J. Indones. Math. Soc.
Vol. 30, No. 02 (2024), pp. 218–235.

HIDDEN MARKOV REPRESENTATION OF MICROCREDIT

Maria Angela A. Giva[1], Jasmin-Mae S. Luy[2*], and Mary Elizabeth R. Segui[3]

Institute of Mathematics, University of the Philippines Diliman
[1] **magiva@up.edu.ph**
[2*] **corresponding author: jsantos@math.upd.edu.ph**
[3] **maryelsegui@gmail.com**

Abstract. Microcredit is a method of lending small amounts of money to low-income individuals who have no access to traditional financial institutions. Upon applying for a loan, an individual may either be able to repay it and be granted a loan again, otherwise s/he demands for a new loan. These events influence certain factors, which can be illustrated through a hidden Markov model (HMM). This study provides a hidden Markov representation of microcredit taking into consideration the borrower's acquisition of small businesses. Model algorithms used in addressing the problems in HMM, such as the Viterbi algorithm, are discussed and implemented via numerical examples.

Key words and Phrases: Microcredit, hidden Markov model, Viterbi algorithm

1. INTRODUCTION

Microfinance is the provision of various financial services to poor or low income individuals who have no access to traditional financial institutions. Such financial service is microcredit, which involves granting of small loans to individuals. These loans are used to support income-generating activities, such as establishing or expanding borrowers' own small businesses, which in turn would enable them to raise their income and improve their standard of living. To cater to the specific needs and preferences of the borrowers, microcredit institutions offer options for both individual and group lending. This study focuses on representing microcredit in the case of an individual borrower.

There are various studies about microcredit and its representation. Diener et al. [3] proposed a simplified Markov chain representation of microcredit which describes the repayment process of an individual borrower. The expected total

2020 Mathematics Subject Classification: 60J20
Received: 16-09-2023, accepted: 23-08-2024.

discounted return of a borrower is computed and the absence of strategic default is analyzed. Diener and Khodr [4] studied the process for the case of two borrowers. Bernardino and Santos [2] extended these results for the case of three borrowers.

Meanwhile, a hidden Markov model (HMM) can be used in describing various processes. Segui [18], for example, utilized Markov and hidden Markov models in predicting browser behavior on a vehicle website. In finance, Kamath and Jahan [11] predicted possible loan defaulters in banks by performing sentiment analysis on Facebook posts and incorporating these into a hidden Markov model to estimate the probabilities of loan default. Li [12] applied HMM in predicting financial market behaviors. Giamperi et al. [6] modelled default in bond portfolios as a hidden Markov process to identify risk factors.

Although HMM serves as a valuable tool in various fields including forecasting and finance, only a few provide such representation in microcredit, using different observable processes. Ntwiga et al. [13] used HMM in identifying credit scores of M-Shwari microcredit clients in Kenya, which utilized socio-demographics, telecommunication characteristics, and account activities. Ntwiga and Weke [15] utilized borrower deposits and withdrawal dynamics of M-Shwari clients in HMM for credit scoring. Ntwiga et al. [14] utilized risk financial activities and credit scores in HMM training, which aids in identifying who among low earners qualify for a microloan. This study aims to contribute to existing literature by extending a Markov model of microcredit for individual lending by Diener et al. [3] to provide a simple theoretical hidden Markov representation. This study introduces acquisition of a small business as the observable component of the lending Markov process. Moreover, the fundamental problems in HMM as well as their corresponding algorithms are studied and illustrated via numerical examples.

2. Markov Chains

A Markov chain is a stochastic model which describes sequences of some random variables where the probability of a future event depends only on the current event.

Definition 2.1. *A discrete-time stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$, or simply $\{X_n\}$, is called a **Markov Chain** if*

$$\mathbb{P}(X_n = j \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = \mathbb{P}(X_n = j \mid X_{n-1} = x_{n-1}),$$

*for $x_0, x_1, \ldots, x_{n-1}, j \in \mathbb{N}_0$; with **transition probability matrix** T of order N, with the (i,j)-th entry given by*

$$t_{i,j} = \mathbb{P}(X_{n+1} = s_j \mid X_n = s_i),$$

where $0 \le t_{i,j} \le 1$ and $\sum_{j=1}^{N} t_{i,j} = 1$ for each row $i \in \{1, \ldots, N\}$.

Diener [3] presented a simple Markov chain representation of microcredit for the case of an individual borrower. In the model, a borrower can either be in a state of demanding for a loan, denoted by D, or in a state of being a beneficiary of a loan, denoted by B. The probability that a borrower in state D will be granted and be a beneficiary of a loan is given by ϕ, otherwise s/he will stay in state D with a probability $1-\phi$. If the borrower is able to repay the loan with probability δ, then s/he can stay in state B and be a beneficiary again. Otherwise, s/he will proceed to state D with probability $1-\delta$ and demand for a loan again. This stochastic sequence can be modelled by a Markov chain $\{X_n\}$, where each X_n is drawn from the state space

$$S \coloneqq \{D, B\}\,. \tag{1}$$

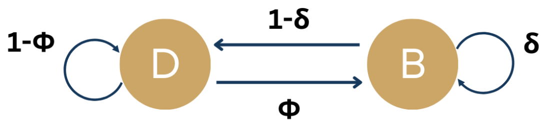

FIGURE 1. Markov Chain representation of microcredit [5]

The transition probability matrix of the Markov chain $\{X_n\}$ is given by

$$T = \begin{bmatrix} 1-\phi & \phi \\ 1-\delta & \delta \end{bmatrix}, \tag{2}$$

where

$$\begin{aligned}
\mathbb{P}(X_{n+1} = D \mid X_n = D) &= 1-\phi \\
\mathbb{P}(X_{n+1} = D \mid X_n = B) &= 1-\delta \\
\mathbb{P}(X_{n+1} = B \mid X_n = D) &= \phi \\
\mathbb{P}(X_{n+1} = B \mid X_n = B) &= \delta.
\end{aligned}$$

Remark 2.2. *The initial state distribution π_0 of the Markov chain $\{X_n\}$ is given by*

$$\pi_0 = \begin{bmatrix} \pi_0(D) & \pi_0(B) \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{3}$$

which follows from the idea that a borrower needs to demand for a loan first before being granted a loan.

