International Journal of Information Technology & Decision Making
Vol. 12, No. 4 (2013) 813–835

DOI: 10.1142/S0219622013500302

FUZZY HIDDEN MARKOV CHAIN FOR WEB APPLICATIONS

R. SUJATHA, T. M. RAJALAXMI* and B. PRABA

Department of Mathematics
SSN College of Engineering
Old Mahabalipuram Road, Chennai, Tamil Nadu, India
**laxmi.raji18@gmail.com*

Hidden Markov model (HMM) has become increasingly popular in the last several years. Real-world problems such as prediction of web navigation are uncertain in nature; in this case, HMM is less appropriate i.e., we cannot assign certain probability values while in fuzzy set theory everything has elasticity. In addition to that, a theory of possibility on fuzzy sets has been developed to handle uncertainity. Thus, we propose a fuzzy hidden Markov chain (FHMC) on possibility space and solve three basic problems of classical HMM in our proposed model to overcome the ambiguous situation. Client's browsing behavior is an interesting aspect in web access. Analysis of this issue can be of great benefit in discovering user's behavior in this way we have applied our proposed model to our institution's website (www.ssn.edu.in) to identify how well a given model matches a given observation sequence, next to find the corresponding state sequence which is the best to explain the given observation sequence and then to attempt to optimize the model parameters so as to describe best how a given observation sequence comes about. The solution of these problems help us to know the authenticity of the website.

Keywords: Triangular fuzzy number (TFN); generalized division of TFN; possibility space; conditional possibility; fuzzy Markov chain; hidden Markov model.

1. Introduction

Hidden Markov model (HMM) is a doubly stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols in such a way that the HMM constitutes of a initial probability, the transition probability and the output symbol observation probability. In HMM, there are three basic problems of interest that must be solved for the model. These problems are namely Evaluation problem, Optimization problem and Training problem.[1]

Nowadays fuzzy approach HMM is developing extensively and widely applied in many areas. New smoothing method based on fuzzy vector quantization (FVQ) is used to avoid more training data in speech recognition.[2,3] Fuzzy Q functions of observed data for discrete HMMs and continuous HMMs are exploited in the field of Speech and Speaker Recognition.[4,5] Fuzzy-C-means HMM is used to decide the status of the traffic with the assumption that the observable events are uncertain.[6]

Detection of intrusion of a website using fuzzy similarity measure instead of the probability measure is discussed in Ref. 7. In Ref. 8, fuzzy clustering function for the emission and transition matrices to analyze the human wrist motion is discussed. Classification of disease for transcranial doppler (TCD) study of the adult intracerebral circulation using fuzzy discrete hidden Markov model (FDHMM) is given in Ref. 9. Breast Cancer identification employing HMM-fuzzy rule approach is discussed in Ref. 10. Classification (hard class and fuzzy class) of satellite image segmentation using hidden fuzzy Markov random fields is given in Ref. 11. Dempster−Shafer fusion in the context of different multisensor Markov models to show that the posterior distribution remains calculable in different general situations and the applications in remote sensing area is discussed in Ref. 12. In Ref. 13, the authors discussed that fuzzy Markov random chains for image segmentation in one hand and on the other hand, they modeled the uncertainty on the observed data using Markovian Bayesian scheme models. It follows that, some authors assumed observations are uncertain,[4−6,9−13] and some authors assumed transition between the hidden states and observations are uncertain.[8] On the basis of their assumption they used fuzzy concepts on HMM.

Our aim is to solve three basic problems of classical HMM to HMM on possibility space, because this space is used to model the incomplete information in a flexible way hence we named HMM on possibility space as fuzzy hidden Markov chain (FHMC) by the existence theorem of possibility space. The word chain is due to the assumption that the states and the time steps are discrete.[14] To capture the real-world fuzziness we have converted the elements of initial possibility distribution, transition possibility between the hidden states and observation possibility of each state in to a special type of fuzzy number called triangular fuzzy number. The operations in the possibility theory is minimum and maximum whereas in the probability theory it is multiplication and addition. The advantage of our proposed model is that it solves the two problems namely Evaluation problem and Optimization problem in a single algorithm, this is due to the operations we have handled which shows that our time consumption is saved. Real-world problems such as web navigation are very challenging to solve. Surfing the website involves traversing the connections among hyperlinked documents. In the literature survey for analyzing the collection of data, questionnaire survey is commonly used to collect opinions and views in the analytic hierarchy process (AHP) but in the AHP, the score items for a comparison matrix in a questionnaire increase drastically if there are more comparisons, which result in longer survey. Therefore, induced bias matrix model (IBMM) is proposed to estimate the missing item scores of the reciprocal pairwise comparison matrix.[15] A Multiple factor hierarchical clustering algorithm for large scale text collection that combines user browsing and retrieval history is given in Ref. 16. Co-word analysis technique is also used to collect the data with the software named CoPalRed.[17] Extracting the web log files using web log analyzer to analyze the user's navigated path obviously creates ambiguity. Hence, classical HMM is less appropriate and to overcome this uncertainty, we have applied our proposed model to our institution website www.ssn.edu.in and we have performed the simulation to analyze the accessibility of the website among the users.

In Sec. 2, we have discussed the preliminaries and HMM, FHMC and three problems of FHMC have explained in Sec. 3, in Sec. 4 illustration and simulation has presented and finally concluded.

2. Basic Concepts and Preliminaries

Fuzzy sets, as its name implies, basically, a theory of graded concepts. Let Γ be the universe of discourse whose generic element is denoted by ω. A fuzzy set $\tilde{A}$ defined on Γ is a mapping from Γ to the unit interval $[0, 1]$, $\mu_{\tilde{A}}(\omega)$ is referred to as the membership function whose value at ω signifies the grade of membership of ω of the fuzzy set $\tilde{A}$ and may vary from 0 to 1. A normalized convex fuzzy set $\tilde{A}$ on Γ whose membership function $\mu_{\tilde{A}}$ is piecewise continuous is called the fuzzy number. The concept of a fuzzy number was introduced by Mizumoto and Tanaka in 1979.[18] The importance of the concept of fuzzy number is still growing due to its application in the frame work of expert systems; roughly speaking a fuzzy number can be considered as a representation for an ill known quantity. A triangular fuzzy number $\tilde{A} = (a_1, a_2, a_3)$ where $a_1 < a_2 < a_3$ is a special type of fuzzy number and it satisfies,[19]

(1) the membership function $\mu_{\tilde{A}}(x) = 1$ at $x = a_2$;
(2) the graph of $y = \mu_{\tilde{A}}(x)$ on $[a_1, a_2]$ is a straight line from $(a_1, 0)$ to $(a_2, 1)$ and also on $[a_2, a_3]$ the graph is a straight line from $(a_2, 1)$ to $(a_3, 0)$;
(3) $\mu_{\tilde{A}}(x) = 0$ for $x \le a_1$ or $x \ge a_3$.

An α-cut of a fuzzy set $\tilde{A}$ denoted by $\tilde{A}_\alpha$ is defined as

$$\tilde{A}_\alpha = [a_1 + \alpha(a_2 - a_1), a_3 - \alpha(a_3 - a_2)] = [a^-, a^+], \quad 0 \le \alpha \le 1. \tag{2.1}$$

0.5 cut of the fuzzy set $\tilde{A}$ is depicted in Fig. 1.

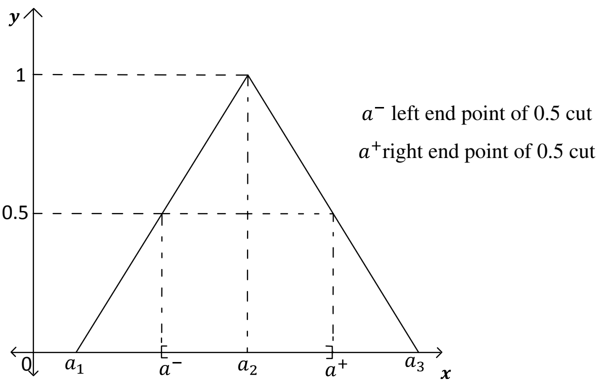

Fig. 1. 0.5 cut.

