

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Stock Market Trend Prediction Using Hidden Markov Model

*Deneshkumar Venugopal,
Senthamarai Kannan Kaliyaperumal
and Sonai Muthu Niraikulathan*

Abstract

In Recent years many forecasting methods have been proposed and implemented for the stock market trend prediction. In this Chapter, the trend analyses of the stock market prediction are presented by using Hidden Markov Model with the one day difference in close value for a particular period. The probability values π gives the trend percentage of the stock prices which is calculated for all the observe sequence and hidden sequences. This chapter helps for decision makers to make decisions in case of uncertainty on the basis of the percentage of probability values obtained from the steady state probability distribution.

Keywords: stock market, HMM, TPM, EPM and trend prediction

1. Introduction

The fundamental idea behind a hidden Markov model is that there is a Markov process we cannot observe that determines the probability distribution for what we do observe. Thus a hidden Markov model is specified by the transition density of the Markov chain and the probability laws that govern what we observe given the state of the Markov chain. Given such a model, we want to estimate any parameters that occur in the model. And also determined the most likely sequence for the hidden process. Finally we may want the probability distribution for the hidden states at every location.

Let y_t represents the observed value of the process at location t for $t = 1, \dots, T$, θ_t the value of the hidden process at location t and let ϕ represents parameters necessary to determine the probability distribution for y_t given θ_t and θ_t given θ_{t-1} . In our applications, y_t will either be an increase or decrease and the hidden process will determine the probability distribution of observing different letters.

Our model is then described by the sets of probability distributions $p(y_t | \theta_t, \phi)$ and $p(\theta_t | \theta_{t-1}, \phi)$. A crucial component of this model is that the y_t are independent given the set of θ_t and θ only depends directly on its neighbors θ_{t-1} and θ_{t+1} . The various distribution in which we are interested are $p(\phi | y_1, \dots, y_T)$, $p(\theta_t | y_1, \dots, y_T)$ for all t and $p(\theta_1, \dots, \theta_T | y_1, \dots, y_T)$. We will adopt a Bayesian perspective, so that we treat θ_t as a random variable [1, 2].

The measure of best is to find the path that has the maximum probability in the HMM, given the sequence X . Recall that the model gives the joint probabilities

$Pr(H, X)$ for all sequence, it also gives the posterior probability $Pr(H, X) = Pr(H, X)/Pr(X)$, for every possible state path H through the model, conditioned on the sequence X with maximum posterior probability [3, 4]. Given that the denominator $Pr(X)$ is constant in the conditional probability formula for a given sequence X , maximizing the posterior probability is equivalent to finding the state path H^* that maximizes the joint probability $Pr(H^*, X)$. Nguyen [5] has determined the optimal number of states for the HMM by using the AIC, BIC and HQ information criteria and also discussed the applications of HMM in stock trading. Hassan and Nath [6] have applied HMM to the airlines stock forecast. HMMs have been used for pattern recognition and classification problems and it was suitable for modeling dynamic systems.

2. Hidden Markov model

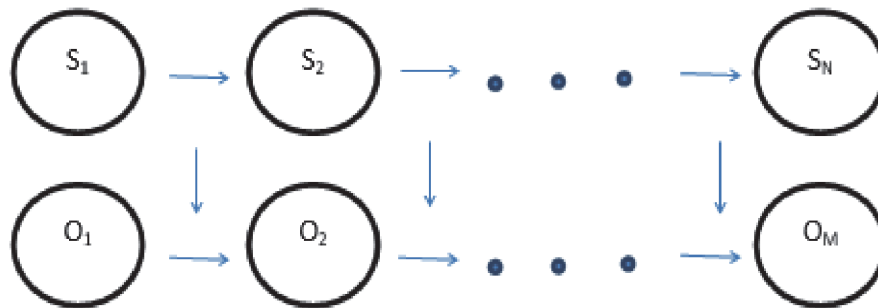
Hidden Markov model (HMM) is a stochastic model which is not directly observable, It describes the observable events that are depends on internal factors. The observable events are represented as symbols, where the invisible factor involved in the observation is represented as a state. HMM is a stochastic model where the system is assumed to be a Markov Process with hidden states and it gives better accuracy than the other models. Using the given input values, the parameters of the HMM (λ) denoted by A, B and π are found out. An HMM is defined as $\lambda = (S, O, A, B, \pi)$ where $S = \{s_1, s_2, \dots, s_N\}$ is a set of N possible states $O = \{o_1, o_2, \dots, o_M\}$ is a set of M possible observation symbols, A is an $N \times N$ state Transition Probability Matrix (TPM), B is an $N \times M$ observation or Emission Probability Matrix (EPM) and Π is an N dimensional initial state probability distribution vector and A, B and π should satisfy the following conditions (**Figure 1**):

