain. Especially in the telephone weather query, telephone timetable query, banking telephone services, telephone stock query, … etc. Investigating into the G06 and G10 of the international patent classification (IPC) can find out that the speech recognition patents are growth from 1988 until 1998. Those patents are researched in Hidden Markov Models in the mostly. So it's implied the conclusion that the speech recognition technology is very ripe after 1998. This is illustrated in Fig. 1.1.

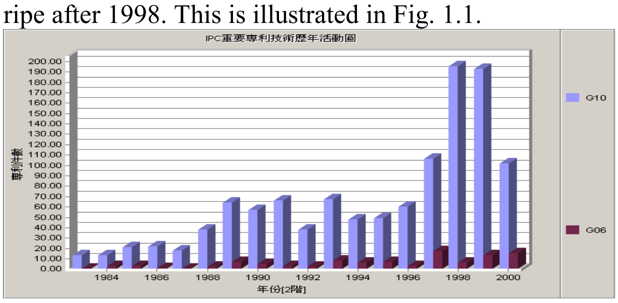

Fig1.1 IPC G10 and G06 analysis graph

As far as we know, the traditional speech recognition system shows the some major components. This is illustrated in Fig. 1.2. The digitized speech signal is first transformed into a set of useful measurements or features at a fixed rate. These measurements are then used to search for the most likely word candidate, making use of constraints imposed by the acoustic, lexical, and language models. Throughout this process, training data are used to determine the values of the model parameters. In this work, the input signal will be transfer into stock day trade wave signal.

In order to facilitate identification of objects from patterns, this work will assume that the set of patterns that can be produced by a component is finite. So this work will make some experiment for try to find out the pattern from the day trade history records of the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIFEX future, TX) from 1998 to 2002. In additionally, according to the law source of the Securities and Futures Commission (SFC) day trade rule in May 1994. The law provides a lawful transaction rule to investor. This work will base on the law for approve the architecture feasibility.

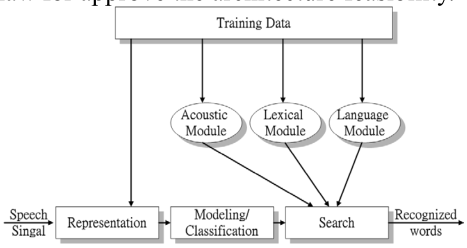

Fig1.2 The speech recognition system architecture

1.2 Research scope of this work

In this work, the research source data is the day trade history records of the TAIFEX futures (TX) from 1998 to 2002. As result of the records are day trades. So it will not be impacted form any factors. No matter policy, psychology, microeconomics, economics, capital, last night USA stock market etc. It's included the opening price, opening quotation, and closing price, closing quotation, volume, transaction price, high price and low price. It's total 866 days trade history records from 21, July, 1998 to 20, February, 2002.The training time is from 21, July, 1998 to 21, July, 2001. The simulation time is from 24, July 2001 to 20, February 2002. Additionally, this work will try to apply the HMM into the recognition model to be the forecasting strategy and compare two traditional invest strategy models. One is the Random walk[9] strategy model that is the benchmark model, and the other is the Investment Trust & Consulting Association Modified Trading model[18].

2. Background

2.1 Overview stock market theory

The trade credit is the generic term for a stock investor's purchase of supplies. In the traditionally, the vendors and suppliers are often willing to sell on credit and this source of working capital financing is very common for both startup and growing businesses. Suppliers know that most small business rely primarily upon a limited number of suppliers and that small businesses typically represent relatively small order risks; as long as the supplier keeps a tight rein on credit terms and receivables, most small businesses are a worthwhile gamble for future business. But the stock trade credit is be provided rule form SFC for those vender's want to buy stock but they have not enough money, or those supplier want to sell stock but they have not hold any stock. The rule has three systems include Purchase on margin, short on sell and day trade.(Purchase on margin ,Short on sell and Day trade)

The day trade is the generic term for a stock investor's to loan stocks or money from Stock Exchange Company and sell or buy stocks in the same day. That is due to one of the lowest risk transaction process. In the other words, it's the lowest return transaction process. In fact the process have a very good operation in the same day. So it can evade the risk of the last night USA stock risk[19]. No matter what's happen, the investor just loss the 7% of capital even which the decision is wrong. So the day trade of the SFC provides a very stability trade system. But the day trade system has three previous conditions as below.[19]

· The volume must over some quotation. That is because the investor can buy and sell into stock exchange market any times.

· The price undulation must over some degree. That is because the investor can buy and sell into stock exchange market any times. So the investor can have at least returns.

· The transaction cost must low or the returns of the investor will be reduce comparative.

Summing up the previous various reasons, the stock market cannot find out the stock to fit the each condition. But in fact, it is not adapting to the stock market day trade conditions. The conditions are more adapting to futures market. As well as, the Taiwan government trend to encourage investor to make uptrend. Just as the margin of finance margin trade is lower than the margin of stock margin trade. Such as the other rules limits to the buy and sell order. So the strategy of the operation is very important. The thesis will experiment in the TAIFEX futures (TX) .

2.2 TAIFEX futures (TX)

The TAIFEX futures (TX) contract is based on the SFC security and future laws. The TAIFEX futures (TX) is the generic term for each stock's volume of circulation, which enter the market, be weighted to calculate the index value. In other words, the big capital stock will impact over the small capital stock. It is mean that TAIFEX futures (TX) contract is based on the volume of circulation-weighted index to be the target. So the relation is very interdependence between the TAIFEX futures (TX) and the volume of circulation. Even the weighted index pulsating will bring the TAIFEX futures (TX) pulsating in the same time. So the investor does not need to study each stock changing. The investor just need to study the TAIFEX futures (TX).

2.3 Overview forecast time series theory

In the section, the thesis wills introduction two major basic time series stock forecast models, Linear forecast model and Non- Linear. forecast model. The section will introduction from basic model. So the section will introduce four linear forecast models and five non-linear forecast modules. The introduction will not only introduce the module theory and also introduce the application.