It is worth noting that contrary to what holds in set theory, $\tilde{A} \cup \tilde{A}^c \neq \Gamma$ and $\tilde{A} \cap \tilde{A}^c \neq \phi$ because it is not certain where $\tilde{A}$ ends and $\tilde{A}^c$ begins. This is the fundamental reason that places probability and fuzzy sets apart and mathematical apparatus of the theory of fuzzy sets provides a natural basis for the theory of possibility. This theory was coined by Zadeh in the late 1970s as an approach to model flexible restrictions constructed from vague pieces of information described by means of fuzzy sets.[20] Possibility theory is maxitive and not additive, i.e., the possibility of a disjunction of events is the maximum of the possibilities of each event

A possibility space is a triple $(\Gamma, \Im, \sigma)$ where[21]:

(i) $\Im$ is a class of all subsets of Γ, i.e., elements of $\Im$ represent the collection of events of interest in that experiment.
(ii) For every $A \in \Im$, the non-negative number $\sigma(A)$ is the possibility that the event A occurs. The map $A \to \sigma(A)$, called a possibility, if $\sigma : \Im \to [0, 1]$, with the following properties:
   (a) $\sigma(\phi) = 0$ and $\sigma(\Gamma) = 1$.
   (b) For an arbitrary collection of sets $A_i \in \Im$, $\sigma(\cup_{i\in I} A_i) = \sup_{i\in I} \sigma(A_i)$.

In the possibility space $(\Gamma, \Im, \sigma)$, given B occurring, we consider the possibility of A, i.e., $\sigma(A|B)$. Suppose $\sigma(B)$ and $\sigma(AB)$ where $\sigma(AB) = \sigma(A \cap B)$ are known. They represent possibilities of B and AB, respectively. We note $B = (B - AB) \cup AB$. If $\sigma(B) = \sigma(AB)$, then it can be said that B achieves its realization on AB. If $\sigma(B) > \sigma(AB)$, then it can be said that B achieves its realization on $B - AB$ rather than on AB. So, given B occurring, if B achieves its realization on AB, then A also occurs. In this way there should be $\sigma(A|B) = 1$. If B achieves its realization on $B - AB$, then the occurrence of B makes no difference on occurrence of AB. Thus $\sigma(A|B) = \sigma(AB|B) = \sigma(AB)$. The conditional possibility of $A \in \Im$ given B denoted by $\sigma(A|B)$ is defined by,[21]

$$\sigma(A|B) = \begin{cases} 1, & \text{if } \sigma(AB) = \sigma(B), \\ \sigma(AB), & \text{if } \sigma(AB) < \sigma(B). \end{cases} \tag{2.2}$$

A possibilistic variable X is a mapping from Γ to an arbitrary universe $\mathbb{U}$ and the possibility distribution function is given by $g(x) = \sigma(X = x) \forall\, x \in \mathbb{U}$, the possibilistic variable X determines a normalized fuzzy set defined on $\mathbb{U}$.[21] A fuzzy Markov chain on the possibility space has the finite number of states $\mathbb{S} = \{1, 2, \ldots, s\}; \mathbb{S} \in \mathbb{U}$ and a possibilistic variables $X = \{X_n; n \in N\}$ an $\mathbb{S}$ valued stochastic process on possibility space and whose possibility measure is σ such that the chain satisfies the Markov property,[22] i.e., for all $j \in \mathbb{S}$ and for each time step $n > 0$ we have,

$$\sigma(X_{n+1} = j | X_0, X_1, \ldots, X_n) = \sigma(X_{n+1} = j | X_n). \tag{2.3}$$

The transition possibility $\tilde{p}_{ij}$ of the system from state i to state j is defined as for each $i, j \in \mathbb{S}$

$$\tilde{p}_{ij} = \sigma(X_{n+1} = j | X_n = i). \tag{2.4}$$

In the case if $\tilde{p}_{ij}$ is independent of time then we can say that the chain is an homogeneous fuzzy Markov chain. Let $\tilde{P} = (\tilde{p}_{ij})$ is an $s \times s$ matrix of transition possibilities and since the possibility theory is maxitive and not additive consequently maximum element of a each row vector of transition matrix is 1 while in probability theory the row sum is 1 (stochastic matrix). Initial possibility vector of the system is denoted by $\tilde{p}^{(0)} = (\tilde{p}_1^{(0)}, \tilde{p}_2^{(0)}, \ldots, \tilde{p}_s^{(0)})$, where $\tilde{p}_i^{(0)} = \sigma(X_0 = i)$ is the possibility of being in the state i initially. To capture the vagueness involved in the system, triangular fuzzy number has been used to the elements of initial, transition possibilities of each state.

2.1. *Hidden Markov model*

A classical HMM is a doubly embedded stochastic process with an underlying process which is a discrete time finite state homogeneous Markov chain that is not observable (it is hidden), but can only be observed through another set of stochastic processes that is a discrete time memory less invariant observations.[1]

Elements of HMM:

(i) s the number of states in the chain, we have denoted the state at time step n as X_n;
(ii) m the number of distinct observation symbols per state, i.e., the discrete output of system. We have denoted the individual symbols as $V = \{v_1, v_2, \ldots, v_m\}$;
(iii) the state transition probability distribution matrix, denoted by $P = (p_{ij})$;
(iv) the observation symbol probability distribution in state j, denoted by $B = (b_j(k))$ where $b_j(k) = P(v_k \text{ at } n | X_n = j)$;
(v) the initial state distribution $p^{(0)} = (p_i^{(0)})$, where $1 \le i, j \le s$ and $1 \le k \le m$.

Given appropriate values of s, m, P, B and $p^{(0)}$, the HMM can be used as a generator to give an observation sequence

$$O = \{o_0, o_1, \ldots, o_{N-1}\}$$

(where each observation o_n is one of the symbol from V, and $N - 1$ is the number of observations in the sequence). We need five elements to specify HMM. A compact notation to indicate the complete parameter set of the model is denoted by $\lambda = (P, B, p^{(0)})$. Given the HMM, there are three basic problems of interest namely:

(1) *Evaluation Problem*: Given the observation sequence $O = \{o_0, o_1, \ldots, o_{N-1}\}$, and a model $\lambda = (P, B, p^{(0)})$ how do we efficiently compute $P(O|\lambda)$, the probability of the observation sequence, for the given model?
(2) *Uncover the hidden part of the model*: Given the observation sequence $O = \{o_0, o_1, \ldots, o_{N-1}\}$, and the model λ how do we choose a corresponding state sequence $X = X_0, \ldots, X_{N-1}$ which is optimal in some meaningful sense (i.e., best "explains" the observations)?

(3) *Training of the Model*: How do we adjust the model parameters $\lambda = (P, B, p^{(0)})$ to maximize $P(O|\lambda)$?

3. Fuzzy Hidden Markov Chain

In this section, we build a FHMC on possibility space in such a way that the initial possibility, transition possibility and observation possibility values are constructed as the Triangular Fuzzy Number to capture the imprecision and solved three basic problems of classical HMM to our proposed model.

FHMC is similar to the HMM where the underlying unobservable process is the fuzzy Markov chain and the observable process is the sequence of outcomes where the observation o_n is independently generated by the state X_n. A formal definition of FHMC is a bivariate discrete process $\{X_n, o_n\}_{n\geq 0}$, where $\{X_n\}$ is a fuzzy Markov chain on possibility space $(\Gamma, \Im, \sigma)$, $\{o_n\}$ is the sequence of observation such that the conditional distribution of o_n only depends on X_n. We have denoted the set in which $\{o_n\}$ takes its value from V.

By Eq. (2.2), the observation symbol possibility distribution in state j, $\tilde{B} = \{\tilde{b}_j(k)\}$,where $\tilde{b}_j(k) = \sigma(v_k \text{ at } n|X_n = j), 1 \leq j \leq s, 1 \leq k \leq m$, is the possibility of individual symbol v_k given that the state is j at step n.

$$\tilde{b}_j(k) = \begin{cases} 1, & \text{if } \sigma(v_k j) = \sigma(j), \\ \sigma(v_k j), & \text{if } \sigma(v_k j) < \sigma(j). \end{cases} \tag{3.1}$$

Thus fuzzy theory replaces probability theory and this leads to a new definition of hidden Markov model parameters denoted by $\tilde{\lambda} = (\tilde{P}, \tilde{B}, \tilde{p}^{(0)})$, where $\tilde{P} = (\tilde{p}_{ij})$ the state transition possibility distribution; $\tilde{B} = \{\tilde{b}_j(k)\}$, the possibility distribution of observation and $\tilde{p}^{(0)} = (\tilde{p}_i^{(0)}), 1 \leq i, j \leq s, 1 \leq k \leq m$, the initial possibility distribution.

We have solved the evaluation problem of FHMC using forward system; the main difference between this forward system and classical one is the operation. Here we have used min$-$max operation instead of multiplication and addition, respectively. We have computed the optimization path of FHMC by modified Viterbi algorithm. The beauty of FHMC is it solves the evaluation problem and computation of optimal path in a single modified Viterbi algorithm itself and this saves our time consumption. Finally we have trained the model parameters of FHMC with the help of backward system and with the generalized division of triangular fuzzy number.[23] To demonstrate how the FHMC works, let us consider an example.