To illustrate, suppose that the probability that a borrower moves from state D to state B is equal to 0.7. Then, from (2), the probability of staying in state D is equal to $1-0.7 = 0.3$. Meanwhile, we let the probability of staying in state B be equal to 0.4. Then the probability of transitioning from state B to state D is equal to $1-0.4 = 0.6$. Hence, the transition probability matrix for the case of our example is given by

$$T = \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix}. \tag{4}$$

Suppose a borrower is in state D in the first period. We wish to identify the probability that the borrower for the next 4 periods is in the state sequence $D - B - B - D$. Formally, the observation sequence, denoted Q, is given by $Q = \{D, D, B, B, D\}$, and we want to solve for $\mathbb{P}(Q)$. This can be expressed and evaluated as

$$\begin{aligned}\mathbb{P}(Q) &= \mathbb{P}(D) \cdot \mathbb{P}(D|D) \cdot \mathbb{P}(B|D) \cdot \mathbb{P}(B|B) \cdot \mathbb{P}(D|B) \\ &= \pi_0(D) \cdot t_{11} \cdot t_{12} \cdot t_{22} \cdot t_{21} \\ &= 1(0.3)(0.7)(0.4)(0.6) \\ &= 0.0504.\end{aligned}$$

Note that such probability can be solved since each state in the considered Markov process corresponds to an observable event. We extend this notion to the case when a Markov process involves an underlying hidden component but can be observed through another stochastic process.

3. Hidden Markov Models

A hidden Markov model (HMM), denoted $\{X_n, Y_n\}_{n\in\mathbb{N}_0}$ is a Markov process composed of two components X_n and Y_n. The first component $\{X_n\}$ is a Markov chain that is characterized by states which are "hidden" or unobservable. On the other hand, the second component $\{Y_n\}$ is a stochastic process, not necessarily a Markov chain, is characterized by states which are observable. A hidden Markov model assumes the existence of states which are not directly observable, and that these states influence the observable components. Hence, by examining a set of observations, we also get to have an insight about the underlying Markov process.

A hidden Markov model is formally defined by the following components [1], [10]:

- The state space $S = \{s_1, s_2, \ldots, s_N\}$, which is the set of N possible hidden states,
- The observation sequence $O = \{y_1, y_2, \ldots, y_k\}$, which is the set of k observations, where each element is taken from the set of q possible observable states or emissions, given by $P = \{p_1, p_2, \ldots, p_q\}$,
- The initial state distribution vector π_0 given by

$$\pi_0 = \begin{bmatrix}\pi_0(s_1) & \pi_0(s_2) & \cdots & \pi_0(s_N)\end{bmatrix},$$

  where each i-th entry represents the probability that the model starts at state s_i,
- The transition probability matrix $T = [t_{i,j}]$ of the hidden Markov component $\{X_n\}$, where each (i,j)-th entry represents the probability of transitioning from state s_i to state s_j, and
- The emission probability matrix $M = [m_{i,j}]$, where each (i,j)-th entry represents the probability of an observation p_j being generated when the hidden Markov process is in state s_i.

A hidden Markov model is characterized by two key assumptions [10], [18]:

- *Markov Assumption*: The probability of a particular hidden state depends only on the previous hidden state. That is,

$$\mathbb{P}(X_n = s \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = s \mid X_{n-1} = x_{n-1}),$$

  where $n \in \mathbb{N}_0$ and $s, x_1, \ldots, x_n \in S$.

- *Output Independence Assumption*: The probability of observing a particular output depends only on the hidden state that produced such observation. That is,

$$\mathbb{P}(Y_n = p \mid X_1 = x_1, \ldots, X_n = x_n, Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}) = \mathbb{P}(Y_n = p \mid X_n = x_n),$$

  where $n \in \mathbb{N}_0$, $x_1, \ldots, x_n \in S$ and $p, y_1, \ldots, y_{n-1} \in P$.

These assumptions lead to the following theorem.

Theorem 3.1. *Let $n, k \in \mathbb{N}_0$, and let $\{X_n, Y_n\}_{n \in \mathbb{N}_0}$ be a hidden Markov model, where $\{X_n\}$ is an unobservable Markov process with state space S, and $\{Y_n\}$ is an observable process of emissions, whose values are in P. If $k \leq n$, then X_{n+1} and Y_k are conditionally independent given X_n, that is,*

$$\mathbb{P}\left(X_{n+1} = s | X_n = x_n, Y_k = y_k\right) = \mathbb{P}\left(X_{n+1} = s | X_n = x_n\right),$$

where $s, x_n \in S$ and $y_k \in P$.

Proof. Let $n, k \in \mathbb{N}_0$ and let $k \leq n$. Consider $\mathbb{P}\left(X_{n+1} = s | X_n = x_n, Y_k = y_k\right)$.

Case 1: Suppose $k = n$. By the definition of conditional probability and by the observation independence assumption, we have

$$\begin{aligned}
\mathbb{P}\left(X_{n+1} = s | X_n = x_n, Y_n = y_n\right) &= \frac{\mathbb{P}\left(X_{n+1} = s, X_n = x_n, Y_n = y_n\right)}{\mathbb{P}\left\{X_n = x_n, Y_n = y_n\right\}} \\
&= \frac{\mathbb{P}\left(X_{n+1} = s, Y_n = y_n | X_n = x_n\right)\mathbb{P}\left(X_n = x_n\right)}{\mathbb{P}\left(Y_n = y_n | X_n = x_n\right)\mathbb{P}\left(X_n = x_n\right)} \\
&= \frac{\mathbb{P}\left(X_{n+1} = s, Y_n = y_n | X_n = x_n\right)}{\mathbb{P}\left(Y_n = y_n | X_n = x_n\right)} \\
&= \frac{\mathbb{P}\left(Y_n = y_n | X_n = x_n\right)\mathbb{P}\left(X_{n+1} = s | X_n = x_n\right)}{\mathbb{P}\left(Y_n = y_n | X_n = x_n\right)} \\
&= \mathbb{P}\left(X_{n+1} = s | X_n = x_n\right).
\end{aligned}$$