$$\sum_{j=1}^N a_{ij} = 1 \text{ where } 1 \leq i \leq N;$$

$$\sum_{j=1}^M b_{ij} = 1 \text{ where } 1 \leq i \leq N;$$

$$\sum_{i=1}^N \pi_i = 1 \text{ where } \pi_i \geq 0$$

Hidden sequence



Observation Sequence.

Figure 1.
Diagram of HMM.

2.1 Evaluation problem

Given the HMM = {A,B, π } and the observation sequence O = o1,o2, ... ,oM, the probability that model λ has generated sequence O is calculated. Often this problem is solved by the Forward Backward Algorithm [7, 8].

2.2 Decoding problem

Given the HMM λ = {A,B, π } and the observation sequence O = o1,o2, ... ,oM, calculate the most likely sequence of hidden states that produced this observation sequence O. Usually this problem is handled by Viterbi Algorithm [7, 8].

2.3 Learning problem

Given some training observation sequences O = o1,o2, ... ,oM, and general structure of HMM (numbers of hidden and visible states), determine HMM parameters λ = {A,B, π } that best fit training data. The most common solution for this problem is Baum-Welch algorithm [9, 10] which is considered as the traditional method for training HMM.

3. Results and discussions

In this chapter, the data has been taken from Yahoofinance.com and the NSE daily close value data for a month of January 2020 period is considered for the analysis.

Here two observing symbols “I” for Increasing states and the symbols “D” for decreasing states have been used. If the differences of close value greater than 0 its observing that the symbol is “f” and If the differences of close value less than 0 its observing that the symbol is “D”. There are six hidden states assumed and are denoted by the symbol S1, S2, S3, S4, S5, S6 are indicates that very low, low, moderate low, moderate high, high and very high respectively. The states are not directly observable.

The situations of the stock market are considered hidden. Given a sequence of observation we can find the hidden state sequence that produced those observations. **Table 1** shows the daily close value of the stock market.

Interval values:

S1 = -9500 to -551.

S2 = -550 to -251.

S3 = -250 to 249.

S4 = 250 to 8500.

The various probability values of TPM, EPM and π for difference in one day, two days, three days, four days, five days, six days close value are calculated as given below (**Table 2**).

Probability values of TPM, EPM, and π for difference in one day close value (**Figure 2** and **Table 3**):

	S1	S2	S3	S4		I	D
S1	0	0	1	0	S1	0	1
S2	0	0	1	0	S2	0	1
S3	0.071	0.071	0.4286	0.4286	S3	0.2849	0.7143
S4	0	0	1	0	S4	0.5	0.5

S. no	Date	Close
1	01/02/2020	41,626.64
2	01/03/2020	41,464.61
3	01/06/2020	40,676.63
4	01/07/2020	40,869.47
5	01/08/2020	40,817.74
6	01/09/2020	41,452.35
7	01/10/2020	41,599.72
8	01/13/2020	41,859.69
9	01/14/2020	41,952.63
10	01/15/2020	41,872.73
11	01/16/2020	41,932.56
12	01/17/2020	41,945.37
13	01/20/2020	41,528.91
14	01/21/2020	41,323.81
15	01/22/2020	41,115.38
16	01/23/2020	41,386.4
17	01/24/2020	41,613.19
18	01/27/2020	41,155.12
19	01/28/2020	40,966.86
20	01/29/2020	41,198.66
21	01/30/2020	40,913.82
22	01/31/2020	40,723.49

Table 1.
Daily close value of NSE.