3.1. *Simple demonstration*

Consider the state space as $\mathbb{S} = \{CSE, IT\}$ and assume the outcomes as Faculty (F), News (N), therefore the set of all outcomes $V = \{v_1, v_2\} = \{F, N\}$. We are interested to find the user's accessibility and the most likely state sequence, hence it is necessary to know the possibilities of using each state in user's navigation, as well as the conditional possibilities of hitting Faculty and News of each state. Mathematically,

(1) States: $\mathbb{S} = \{CSE, IT\} = \{1, 2\}$.
(2) The set of all outcomes of each state: $V = \{v_1, v_2\} = \{F, N\}$.
(3) The sequence of the navigation: $X = (x_1, x_2)$ and $x_1, x_2 \in \mathbb{S}$. We have noticed that there are $2^2 = 4$ possible sequences: $X_1 = (1, 1)$; $X_2 = (1, 2)$; $X_3 = (2, 1)$ and $X_4 = (2, 2)$.
(4) The sequence of outcomes: $O = (o_0, o_1)$. We have noticed that each observation would take one of the outcomes, i.e., $o_n \in V$ ($o_n = F$ or N), $n = 0, 1$.
(5) By Eq. (2.4), the possibilities of using Departments CSE and IT in the sequence is given by

$$\tilde{P} = \begin{bmatrix} \tilde{p}_{11} & \tilde{p}_{12} \\ \tilde{p}_{21} & \tilde{p}_{22} \end{bmatrix},$$

where $\tilde{p}_{ij}$ represents the possibility that state i is used first followed by state j.
(6) The possibility of hitting a Faculty and a News for CSE: $\tilde{p}_1(F) = \tilde{\rho}_1$ and $\tilde{p}_1(N) = \tilde{\rho}_2$. Similarly for IT: $\tilde{p}_2(F) = \tilde{\rho}_3$ and $\tilde{p}_2(N) = \tilde{\rho}_4$. This information can be put into a vector:

$$\tilde{B} = [\tilde{b}_1(F) \quad \tilde{b}_1(N) \quad \tilde{b}_2(F) \quad \tilde{b}_2(N)] = [\tilde{\rho}_1 \quad \tilde{\rho}_2 \quad \tilde{\rho}_3 \quad \tilde{\rho}_4],$$

which we can get from Eq. (3.1).
(7) The possibilities of CSE and IT being used at initially: $\tilde{p}^{(0)} = [\tilde{p}_1^{(0)}, \tilde{p}_2^{(0)}]$.

If the model parameters, $\tilde{\lambda} = \{\tilde{P}, \tilde{B}, \tilde{p}^{(0)}\}$, are known, we can find the conditional possibility of the outcome given the observation sequences. For example, assuming that the observation is $O = (N, F)$,

$$\begin{aligned}
\sigma(O|X_1, \tilde{\lambda}) &= \min\{\tilde{b}_1(N), \tilde{b}_1(F)\} = \min\{\tilde{\rho}_1, \tilde{\rho}_2\},\\
\sigma(O|X_2, \tilde{\lambda}) &= \min\{\tilde{b}_1(N), \tilde{b}_2(F)\} = \min\{\tilde{\rho}_2, \tilde{\rho}_3\},\\
\sigma(O|X_3, \tilde{\lambda}) &= \min\{\tilde{b}_2(N), \tilde{b}_1(F)\} = \min\{\tilde{\rho}_4, \tilde{\rho}_1\},\\
\sigma(O|X_4, \tilde{\lambda}) &= \min\{\tilde{b}_2(N), \tilde{b}_2(F)\} = \min\{\tilde{\rho}_4, \tilde{\rho}_3\}.
\end{aligned}$$

On the other hand, the conditional possibility of the user navigation sequence given the model parameter $\tilde{\lambda}$ is given by

$$\begin{aligned}
\sigma(X_1|\tilde{\lambda}) &= \min\{\tilde{p}_1^{(0)}, \tilde{p}_{11}\},\\
\sigma(X_2|\tilde{\lambda}) &= \min\{\tilde{p}_1^{(0)}, \tilde{p}_{12}\},\\
\sigma(X_3|\tilde{\lambda}) &= \min\{\tilde{p}_2^{(0)}, \tilde{p}_{21}\},\\
\sigma(X_4|\tilde{\lambda}) &= \min\{\tilde{p}_2^{(0)}, \tilde{p}_{22}\}.
\end{aligned}$$

By Hisdal inequality,[24] the possibilities of the observation sequences and state sequences occur simultaneously is given by,

$$\sigma(O, X|\tilde{\lambda}) = \min\{\sigma(O|X, \tilde{\lambda}), \sigma(X|\tilde{\lambda})\}$$

or

$$\sigma(O, X_1|\tilde{\lambda}) = \min\{\sigma(O|X_1, \tilde{\lambda}), \sigma(X_1|\tilde{\lambda})\} = \min\{\tilde{p}_1^{(0)}, \tilde{\rho}_1, \tilde{p}_{11}, \tilde{\rho}_2\},$$
$$\sigma(O, X_2|\tilde{\lambda}) = \min\{\sigma(O|X_2, \tilde{\lambda}), \sigma(X_2|\tilde{\lambda})\} = \min\{\tilde{p}_1^{(0)}, \tilde{\rho}_2, \tilde{p}_{12}, \tilde{\rho}_3\},$$
$$\sigma(O, X_3|\tilde{\lambda}) = \min\{\sigma(O|X_3, \tilde{\lambda}), \sigma(X_3|\tilde{\lambda})\} = \min\{\tilde{p}_2^{(0)}, \tilde{\rho}_4, \tilde{p}_{21}, \tilde{\rho}_1\},$$
$$\sigma(O, X_4|\tilde{\lambda}) = \min\{\sigma(O|X_4, \tilde{\lambda}), \sigma(X_4|\tilde{\lambda})\} = \min\{\tilde{p}_2^{(0)}, \tilde{\rho}_4, \tilde{p}_{22}, \tilde{\rho}_3\}.$$

Clearly, the most likelihood sequence of navigation is the one that has the maximum value, which is not difficult to find now. For example, see Table 1, given observation sequence $O = (N, F)$. The most likelihood sequence of navigation is $X_4 = (2, 2) = (IT, IT)$, maximum TFN value among these can be done by comparison of TFN.[25] This number gives the exact approximation of lower and upper range of the value rather than the single probability value in HMM and hence HMM is less suited for this example. If the model parameters, $\tilde{\lambda}$, are unknown, then we can find them by maximizing the total possibility formula of the observation,[21]

$$\begin{aligned}\sigma(O|\tilde{\lambda}) &= \max_i\{\min[\sigma(O|X_i, \tilde{\lambda}), \sigma(X_i|\tilde{\lambda})]\} \\ &= \max_i \sigma(O, X_i|\tilde{\lambda}).\end{aligned} \tag{3.2}$$

In general, we need to compute $\sigma(O|\tilde{\lambda})$ the possibility of the observation sequence $O = \{o_0, o_1, \ldots, o_{N-1}\}$, given the model $\tilde{\lambda}$. The most easy technique of doing this is by enumerating each likely state sequence of length $N - 1$ as in the previous demonstration. Consider one such fixed state sequence $X = X_0, X_1, \ldots, X_{N-1}$ where X_0 is the initial state. The possibility of the observation sequence O for the above state

Table 1. Example of FHMC.