Case 2: Suppose $k < n$ and let $x_k \in S$. Because $\{X_n\}$ is a Markov chain, it follows that

$$\begin{aligned}
\mathbb{P}\left(X_{n+1}=s|X_n=x_n,Y_k=y_k\right) &= \mathbb{P}\left(X_{n+1}=s|X_n=x_n,X_k=x_k,Y_k=y_k\right)\\
&= \frac{\mathbb{P}\left(X_{n+1}=s,X_n=x_n,X_k=x_k,Y_k=y_k\right)}{\mathbb{P}\left(X_n=x_n,X_k=x_k,Y_k=y_k\right)}\\
&= \frac{\mathbb{P}\left(X_{n+1}=s,X_n=x_n,Y_k=y_k|X_k=x_k\right)}{\mathbb{P}\left(X_n=x_n,Y_k=y_k|X_k=x_k\right)}\\
&= \frac{\mathbb{P}\left(Y_k=y_k|X_k=x_k\right)\mathbb{P}\left(X_{n+1}=s,X_n=x_n|X_k=x_k\right)}{\mathbb{P}\left(Y_k=y_k|X_k=x_k\right)\mathbb{P}\left(X_n=x_n|X_k=x_k\right)}\\
&= \frac{\mathbb{P}\left(X_{n+1}=s,X_n=x_n,X_k=x_k\right)\mathbb{P}\left(X_k=x_k\right)}{\mathbb{P}\left(X_n=x_n,X_k=x_k\right)\mathbb{P}\left(X_k=x_k\right)}\\
&= \mathbb{P}\left(X_{n+1}=s|X_n=x_n\right).
\end{aligned}$$

□

The event that a borrower is in a state given in (1), of being in demand or beneficiary of a loan, may not be directly observed. It may be due to data privacy, borrowers' personal reasons, or client records unavailability. But these events definitely influence various factors in the lending process. One of the main goals of microcredit is to provide borrowers with capital for income-generating activities, such as starting or expanding their own business, which would eventually aid in loan repayment. Thus, the acquisition of a new small business is an event influenced by the state of a borrower. Such small business, which are directly observable, could be street food carts, mini *sari-sari* stores and bakeries, hair and nail salon, alterations shop, cleaning services, among others. If a borrower is in state D, the probability that s/he has not acquired a new small business is given by λ, while the probability that s/he has acquired a new small business is given by $1-\lambda$. On the other hand, if the borrower is in state B, the probability that s/he has not acquired a new small business is given by ε, while the probability that s/he has acquired a new small business is given by $1-\varepsilon$. Such process can be represented by a hidden Markov model as follows.

Proposition 3.2. *A hidden Markov representation of microcredit repayment is given by*

$$\{X_n, Y_n\}_{n\in\mathbb{N}_0}, X_n \in S \coloneqq \{D, B\}, Y_n \in P \coloneqq \{0, 1\},$$

where D denotes the state of a borrower being in demand of a loan, B denotes the state of being a beneficiary of a loan, 0 denotes the event that a borrower has no new small business, and 1 denotes the event that a borrower has acquired a new small business, such that

$$\begin{aligned}
&\mathbb{P}(Y_n = 0 \mid X_n = D) = \lambda\\
&\mathbb{P}(Y_n = 1 \mid X_n = D) = 1-\lambda\\
&\mathbb{P}(Y_n = 0 \mid X_n = B) = \varepsilon\\
&\mathbb{P}(Y_n = 1 \mid X_n = B) = 1-\varepsilon.
\end{aligned}$$

The transition probability matrix T of the hidden Markov model is given by (2) while the emission probability matrix M is given by

$$M = \begin{bmatrix} \lambda & 1-\lambda \\ \varepsilon & 1-\varepsilon \end{bmatrix}. \tag{5}$$

Remark 3.3. *The initial state distribution* π_0 *of the hidden Markov representation* $\{X_n, Y_n\}_{n\in\mathbb{N}_0}$ *is the initial state distribution of the unobservable Markov component* $\{X_n\}_{n\in\mathbb{N}_0}$ *given in* (3)*.*

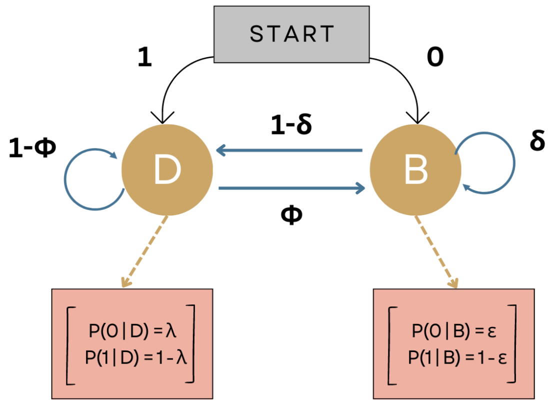

FIGURE 2. Hidden Markov chain representation of microcredit

4. HMM PROBLEMS AND ALGORITHMS

A hidden Markov model is characterized by three fundamental problems:

- *Likelihood*: Given a hidden Markov model with transition and emission probability matrices T and M, respectively, determine the likelihood of an observation sequence O.
- *Learning*: Given an observation sequence O and a hidden state space S, learn the hidden Markov model's transition and emission probability matrices T and M, respectively.
- *Decoding*: Given a hidden Markov model with transition and emission probability matrices T and M, respectively, and an observation sequence O, determine the most probable hidden state sequence Q.

Each of these problems can be solved using known algorithms. We discuss these algorithms following [10], with complete details found in [8] and [18].

