

Integrated Assessment in Prediction Using Markov Model

Zhouhao Bao*

Lafayette College, 111 Quad Drive, Easton, PA, USA 18042

* Corresponding author: baoz@lafayette.edu

Abstract. Markov model, usually used in the prediction of market failures, is useful in random changing systems. In recent years, the Markov model attracts interest not only in the data science area but also in the economic area. This paper studies specifically the role of the Markov model in forecasting supply and demand side effects. Among the issues considered are the Markov chain, Hidden Markov model, Markov decision process, Tolerate Markov model, and Markov-chain forecasting models. Possible issues and improvements of the Markov model and its applications to specific cases in various fields are also briefly discussed. The result of the study shows that the Markov model can effectively simulate and forecast the supply and demand side effects. The result of this research help builds up the fundamental ideas and structure of the Markov model. Also, the result can help policymakers as references and be used to predict future supply and demand side effects in various data and economic fields, which could be beneficial by avoiding possible incoming economic failures.

Keywords: Markov Model, Hidden Markov Model, Markov Chain, and Forecasting.

1. Introduction

Markov model, issued by Russian mathematician Andrey Markov, is a process that states that the probability of each event is dependent solely on the current status of events and is irrelevant to past events. In other words, this can be understood as what is going to happen in a sequence depending only on the current status. Therefore, a Markov chain actually represents an infinite sequence of events that happens to give a discrete time sequence. There are two different types of Markov chain: discrete-time Markov chain and continuous-time Markov chain. The difference between these two types of Markov chains relies on the difference in measurability. A discrete-time Markov chain, just like the name, has a measurable state space. However, the continuous-time Markov chain contains any process which has Markov property [1]. A classic example of a Markov chain is “Drunkard’s walk”, which states that the probability of a transition walk along the number line is independent of each other and depends solely on its current position instead of any other patterns it took before. Moreover, in addition, there is a Hidden Markov model, which is a Markov model but the Markov process, say A, is hidden and unobservable. Also, in the Hidden Markov model, it is required to have an observable process, say B, in which the consequences of B are influenced by the consequences of A. The model enables us to observe process A, which used to be unobservable, by looking through B. An example of a Hidden Markov model is a “weather guessing game”. Let’s say there are two persons, X and Y. Person Y talks about what he did to person X every day. Person Y will do different things according to different weather conditions: he will clean the house if it’s rainy outside, go to the park if it’s sunny, and clean the snow if it’s snowing. Person X has no idea what the weather is like outside (he cannot observe the outcome of the event directly), but he can communicate with person Y and realize the weather according to what person Y is doing since there is a certain chance that he will perform one of the three activities [2]. This entire system is an example of a Hidden Markov model.

There are many real-life problems and datasets, which can be solved or improved by the Hidden Markov model. For example, in the railway safety area, using casualty data related to railway accidents from 2010-2020 in China to predict the casualty number from the year 2021 to 2023, the Hidden Markov model is proved to have the advantages of improving the accuracy of the prediction model by 22% in 2019 and 50% in 2020 [3]. Moreover, the Hidden Markov model is proven to have

more efficiency in predicting ocean accidents. By the year 2022, the prediction efficiency and accuracy are proved to be 89%, which was 27.5% higher than previous predictions [4].

However, despite the increasing number of studies of the Markov model, there are fewer studies of the Markov model related to predicting the supply and demand side. Predicting future possibilities, which is a crucial part of the economy, is drawing more and more attention from both policymakers and central banks in recent years. Therefore, studying the Markov model in prediction to predict future possible situations is important. In this paper, the Hidden Markov model is studied by structuring, analyzing, and summarizing through various dimensions underlying different frameworks. The paper first studies the structure of the Hidden Markov model. Then, compare the results under the model's prediction to form results and analyze these results. Finally, the paper list limitations and possible ways to improve the model and forms a conclusion.