Probability values of TPM, EPM, and π for difference in two day close value (Figure 3 and Table 4).

$$\begin{bmatrix} & S1 & S2 & S3 & S4 \\ S1 & 0 & 0 & 1 & 0 \\ S2 & 0 & 0 & 1 & 0 \\ S3 & 0.0111 & 0 & 0.5555 & 0.3333 \\ S4 & 0 & 0.3333 & 0.5 & 0.1667 \end{bmatrix} \quad \begin{bmatrix} & I & D \\ S1 & 0.5 & 0.5 \\ S2 & 0.5 & 0.5 \\ S3 & 0.4444 & 0.5556 \\ S4 & 1 & 0 \end{bmatrix}$$

Probability values of TPM, EPM, and π for difference in three day close value (Figure 4 and Table 5):

$$\begin{bmatrix} & S1 & S2 & S3 & S4 \\ S1 & 0 & 0 & 0 & 1 \\ S2 & 0 & 0 & 0.75 & 0.25 \\ S3 & 0 & 0.6 & 0.2 & 0.2 \\ S4 & 0.5 & 0.2 & 0.2 & 0.2 \end{bmatrix} \quad \begin{bmatrix} & I & D \\ S1 & 0 & 1 \\ S2 & 1 & 1 \\ S3 & 0.6 & 0.4 \\ S4 & 1 & 0 \end{bmatrix}$$

S. no	c.v	D in 1 day CV	o.s	D in 2 days CV	o.s	D in3 days CV	o.s	D in 4 day CV	o.s	D in 5 day CV	o.s	D in6 days CV	o.s
1	41,626.64												
2	41,464.61	162.03	I										
3	40,676.63	787.98	I	-625.95	D								
	40,869.47	-192.84	D	980.82	I	-1606.77	D						
5	40,817.74	51.73	I	-244.57	D	1225.39	I	-2882.16	D				
6	41,452.35	-634.61	D	686.84	I	-930.91	D	2156.3	I	-4988.46	D		
7	41,599.72	-147.37	D	-487.24	D	1173.58	I	2104.49	I	4260.79	I	-9249.25	D
8	41,759.69	-259.97	D	112.6	I	-599.84	D	1773.42	I	-3877.91	D	8138.7	I
9	41,952.63	-92.94	D	-167.03	D	279.63	I	-879.47	D	2652.89	I	-6530.8	D
10	41,872.73	79.9	I	-172.84	D	5.81	I	273.82	I	-1153.28	D	3806.18	I
11	41,932.56	-59.83	D	139.73	I	-312.57	D	318.38	I	-44.56	D	-1108.73	D
12	41,945.37	-12.81	D	-47.02	D	-92.71	D	405.28	I	-86.9	D	42.34	I
13	41,528.91	416.46	I	403.65	I	-450.67	D	357.96	I	47.32	I	-134.22	D
14	41,323.81	205.1	I	211.36	I	192.22	I	-642.96	D	1000.92	I	-953.6	D
15	41,115.38	208.43	I	-3.33	D	214.69	I	-22.4	D	-620.56	D	1621.48	I
16	41,386.4	-271.02	D	479.45	I	-482.78	D	697.47	I	-719.87	D	99.31	I
17	41,613.19	-226.79	D	-44.23	D	523.68	I	-1006.46	D	1703.93	I	-2423.8	D
18	41,155.12	458.01	I	-684.86	D	640.63	I	-116.95	D	-889.51	D	2593.44	I
19	40,966.86	188.26	I	269.81	I	-415.05	D	1055.68	I	938.73	I	-1828.24	D
20	41,198.66	-231.8	D	420.06	I	-150.25	D	-264.8	D	1320.48	I	-381.75	D
21	40,913.82	284.84	I	-516.64	I	936.7	I	-1086.95	D	822.15	I	498.33	I
22	40,723.49	190.33	I	94.51	I	-611.15	D	1547.85	I	-2634.8	D	3456.95	I

Table 2.
Daily close value for finding differences in one day, two day, three days, four days, five days, six days close value.

	S1		S2		S3		S4	
	I	D	I	D	I	D	I	D
S1	0	0	0	0	1	0	0	0
S2	0	0	0	0	1	0	0	0
S3	0.071	0	0.071	0	0.1429	0.2857	0	0.4286
S4	0	0	0	0.8	0.2	0	0	0

Table 3.
Transitions with probability values for one day close value.