4.1. **Forward Algorithm.**

The likelihood problem aims to to evaluate the probability of a given emission sequence O being observed from a hidden Markov model $\{X_n, Y_n\}$. Given an observation sequence $O = \{y_1, y_2, \ldots, y_k\}$, there are various possible hidden state sequences Q that must have produced such observation. Hence, to compute for $\mathbb{P}(O)$, the sum of the joint probability of O and Q over all possible hidden state sequences Q is computed. That is, by the definition of conditional probability and by the two key assumptions of HMM,

$$\begin{aligned} \mathbb{P}(O) &= \sum_Q \mathbb{P}(O, Q) \\ &= \sum_Q \mathbb{P}(O|Q)\,\mathbb{P}(Q) \\ &= \sum_Q \left[\left(\prod_{i=1}^{k} \mathbb{P}(Y_i = y_i | X_i = x_i)\right)\left(\prod_{i=1}^{k} \mathbb{P}(X_i = x_i | X_{i-1} = x_{i-1})\right)\right]. \end{aligned} \tag{6}$$

For a hidden Markov model $\{X_n, Y_n\}$ with a state space of N elements and an observation sequence of length k, there are N^k possible hidden state sequences Q. Thus, when dealing with HMM with a large state space and a long observation sequence, the number of possible hidden state sequences would also be very large, and so computing for the likelihood of O using (6) would not be practical. This is addressed by the forward algorithm. To implement this, we start with the following definition taken from [18].

Definition 4.1. *Let $\{X_n, Y_n\}$ be a hidden Markov model with state space $S = \{s_1, \ldots, s_N\}$ and an observation sequence $O = \{y_1, \ldots, y_k\}$. The forward path probability of the i-th state at time j, denoted $f_{i,j}$, is the probability of being in state s_i after seeing the first j observations. That is,*

$$f_{i,j} = \mathbb{P}(Y_1 = y_1, Y_2 = y_2, \ldots, Y_j = y_j, X_j = s_i). \tag{7}$$

For the algorithm, we construct a forward probability matrix $F = [f_{i,j}]$, where each (i, j)-th entry represents the forward probability of the i-th state at time j. Let $z_j \in \{1 \ldots, q\}$ such that $y_j = p_{z_j}$ for all $y_j \in O$ and $i \in \{1 \ldots, N\}$. By definition of conditional probability, each entry in the first column of F can be written

$$\begin{aligned} f_{i,1} &= \mathbb{P}(Y_1 = y_1, X_1 = s_i) \\ &= \mathbb{P}(Y_1 = p_{z_1} \mid X_1 = s_i)\mathbb{P}(X_1 = s_i) \\ &= m_{i,z_1}\pi_0(s_i). \end{aligned} \tag{8}$$

To obtain the succeeding columns of F, we let $i \in \{1\ldots,N\}$ and $j \in \{2,\ldots,k\}$. By definition of conditional probability and the law of total probability,

$$\begin{aligned}
f_{i,j} &= \mathbb{P}(Y_1 = y_1,\ldots,Y_j = y_j, X_j = s_i)\\
&= \mathbb{P}(Y_j = y_j|X_j = s_i)\sum_{l=1}^{N}\mathbb{P}(Y_1 = y_1,\ldots,Y_{j-1} = y_{j-1}, X_j = s_i, X_{j-1} = s_l)\\
&= \mathbb{P}(Y_j = y_j|X_j = s_i)\sum_{l=1}^{N}\mathbb{P}(Y_1 = y_1,\ldots,Y_{j-1} = y_{j-1}, X_j = s_i|X_{j-1} = s_l)\,\mathbb{P}(X_{j-1} = s_l)\\
&= \mathbb{P}(Y_j = p_{z_j}|X_j = s_i)\sum_{l=1}^{N}\mathbb{P}(X_j = s_i|X_{j-1} = s_l)\,\mathbb{P}(Y_1 = y_1,\ldots,Y_{j-1} = y_{j-1}, X_{j-1} = s_l)\\
&= m_{i,z_j}\sum_{l=1}^{N} t_{l,i}\, f_{l,j-1}.
\end{aligned}$$

Finally, note that

$$\begin{aligned}
\mathbb{P}(O) &= \mathbb{P}(Y_1 = y_1,\ldots,Y_k = y_k)\\
&= \sum_{i=1}^{N}\mathbb{P}(Y_1 = y_1,\ldots,Y_k = y_k, X_k = s_i)\\
&= \sum_{i=1}^{N} f_{i,k}.
\end{aligned}$$

Thus, $\mathbb{P}(O)$ is obtained by taking the sum of the entries of the last column of F. This solves the likelihood problem.

4.2. **Forward-Backward Algorithm.**

The learning problem aims to estimate the model's transition and emission probability matrices that would best explain a given observation sequence [7]. That is, for a hidden Markov model $\{X_n, Y_n\}$ with an observation sequence O, we need to find the model's transition and emission probability matrices T and M, respectively, that would maximize the likelihood of O. However, there is no known algorithm yet that can estimate matrices that maximize $\mathbb{P}(O)$. Instead, the forward-backward algorithm is implemented such that the resulting $\mathbb{P}(O)$ is locally maximized [16]. As the name suggests, the algorithm utilizes both forward and backward probabilities. The backward probability is defined as follows.

Definition 4.2. *Let $\{X_n, Y_n\}$ be a hidden Markov model with state space $S = \{s_1,\ldots,s_N\}$ and $O = \{y_1,\ldots,y_k\}$ be a sequence of observations. The backward probability of the i-th state at time j, denoted $b_{i,j}$, is the probability of observing the emissions from time $j+1$ to the end, given that the model is in state s_i at time j. That is,*

$$b_{i,j} = \mathbb{P}(Y_{j+1} = y_{j+1}, Y_{j+2} = y_{j+2},\ldots,Y_k = y_k \mid X_j = s_i). \tag{9}$$

Let $i \in \{1, \ldots, N\}$ and at $j = k$, let $b_{i,k} = 1$ [10]. For $j \in \{k-1, \ldots, 1\}$, the backward probability $b_{i,j}$ can be expressed as follows.