2.4 Overview Modified Trading method theory

In the current method of the Investment Trust and Consulting Association, the trader uses the population method of modified trading on day trade. About the method is follow as the experiment of the Investment Trust and Consulting Association. So this work will to implementation the theory in an algorithm. The method can transfer to the below table. The method catches three statuses from three-time point. One is 9:05 and the other is 9:10, and other is 9:15. The three statuses can combine to a status series.

The Modified Trading method applies to all eight Full and Partial Gap scenarios above. The only difference is instead of waiting until the price breaks above the high; you enter the trade in the middle of the rebound. The other requirement for this method is that the stock should be trading on at least twice the average volume for the last five days. This method is only recommended for those individuals who are proficient with the eight strategies above, and have fast trade execution systems. Since heavy volume trading can experience quick reversals, mental stops are usually used instead of hard stops.

3. Hidden Markov Model

3.1 Overview Hidden Markov Model

The Hidden Markov Model has been published form 1960s[5]. This model is based on the probability and statistics theory. The Hidden Markov Model is a finite set of states, each of which is associated with a (generally multidimensional) probability distribution. The model is the population method on the speech recognition domain. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are ``hidden'' to the outside; hence the name Hidden Markov Model. This work hope to applied the model in the stock day trade forecast.

As well the HMM has two major basic model which one is Continuous HMM(CHMM) and the other is Discrete HMM(DHMM) by the observation probability. As a result of the stock daily record is coming from matched pricing between buy and sell. So the day trade records are one of the discrete observation probabilities. The thesis is picking up the discrete HMM to experiment.

An HMM is a Markov chain and a double-layer random probability system, where each state generates an observation. Only see the observations, and the goal is to infer the hidden state sequence. HMMs are very useful for time-series modeling, since the discrete state-space can be used to approximate many non-linear, non-Gaussian systems. The hidden layer cannot be observation directly. But the hidden layer can be observed from the past random process output series sets. It is mean that the Markov chain includes the hidden layer and observation output layer.

In the thesis, the HMM is assumeed that the day trade signal was produced status series (hidden layer) by a random process. Then produce an observation sequence by the status series and the other random process. The generally HMM is a Left-Right Hidden Markov Model. The HMM's architecture is a double stochastic process. This is illustrated in Fig. 3.1.

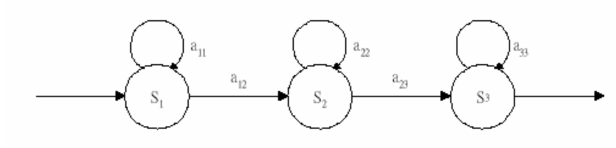

Fig 3.1 the three status Left-Right HMM

However the un-observed layer is a finite set markov chain. Its state probability distribution and states transaction are decided on between the state initial probability vector and state transition probability matrix. But the observed layer is the output of the output layer, which based on each probability distribution.

In theory, the HMM can be a multi layer. Although the result of the multi layer Markov chain is better then the single layer Markov chain. But in opposition, the model is more complex and operation.

However, the work apply the HMM into stock forecast recognition domain. It have to resolve some problem first as the below. (1)How to split each signal? (2)How to like the state sequence? (3)How much is the each status probability?(4)How to detect the next status?

About the problem, the first three problems are belonging to tanning problem. The (4) problem is the recognition process. The (1) belongs to the HMM pre-process problem. The (2) (3) belong to the building the HMM problem. The (4) is the short path problem. In the speech recognition domain, the popular problem is the Vitergbi algorithm.

3.2 HMM Parameter

The HMM consisted of a triple observed probability parameters(A,B, π). So the mathematical model of HMM can been represented to $\lambda=(A, B, \pi)$.

Each probability in the state transition matrix and in the confusion matrix is time independent - that is, the matrices do not change in time as the system evolves. In practice, this is one of the most unrealistic assumptions of Markov models about real processes. So the λ is the Markov Model. In order to define an HMM completely, following elements are needed.

In order to define an HMM completely, following elements are needed.

· The number of states of the model, *N*.

· The number of observation symbols in the alphabet, *M*. If the observations are continuous then *M* is infinite.

· A set of state transition probabilities $A=\{a_{ij}\}$.

$a_{ij}=p\{q_{t+1}=j|q_t=i\}, 1\le i,j\le N$ where q_t denotes the current state.

Transition probabilities should satisfy the normal stochastic constraints,

$$a_{ij}\ge 0, \qquad 1\le i,j\le N \tag{3-2-1}$$

and
$$\sum_{j=1}^{N} a_{ij}\ge 0, \qquad 1\le i,j\le N \tag{3-2-2}$$

· A probability distribution in each of the states, $B=\{b_j(k)\}$

$$b_j(k) = p\{a_t = v_k | q_t = j\}, 1 \le j \le N, 1 \le k \le N \quad (3\text{-}2\text{-}3)$$

Where V_k denotes the k_{th} observation symbol in the alphabed and O_k the current parameter vector.

・Following stochastic constraints must be satisfied.

$$b_j(k) \ge 0, \qquad 1 \le j \le N, 1 \le k \le M \quad (3\text{-}2\text{-}4)$$

and $\sum_{j=1}^{N} b_j(k) = 1, \qquad 1 \le j \le N \quad (3\text{-}2\text{-}5)$

・The initial state distribution, $\pi = \{\pi_i\}$

Where, $\pi_i = p\{q_t = i\}, \ 1 \le i \le N$

Therefore fuction can use the compact notation

$$\lambda = \{A, B, \pi\} \quad (3\text{-}2\text{-}6)$$

to denote an HMM with discrete probability distributions.

3.3 Vector quantification

In this paper, the recognition system must to do a characteristic value first. [29] Otherwise the system will have a very complex calculate and data. So the vector quantification will calculates the classification by the similar classes. But the general quantification will level down the recognition rate. So this work will apply the vector quantification to do the pre-process.