FHMC
Initial possibility vector
$[(1,1,1) \quad (0.8, 0.85, 0.9)]$.
Transition possibility matrix
$\begin{pmatrix} (1,1,1) & (0.5, 0.55, 0.6) \\ (0.6, 0.67, 0.79) & (1,1,1) \end{pmatrix}$.
Output symbol observation possibility
$[(0.4, 0.5, 0.65), \quad (1,1,1), \quad (0.6, 0.7, 0.79), \quad (1,1,1)]$.
$\sigma(O, X_1|\tilde{\lambda})$ $(0.4, 0.5, 0.65)$,
$\sigma(O, X_2|\tilde{\lambda})$ $(0.5, 0.55, 0.6)$,
$\sigma(O, X_3|\tilde{\lambda})$ $(0.4, 0.5, 0.65)$,
$\sigma(O, X_4|\tilde{\lambda})$ $(0.6, 0.7, 0.79)$.
Most navigated path
$IT - IT$
$\sigma(O|\tilde{\lambda}) = (0.6, 0.7, 0.79)$

sequence is

$$\sigma(O|X, \tilde{\lambda}) = \min_{0 \le n \le N-1} \sigma(o_n|X_n, \tilde{\lambda}), \tag{3.3}$$

where we have assumed statistical independence of observations. Thus we get

$$\sigma(O|X, \tilde{\lambda}) = \min\{\tilde{b}_{X_0}(o_0), \tilde{b}_{X_1}(o_1), \ldots, \tilde{b}_{X_{N-1}}(o_{N-1})\}. \tag{3.4}$$

The possibility of such a state sequence X can be written as

$$\sigma(X|\tilde{\lambda}) = \min\{\tilde{\pi}_{X_0}, \tilde{p}_{X_0X_1}, \tilde{p}_{X_1X_2}, \ldots, \tilde{p}_{X_{N-2}X_{N-1}}\}, \tag{3.5}$$

by Hisdal inequality,

$$\sigma(O, X|\tilde{\lambda}) = \min\{\sigma(O|X, \tilde{\lambda}), \sigma(X|\tilde{\lambda})\}. \tag{3.6}$$

The possibility of observation given the model is obtained by maximizing this joint possibility over all likely state sequences X giving

$$\begin{aligned}\sigma(O|\tilde{\lambda}) &= \max_{\text{all X}}\{\sigma(O, X|\tilde{\lambda})\} \\ &= \max_{\mathrm{X}_0,\mathrm{X}_1,\ldots,\mathrm{X}_{N-1}} \{\min[\tilde{\pi}_{X_0}, \tilde{b}_{X_0}(o_0), \tilde{p}_{X_0X_1}, \tilde{b}_{X_1}(o_1), \ldots, \tilde{p}_{X_{N-2}X_{N-1}}, \tilde{b}_{X_{N-1}}(o_{N-1})]\}.\end{aligned}$$

The calculation of $\sigma(O|\tilde{\lambda})$ involves on the order of $2(N-1)s^{(N-1)}$ calculations, because there are s possible states which can be reached at every $n = 0, 1, \ldots, N-1$ and for each such state sequence about $2(N-1)$ calculations are required. This calculation is computationally unfeasible, even for small values of s and $N-1$. Clearly a more efficient algorithm is required.

3.2. *Solution for evaluation problem*

Forward system

Consider the forward variable $\alpha_n(i)$ as

$$\tilde{\alpha}_n(i) = \sigma(o_0, o_1, \ldots, o_n, X_n = i|\tilde{\lambda}), \tag{3.7}$$

i.e., the possibility of the partial observation sequence, $o_0, o_1, \ldots, o_n$ (until step n) and state i at step n, given the model $\tilde{\lambda}$. We have solved $\tilde{\alpha}_n(i)$ inductively as follows:

(1) Initialization

$$\tilde{\alpha}_0(i) = \min[\tilde{p}_i^{(0)}, \tilde{b}_i(o_0)] \quad 1 \le i \le s.$$

(2) Induction

$$\tilde{\alpha}_{n+1}(j) = \min\left\{\left[\max_{1 \le i \le s}[\min(\tilde{\alpha}_n(i), \tilde{p}_{ij})]\right], \tilde{b}_j(o_{n+1})\right\}, \quad 0 \le n \le N-2, \ \ 1 \le j \le s,$$

see Fig. 2.

(3) Termination

$$\sigma(O|\tilde{\lambda}) = \max_{1 \le i \le s}[\tilde{\alpha}_{N-1}(i)].$$

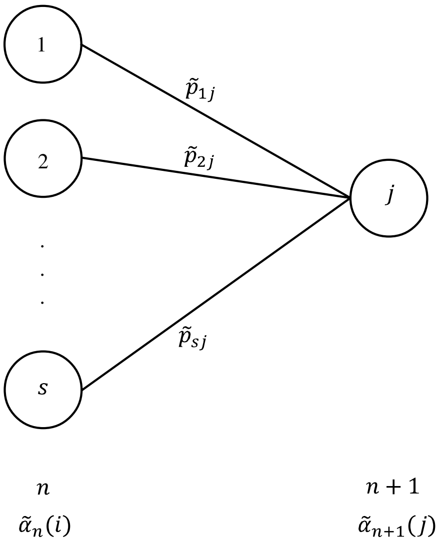

Fig. 2. A schematic representation of forward system.

Since $\tilde{\alpha}_n(i)$ is the possibility of the joint event that $o_0, o_1, \ldots, o_n$ are observed and the state i at time step n, the expression $\min(\tilde{\alpha}_n(i), \tilde{p}_{ij})$ is then the possibility of the joint event that $o_0, o_1, \ldots, o_n$ are observed and state j is reached at step $n+1$ via state i at step n. Maximizing this expression over all s possible states $i, 1 \leq i \leq s$ at step n results in the possibility of j at step $n+1$ with all accompanying previous partial observation. Once this is done and j is known it is easy to calculate $\tilde{\alpha}_{n+1}(j)$. This shows that the computation involved for $\tilde{\alpha}_n(j), 0 \leq n \leq N-1, 1 \leq j \leq s$ is on the order of $s^2(N-1)$.

Backward system

Backward possibilities of the system can be used to reestimate the parameters of the system. Backward variable $\tilde{\beta}_n(i)$ on possibility space is defined as

$$\tilde{\beta}_n(i) = \sigma(o_{n+1}, o_{n+2}, \ldots, o_{N-1} | X_n = i, \tilde{\lambda}), \tag{3.8}$$

i.e., the possibility of the partial observation sequence from $n+1$ to the end, given state i at step n and the model $\tilde{\lambda}$. We can solve for $\tilde{\beta}_n(i)$ inductively, as follows:

(1) Initialization:

$$\tilde{\beta}_{N-1}(i) = (1, 1, 1) \quad 1 \leq i \leq s.$$

(2) Induction:

$$\tilde{\beta}_n(i) = \max_{1 \leq j \leq s} \{\min[\tilde{p}_{ij}, \tilde{b}_j(o_{n+1}), \tilde{\beta}_{n+1}(j)]\}, \quad N-2 \leq n \leq 0 \;\; 1 \leq i \leq s,$$

see Fig. 3.

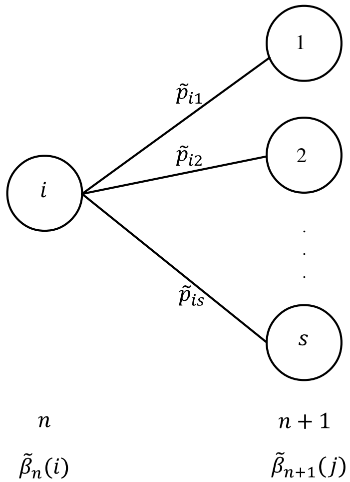

Fig. 3. A schematic representation of backward system.

Step (2) shows that in order to have been in state i at time step n and to account for the observation sequence from time step $n+1$ on, we have to consider all possible states j at step $n+1$ accounting for the transition from i to j (the $\tilde{p}_{ij}$ term) as well as the observation o_{n+1} in state j (the $\tilde{b}_j(o_{n+1})$ term) and then account for the remaining partial observation sequence from state j (the $\tilde{\beta}_{n+1}(j)$ term). Again, the computation of $\tilde{\beta}_n(i), 0 \le n \le N-1, 1 \le i \le s$ requires the order of $s^2(N-1)$ calculations.