2. Methodology

Markov model, issued by Russian mathematician Andrey Markov, is a process that states that the probability of each individual event is dependent solely on the status of events and is irrelevant to past events. In addition to that, the Hidden Markov model, which is a Markov model with an unobservable item, is analyzed more widely. Let X and Y be the discrete process with $n \geq 1$. If $P(Y_n \in A | X = x_1, \dots, X_n = x_n) = P(Y_n \in A | X_n = x_n)$, for every $n \geq 1$, X and Y is a Hidden Markov model. When $P(Y_{t_0} \in A | \{X_t \in B_t\}) = P(Y_{t_0} \in A | X_{t_0} \in B_{t_0})$, X_t and Y_t is a Hidden Markov model.

The Hidden space is assumed to consist of a total number of X values, of which the distributional type is categorical. This means for each of the X possible values that a Hidden variable is within, there is a transition probability to each of the X possible states of the Hidden variable $t + 1$, for a total of X^2 transition probabilities. The sum of transition probabilities for transitions from any given value must sum to 1. Thus, the $X * X$ matrix of transition probabilities is a Markov matrix. The probability of an observed Markov sequence is given by $Y = y(0), y(1), \dots, y(L - 1)$, which L is the length of the sequence, $P(Y) = \sum_X P(Y|X)P(X)$, where the sum $X = x(0), x(1), \dots, x(L - 1)$. Despite these models, Tolerate Markov model is Markov model relies on probability. It assigns different conditional probabilities, considering the last symbol, from the start of the sequence. The Hidden Markov model has a wide range of applications in various fields. For example, there is the application of the Hidden Markov model in the accident analysis area, including assessment of traffic accidents and biological virus research fields.

Also, as the world develops, the number of cars is increasing more rapidly. Despite the convenience brought by the increase in the number of cars, the pollution problem and the traffic problems (especially in large international cities) are being more obvious. Therefore, realizing the issue and predicting the future trend of both car accidents and car numbers is essential. Thus, a study conducted in China suggests that a modified Markov model can predict the traffic situation in Shanghai, China. According to the study result, the multi-stages Markov model enhances the efficiency and response time for traffic jam predictions, which could be provided as a reference to study possible solutions. In an analysis of the pros and cons of the Markov model, the efficiency and the accuracy of the model prediction are often used. From various reports, the analysis of efficiency usually includes the time period, and the accuracy analysis usually includes a large number of samples from a certain time period within a specific range. Within 6 studies related to the Hidden Markov model or modified Markov model, all of them mention the advantages of increasing efficiency using various samples, and 4 of them mention the increase in accuracy as well [4].

3. Hidden Markov model related to railway fertilities prediction

There is a Hidden Markov model, which is a modified Markov model but one of the Markov models is hidden and unobservable [5]. In the Hidden Markov model, it is required to have an

observable process, in which the consequences it is influenced by the consequences of the Hidden process. The model enables us to observe the hidden process, which used to be unobservable, by seeing the observable one. An example of a Hidden Markov model is a “weather guessing game”. Let’s say there are two persons, X and Y. Person Y talks about what he did to person X every day. Person Y will do different things according to different weather conditions: he will clean the house if it’s rainy outside, go to the park if it’s sunny, and clean the snow if it’s snowing. Person X has no idea what the weather is like outside (he cannot observe the outcome of the event directly), but he can communicate with person Y and realize the weather according to what person Y is doing since there is a certain chance that he will perform one of the three activities [6].

The paper, studies and predicts railway fertilities around the 21st century. In order to reflect the safety status and future trends of railway operation, the Hidden Markov model is applied to predict dynamic and random accidents. The paper uses the data on railway accidents, which causes death, from the year 2010 to the year 2020, which shows the safety status of the railway [7]. From the paper, after analysis of the dataset, the accuracy of the predicted value in the year 2019 increased by 22% more than before, and the accuracy of the predicted value in 2020 increased by 50%. It is predicted that from 2021 to 2023, the rate of fatalities declines rapidly in the selected area of China, where the railway accident rate will remain at about 8%, and the railway operation will be in great condition [8]. From the Hidden Markov model analysis in the future, the safety operation can be improved by expanding daily measures and examination and increasing punishment. However, there is uncertainty as the author suggests in his paper, that because of the technological development, the time period in the original Hidden Markov model should be changed since the current situation influences the future condition and the current situation is changing rapidly [9]. The prediction of the model will be accurate even if the condition has changed, according to the author.