$$
\begin{aligned}
b_{i,j} &= \frac{\mathbb{P}\left(Y_{j+1} = y_{j+1}, Y_{j+2} = y_{j+2}, \ldots, Y_k = y_k, X_j = s_i\right)}{\mathbb{P}\left(X_j = s_i\right)} \\
&= \sum_{l=1}^{N} \mathbb{P}\left(Y_{j+1} = y_{j+1}, Y_{j+2} = y_{j+2}, \ldots, Y_k = y_k | X_j = s_i, X_{j+1} = s_l\right) \mathbb{P}\left(X_{j+1} = s_l | X_j = s_i\right) \\
&= \sum_{l=1}^{N} \mathbb{P}\left(Y_{j+1} = y_{j+1} | X_{j+1} = s_l\right) \mathbb{P}\left(Y_{j+2} = y_{j+2}, \ldots, Y_k = y_k | X_{j+1} = s_l\right) \mathbb{P}\left(X_{j+1} = s_l | X_j = s_i\right) \\
&= \sum_{l=1}^{N} \mathbb{P}\left(X_{j+1} = s_l | X_j = s_i\right) \mathbb{P}\left(Y_{j+1} = y_{j+1} | X_{j+1} = s_l\right) b_{l,j+1} \\
&= \sum_{l=1}^{N} t_{i,l} m_{l,z_{j+1}} b_{l,j+1}.
\end{aligned}
$$

Let $\{X_n, Y_n\}$ be a hidden Markov model and let O be a sequence of observations. Moreover, let $f_{i,j}$ and $b_{i,j}$ be the forward and backward probabilities as given in (7) and (9), respectively. The forward-backward algorithm starts with initial guesses for the model's transition and emission probability matrices, denoted $\hat{T} = [\hat{t}_{i,j}]$ and $\hat{M} = [\hat{m}_{i,j}]$. Thorough computation [18] yields new estimates for the transition and emission probability matrices, that is, for $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, k\}$,

$$
\bar{T} = [\bar{t}_{i,j}] = \left[\frac{\displaystyle\sum_{t=1}^{k-1} f_{i,t} \hat{t}_{i,j} \hat{m}_{j,z_{t+1}} b_{j,t+1}}{\displaystyle\sum_{t=1}^{k-1} \sum_{r=1}^{N} f_{i,t} \hat{t}_{i,r} \hat{m}_{r,z_{t+1}} b_{r,t+1}} \right] \quad \text{and} \quad \bar{M} = \left[\frac{\displaystyle\sum_{t=1 | y_t = p_j}^{k} f_{i,t} b_{i,t}}{\displaystyle\sum_{t=1}^{k} f_{i,t} b_{i,t}} \right].
$$

The algorithm proceeds by iteratively obtaining new estimates for the transition and emission probability matrices wherein each iteration increases the likelihood of the observation state sequence O. The algorithm only stops when the resulting $\mathbb{P}(O)$, obtained via the forward algorithm, converges to a local maximum or when a certain number of iterations is reached [16]. Upon convergence, we take the newly obtained matrices as the model's transition and emission probability matrices.

4.3. **Viterbi Algorithm.**

The decoding problem aims to obtain the most probable hidden state sequence Q given an observation sequence O. That is, if $O = \{y_1, y_2, \ldots, y_k\}$ is a sequence of emissions observed from a hidden Markov process $\{X_n, Y_n\}$, we need to find the particular hidden state sequence $Q = \{q_1, q_2, \ldots, q_k\}$ such that the probability $\mathbb{P}(Q|O)$ is maximum. Since $\mathbb{P}(O)$ does not directly depend on the hidden state sequence Q, we can just find the particular sequence of hidden states that maximizes $\mathbb{P}(Q, O)$ [17]. That is,

$$\arg\max_Q \mathbb{P}(Q, O) = \arg\max_Q \left[\prod_{i=1}^{k} \mathbb{P}(Y_i = y_i \mid X_i = x_i) \prod_{i=1}^{k} \mathbb{P}(X_i = x_i \mid X_{i-1} = x_{i-1})\right].$$

Hence, an approach to the problem would be to compute $\mathbb{P}(Q, O)$ for all possible hidden state sequences Q and choose the one which produces the maximum probability. However, similar with the likelihood problem, this method is inefficient since we might deal with a large number of possible hidden state sequences. To address this, a decoding algorithm known as the Viterbi algorithm is utilized.

Definition 4.3. *Let $\{X_n, Y_n\}$ be a hidden Markov model with state space $S = \{s_1, s_2, \ldots, s_N\}$, $O = \{y_1, y_2, \ldots, y_k\}$ be a sequence of emissions, and $Q = \{x_1, x_2, \ldots, x_k\}$ be a sequence of hidden states. The Viterbi path probability of the i-th state at time j is defined*

$$\max_{x_1, x_2, \ldots, x_{j-1}} \mathbb{P}(X_1 = x_1, \ldots, X_{j-1} = x_{j-1}, Y_1 = y_1, Y_2 = y_2, \ldots, Y_j = y_j, X_j = s_i).$$

Similar to the forward algorithm, the Viterbi algorithm starts by constructing an $N \times k$ matrix $V = [v_{i,j}]$, whose (i, j)-th entry is the Viterbi path probability of the i-th state at time j. Note that this is the probability that the hidden Markov process is in state s_i after passing through the most probable state sequence $x_1, \ldots, x_{j-1}$, after seeing the first j observations [10]. Let $i \in \{1, \ldots, N\}$ and let $z_j \in \{1, \ldots, q\}$ such that $y_j = p_{z_j}$ for all $y_j \in O$. Since there is no preceding state at time $j = 1$, then from (8), $v_{i,1}$ is given by

$$v_{i,1} = \mathbb{P}(Y_1 = y_1, X_1 = s_i) = m_{i,z_1} \pi_0(s_i).$$

For the succeeding columns of V, we let $i \in \{1, \ldots, N\}$, $j \in \{2, \ldots, k\}$, and $g_j \in \{1, \ldots, N\}$ such that $x_j = s_{g_j}$ for all $x_j \in Q$. Then, by definition of conditional probability and Theorem 3.1, it can be shown that

$$v_{i,j} = \max_{l=1}^{N} m_{i,z_j} t_{l,i} v_{l,j-1}.$$

Now, note that each entry on the last column of V is given by

$$v_{i,k} = \max_{x_1, x_2, \ldots, x_{k-1}} \mathbb{P}(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, Y_1 = y_1, \ldots, Y_k = y_k, X_k = s_i).$$