The code book is been produced by the vector quantification of all stock day trade training dara.. The vector quantification will been selected from the stock day trade traing dara by the geometric distance clustering. The clustering will group by the closing group. Then it will find out the must representative vector (A mean of the culster) to do the code word. Ex: Assuming that the two vector quantification are

$$X = [x_1, x_2, \ldots\ldots, x_n]^T \quad (3\text{-}3\text{-}1)$$

and

$$Y = [y_1, y_2, \ldots\ldots, y_n]^T \quad (3\text{-}3\text{-}2)$$

so the distance define is as below

$$d(X, Y) = \sqrt{\sum_{i=1}^{n} [x_i - y_i]^2} \quad (3\text{-}3\text{-}3)$$

This work will need N vector quantification to produce M code word. The population is the K-mean Clustering algorithm. The algorithm lists as below steps.

Step 1: Select an initial partition with K clusters

- An initial partition can be formed by first specifying a set of K seed points which can be the first K patterns or K patterns chosen randomly from the pattern matrix
- Different initial partitions can lead to different final clustering because algorithms based on square-error can converge to local minima. This is especially true if the clusters are not separated well.

Step 2: Generate a new partition by assigning each pattern to its closest cluster center.

Step 3: Computer new cluster centers as the centroid of clusters

- A set of K patterns that are well separated form each other can be obtained by taking the Centerior of the data as the first seed point and selection successive seed points which are at least a certain distance away from the seed points already chosen.
- Square-Error

Step 4: Repeat steps 2 and 3 until an optimum value of the criterion function is found

- Partitions are updated by reassigning patterns to clusters in an attempt to reduce the square-error.
- The Euclidean metric is the most common metric for computing the distance between pattern and a cluster centers

Step 5: Adjusts the number of clusters by merging and splitting existing clusters or by removing small, or outlier, clusters.

- Some clustering algorithms can create new clusters or merge existing clusters if certain conditions are met.

3.4 Building Hidden Markov Model

In this research, the stock day trade records will be representatived vector. pre-process. So the each day will be transfed to the vector series. It represents the each day vector quantification code book. Then the system can forecast the closing pric by the code book seraching.

But the system must been trained the history day trade data before forecast. The training process can get the each model prameter λ_1, λ_2,…… λ_v. Assuming that the λ is known number and the $O = \{o_1, o_2, \ldots\ldots, o_T\}$ is the output vector quantification series(T=length). The system will generate the $q = \{q_1, q_2, \ldots\ldots, q_T\}$ which is the input vector quantification series. So the probability is as below.

$$P(\frac{O}{\lambda}) = \sum_q p(O, \frac{q}{\lambda}) = \sum_{q_1, q_2, \ldots, q_T} 1 \cdot b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \ldots . a_{q_{T-1} q_T} b_{q_T}(o_T)$$

To assume that the HMM have N status, and the HMM can transfer from any one state to anothers. So the status transfer series have N^T .If the T is very big ,the system will have very heavy Big-O. So the probability of the HMM vector quantification series can calcuate form the Forward procedure and Backward procedure.

3.4.1 Forward procedure

Definition: In t time points, there are status in the i status. And the observed probability is the $o_1, o_2, \ldots\ldots, o_t$, and the probability of the series is $\alpha_t(i)$.

$$\alpha_t(i) = P(O_1 O_2 \ldots . O_t, q_t = i | \lambda)$$

$$P(O_1 O_2 \ldots . O_t, q_t = i | \lambda) = \sum_{i=1}^{N} P(O_1 O_2 \ldots . O_t, q_t = i | \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

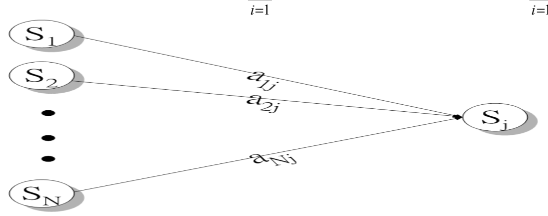

Fig 3.2 The Forward procedure

Algorithm step:

Step 1: Initial setup

$$\alpha_t(i) = \pi_i b_i(O_1), 1 \le i \le N$$

Step 2: Recursive

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_i(i) a_{ij}\right] b_j(O_{t+1}), \begin{matrix} 1 \le i \le T-1 \\ 1 \le j \le N \end{matrix}$$

Step 3: Final

$$P(\frac{O}{\lambda}) = \sum_{i=1}^{N} \alpha_t(i)$$

3.4.2 Backward procedure

Definition: In t time points, there are λ in the i status. And the observed probability is the $o_{t+1}, o_{t+2}, \ldots\ldots, o_T$ from t+1 to T, and the probability of the series is $\beta_t(i)$.

$$\beta_t(i) = P(O_{t+1} O_{t+2} \ldots\ldots O_T, q_t = i | \lambda)$$

$$P(O_{t+1} O_{t+2} \ldots\ldots O_T, q_t = i | \lambda) = \sum_{i=1}^{N} P(O_{t+1} O_{t+2} \ldots\ldots O_T, q_t = i | \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

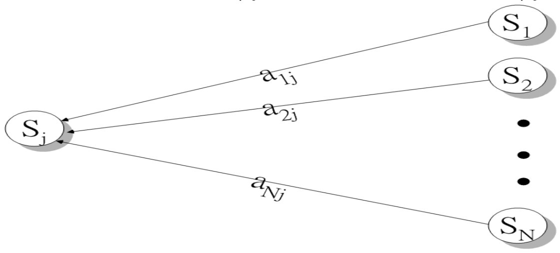

Fig 3.3 The Backward procedure

Algorithm step:

Step 1: Initial setup

$$\beta_T(i) = 1, 1 \le i \le N$$

Step 2: Recursive

$$\beta_t(i) = \left[\sum_{i=1}^{N} a_{ij} b_i(o_{t+1})\right] \beta_{t+1}(j), \begin{matrix} t = T-1, T-2, \ldots 1 \\ 1 \le i \le N \end{matrix}$$

Step 3: Final

$$P(\frac{O}{\lambda}) = \sum_{i=1}^{N} \beta_t(i)$$

3.5 Estimate the HMM

One great advantage of HMMs is that they can be estimated from sequences, without having to align the sequences first. The sequences used to estimate or train the model are called the training sequences, and any reserved sequences used to evaluate the model are called the test sequences. The model estimation is done with the forward-backward algorithm, also known as the Baum-Welch algorithm, which is described in [4]. It is an iterative algorithm that maximizes the likelihood of the training sequences. So assuming that two vector quantifications $P(O/\lambda), P(O/\bar{\lambda})$ to represent the likehood of O. Then have the $P(O/\bar{\lambda}) \ge P(O/\lambda)$ relateion between the two vector quantifications. Then use the substitution methods to find out the new parametne λ.