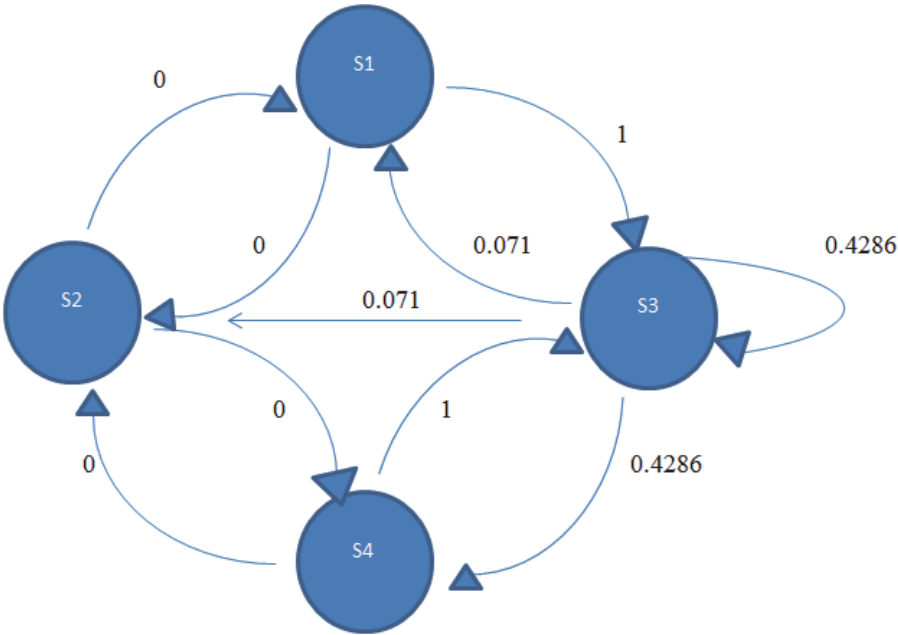


Figure 2.
Diagram of TPM day 1.

	S1		S2		S3		S4	
	I	D	I	D	I	D	I	D
S1	0	0	0	0	0	0	0.5	0.5
S2	0	0	0	0	0.5	0.5	0	0
S3	0	0.111	0	0	0.3333	0.2222	0.1111	0.2222
S4	0	0	0.3333	0	0.5	0	0.1667	0

Table 4.
Transition table with probability values for difference in two day close value.

Probability values of TPM, EPM and π for difference in four days close value (Figure 5 and Table 6):