Taking the maximum of the entries on the last column of V, we have

$$\begin{aligned} \max_{i=1}^{N} v_{i,k} &= \max_{i=1}^{N} \left[\max_{x_1, x_2, \ldots, x_{k-1}} \mathbb{P}(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, Y_1 = y_1, \ldots, Y_k = y_k, X_k = s_i)\right] \\ &= \max_{x_1, x_2, \ldots, x_k} \mathbb{P}(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, Y_1 = y_1, \ldots, Y_k = y_k, X_k = x_k) \\ &= \max_Q \mathbb{P}(Q, O). \end{aligned} \tag{10}$$

Now, to retrieve the most probable state sequence Q, we first need to keep track of the state that maximizes $v_{i,j}$ [16]. To do this, we construct an $N \times k$ matrix $W = [w_{i,j}]$, whose (i,j)-th entry is the index of the state at time $j-1$ that maximizes $v_{i,j}$, given by

$$w_{i,j} = \arg\max_{g_{j-1}} \Big[\mathbb{P}(Y_j = y_j | X_j = s_i)\, \mathbb{P}(X_j = s_i | X_{j-1} = s_{g_{j-1}})\, v_{g_{j-1},j-1} \Big].$$

Let $i \in \{1, \dots, N\}$. For the first column of W, we let $w_{i,1} = 0$ since there is no preceding state at time $j = 1$. For the succeeding columns, given $i \in \{1 \dots, N\}$ and $j \in \{2 \dots, k\}$, $w_{i,j}$ it can be shown that

$$w_{i,j} = \arg\max_{l=1}^{N} m_{i,z_j} t_{l,i} v_{l,j-1}.$$

To proceed with the algorithm, let $Q^* = \{x_1^*, x_2^*, \dots, x_k^*\}$ be the most probable hidden state sequence. From [16], the final state in the state sequence Q^* is given by the equation $x_k^* = s_{g_k^*}$ where

$$g_k^* = \arg\max_{i=1}^{N} v_{i,k}. \tag{11}$$

The prior states in Q^* can be obtained using the equation $x_{t-1}^* = s_{g_{t-1}^*}$ where $g_{t-1}^* = w_{g_t^*,t}$. That is, for $t \in \{k, \dots, 2\}$,

$$g_{t-1}^* = \arg\max_{g_{t-1}} \Big[\mathbb{P}(Y_t = y_t | X_t = s_{g_t^*})\, \mathbb{P}(X_t = s_{g_t^*} | X_{t-1} = s_{g_{t-1}})\, v_{g_{t-1},t-1} \Big]. \tag{12}$$

From (10) and (11), we have that

$$\max_{Q} \mathbb{P}(Q, O) = \max_{i=1}^{N} v_{i,k} = v_{g_k^*,k}. \tag{13}$$

Moreover, from (12), we have

$$\begin{aligned} v_{g_t^*,t} &= \max_{g_{t-1}} \Big[\mathbb{P}(Y_t = y_t | X_t = s_{g_t^*})\, \mathbb{P}(X_t = s_{g_t^*} | X_{t-1} = s_{g_{t-1}})\, v_{g_{t-1},t-1} \Big] \\ &= \mathbb{P}(Y_t = y_t | X_t = s_{g_t^*})\, \mathbb{P}(X_t = s_{g_t^*} | X_{t-1} = s_{g_{t-1}^*})\, v_{g_{t-1}^*,t-1}, \end{aligned} \tag{14}$$

where $t \in \{k, \dots, 2\}$. Utilizing (13) and (14), we have

$$\begin{aligned}
\max_Q \mathbb{P}(Q, O) &= v_{g_k^*, k} \\
&= v_{g_{k-1}^*, k-1}\, \mathbb{P}(Y_k = y_k | X_k = s_{g_k^*})\, \mathbb{P}(X_k = s_{g_k^*} | X_{k-1} = s_{g_{k-1}^*}) \\
&= v_{g_{k-2}^*, k-2}\, \mathbb{P}(Y_{k-1} = y_k - 1 | X_{k-1} = s_{g_k^*})\, \mathbb{P}(X_{k-1} = s_{g_{k-1}^*} | X_{k-2} = s_{g_{k-2}^*}) \times \\
&\qquad\qquad \mathbb{P}(Y_k = y_k | X_k = s_{g_k^*})\, \mathbb{P}(X_k = s_{g_k^*} | X_{k-1} = s_{g_{k-1}^*}) \\
&\;\;\vdots \\
&= v_{g_1^*, 1} \prod_{l=2}^{k} \mathbb{P}(Y_l = y_l | X_l = s_{g_l^*})\, \mathbb{P}(X_l = s_{g_l^*} | X_{l-1} = s_{g_{l-1}^*}) \\
&= \mathbb{P}(Y_1 = y_1 \mid X_1 = s_{g_1^*}) \mathbb{P}(X_1 = s_{g_1^*}) \prod_{l=2}^{k} \mathbb{P}(Y_l = y_l | X_l = s_{g_l^*})\, \mathbb{P}(X_l = s_{g_l^*} | X_{l-1} = s_{g_{l-1}^*}) \\
&= P(Q^*, O).
\end{aligned}$$

This shows that the joint probability of Q and O is at maximum when $Q = Q^*$. This implies that Q^* is the most probable state sequence given an observation sequence O. This solves the decoding problem.

5. Numerical Illustrations

Let $\{X_n, Y_n\}$ be a hidden Markov representation of microcredit given in Proposition 3.2. The hidden state space is given by $S = \{s_1, s_2\} = \{D, B\}$ and the set of possible emissions is given by $P = \{p_1, p_2\} = \{0, 1\}$. From Remark 3.3, the initial state distribution of $\{X_n, Y_n\}$ is given by $\pi_0 = \begin{bmatrix}\pi_0(D) & \pi_0(B)\end{bmatrix} = \begin{bmatrix}1 & 0\end{bmatrix}$. To illustrate, let the transition probability matrix be the same as given in (15), that is,

$$T = \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix}. \tag{15}$$

Moreover, if the borrower is in state D, we let the probability of not having any new small business be equal to 0.5. Then, from (5), the probability that s/he acquires a new small business given that s/he is in state D is 0.5. Meanwhile, if the borrower is in state B, we let the probability of not having any new small business be equal to 0.4. Then, from (5), the probability that s/he acquires a new small business given that s/he is in state B is 0.6. Thus, the emission probability matrix of our example is given by

$$M = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{bmatrix}. \tag{16}$$

Figure 3 illustrates the hidden Markov model $\{X_n, Y_n\}$.