Step 1: In t time points, there λ transfer from the i status to j status. The λ probability is $\zeta_t(i,j)$.

$$\zeta_t(i,j) = P(q_t = i, q_{t+1} = j | O, \lambda)$$

$$= \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O/\lambda)}$$

$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O/\lambda)}$$

$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{J=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}$$

Step 2: In t time points, there λ transfer from the i status to i status. The λ probability is $\gamma_t(i)$.

$$\gamma_t(i) = \sum_{j=1}^{N} \zeta_t(i,j)$$

$$= \frac{\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{p(o/\lambda)}$$

Step 3: The new transfer function will be changed to $\bar{a}_{ij}$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \zeta_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$= \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{T-1} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}$$

$$\bar{b}_{ij} = \frac{\sum_{\substack{t=1 \\ s.t.o_t = v_k}}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(j)}$$

$$= \frac{\sum_{\substack{t=1 \\ s.t.o_t = v_k}}^{T} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{T} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}$$

Step 4: α, β scaling

$$\alpha_t(i)a_{ij}b_j(O_{t+1}) = \frac{\tilde{\alpha}_t(i)}{\prod_{r=1}^{i} c_r} a_{ij}b_j(O_{t+1}) \frac{\tilde{\beta}_{t+1}(j)}{\prod_{r=t+1}^{T} c_r}$$

$$= \frac{\tilde{\alpha}_t(i)a_{ij}b_j(O_{t+1})\tilde{\beta}_{t+1}(j)}{\prod_{r=t+1}^{T} c_r}$$

Step 5: So the $\overline{\alpha}_{ij}, \overline{b}_j(k)$ changed to.

$$\overline{\alpha}_{ij} = \frac{\sum_{t=1}^{T-1} \tilde{\alpha}_t(i)a_{ij}b_j(O_{t+1})\tilde{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N} \tilde{\alpha}_t(i)a_{ij}b_j(O_{t+1})\tilde{\beta}_{t+1}(j)}$$

$$\overline{b}_j(k) = \frac{\sum_{\substack{t=1 \\ s.t.o_t=v_k}}^{T} \sum_{j=1}^{N} \tilde{\alpha}_t(i)a_{ij}b_j(O_{t+1})\tilde{\beta}_{t+1}(j)}{\sum_{t=1}^{T}\sum_{j=1}^{N} \tilde{\alpha}_t(i)a_{ij}b_j(O_{t+1})\tilde{\beta}_{t+1}(j)}$$

3.6 Viterbi algorithm

Summary pervious process, this work will have a very complex algorithm to find out the correct $P(O_1O_2......O_T|M)$. So the Viterbi algorithm is been used to find out status transfer series by the stock day trade vector quantification. So the algorithm can closing the $P(O_1O_2......O_T|M)$ in the optimation algorithm.[12]

Assuming that the vector quantification is as below.

Vector quantification : $O = (O_1O_2......O_T)$ (3-6-1)

This initializes the probability calculations by taking the product of the initial hidden state probabilities with the associated observation probabilities. The max probability is as below.

$$\delta_t(i) = \max_{q_1,q_2,....q_t} P(q_1, q_2, ...q_{t-1}, q_t = i, O|\lambda) \quad (3\text{-}6\text{-}2)$$

According to the recusirve algorithm, So the max probability is as below during the status j at the time point t+1.

$$\delta_{t+1}(j) = [\max_i \delta_t(i)a_{ij}] \cdot b_j(o_{t+1}) \quad (3\text{-}6\text{-}3)$$

Step 1: Formal definition of algorithm

The algorithm may be summarized formally as:

$$\delta_1(i) = \pi_i b_i(o_1), 1 \le i \le N \quad (3\text{-}6\text{-}4)$$

$$\phi_1(i) = 0 \quad (3\text{-}6\text{-}5)$$

Step 2: Recursive

$$\delta_t(i) = \max_{1\le j\le N}[\delta_{t-1}(i)a_{ij}]b_i(o_1), 2 \le t \le T, 1 \le j \le N \quad (3\text{-}6\text{-}6)$$

$$\psi_t(j) = \arg\max_{1\le j\le N}[\delta_{t-1}(i)a_{ij}]b_i(o_1), 2 \le t \le T, 1 \le j \le N \quad (3\text{-}6\text{-}7)$$

Step 3: End criteria

$$P^* = \max_{1\le j\le N}[\delta_T(i)] \quad (3\text{-}6\text{-}8)$$

$$q_T^* = \arg\max_{1\le j\le N}[\delta_T(i)] \quad (3\text{-}6\text{-}9)$$

Step 4: Path look back

$$q^*_t = \psi_{t+1}(q^*_{t+1}), t = T-1, T-2, ..., 1 \quad (3\text{-}6\text{-}10)$$

Symbol define:

$\delta_t(i)$: The path have the max probability at time point t.

$\psi_t(i)$: The satus is that before current status.

P^* : The result is that be calculated by Viterbi algorithm.

q^* : It is the optimal status time series. $(q_1^*, q_2^*, ...q_T^*)$

Summarized pervious, the Viterbi algorithm is follow the initial probability $\delta_t(i)$. Then it can obtain a max probability series by rescursive algorithm.(3-6-8) The Viterbi algorithm can obtain the max probability status time series. (3-6-10). So the algorithm can reduce the algorithm search time. This is illustrated in Fig. 3.4.