3.3. *Solution to find the optimal path*

There are several ways to solve this problem, namely to find the 'optimal' state sequence associated with the given observation sequence. Solution to the above problem is to maximize $\sigma(X|O,\tilde{\lambda})$ which is equivalent to maximizing $\sigma(X,O|\tilde{\lambda})$. To find the single best state sequence, $X = \{X_0, X_1, \ldots, X_{N-1}\}$ for the given observation sequence $O = \{o_0, o_1, \ldots, o_{N-1}\}$ and the model, we need to define the quantity,

$$\tilde{\gamma}_n(i) = \max_{X_0, X_1, \ldots, X_{n-1}} \sigma(X_0, X_1, \ldots, X_n = i, o_0, o_1, \ldots, o_n|\tilde{\lambda}), \tag{3.9}$$

i.e., $\tilde{\gamma}_n(i)$ is the highest possibility along a single path, at time step n, which accounts for the first n observations and ends in state i. By induction we have

$$\tilde{\gamma}_{n+1}(j) = \min\left\{\left[\max_i[\min(\tilde{\gamma}_n(i), \tilde{p}_{ij})]\right], \tilde{b}_j(o_{n+1})\right\}, \quad i, j \in \mathbb{S}, \ 0 \le n \le N-2, \tag{3.10}$$

$$\begin{aligned}\tilde{\gamma}_0(i) &= \sigma(X_0 = i, o_0|\tilde{\lambda}) \\ &= \min\{\sigma(o_0|X_0 = i, \tilde{\lambda}), \sigma(X_0 = i|\tilde{\lambda})\}(\text{by Hisdal inequality}) \\ &= \min\{\tilde{b}_i(o_0), \tilde{p}_i^{(0)}\}.\end{aligned}$$

To retrieve the state sequence, we need to keep track of the argument which maximized Eq. (3.10),[26] for each n and j. We have done this via the array $\varphi_n(j)$. The complete procedure for finding the best state sequence can now be stated as follows:

Modified Viterbi Algorithm

(1) Initialization:

$$\tilde{\gamma}_0(i) = \min[\tilde{p}_i^{(0)}, \tilde{b}_i(o_0)], \quad 1 \leq i \leq s,$$
$$\varphi_0(i) = 0.$$

(2) Recursion:

$$\tilde{\gamma}_{n+1}(j) = \min\left\{\left[\max_{1\leq i\leq s}[\min(\tilde{\gamma}_n(i), \tilde{p}_{ij})]\right], \tilde{b}_j(o_{n+1})\right\}, \quad 0 \leq n \leq N-2, \;\; 1 \leq j \leq s,$$
$$\varphi_{n+1}(j) = \arg\max_{1\leq i\leq s}[\min(\tilde{\gamma}_n(i), \tilde{p}_{ij})], \quad 0 \leq n \leq N-1, \;\; 1 \leq j \leq s.$$

(3) Termination

$$P^* = \max_{1\leq i\leq s}[\tilde{\gamma}_{N-1}(i)],$$
$$X_N^* = \arg\max_{1\leq i\leq s}[\tilde{\gamma}_{N-1}(i)].$$

(4) Path (state sequence) backtracking:

$$X_n^* = \varphi_{n+1}(X_{n+1}^*), \quad n = N-2, \;\; N-3, \ldots, 0.$$

We can note that the modified Viterbi algorithm is same as the forward system except for the backtracking. Hence one can easily find $P^* = \sigma(O|\tilde{\lambda})$ and the optimal path in this algorithm itself.

3.4. *Solution of parameter reestimation*

The third problem of FHMC of the model is to determine a technique to adjust the model parameters $\tilde{\lambda} = (\tilde{P}, \tilde{B}, \tilde{p}^{(0)})$ such that it maximize the possibility of the observation sequence given the model. We have obtained the solution for this problem using iterative procedure.

In order to describe the procedure for reestimation of FHMC parameters, we have to define $\tilde{\xi}_n(i,j)$, the possibility of being in state i at step n, and state j at step $n+1$, given the model and the observation sequence, i.e.,

$$\tilde{\xi}_n(i,j) = \sigma(X_n = i, X_{n+1} = j \mid O, \tilde{\lambda}). \tag{3.11}$$

It is clear that, from the definitions of the forward and backward variables, we can write $\tilde{\xi}_n(i,j)$ in the form

$$\tilde{\xi}_n(i,j) = \frac{\min\{\tilde{\alpha}_n(i), \tilde{p}_{ij}, \tilde{b}_j(o_{n+1}), \tilde{\beta}_{n+1}(j)\}}{\max_{1\leq i\leq s}\{\max_{1\leq j\leq s}[\min(\tilde{\alpha}_n(i), \tilde{p}_{ij}, \tilde{b}_j(o_{n+1}), \tilde{\beta}_{n+1}(j))]\}}. \tag{3.12}$$

Here we have employed the generalized division of triangular fuzzy number where the denominator gives the desired possibility measure and we have defined $\tilde{\zeta}_n(i)$ as the possibility of being in state i at step n, given the observation sequence and the model; hence we can relate $\tilde{\zeta}_n(i)$ to $\tilde{\xi}_n(i,j)$ by maximizing over j, giving

$$\tilde{\zeta}_n(i) = \max_{1\le j\le s} \tilde{\xi}_n(i,j). \tag{3.13}$$

If we take the max of $\tilde{\zeta}_n(i)$ over the step index n, we obtain a quantity which can be interpreted as the expected number of times that state i is visited, or equivalently, the expected number of transitions made from state i by excluding the step $n = N-1$ from the maximization. Likewise, maximization of $\tilde{\xi}_n(i,j)$ over n, i.e., from $n = 0$ to $n = N-2$ can be interpreted as the expected number of transitions from state i to state j. Explicitly

$$\max_{0\le n\le N-2} \tilde{\zeta}_n(i) : \text{Expected number of transitions from state } i, \tag{3.14}$$

$$\max_{0\le n\le N-2} \tilde{\xi}_n(i,j) : \text{Expected number of transitions from state } i \text{ to state } j. \tag{3.15}$$

By means of the above formulas we have given a system of reestimation of the parameters of a FHMC. A set of reasonable reestimation formulas for $\tilde{p}^{(0)}$, $\tilde{P}$ and $\tilde{B}$ are

$$\bar{\tilde{p}}_i^{(0)} = \text{expected number of times in state } i \text{ at step}(n=0) = \tilde{\zeta}_0(i), \tag{3.16}$$

$$\begin{aligned} \bar{\tilde{p}}_{ij} &= \frac{\text{expected number of transitions from state } i \text{ to state } j}{\text{expected number of transitions from state } i} \\ &= \frac{\max_{0\le n\le N-2}\tilde{\xi}_n(i,j)}{\max_{0\le n\le N-2}\tilde{\zeta}_n(i)}, \end{aligned} \tag{3.17}$$

$$\begin{aligned} \bar{\tilde{b}}_j(k) &= \frac{\text{expected number of times in state } j \text{ and observing symbol } v_k}{\text{expected number of times in state } j} \\ &= \frac{\max_{0\le n\le N-2, \ni\, o_n=v_k}\tilde{\zeta}_n(j)}{\max_{0\le n\le N-2}\tilde{\zeta}_n(j)}. \end{aligned} \tag{3.18}$$

By using the current model $\tilde{\lambda} = (\tilde{P}, \tilde{B}, \tilde{p}^{(0)})$ to compute the right-hand sides of Eqs. (3.16)–(3.18) then we get the reestimated model as $\bar{\tilde{\lambda}} = (\bar{\tilde{P}}, \bar{\tilde{b}}, \bar{\tilde{p}}^{(0)})$.

On the basis of the above procedure, if we iteratively use $\bar{\tilde{\lambda}}$ in a place of λ and repeat the reestimation calculation, we then can improve the possibility of O being observed from the model until some limiting point is reached. The final result of this reestimation procedure is called a maximum likelihood estimate of FHMC.

4. Illustration

World Wide Web is a large, distributed hypertext repository of information, whose users navigate through links and view through browsers. These links again have some other new links. User hits the particulars of their own interest.

Web log files contains the data regarding user's navigation of the website. Extracting the log files using web log analyzer even for one day can have the enormous user's navigated path and this path also contains the meaningless path which obviously creates the uncertainty. For this reason we cannot assign a certain probability values to the model. Hence in this situation our model suits well and in this section we have illustrated our proposed model to our institution website (www.ssn.edu.in) by assuming that the hidden states are the departments and observations are the department attributes. Solution of the evaluation problem reveals that for a given observation sequence how much possibility that the website is accessed by the users on that particular day help us to know the familiarity of the department among the users. Optimization of the users navigated path enables that the corresponding state sequence is the best to explain the given observation. Finally, we trained the model parameter to maximize the possibility of given observation sequence.

The style of our institution website is depicted in Fig. 4.