4. Markov model in Traffic Condition Prediction Model

There is the application of the Hidden Markov model in the accident analysis area, including assessment of traffic accidents and biological virus research fields [10]. As the world develops, the number of cars is increasing more rapidly. This is because cars can bring a lot of benefits and reduce the transitional cost of people when travel and from work to work [11]. Despite the convenience brought by the increase in the number of cars, the pollution problem and the traffic jam problems (especially in large international cities) are being more obvious. From the data, the average car density is increasing and exceeding the limit by 600% than in 10 years ago in cities like Beijing and Shanghai [12]. Therefore, realizing the issue and predicting the future trend of both car accidents and car numbers is essential.

Thus, a study conducted in China suggests that a modified Markov model can predict the traffic situation in Shanghai, China. According to the study result, the multi-stages Markov model enhances the efficiency and response time for traffic jam predictions, which could be provided as a reference to study possible solutions. To be more specific, with the old model, the traffic condition was predicted with a low accuracy of 32% and the efficiency of the model was low since it needs a lot of data collection from both past to present. The process of collecting raw data was both time-consuming and money-consuming. Using the Hidden Markov model, the cost of making such prediction was reduced significantly and the efficiency is increased since we no longer need the data from the past by the definition of the Hidden Markov model. Also, the Hidden Markov model ignores the elements, such as the cause of the accident and the randomness of the accidents since we can assess these data by just seeing the outcome. In an analysis of the pros and cons of the Markov model, the efficiency and the accuracy of the model prediction are often used.

From various reports, the analysis of efficiency usually includes the time period, and the accuracy analysis usually includes a large number of samples from a certain time period within a specific range. Within 6 studies related to the Hidden Markov model or modified Markov model, all of them mention

the advantages of increasing efficiency using various samples, and 4 of them mention the increase in accuracy as well.

5. Limitations and future outlook

Markov model, which was issued by Russian mathematician Andrey Markov, is a process that states that the probability of each individual event is dependent solely on the current status of events and is irrelevant to past events. In other words, this can be understood as what is going to happen in a sequence depending only on the current status. The Hidden Markov model has certain limitations as well. First, the Markov model, including the Hidden Markov model, will be more complicated when more factors and more targets are considered. This complexity is problematic because of various reasons. The complexity will cause the cost of the prediction and the model to increase rapidly. The cost of conducting a new model is a significant factor to consider because of the opportunity cost, which this amount of money could be used for other areas. Also, the complexity is problematic because some probabilities, which are used in the analysis of the model, are time-sensitive. Therefore, if the time-sensitive factors are influenced, the prediction of the model will have more error and uncertainty. Moreover, the size of the cohorts is difficult to measure and determine. Because the required sample size is essential to calculate and included in the Markov model, is difficult to measure, and is included the model also creates more errors [13].

Despite the drawbacks and uncertainties of the Markov model, the future of the Markov model in prediction is bright [14]. Since nowadays there are more factors to consider when doing predictions and many of the factors are unknown, the Markov model provides a way to analyze and predict the future in an easier but more accurate way. Moreover, the Markov model is actually developing itself, from the original Markov model to a Hidden Markov model, the model can be modified and fit into different scenarios [15]. This creates convenience and more possibilities in the future.