S1

S2

S3

S4

S1

S2

S3

S4

0.3858

0.1429

0.1429

0.4286

0.5

0

0.5

0

0

0

1

0

0.4286

0.1429

0

0.4286

I

D

S1

S2

S3

S4

0.1429

0.9573

0.5

0.5

0

1

1

0

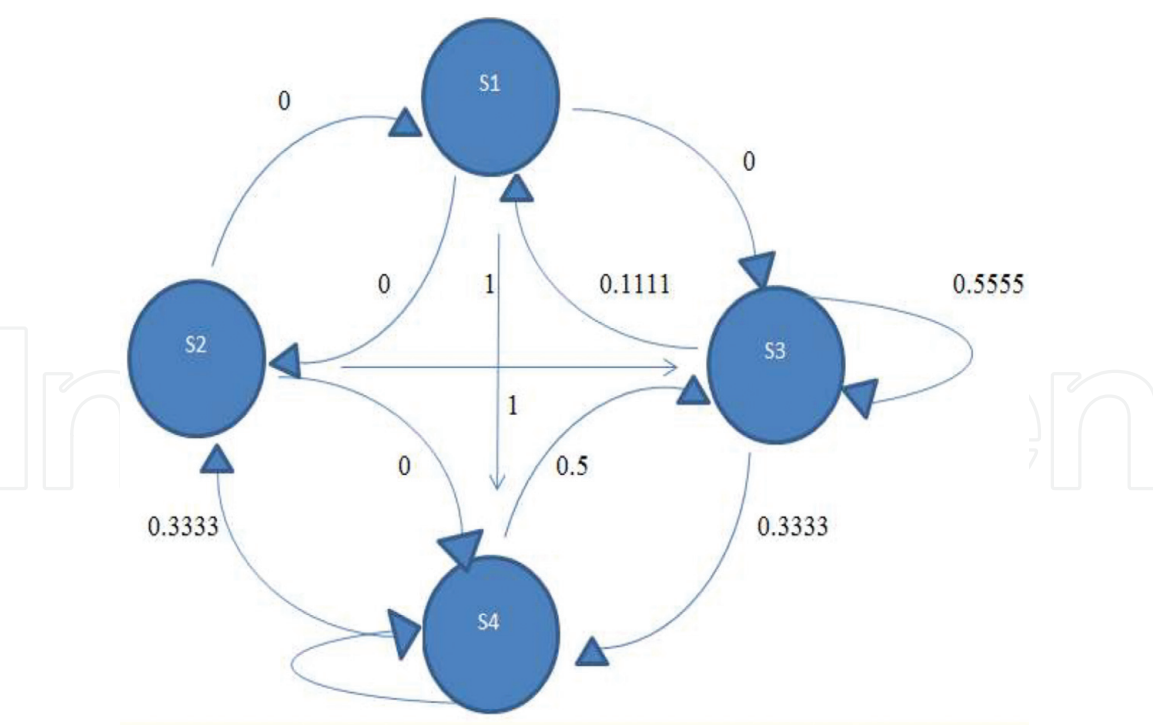


Figure 3.
Diagram of TPM day 2.

	S1		S2		S3		S4	
	I	D	I	D	I	D	I	D
S1	0	0	0	0	0	0	0	1
S2	0	0	0	0	0	0.75	0	0.25
S3	0	0	0.4	0.2	0.2	0	0	0.2
S4	0.5	0	0.2	0	0.2	0	0.2	0

Table 5.
Transition table with probability values for difference in three day close value.

Probability values of TPM, EPM and π for difference in five days close value (Figure 6 and Table 7):

	S1	S2	S3	S4
S1	0.1667	0	0.1667	0.6667
S2	0	0	0	0
S3	0	0	0.6667	0.3333
S4	0.7143	0	0	0.6667

	I	D
S1	0	1
S2	0	1
S3	0.3333	0.6667
S4	1	0

Probability values of TPM, EPM and π for difference in six days close value (Figure 7 and Table 8):

	S1	S2	S3	S4
S1	0	0.2	0.2	0.6
S2	0	0	0	1
S3	0.6667	0	0.3333	0
S4	0.5	0	0.25	0.25

	I	D
S1	0	1
S2	0	1
S3	0.667	0.3333
S4	1	0

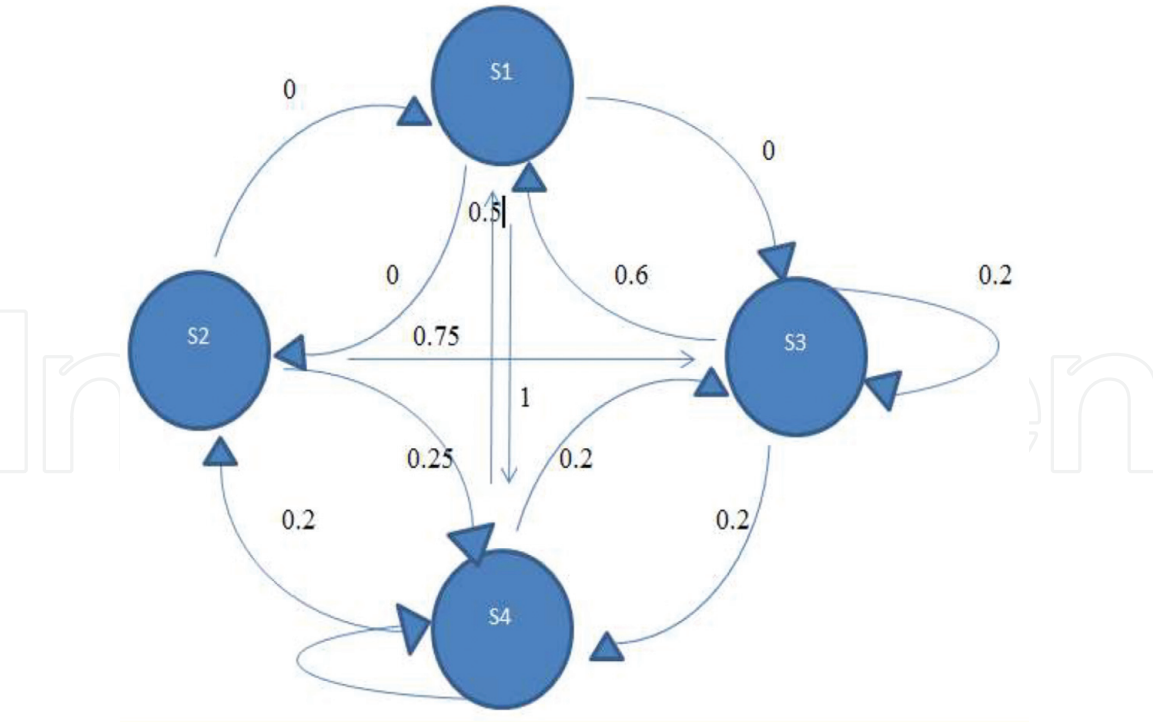


Figure 4.
Diagram of TPM day 3.