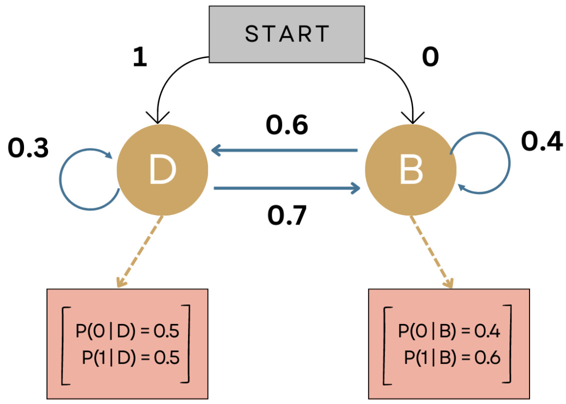

FIGURE 3. An example of a hidden Markov representation of microcredit

Now suppose we have an observation sequence $O = \{0, 0, 1, 1, 0\}$. That is, at times $1, 2$, and 5, the borrower has not acquired any new small business, and at times 3 and 4, the borrower has acquired a new small business. To solve the fundamental problems, we implement their corresponding model algorithms using MATLAB. [8]

For the likelihood problem, we take as input the observation sequence O, the model's transition and emission probability matrices, and the initial state distribution. Implementing the forward algorithm yields

$$F = \begin{bmatrix} 0.5 & 0.075 & 0.0427 & 0.0263 & 0.0158 \\ 0 & 0.105 & 0.0662 & 0.0395 & 0.0103 \end{bmatrix}.$$

Note that each column of matrix F represents the partial probability of O at each hidden state for each time step. Now, to determine the likelihood of O we just add the entries on the last column of F. Hence, the likelihood of the sequence $O = \{0, 0, 1, 1, 0\}$ is

$$\mathbb{P}(O) = 0.0158 + 0.0103 \approx 0.026.$$

For the learning problem, we take the matrices given in (15) and (16) as our initial estimates for the transition and emission probability matrices, respectively. The forward-backward algorithm will then continue to re-estimate these matrices until a certain number of iterations is reached or when the difference between the consecutive resulting $\mathbb{P}(O)$ fall below a specified tolerance. The maximum number of iterations is set at $10,000$ and tolerance at 10^{-20}. After 217 iterations, the newly obtained estimates are given by

$$T = \begin{bmatrix} 0.2764 & 0.7236 \\ \approx 0 & \approx 1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} \approx 1 & \approx 0 \\ 0.4472 & 0.5528 \end{bmatrix}. \tag{17}$$

From T, we can infer that if a borrower is in state B, it is highly likely that s/he will stay in state B and that it is almost unlikely that s/he will transition back to state D. Similarly, it can be implied from matrix M that if a borrower is in state D, it is almost unlikely that s/he acquires a new small business. However, we can see that these does not agree with our previously obtained results. Frazzoli [7] noted that to produce better representations for the matrices T and M, we need to provide a larger observation data. Hence, such conflict probably emerged since we only have few observations for our example.

Finally, for the decoding problem, we take as input the observation sequence O, the model's transition and emission probability matrices, and the initial state distribution. By implementing the Viterbi algorithm, the corresponding most likely hidden state sequence is given by

$$Q = D - B - D - B - D,$$

wherein the joint probability of O and Q is found to be approximately equal to 0.0046. From the obtained sequence Q, we can infer that the borrower is most likely in the state of demanding for a loan when s/he acquired his/her first small business; and when the borrower has acquired another small business, s/he is most likely in the state of being a beneficiary of a loan.

Now suppose there is an equal probability of being in each of the states. That is, if the borrower is in state D, we let the probability of staying in state D and the probability of transitioning to state B be both equal to 0.5. Similarly, if the borrower is in state B, let the probability of transitioning to either state B or D be equal to 0.5. Moreover, if the borrower is in either state D or B, we let the probability of having or not having any new small business be equal to 0.5. Then

$$T = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}. \tag{18}$$

Given the same observation sequence $O = \{0, 0, 1, 1, 0\}$ and the transition and emission probability matrices in (18), we get

$$F = \begin{bmatrix} 0.5 & 0.125 & 0.0625 & 0.0312 & 0.0156 \\ 0 & 0.125 & 0.0625 & 0.0312 & 0.0156 \end{bmatrix},$$

which yields $\mathbb{P}(O) = 0.0156 + 0.0156 \approx 0.0312$.

Given the initial estimates for T and M in (18), 231 iterations yield the best estimates given by

$$T = \begin{bmatrix} 0.2764 & 0.7236 \\ \approx 0 & \approx 1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} \approx 1 & \approx 0 \\ 0.4472 & 0.5528 \end{bmatrix}, \tag{19}$$

with entry values same as that of (17).

Finally, the resulting most likely hidden state sequence is given by

$$Q = D - D - D - D - D,$$

wherein the joint probability of O and Q is found to be approximately equal to 0.002. The obtained sequence Q implies that the borrower is most likely in the state of being in demand of a loan at all time periods. That is, the borrower didn't transition to being a beneficiary of a loan even after acquiring of a new small business.