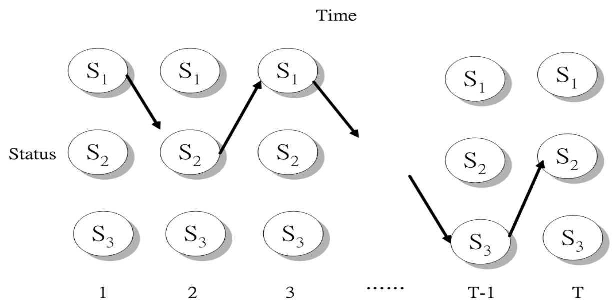

Fig. 3.4 The optimal algorithm of the Viterbi algorithm

4. Implementation

4.1 Research scope

This paper will build a HMM architecture to forecast the TAIFEX futures (TX) closing price by the previous training. The data source is come from the TEJ database. This paper is forecasting the TAIFEX futures (TX) closing price. So this paper will to build an each five minutes day trade records first. This paper will to forecast the TAIFEX futures (TX) behavior in one day.

The source data of this work is as below list.

(1)The records of the each five minutes TAIFEX futures
(2)The volume of the each five minutes
(3)The high price of the each five minutes
(4)The low price of the each five minutes
(5)The opening price
(6)The closing price

The source data has 41439 records of the each five minutes TAIFX futures (TX). The source data can transfer to total 866 days trade history records from 21, July, 1998 to 20, February, 2002.The training time is from 21, July, 1998 to 21, July, 2000.So the training data has 20946 records. (Table 4.1) The simulation time is from 24, July 2001 to 20, February 2002. So the simulation data has 20493 records.

4.2 Simulation architecture

In this section, this work will to introduce the HMM method in detail that is because the model is more complex then the other two methods. The simulation architecture is illustrated in Fig. 4.1.

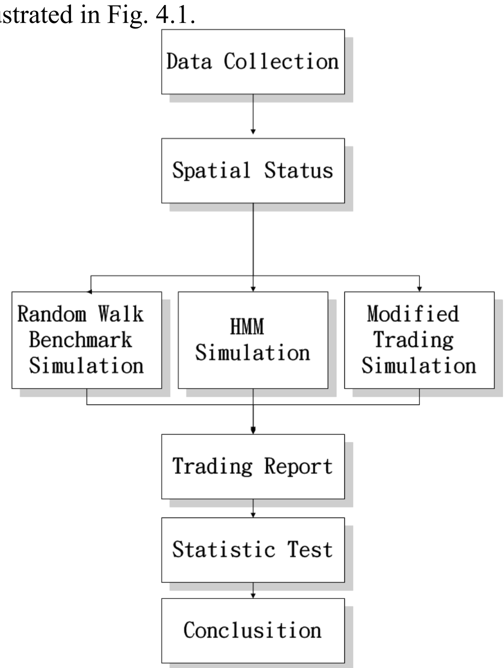

Fig. 4.1. Simulation Architecture

4.3 Simulation model

In this section, this paper will to introduce the two simulation methods in the next two small sections. The first section is the benchmark method -Random Walk method. The second section is the method - Modified Trading method. The next section is the HMM. This work will to introduce the HMM method in detail that is because the model is more complex then the other two methods.

4.3.1Random Walk model

Robbers publish this Random Walk model is in 1967. The model has not intercept item. So the revenue of the forecasting mean(μ) should be zero. The mean is that the random variable is the zero expected value. The Random Walk model is sampling form the simplex with probability density proportional to

$$Z = \frac{x-\mu}{\sigma}$$

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}; -\infty < Z < \infty$$

$$E(z) = E(\frac{x-\mu}{\sigma}) = \frac{1}{\sigma} E(x-\mu) = 0$$

The model is the weak-form efficient market hypothesis in the financial forecast domain. This work will compare HMM model to this financial benchmark model.

4.3.2 Modified Trading model

The Modified Trading Method applies to all eight Full and Partial Gap scenarios above. The only difference is instead of waiting until the price above the high or below the low for a short. The method has fast trade execution systems. Since heavy volume trading can experience quick reversals, mental stops are usually used instead of hard stops. In simple terms, the Gap Trading Strategies are a rigorously defined trading system that uses specific criteria to enter and exit. Trailing stops are defined to limit loss and protect profits. The simplest method is also being applied into Taiwan. All eight of the Gap Trading Strategies can be applied to day trade.

About the method is consists of 3 time points. That is the 10:30, 10:35, and 10:45 respectively. But it is applied in the Taiwan stock market. The method should be mapping into the 9:00, 9:05, and 9:10. That the intrados chart is graphed slightly differently, showing a composite of the trades occurring every 5 minutes. The figure is illustrated in Fig. 4.2. The figure is that Akamai Technologies stock each 5 minutes char. Although a trade-by-trade charting tool would show the exact bid and ask spread, this tutorial will refer to the left-side horizontal line of each bar as the open, and the right side horizontal bar as the close of each trade.

4.4 HMM model

4.4.1 HMM template training model

This work stock day trade training architecture has two major units. One is the parameter analysis unit and the other is the HMM template unit. The architecture is illustrated in Fig. 4.3

(1) Parameter Analysis unit

This work will sample the stock day trade time series. This work will to record the all day trade time series signal. Then scattered about each 5 minutes. So separated the each 5 minutes. So the original forecasting architecture has two major units. One is the parameter analysis unit and the other is the HMM database unit.