	S1		S2		S3		S4	
	I	D	I	D	I	D	I	D
S1	0.1429	0.2429	0	0.1429	0	0.1429	0	0.4286
S2	0.5	0	0	0	0	0.5	0	0
S3	0	0	0	0	0	1	0	0
S4	0.4286	0	0.1429	0	0	0	0.4286	0

Table 6.
Transition table with probability values for difference in four day close value.

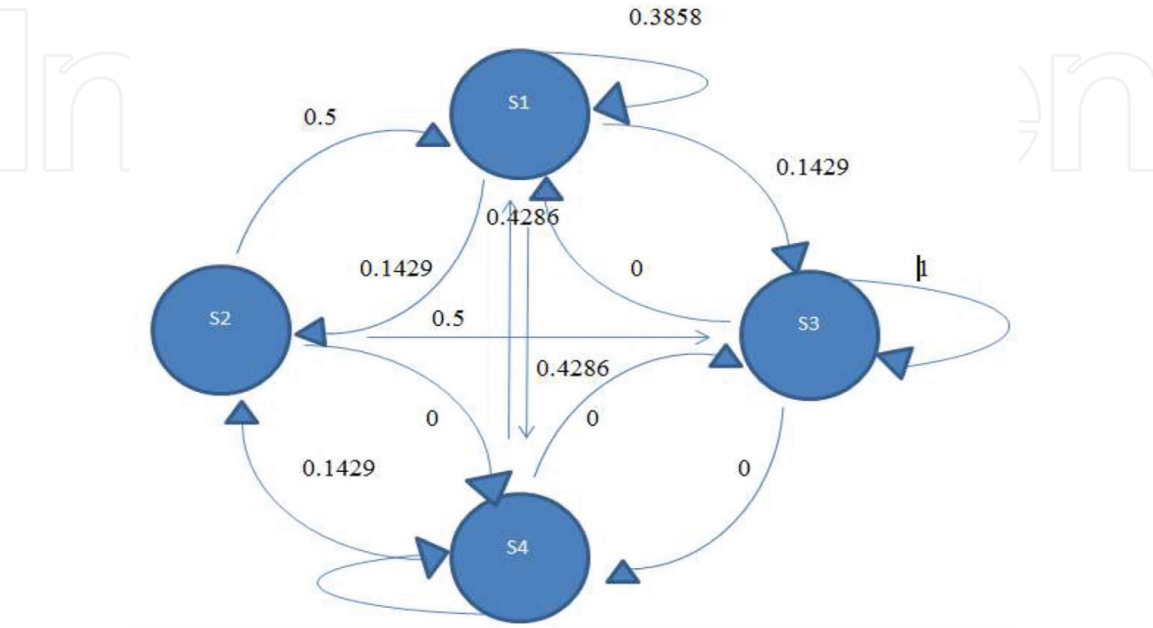


Figure 5.
Diagram of TPM day 4.

	S1		S2		S3		S4	
	I	D	I	D	I	D	I	D
S1	0	0.1667	0	0	0	0.1667	0	0.6667
S2	0	0	0	0	0	0	0	0
S3	0	0	0	0	0	0.6667	0.3333	0
S4	0.7143	0	0	0	0	0	0.2857	0

Table 7.
Transition table with probability values for difference in five day close value.