Let the state space $\mathbb{S}$ be the set of all departments

$$\mathbb{S} = \{EEE, MECH, BME\}, \tag{4.1}$$

and the attributes of each department is

$$V = \{About\ the\ Department, Faculty, News\} = \{v_A, v_F, v_N\}. \tag{4.2}$$

When we sketched out the data, we have experienced the sample observation sequence as

$$O = o_0^F, o_1^A, o_2^A, o_3^F, o_4^N, o_5^A, o_6^N, o_7^F, o_8^F, o_9^A, o_{10}^N, o_{11}^N, o_{12}^N, o_{13}^A, o_{14}^F, o_{15}^F, o_{16}^F. \tag{4.3}$$

The above sequence shows that the observation is the faculty (F) at time step 0, and the observation is the About the Department (A) at time step 1, etc. From the path by extracting the web log files using web log analyzer we have computed the initial possibility vector $\tilde{p}^{(0)}$, transition possibility $\tilde{P}$ between the hidden states and the observation possibility $\tilde{B}$ for each state and finally we have converted all the

Fig. 4. Style of the website.

possibility values into the triangular fuzzy number. The values are given below

$$\tilde{p}^{(0)} = [\,(0.58, 0.65, 0.72) \quad (0.39, 0.48, 0.57) \quad (1.00, 1.00, 1.00)\,],$$

$$\widetilde{P} = \begin{array}{c} \\ EEE \\ MECH \\ BME \end{array} \begin{array}{c} \begin{array}{ccc} EEE & MECH & BME \end{array} \\ \begin{bmatrix} (1.00, 1.00, 1.00) & (0.35, 0.49, 0.61) & (0.24, 0.35, 0.46) \\ (0.28, 0.39, 0.48) & (1.00, 1.00, 1.00) & (0.36, 0.50, 0.63) \\ (0.38, 0.52, 0.65) & (0.30, 0.45, 0.60) & (1.00, 1.00, 1.00) \end{bmatrix} \end{array},$$

$$\begin{array}{rl} & \quad\ \tilde{b}(A) \qquad\qquad\quad \tilde{b}(F) \qquad\qquad\quad \tilde{b}(N) \\ \tilde{B}_{EEE} = & [\,(0.32, 0.46, 0.60) \quad (1.00, 1.00, 1.00) \quad (0.27, 0.38, 0.50)\,], \\ \tilde{B}_{MECH} = & [\,(1.00, 1.00, 1.00) \quad (0.32, 0.47, 0.63) \quad (0.25, 0.40, 0.56)\,], \\ \tilde{B}_{BME} = & [\,(0.14, 0.27, 0.40) \quad (0.30, 0.43, 0.56) \quad (1.00, 1.00, 1.00)\,]. \end{array}$$

Enhancing the modified Viterbi algorithm we obtained the result as follows:

$$\tilde{\gamma}_0(EEE) = (0.58, 0.65, 0.72) \text{ by (A.1)},$$
$$\tilde{\gamma}_0(MECH) = (0.32, 0.47, 0.63) \text{ by (A.2)},$$
$$\tilde{\gamma}_0(BME) = (0.30, 0.43, 0.56) \text{ by (A.3)},$$

$\varphi_0(EEE) = 0, \varphi_0(MECH) = 0, \varphi_0(BME) = 0$ (by definition). Changing the time step from $n = 0$ to $n = 1$ we get,

$$\tilde{\gamma}_1(EEE) = (0.32, 0.46, 0.60) \text{ by (A.4)},$$
$$\varphi_1(EEE) = EEE \text{ (by the definition of Argmax) by (A.5)},$$

similarly for *MECH*,

$$\tilde{\gamma}_1(MECH) = (0.35, 0.49, 0.61) \text{ by (A.6)},$$
$$\varphi_1(MECH) = EEE \text{ by (A.7)},$$

and for *BME*,

$$\tilde{\gamma}_1(BME) = (0.14, 0.27, 0.40) \text{ by (A.8)},$$
$$\varphi_1(BME) = MECH \text{ by (A.9)}.$$

Then by Induction

$$\tilde{\gamma}_{16}(EEE) = (0.30, 0.45, 0.60),$$
$$\varphi_{16}(EEE) = BME,$$
$$\tilde{\gamma}_{16}(MECH) = (0.30, 0.45, 0.60),$$
$$\varphi_{16}(MECH) = BME,$$
$$\tilde{\gamma}_{16}(BME) = (0.30, 0.43, 0.56),$$
$$\varphi_{16}(BME) = BME.$$

$$\tilde{P}^* = \max[\tilde{\gamma}_{16}(EEE), \tilde{\gamma}_{16}(MECH), \tilde{\gamma}_{16}(BME)] = (0.30, 0.45, 0.60) = \sigma(O|\tilde{\lambda}). \tag{4.4}$$

Possibility of given observation sequence and the model is obtained as $(0.30, 0.45, 0.60)$. State sequence which is the best to explain the given observation sequence is computed as follows:

$$
\begin{aligned}
X_{16}^* &= BME,\\
X_{15}^* &= \varphi_{16}(BME) = BME,\\
X_{14}^* &= \varphi_{15}(BME) = BME,\\
\vdots &= \vdots\\
X_5^* &= \varphi_6(BME) = BME,\\
X_4^* &= \varphi_5(BME) = MECH,\\
X_3^* &= \varphi_4(MECH) = MECH,\\
X_2^* &= \varphi_3(MECH) = EEE,\\
X_1^* &= \varphi_2(EEE) = EEE,\\
X_0^* &= \varphi_1(EEE) = EEE.
\end{aligned}
$$

The optimal path is depicted in Fig. 5.

Simulation

To evaluate the performance of parameter estimation on FHMC, we have executed the experiment on iterative use of $\overline{\tilde{\lambda}}$ in the place of $\tilde{\lambda}$. Triangular fuzzy number (lower, middle and upper) values of $\tilde{\alpha}_n(i)$,$\tilde{\beta}_n(i)$ for the initial and final iteration for the states *EEE, MECH* and *BME* are depicted in Fig. 6.

In Fig. 6, initial iteration of the $\tilde{\alpha}_n(i)$ values for the states *EEE, MECH* and *BME* are given in the left-hand side (LHS) and final iteration values of $\tilde{\alpha}_n(i)$ for the corresponding states are given in the right-hand side (RHS). By comparing both sides of the figure one can easily notice that the TFN values for the state *EEE* and *MECH* are improved from [0.25, 0.72] to [0.83, 1.00] and for the *BME* it is improved from [0.14, 0.63] to [0.467, 1.00] which shows that our model has performed well. In particular the states *EEE* and *MECH* in the RHS of the figure converge to the possibility value of (1.00, 1.00, 1.00) for the same 12 time steps.

In Fig. 7, backward possibility $\tilde{\beta}_n(i)$ for the initial iteration of the states *EEE*, *MECH* and *BME* are given in the LHS and in the RHS, final iteration of the $\tilde{\beta}_n(i)$ for the corresponding states has given. Here also we can notice that for the state *EEE*,

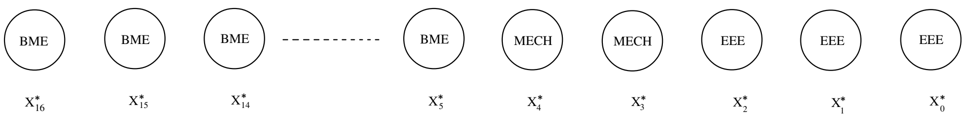

Fig. 5. A schematic representation of optimal state sequence.

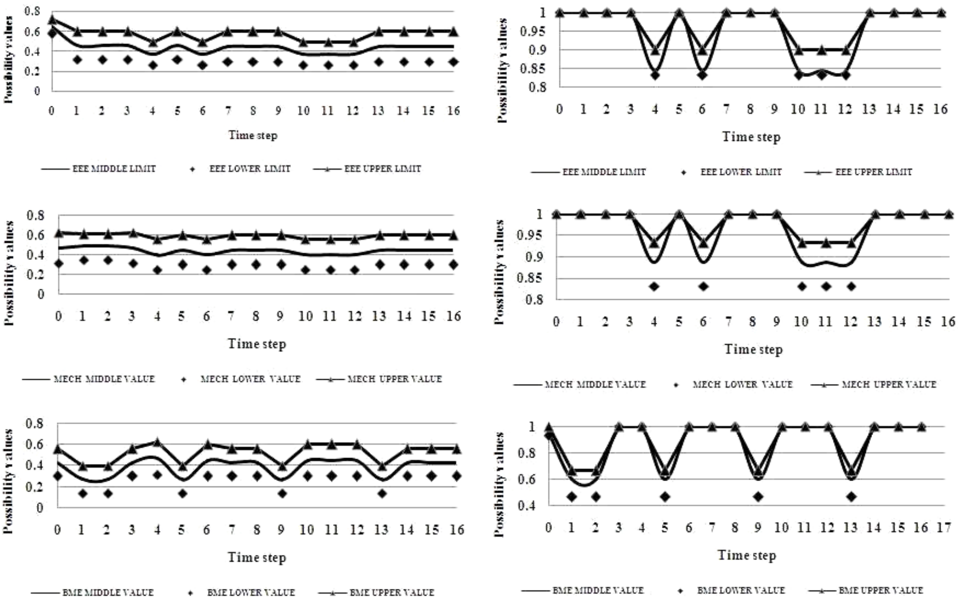

Fig. 6. A schematic representation of initial and final iteration values of $\tilde{\alpha}_n(i)$ for the states EEE, MECH and BME.