(2) HMM template database unit

This unit will store the stock day trade time series. The database purpose is storing the each day trade time series for HMM template model. The architecture is illustrated in Fig. 4.3

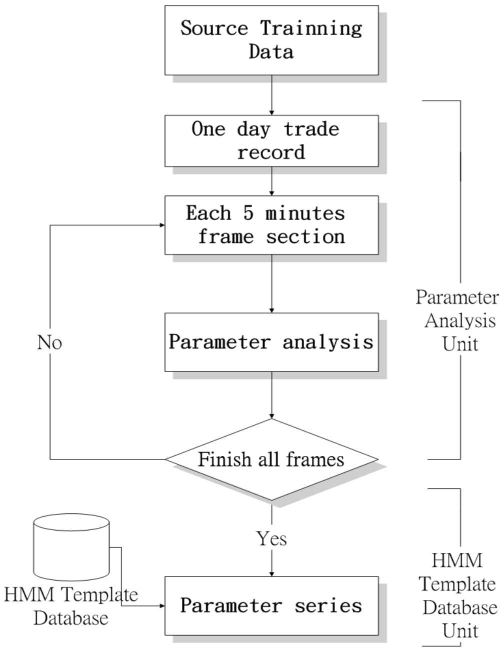

Fig. 4.3.The HMM training architecture

5. Experiment

5.1 Research scope boldface

The source data has 41439 records of the each 5 minutes TAIFX futures (TX). The source data can transfer to total 866 days trade history records from 21, July, 1998 to 20, February, 2002.The training time is from 21, July, 1998 to 21, July, 2000.So the training data has 20946 records and 538 days. The simulation time is from 22, July 2001 to 20, February 2002. So the simulation data has 20493 records and 326 days. So the experiment result will have 326 days trade record. This work will follow each experiment simulation report to do a comprehensive survey.

5.2 Random Walk experiment

The Random Walk method simulation follows previous data scope to simulation. Then the method result is the below table 5.1. The result is for statistic test. So the result will follow the below three functions 5-2-1, 5-2-2, and 5-2-3 to calculate result. The simulation time is from 22, July 2001 to 20, February 2002.

$$Sum = \sum_{Day=1}^{n} \mathrm{Re}\,sult \qquad (5\text{-}2\text{-}1)$$

$$\mu = \frac{\sum_{Day=1}^{n} \mathrm{Re}\,sult}{n} \qquad (5\text{-}2\text{-}2)$$

$$Variance = \sigma^2 = \frac{\sum_{Day=1}^{n} (\mathrm{Re}sult - \mu)^2}{n} \qquad (5\text{-}2\text{-}3)$$

Simulation Result	
Method	RW
Sum	-3435
No deal	0
Success	142
Fail	184
Mean	-10.53680982
Var	19635.98484
N	326
Max	412
Min	-456

Table 5.1 Random Walk simulations Result

5.3 Modified Trading experiment

The Modified Trading method simulation is follow previous data scope to simulation. Then the method result is the below table 5.2. The result is for statistic test. So the result will follow the previous three functions 5-2-1, 5-2-2, and 5-2-3 to calculate result. The simulation time is from 22, July 2001 to 20, February 2002.

Simulation Result	
Method	MT
Sum	3611
No deal	125
Success	107
Fail	94
Mean	11.07668712
Var	12873.59841
N	326
Max	456
Min	-363

Table 5.2 The Modified Trading simulations Result

5.4 HMM experiment

This work obtained a TAIFEX futures (TX) status transfer probability matrix as the Table 5.3. In the table, the two status (HH and LL) are the maximum of the night status probability matrix. So if the relative high status is occurring in the second time points, then the 3rd time point will be relative high status which is 50.1034% of probability.

Statue Before / Statue Now	HH	HN	HL	NH	NN	NL	LH	LN	LL
HH	**0.501034**	0.253623	0.245343	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
HN	0.000000	0.000000	0.000000	0.342752	0.345230	0.312018	0.000000	0.000000	0.000000
HL	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.352833	0.348325	0.298842
NH	0.405234	0.312751	0.282015	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
NN	0.000000	0.000000	0.000000	0.213502	0.401256	0.385242	0.000000	0.000000	0.000000
NL	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.194253	0.392130	0.413617
LH	0.364223	0.392130	0.243647	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
LN	0.000000	0.000000	0.000000	0.214453	0.372223	0.413324	0.000000	0.000000	0.000000
LL	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.186273	0.332130	**0.481597**

Table 5.3 TX status transfer probability matrix

The HMM method simulation is follow previous data scope to simulation. Then the method result is the below table 5.4. The result is for statistic test. So the result will follow the previous three functions 5-2-1, 5-2-2, and 5-2-3 to calculate result. The simulation time is from 22, July 2001 to 20, February 2002.

Simulation Result	
Method	HMM
Sum	8488
No deal	112
Success	126
Fail	88
Mean	26.03680982
Var	12962.5692
N	326
Max	411
Min	-412

Table 5.4 The HMM simulations Result

5.5 Statistic tests

This work obtained a result of the TAIFEX futures (TX) each simulation methods. The result is been ordered into the table5.5. So this section will to do some statistic test to verify the result. This work will use the difference variance tests and the difference of means tests.

Simulation Result			
Method	RW	MT	HMM
Sum	-3435	3611	8488
No deal	0	125	112
Success	142	107	126
Fail	184	94	88
Mean	-10.53680982	11.07668712	26.03680982
Var	19635.98484	12873.59841	12962.5692
N	326	326	326
Max	412	456	411
Min	-456	-363	-412

Table 5.5 The each method Result

(1)Difference variance tests:

For match pair, the output generated includes the value of the F statistic for the null hypothesis that the variance of the differences is equal to zero. In the other words, that the variance of the one simulation result will beater then the other. This work will use the Right-tailed test to verify. So the test calculates as below.

$$H_0: \ \sigma^2{}_{MT} \geq \sigma^2{}_{RW} \tag{5-4-1}$$

$$H_1: \ \sigma^2{}_{MT} < \sigma^2{}_{RW} \tag{5-4-2}$$

$$F^0 = \frac{\hat{S}^2{}_{MT}}{\hat{S}^2{}_{RW}} = \frac{1287359841}{1963898484} = 0.65561 < F_{0.05}(\infty,\infty) = 1 \tag{5-4-3}$$

So according to the staistic difference of means tests, this work must to reject H_o. In othere words, the tests obtained the significance result of this, which the Modified Trading method is beater then the Random Walk method.