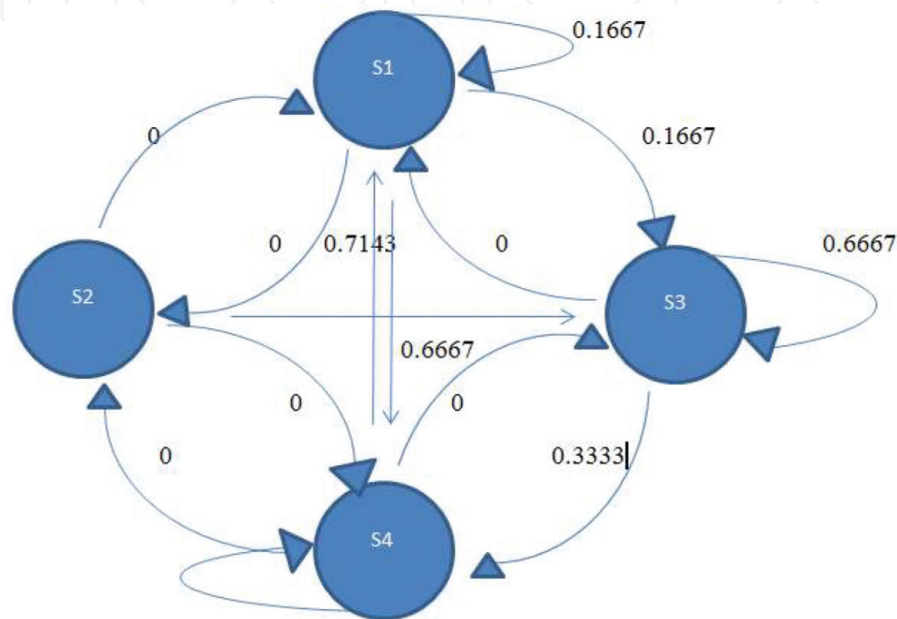


Figure 6.
Diagram of TPM day 5.

	S1		S2		S3		S4	
	I	D	I	D	I	D	I	D
S1	0	0	0	0.2	0	0.2	0	0.6
S2	0	0	0	0	0	0	0	1
S3	0.3333	0.3333	0	0	0.3333	0	0	0
S4	0.5	0	0	0	0.25	0	0.25	0

Table 8.
Transition table with probability values for difference in six day close value.

The various transitions probability values for difference in one day to six days close values are displayed in **Figure 2** to **Figure 7** respectively.

Optimum Sequence of States:

To generate a random sequence of emission symbols and states are calculated by using the function “Hmngenerate”. The HMM matlab toolbox syntax is: [Sequence, States] = Hmngenerate(L,TPM,EPM). The length of both sequence and state to be generated is denoted by L [11]. The fitness function used for finding the fitted value of sequence of states is defined by

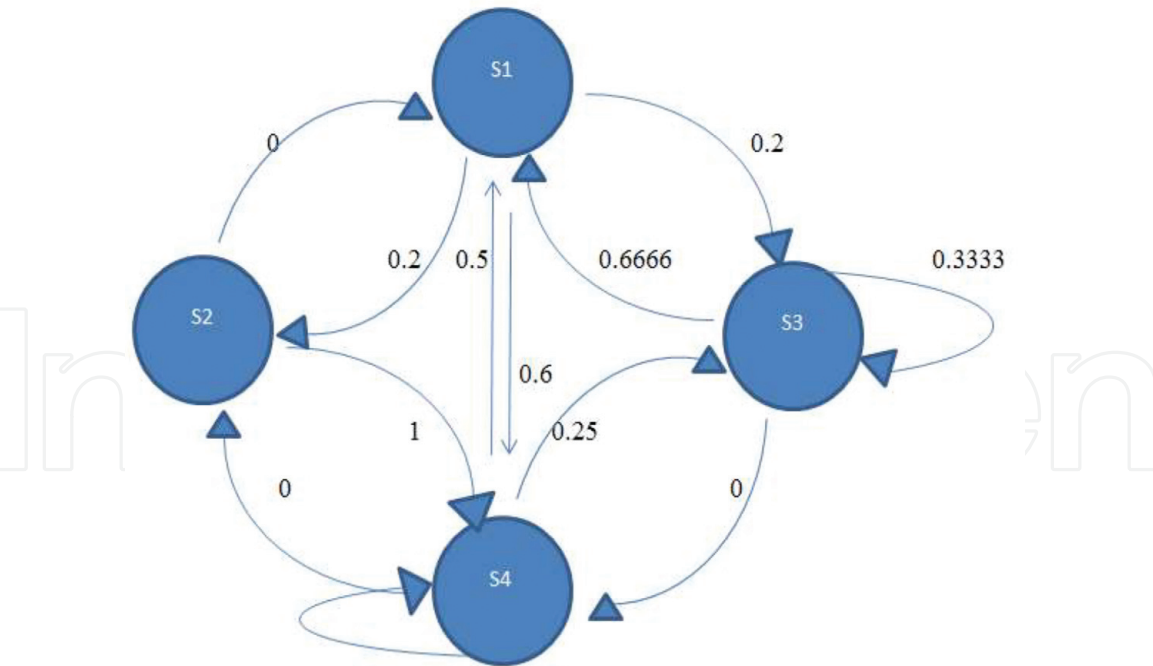


Figure 7.
Diagram of TPM day 6.

$$(\text{Fitness} =) \frac{1}{\sum \text{compare}(i,j)}$$

Using the iterative procedure, for each TPM and EPM framed we get an optimum sequence of states generated.