6. Conclusion

The aim of the paper is to study the Hidden Markov model by structuring, analyzing, and summarizing through various dimensions underlying different frameworks. The paper first studies the structure of the Hidden Markov model. Then, the process of setting up the Hidden Markov model and the applications of Hidden Markov model in various fields using the data from the current situation to the prediction of future results. In conclusion, the Markov model, including the hidden Markov model is effective and efficient in many fields, including railway fertilities prediction and traffic condition prediction model. To be more specific, the accuracy of the predicted value in the year 2019 increased by 22% more than before, and the accuracy of the predicted value in 2020 increased by 50%. Despite the study of Hidden Markov model, the paper also gets a conclusion after examining the results and getting some possible suggestions in order to improve the model. Although there are drawbacks, such as high uncertainty created by uncertain sample size and complexity, the Markov model has a bright future and can be used in various fields which promotes efficiency and accuracy in prediction. The meaning of this research paper is the result can help build up the fundamental ideas and structure of the Markov model for future use in reality. Also, the result can help policymakers as references and be used to predict future supply and demand side effects in various data and economic fields with a certain amount of accuracy, which could be beneficial by avoiding possible incoming economic failures.

References

- [1] Xin Fang, et al." Prediction model of railway accident death toll based on grey Markov chain " *Transportation technology and economics* 24.04(2022):37-42.
- [2] Huayou, Wang, et al." Prediction of civil aviation accident syndrome number based on "regression + Markov" combination model." *Systems Engineering and Electronics*.

- [3] Shipeng, Wang, et al." Prediction of Marine dangerous accidents in Shandong Province based on Grey Markov model." *Journal of Shandong Jiaotong University* 30.03(2022):40-45.
- [4] Jiahao, Wang, Haijun Zhang, and Hai Fang." Cost-benefit analysis of rotavirus vaccine inclusion in national immunization programmes based on decision tree-Markov model." *Vaccines and immunization in China* 28.03(2022):309-316.
- [5] Xin Wang, et al." Research on group relationship prediction of railway travel based on Hidden Markov model." *Computer Science* 49.S1(2022):247-255.
- [6] Mingsheng, Zhou, and Shuyang Liu." A traffic condition prediction model based on improved Markov chain." *Application of digital technology* 48.05(2022):27-30+36.
- [7] Jiamin, Wang, et al." Effects of diabetes screening on prevention of cardiovascular disease in community population based on Markov model." *Journal of Peking University (Medical Science)* 54.03(2022):450-457.
- [8] Junchao, Huang, and Yong Hu." Research on equipment failure prediction based on Grey Markov model." *Logistics technology* 45.04(2022):163-166.
- [9] Kun Wang, Xiujuan Sheng, and Hefei, Liu." Implicit inhomogeneous Markov multiple linear regression model and its Bayesian estimation." *Statistics and Decision making* 38.02(2022):16-20.
- [10] Rebello, Sinda, Hongyang Yu, and Lin Ma. "An integrated approach for system functional reliability assessment using Dynamic Bayesian Network and Hidden Markov Model." *Reliability Engineering & System Safety* 180 (2018): 124-135.
- [11] Cheng, Peng, et al. "Asynchronous fault detection filtering for piecewise homogenous Markov jump linear systems via a dual hidden Markov model." *Mechanical Systems and Signal Processing* 151 (2021): 107353.
- [12] Zhou, Liang, et al. "Multi-scenario simulation of urban land change in Shanghai by random forest and CA-Markov model." *Sustainable Cities and Society* 55 (2020): 102045.
- [13] Cheng, Jun, et al. "Hidden Markov model-based nonfragile state estimation of switched neural network with probabilistic quantized outputs." *IEEE Transactions on cybernetics* 50.5 (2019): 1900-1909.
- [14] Kouadri, Abdelmalek, et al. "Hidden Markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems." *Renewable Energy* 150 (2020): 598-606.
- [15] Sesia, Matteo, Chiara Sabatti, and Emmanuel J. Candès. "Gene hunting with hidden Markov model knockoffs." *Biometrika* 106.1 (2019): 1-18.