(2) Difference of means tests

For match pair, the output generated includes the value of the T statistic for the null hypothesis that the mean of the differences is equal to zero. In the other words, that the mean of the one simulation result will beater then the other. This work will use the Left-tailed test to verify. So the test calculates as below.

$$H_0: \ \mu_{RW} \geq \mu_{MT} \tag{5-4-4}$$

$$H_1: \ \mu_{RW} < \mu_{MT} \tag{5-4-5}$$

$$t^0 = \frac{\overline{d} - 0}{\frac{\hat{S}_d}{\sqrt{n}}} = \frac{-21.635}{\frac{113.636}{18.055}} = -3.43 < t_{(0.05;326)} = -1.645 \tag{5-4-6}$$

So according to the staistic difference of means tests, this work must to reject H_o. In othere words, the tests obtained the significance result of this, which the Modified Trading method is beater then the Random Walk method.

$$H_0: \ \mu_{MT} \geq \mu_{HMM} \tag{5-4-7}$$

$$H_1: \ \mu_{HMM} < \mu_{MT} \tag{5-4-8}$$

$$t^0 = \frac{\overline{d} - 0}{\frac{\hat{S}_d}{\sqrt{n}}} = \frac{-14.96}{\frac{114.028}{18.055}} = -2.368 < t_{(0.05;326)} = -1.645 \tag{5-4-9}$$

So according to the staistic difference of means tests, this work must to reject H_o. In othere words, the tests obtained the significance result of this, which the HMM method is bater then the Modified Trading method.

5.6 Time series analysis of regression

The Time Series Forecast is determined by calculating a linear regression trend line using the Least Squares method. The least squares fit technique fits a trend line to the data in the chart by minimizing the distance between the data points and the linear regression trend line. The HMM method have the best revenue. But the result can't be sure the revenue will very good in future. So the analysis of regression can make sure the result. So the Least Squares method calculate as below.

$Unit: day$

$Origin: 2000/7/20$

$$Y_T = a + bX \tag{5-5-10}$$

$$a = \frac{\begin{vmatrix} \sum Y & \sum X \\ \sum XY & \sum X^2 \end{vmatrix}}{\begin{vmatrix} n & \sum X \\ \sum X & \sum X^2 \end{vmatrix}} = \frac{\sum Y \sum X^2 - \sum X \sum XY}{n\sum X^2 - \left(\sum X\right)^2} = 44.01153 \tag{5-5-12}$$

$$b = \frac{\begin{vmatrix} n & \sum Y \\ \sum X & \sum XY \end{vmatrix}}{\begin{vmatrix} n & \sum X \\ \sum X & \sum X^2 \end{vmatrix}} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} = 0.019704 \qquad (5\text{-}5\text{-}11)$$

So from the slope of the HMM method will have a plus revenue in the long-term trend.

5.7 Statistic Test conclusions

Summarized the previous section, this work have some conclusion. That HMM have the best revnue. Because the HMM method is bater then the Modified Trading method and the Modified Trading method is bater then the Random Walk method.

The HMM method have the best revenue. But in the risk view, the HMM just only have the same risk with the Modified Trading method. So this work will arrange the relation in the below table 5.7 and table 5.8.

	RW	MT	HMM
RW		No	No
MT	Yes		No
HMM	Yes	Yes	

Table 5.7 The each method revenue compare

	RW	MT	HMM
RW		No	No
MT	Yes		Equal
HMM	Yes	Equal	

Table 5.8 The each method risk compare

6. Conclusion

6.1 Thesis Conclusion

The HMM model is a very population model in the speech recognition domain. This work approve the HMM model have a very well forecasting in the stock recognition domain. Especially, the TAIFEX futures from the opening price recognized the closing price.

However, this work has a very conspicuous by the statistics testing of significance. The HMM model is beater then Random Walk model and the Modified Trading model. But this work only has a blemish in an otherwise perfect thing, which the risk is same with the Modified Trading model. Additionally, this work approve that long-term trend of the HMM model will be a positive slope line. In other words, the revenue will be increasing.

In clusion, the HMM model is a feasible and stable model in the TAIFEX futures day trading. The HMM model will have positive revenue in the long-term time series.

6.2 Discussion

This work has a very conspicuous result by the statistics testing of significance and the approving of the time series method. But this section will draw a conclusion to collect factors of back. This work has some conclusions as below list.

(1)The HMM model is based on the traditional Bayes' theorem. So the model can calculate the probability of the each status delivery.

(2) The HMM model can recognize the complex speech domain. So this method can easy to recognize three status of the stock market even to forecast the future.

(3) The TAIFEX futures have similar characters with the speech frequency.

(4) The pass papers just apply the HMM model in American stock market and those papers just apply into the next day trading.

(5) The TAIFEX futures have some special characters just in Taiwan.

6.3 Future directions

Even if, this work has very conspicuous result. But the HMM model have a risk just like the Modified Trading method. So in the future, this feet of clay can be fixed by the invest strategy just like some speech recognition system be fixed some defect s form the system architecture. So this work has some suggestions as below list.

(1) This work just applies the model in the TAIFEX future. In the future, the experiment can apply into the other stocks.

(2) The HMM model has a risk. In the future, the experiment can try to find out the optimal invest strategy to evade the risk.

(3) The HMM model is been applied into the neural network in the pass patents. In the future, the experiment can try to lead in the neural network.

(4) This works just splitter three statuses. In the future, the experiment can try to use the really K-means algorithm. In the other words, the experiment can try to use the binary splitter status.

(5) This work just used one Lot to order the TAIFEX future. In the future, the experiment can try to find out the optimal number of Lot.

(6)This work just applied the method in the front process. In the future, the experiment can try into the rear process. Just like search algorithm or strategy decision.