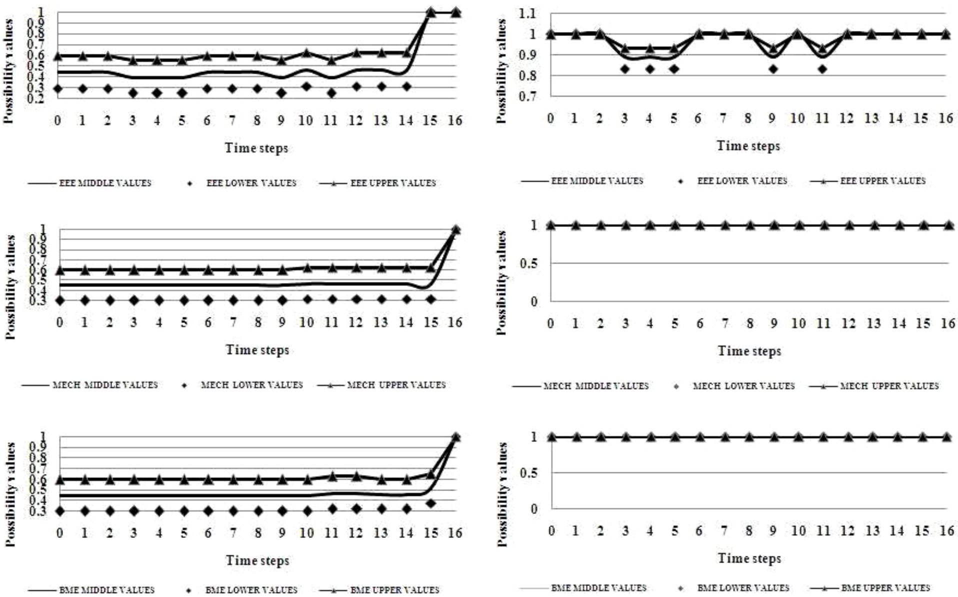

Fig. 7. A schematic representation of initial and final iteration values of $\tilde{\beta}_n(i)$ for the states EEE, MECH and BME.

the possibility values are improved from [0.25, 1] into [0.83, 1], and for the states *MECH* and *BME* the values are improved from [0.3, 1] into 1.

From the simulation results, we have obtained the maximum likelihood value of the model as $\overline{\tilde{P}}^* = (1.00, 1.00, 1.00)$ i.e., the possibility of given observation sequence is maximized from $(0.30, 0.45, 0.60)$ to $(1.00, 1.00, 1.00)$ which shows that the departments are more authentic among the users. Using Eqs. (3.16)−(3.18) the reestimated values are obtained which is given below and calculations are given in appendices A, B and C.

$$\overline{\tilde{p}}^{(0)} = [(1.00, 1.00, 1.00) \quad (1.00, 1.00, 1.00) \quad (0.93, 0.95, 1.00)],$$

$$\overline{\tilde{P}} = \begin{array}{c} \\ EEE \\ MECH \\ BME \end{array} \begin{array}{c} \begin{array}{ccc} \quad EEE \quad & \quad MECH \quad & \quad BME \quad \end{array} \\ \begin{bmatrix} (1.00, 1.00, 1.00) & (1.00, 1.00, 1.00) & (0.77, 0.78, 0.80) \\ (0.80, 0.87, 0.93) & (1.00, 1.00, 1.00) & (1.00, 1.00, 1.00) \\ (1.00, 1.00, 1.00) & (1.00, 1.00, 1.00) & (1.00, 1.00, 1.00) \end{bmatrix} \end{array},$$

$$\begin{aligned} & \qquad\qquad \tilde{b}(A) \qquad\qquad\qquad \tilde{b}(F) \qquad\qquad\qquad \tilde{b}(N) \\ \overline{\tilde{B}}_{EEE} &= [\,(1.00, 1.00, 1.00) \quad (1.00, 1.00, 1.00) \quad (0.83, 0.84, 0.90)\,], \\ \overline{\tilde{B}}_{MECH} &= [\,(1.00, 1.00, 1.00) \quad (1.00, 1.00, 1.00) \quad (0.83, 0.88, 0.93)\,], \\ \overline{\tilde{B}}_{BME} &= [\,(0.47, 0.60, 0.66) \quad (1.00, 1.00, 1.00) \quad (1.00, 1.00, 1.00)\,]. \end{aligned}$$

5. Conclusion

Real-world applications have uncertainty in it due to the imprecision in the data. Fuzzy numbers have the capability to overcome this situation consequently the mathematical apparatus of the theory of fuzzy sets provides a natural basis for the theory of possibility, hence we have proposed fuzzy hidden Markov chain on possibility space and solved three basic problems of classical HMM to our proposed model. The algorithm which we have adapted namely modified Viterbi algorithm itself gives the solution for evaluation problem and optimization problem, this shows that the algorithm reduces our time consumption. Finally we have trained the problem to maximize the model parameter. We applied our proposed model to our institution website and also performed simulation. The simulation results shows that our model is more authentic and more user friendly.

Appendix A. Calculation of Evaluation Problem and to Find the Optimal Path

$$\begin{aligned} \tilde{\gamma}_0(EEE) &= \sigma(EEE, o_0|\tilde{\lambda}) \\ &= \min[\tilde{p}^{(0)}_{EEE}, \tilde{b}_{EEE}(o_0) = F] \\ &= \min[(0.58, 0.65, 0.72), (1.00, 1.00, 1.00)] \\ &= (0.58, 0.65, 0.72), \text{(by comparison of TFN)}, \end{aligned} \tag{A.1}$$

$$\begin{aligned}
\tilde{\gamma}_0(MECH) &= \sigma(MECH, o_0|\tilde{\lambda}) \\
&= \min[\tilde{p}^{(0)}_{MECH}, \tilde{b}_{MECH}(o_0) = F] \\
&= \min[(0.39, 0.48, 0.57), (0.32, 0.47, 0.63)] \\
&= (0.32, 0.47, 0.63), \qquad \text{(A.2)} \\
\tilde{\gamma}_0(BME) &= \sigma(BME, o_0|\tilde{\lambda}) \\
&= \min[\tilde{p}^{(0)}_{BME}, \tilde{b}_{BME}(o_0) = F] \\
&= \min[(1.00, 1.00, 1.00), (0.30, 0.43, 0.56)] \\
&= (0.30, 0.43, 0.56), \qquad \text{(A.3)} \\
\tilde{\gamma}_1(j) &= \min\left\{\{\max_{i\in\mathbb{S}}[\min(\tilde{\gamma}_0(i), \tilde{p}_{ij})], \tilde{b}_j(o_1)\}\right\},
\end{aligned}$$

$$\begin{aligned}
&\tilde{\gamma}_1(EEE) \\
&= \min\left\{\left[\max\left\{\begin{array}{c}\min(\tilde{\gamma}_0(EEE), \tilde{p}_{EEE\ EEE}) \\ \min(\tilde{\gamma}_0(MECH), \tilde{p}_{MECH\ EEE}) \\ \min(\tilde{\gamma}_0(BME), \tilde{p}_{BME\ EEE})\end{array}\right\}\right], \tilde{b}_{EEE}(o_1) = A\right\} \\
&= \min\left\{\left[\max\left\{\begin{array}{l}\min((0.58, 0.65, 0.72), (1.00, 1.00, 1.00)) \\ \min((0.32, 0.47, 0.63), (0.28, 0.39, 0.48)) \\ \min((0.30, 0.43, 0.56), (0.38, 0.52, 0.65))\end{array}\right\}\right], (0.32, 0.46, 0.60)\right\} \\
&= \min\left\{\left[\max\left\{\begin{array}{l}(0.58, 0.65, 0.72) \\ (0.28, 0.39, 0.48) \\ (0.30, 0.43, 0.56)\end{array}\right\}\right], (0.32, 0.46, 0.60)\right\} \\
&= (0.30, 0.46, 0.60), \qquad \text{(A.4)}
\end{aligned}$$

$$\varphi_1(EEE) = EEE, \text{Similarly for other states,} \qquad \text{(A.5)}$$

$$\begin{aligned}
\tilde{\gamma}_1(MECH) &= \min\left\{\left[\max\left\{\begin{array}{c}\min(\tilde{\gamma}_0(EEE), \tilde{p}_{EEE\ MECH}) \\ \min(\tilde{\gamma}_0(MECH), \tilde{p}_{MECH\ MECH}) \\ \min(\tilde{\gamma}_0(BME), \tilde{p}_{BME\ MECH})\end{array}\right\}\right], \tilde{b}_{MECH}(o_1)\right\} \\
&= (0.35, 0.49, 0.61), \qquad \text{(A.6)} \\
\varphi_1(MECH) &= EEE, \qquad \text{(A.7)} \\
\tilde{\gamma}_1(BME) &= (0.14, 0.27, 0.40), \qquad \text{(A.8)} \\
\varphi_1(BME) &= MECH. \qquad \text{(A.9)}
\end{aligned}$$