On the other hand, suppose that there is a significantly higher probability of being in one state than the other. That is, if the borrower is in state D, we let the probability of staying in state D be equal to 0.1 and the probability of transitioning to state B be 0.9. Similarly, if the borrower is in state B, we let the probability of staying in state B be equal to 0.9 and the probability of transitioning back to state D be equal to 0.1. Moreover, if the borrower is in state D, we let the probability of not having any new small business be equal to 0.9 and the probability of having a new small business be equal to 0.1. On the other hand, if the borrower is in state B, we let the probability of not having any new small business be equal to 0.1 and the probability having a new small business be equal to 0.9. Thus,

$$T = \begin{bmatrix} 0.1 & 0.9 \\ 0.1 & 0.9 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}. \tag{20}$$

Given the same observation sequence $O = \{0, 0, 1, 1, 0\}$ and matrices T and M in (20), we get

$$F = \begin{bmatrix} 0.9 & 0.081 & 0.0016 & 0.0013 & 0.0098 \\ 0 & 0.0810 & 0.1312 & 0.1076 & 0.0098 \end{bmatrix},$$

which gives $\mathbb{P}(O) = 0.0098 + 0.0098 \approx 0.0196$.

Given the initial estimates for T and M in (20), 196 iterations yield the best estimates given by

$$T = \begin{bmatrix} 0.2764 & 0.7236 \\ \approx 0 & \approx 1 \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} \approx 1 & \approx 0 \\ 0.4472 & 0.5528 \end{bmatrix}, \tag{21}$$

which again have the same entries as that of the matrices given in (17) and (19). This implies that even though we used different initial estimates of T and M, the resulting final estimates would still be the same given that we have the same observation sequence.

Finally, the resulting most likely hidden state sequence is given by

$$Q = D - D - B - B - D,$$

wherein the joint probability of O and Q is found to be approximately equal to 0.005. The obtained sequence Q implies that the borrower is most likely in the state of demanding for a loan during the time that s/he has not acquired any small business and in the state of being a beneficiary of a loan when s/he acquired a new small business. We can see that these implications indeed reflect the transition probability assumptions that we set in the example.

6. Conclusion and Recommendations

This study developed, studied, and illustrated a hidden Markov chain representation of microcredit where the hidden component considered is loan benefit, while the observable component is the borrower's acquisition of a small business.

The fundamental problems that are of interest when dealing with HMM are thoroughly discussed and we have shown that these problems can be efficiently addressed using known algorithms. Then, by giving numerical examples of our hidden Markov representation, we were able to illustrate the implementation of these algorithms which can provide valuable insights regarding the microcredit process. In particular, by implementing the Viterbi algorithm for the decoding problem, we have shown that one can infer whether a borrower have been in demand or a beneficiary of a loan during the time of acquisition of a small business. This could help assess the effectiveness of microcredit in fostering entrepreneurial activities, which is one of its main objectives.

For future research, results may be improved by considering a larger set of data. One can also explore other factors in microlending, such as credit exclusion for a more detailed representation of the underlying microcredit process. Finally, one can consider other possible observable sequences that are influenced by the states of the hiddden Markov process.

REFERENCES

[1] Awad, M. and Khanna, R., "Efficient learning machines: theories, concepts, and applications for engineers and system designers", *Springer nature*, (2015).

[2] Bernardino, T. T. and Santos, J.M., "Markov chain representation of individual and group lending in microcredit", *AIP Conference Proceedings* 2192(1) (2019), 060004-1 - 060004-13. https://doi.org/10.1063/1.5139150.

[3] Diener, F., Diener, M., Khodr, O., and Protter, P., "Mathematical models for microlending", *Proceedings of the 16th Mathematical Conference of the Bangladesh Mathematical Society* Volume 1 (2009).

[4] Diener, F. and Khodr, O., "Mathematical models for individual and group lending in microfinance", Université de Nice, Laboratoire de Mathématiques (2008), https://math.univ-cotedazur.fr/ diener/Mifi/papierKodrDiener.pdf.

[5] Dulay, S., "Markov chain models of microcredit and the application of the perron-frobenius theorem", *BS Thesis*, Institute of Mathematics, University of the Philippines Diliman (2018).

[6] Giampieri, G., Davis, M., and Crowder, M. "Analysis of default data using hidden Markov models", *Quantitative Finance*, 5(1) (2005), 27–34. https://doi.org/10.1080/14697680500039951

[7] Frazzoli, E., "Principles of Autonomy and Decision Making", *University Lecture*, Massachusetts Institute of Technology (2010).

[8] Giva, M.A., "Hidden markov representation of microcredit", *BS Thesis*, Institute of Mathematics, University of the Philippines, Diliman (2023).

[9] Hassan, R., and Nath, B., "Stock market forecasting using hidden Markov model: A new approach", *Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications* (2003).

[10] Jurafsky, D. and Martin, J., "Speech and language processing", https://web.stanford.edu/ jurafsky/slp3/A.pdf, Accessed: 03-05-2023.

[11] Kamath, H. and Jahan, N. F., "Using hidden markov model to monitor possible loan defaults in banks", *International Journal of Economics & Business Administration* Vol. 0(4) (2020), 1097-1107.

[12] Li, N., "Hidden markov model and financial application" *Master's thesis*, The University of Texas at Austin (2016).

[13] Ntwiga, D.B., Ogutu, C., and Kirumbu, M.K., "Inclusion of peer group and individual low-income earners in m-shwari micro-credit lending: a hidden markov model approach", *International Journal of Electronic Finance* 9(2) (2018), 121-133.

[14] Ntwiga, D.B., Ogutu, C., Kirumbu, M.K., and Weke, P., "A Hidden Markov Model of Risk Classification among the Low Income Earners", *Journal of Finance and Economics*, 6(6) (2018), 242-249. https://doi.org/10.12691/jfe-6-6-6

[15] Ntwiga, D.B. and Weke, P., "Credit Scoring for M-Shwari using Hidden Markov Model", *European Scientific Journal, ESJ,* 12(15) (2016), 176. https://doi.org/10.19044/esj.2016.v12n15p176

[16] Rabiner, L., "A tutorial on hidden markov models and selected applications in speech recognition" *Proceedings of the IEEE* 77(2) (1989), 257–286.

[17] Ramage, D., "Hidden markov models fundamentals", *CS229 Section Notes* (2007), https://cs229.stanford.edu/section/cs229-hmm.pdf.

[18] Segui, M. E., "Predicting a browser's next online behavior using markov and hidden markov models with preference clustering", *Master's Thesis*, Institute of Mathematics, University of the Philippines, Diliman (2023).