References

[1].Anderson, W.N. Jr., Kleindorfer, G.B., Kleindorfer, P.R., Woodroofe, M.B. "Consistent estimates of the parameters of a linear system." Ann.Math. Stat.40, 2064-2075, 1969

[2].Adlar Jeewook Kim, ," Input/Output Hidden Markov Models for Modeling Stock Order Flows", M.I.T. A.I. Lab Tech. Rep., No. 1370, Jan,2001

[3].A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang,"Phoneme recognition using time-delay neural networks," IEEE Trans. Acoust., Speech, Signal Processing, vol. 37,pp.328-339,1989

[4].A.Poritz, “Hidden Markov Models: A guided tour,” in Proc. Int. Conf. Acoust., Speech, Signal Processing, pp.7-13,1988

[5].Acero A., Stern R.M., “Environmental robustness in automatic speech recogtnition,” Proc ICASSp-90, pp 849-952, Albuq., NM,1990

[6].Admati, A.R. and P. Pfleidere,”The Theory of Intraday Patterns: Volume and Price Variability,” Review of Financial Studies, vol.1,pp.3-40,1988

[7].Bahl L.R., Jelinek F., “A maximum likelihood approach to continuous speech recognition,” IEEE Trans. Pattern Analysis Machine Intelligence,5, ,1983

[8].Bahl L.R., Brown P.F., De Souza P.V., Mercer R.L., “A tree-based statistical language model for natural language speech recognition,” IEEE Trans. On Acoustics Speech Signal Process, 37, No. 7.pp.1001-1008,1989

[9].Bollerslev, T., “On the Correlation Structure for the Generalised Autoregressive Conditional Heteroskedastic Process,” *Journal of Time Series Analysis*, 9, pp. 121-32. 1988

[10].Balasubramanian V.,”Equivalence and reduction of Hidden Markov Models”, M.I.T. A.I. Lab Tech. Rep., No. 1370, Jan,1993

[11]. Bo-Sung Kim, Bok-Gue Park, Jun-Dong Cho, Young-Hoon Chang, “Low Power Viterbi Search Architecture using Inverse Hidden Markov Model,”pp. 724 -732.,2000

[12].Claudio Becchetti and L. P. Ricotti, “Speech Recognition Theory and C++ Implementation,” John Wiley & Sons Ltd, 1999.

[13].Roberts, H., "Statistical versus Clinical Prediction of the Stock Market," unpublished manuscript, Center for Research in Security Prices, University of Chicago,May, 1967

[14].Robinson, P., "Time Series with Strong Dependence," in C. A. Sims (ed.), *Advance sin Econometrics: Sixth World Congress, Vol. I*, Cambridge University Press, pp. 47-95.,1994

[15].Turner, C. M. , R. Startz and C. R. Nelson, "A Markov model of heteroscedasticity, risk and learning in the stock market," *Journal of Financial Economics*, 25, pp. 3-22.,1989

[16].Steinhaus, S. "Comparison of mathematical programs for data analysis," University of Frankfurt, Germany, http://www.informatik.uni-frankfurt.de/ , 1999

[17].Starck, J.L., Murtagh, F., and A. Bijaoui Image and Data Analysis: The Multiscale Approach, Cambridge University Press,1998

[18]. Chen, S. H. and C. W. Tan "Estimating the Complexity Function of Financial Time Series," *The Journal of Management and Economics*, Oct.-Nov., Vol. 3.,1999

[19].Thomas Hellstrom, Kenneth, Kolmstrm,”Predictable Patterns in Stock Returns” Department of Mathematics and Physics, Malardalen University Press,1998

[20].Bagella M., Becchetti, L., Carpentieri, A., “The first shall be last. Do contraian strategy premia violate financial market efficency,” mimeo,1988

[21].D. Easley, N. Kiefer, and M. O’Hara. The information content of the trading process. Journal of Empirical Finance,vol., 4 pp.159-186, 1997.

[22].L. Rabiner.” A tutorial on hidden markov models and selected applications in speech recognition.,“ In Proceeding of IEEE, Vol. 77, No. 2, pages 257–286, 1989.

[23].L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” AT&T Bell Syst. Tech. Journal, vol. 62, no. 4, April,1983

[24].L.B. Jackson, Digital Filters and Signal Processing with MATLAB Exercises, Kluwer Academic Publishers, 1995.

[25].Y. Bengio. “Markovian models for sequential data,” Neural Computing Surveys, 2:129–162, 1999.

[26].Y. Bengio and P. Frasconi.”Input/output hmms for sequence processing,” IEEE Transactions on Neural Networks, 7(5) pp.1231-1249, 1996.

[27].J.G. Wilpon, et al., “Automatic Recognition of Keyword in Unconstrained Speech Using Hidden Markov Models,” IEEE ASSP Magazine, Vol. 38, pp. 1870-1878, 1990.

[28].J. R. Deller, “Discrete-Time Processing of Speech Signals,” Macmillan,1993.

[29]. J.G. Bergh, F.K. Soong, and L.R. Rabiner, “Incorporation of Temporal Structure Into a Vector-Quantization-Based Preprocessor for Speaker-Independent, Isolated-Word Recognition,” AT & T Technical Journal, Vol. 64, No. 5, May-June 1985.

[30]. G. Wilpon and L.R. Rabiner, “A Modified K-Means Clustering Algorithm for Use Isolated Word Recognition,” IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. 33, No. 3, pp. 587-594, June 1985.

[31]. L¨utkepohl, H.: Introduction to multiple time series analysis. Berlin, Heidelberg, New York:,Springer 1991

[32]. Liptser, R.S., Shiryayev, A.N.: Statistics of random processes. Vols. I and II. Berlin, Heidelberg, New York: Springer 1977

[33].Kaas, Robert E and Adrian E. Raftery.“Bayes Factors,” Technical Report 254,Department of Statistics, University of Washington., 1995

[34].Lin, Tsungan and Bill G. Horne. “Learning Long-Term Dependencies in NARXRecurrent Neural Networks.” *1996 IEEE Conference on Neural Networks*. pp. 1329-1338,1996

[35]. Neal, Radford M. “Monte Carlo implementation of Gaussian process models for Bayesian regression and classification,” Technical Report No. 9702, Dept. Of Statistics, University of Toronto. ftp://ftp.cs.toronto.edu/pub/radford, 1997

[36]. Valtchev, V., S. Kapadia, S. J. Young. “Recurrent Input Transformation for Hidden Markov Models.” Cambridge University Engineering Department Technical Report.,1995