The length of the sequence taken as $L = 4$ and the optimum sequence of states obtained from the all six day's differences with TPM and EPM is given in the below and here 'ε' is the start symbol.

1.	ε	→	I S4	→	D S4	→	I S3	→	D S4
2.	ε	→	D S1	→	I S4	→	I S4	→	D S3
3.	ε	→	I S4	→	I s2	→	D S3	→	D S1
4.	ε	→	D S1	→	D S4	→	I S3	→	D S4
5.	ε	→	I S3	→	I S3	→	D S2	→	D S4
6.	ε	→	I S4	→	D S1	→	I S3	→	D S4

Here, the one day difference of TPM and EPM has the shortest path. So the best optimum sequence is found from one day difference in close value. Using the fitness function we compute the fitness value for each of the optimum sequences of states obtained (**Table 9**).

In column four the highest value is the fitness value and the better is the performance of the particular sequence.

S. no.	Comparison of six optimum sequence of states	Calculated value	Fitness = $\frac{1}{\sum comparison(i,j)}$
1	(1,2) + (1,3) + (1,4)	1	1
2	(2,1) + (2,3) + (2,4)	1.7	0.588
3	(3,1) + (3,2) + (3,4)	2.425	0.412
4	(4,1) + (4,2) + (4,3)	3.15	0.32

Table 9.
Comparison of six optimum state sequences.

4. Conclusion

Stock prediction is challenging due to its randomness. Hidden Markov Model can be used for stock prediction by finding hidden patterns. Here the Hidden Markov model easily recognized four states of the stock market and also it was used to predict the future values. The highest value in the Optimum State Sequences is the better performance of the particular sequence. Hidden states and sequences have been generated to easily identify the level of the sequence whether the next day value is increasing. And also identified whether the increasing level is moderate high or high or very high and also decreasing level whether moderate low or low or very low. This model will be very much useful for short term as well as long term investors.

Author details

Deneshkumar Venugopal*, Senthamarai Kannan Kaliyaperumal
and Sonai Muthu Niraikulathan
Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli,
Abishekapatti, Tamil Nadu, India

*Address all correspondence to: vdenesh@msuniv.ac.in

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Medhi J. Stochastic processes. New Age International; 1994.
- [2] Reilly C. Statistics in human genetics and molecular biology. CRC Press; 2009 Jun 19.
- [3] Brejová B, Brown DG, Vinař T. ADVANCES IN HIDDEN MARKOV MODELS FOR SEQUENCE. Bioinformatics Algorithms: Techniques and Applications. 2008 Feb 15;3:55.
- [4] Gupta A, Dhingra B. Stock market prediction using hidden Markov models. In 2012 Students Conference on Engineering and Systems 2012 Mar 16 (pp. 1-4). IEEE.
- [5] Nguyen N. Hidden Markov model for stock trading. International Journal of Financial Studies. 2018 Jun;6(2):36.
- [6] Hassan MR, Nath B. Stock market forecasting using hidden Markov model: a new approach. In 5th International Conference on Intelligent Systems Design and Applications (ISDA'05) 2005 Sep 8 (pp. 192-196). IEEE.
- [7] Rabiner L. Theory and implementation of hidden Markov models. Fundamentals of speech recognition. 1993.
- [8] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989 Feb;77(2): 257-286.
- [9] Lloyd RW. Hidden Markov Models and the Baum-Welch Algorithm. IEEE Information Theory Society Newsletter. 2003 Dec;53(4).
- [10] Mandoiu I, Zelikovsky A. Bioinformatics algorithms: techniques and applications. John Wiley & Sons; 2008 Feb 25.
- [11] Murphy K. HMM toolbox for Matlab. Internet: <http://www.cs.ubc.ca/murphyk/Software/HMM/hmm.html>, [Oct. 29, 2011]. 1998.