Appendix B. Finding Values for Backward Variable

$$\tilde{\beta}_{16}(EEE) = (1,1,1), \tag{B.1}$$

$$\tilde{\beta}_{16}(MECH) = (1,1,1), \tag{B.2}$$

$$\tilde{\beta}_{16}(BME) = (1,1,1). \tag{B.3}$$

$$\tilde{\beta}_n(i) = \max\{\min(\tilde{p}_{ij}, \tilde{b}_j(o_{n+1}), \tilde{\beta}_{n+1}(j))\}, \tag{B.4}$$

i.e.,

$$\tilde{\beta}_{15}(EEE) = \max\left\{\begin{array}{c} \min[\tilde{p}_{EEE\ EEE}, \tilde{b}_{EEE}(o_{16}), \tilde{\beta}_{16}(EEE)], \\ \min[\tilde{p}_{EEE\ MECH}, \tilde{b}_{MECH}(o_{16}), \tilde{\beta}_{16}(MECH)], \\ \min[\tilde{p}_{EEE\ BME}, \tilde{b}_{BME}(o_{16}), \tilde{\beta}_{16}(BME)] \end{array}\right\}, \tag{B.5}$$

$$\begin{aligned}
\tilde{\beta}_{15}(EEE) &= \sigma(o_{16}|EEE, EEE, \tilde{\lambda}) \\
&= \min\{\tilde{p}_{EEE\ EEE}, \tilde{b}_{EEE}(o_{16}), \tilde{\beta}_{16}(EEE)\} \\
&= \min\{(1.00, 1.00, 1.00), (1.00, 1.00, 1.00), (1.00, 1.00, 1.00)\} \\
&= (1.00, 1.00, 1.00) \qquad \text{(B.6)} \\
\text{or} \\
&= \sigma(o_{16}|EEE, MECH, \tilde{\lambda}) \\
&= \min\{\tilde{p}_{EEE\ MECH}, \tilde{b}_{MECH}(o_{16}), \tilde{\beta}_{16}(MECH)\} \\
&= \min\{(0.35, 0.49, 0.61), (0.32, 0.47, 0.63), (1.00, 1.00, 1.00)\} \\
&= (0.32, 0.47, 0.63) \qquad \text{(B.7)} \\
\text{or} \\
&= \sigma(o_{16}|EEE, BME, \tilde{\lambda}) \\
&= \min\{\tilde{p}_{EEE\ BME}, \tilde{b}_{BME}(o_{16}), \tilde{\beta}_{16}(BME)\} \\
&= \min\{(0.24, 0.35, 0.46), (0.30, 0.43, 0.56), (1.00, 1.00, 1.00)\} \\
&= (0.24, 0.35, 0.46), \qquad \text{(B.8)} \\
\max[(\text{B.6}), (\text{B.7}), (\text{B.8})] &= (1.00, 1.00, 1.00), \\
\therefore \tilde{\beta}_{15}(EEE) &= (1.00, 1.00, 1.00). \qquad \text{(B.9)}
\end{aligned}$$

Similarly for other states,

$$\tilde{\beta}_{15}(MECH) = (0.32, 0.47, 0.63), \tag{B.10}$$

$$\tilde{\beta}_{15}(BME) = (0.38, 0.52, 0.65). \tag{B.11}$$

By induction,

$$\tilde{\beta}_0(EEE) = (0.30, 0.45, 0.60), \tag{B.12}$$

$$\tilde{\beta}_0(MECH) = (0.30, 0.45, 0.60), \tag{B.13}$$

$$\tilde{\beta}_0(BME) = (0.30, 0.45, 0.60). \tag{B.14}$$

Appendix C. Calculations of $\tilde{\xi}_n(i,j)$, $\tilde{\zeta}_n(i)$

$$\tilde{\xi}_n(i,j) = \frac{\min\{\tilde{\alpha}_n(i), \tilde{p}_{ij}, \tilde{b}_j(o_{n+1}), \tilde{\beta}_{n+1}(j)\}}{\max_{1\le i\le s}\{\max_{1\le j\le s}[\min(\tilde{\alpha}_n(i), \tilde{p}_{ij}, \tilde{b}_j(o_{n+1}), \tilde{\beta}_{n+1}(j))]\}},$$

$$\tilde{\xi}_0(E,E) = \frac{\min\{\tilde{\alpha}_0(E), \tilde{p}_{EE}, \tilde{b}_E(o_1 = A), \tilde{\beta}_1(E)\}}{\max\left\{\begin{array}{l} \max\begin{bmatrix} \min(\tilde{\alpha}_0(E), \tilde{p}_{EE}, \tilde{b}_E(o_1), \tilde{\beta}_1(E)) \\ \min(\tilde{\alpha}_0(E), \tilde{p}_{EM}, \tilde{b}_M(o_1), \tilde{\beta}_1(M)) \\ \min(\tilde{\alpha}_0(E), \tilde{p}_{EB}, \tilde{b}_B(o_1), \tilde{\beta}_1(B)) \end{bmatrix} \\ \max\begin{bmatrix} \min(\tilde{\alpha}_0(M), \tilde{p}_{ME}, \tilde{b}_E(o_1), \tilde{\beta}_1(E)) \\ \min(\tilde{\alpha}_0(M), \tilde{p}_{MM}, \tilde{b}_M(o_1), \tilde{\beta}_1(M)) \\ \min(\tilde{\alpha}_0(M), \tilde{p}_{MB}, \tilde{b}_B(o_1), \tilde{\beta}_1(B)) \end{bmatrix} \\ \max\begin{bmatrix} \min(\tilde{\alpha}_0(B), \tilde{p}_{BE}, \tilde{b}_E(o_1), \tilde{\beta}_1(E)) \\ \min(\tilde{\alpha}_0(B), \tilde{p}_{BM}, \tilde{b}_M(o_1), \tilde{\beta}_1(M)) \\ \min(\tilde{\alpha}_0(B), \tilde{p}_{BB}, \tilde{b}_B(o_1), \tilde{\beta}_1(B)) \end{bmatrix} \end{array}\right\}},$$

where E, M and B are EEE, $MECH$ and BME, respectively.
Numerator:

$$\begin{aligned}&\min\{(0.58, 0.65, 0.72), (1.00, 1.00, 1.00), (0.32, 0.46, 0.60), (0.30, 0.45, 0.60)\}\\ &\quad = (0.30, 0.45, 0.60).\end{aligned} \tag{C.1}$$

Denominator:

$$\max\{(0.30, 0.45, 0.60), (0.30, 0.45, 0.60), (0.30, 0.43, 0.56)\}$$

$$= (0.30, 0.45, 0.60). \tag{C.2}$$

$$\therefore \tilde{\xi}_0(EEE, EEE) = (1.00, 1.00, 1.00). \tag{C.3}$$

Let us see the calculation for $\tilde{\xi}_0(EEE, BME)$
Numerator:

$$\begin{aligned}&\min\{(0.58, 0.65, 0.72), (0.24, 0.35, 0.46), (0.14, 0.27, 0.40), (0.30, 0.45, 0.60)\},\\ &\quad = (0.14, 0.27, 0.40) = A(say).\end{aligned} \tag{C.4}$$

Denominator:

$$\begin{aligned}&\max\{(0.30, 0.45, 0.60), (0.30, 0.45, 0.60), (0.30, 0.43, 0.56)\},\\ &\quad = (0.30, 0.45, 0.60) = B(say).\end{aligned} \tag{C.5}$$

The $0 - cut$ is given by Eq. (2.1) and therefore $\tilde{A}_0 = [0.14, 0.40]$, $\tilde{B}_0 = [0.30, 0.60]$ and $a^-b^+ = 0.084$ and $a^+b^- = 0.12$. This clearly shows, that $a^-b^+ < a^+b^-$. Hence by the generalized division of TFN $c^- = \frac{a^-}{b^-} = 0.46$, $c^+ = \frac{a^+}{b^+} = 0.66$ and the $1 - cut$ is given by $\tilde{A}_1 = [0.27, 0.27]$,$\tilde{B}_1 = [0.45, 0.45]$ and $a^-b^+ = 0.1215 = a^+b^-$ therefore the resultant triangular fuzzy number is $\tilde{\xi}_0(EEE, BME) = (0.46, 0.60, 0.66)$. Similarly we can compute remaining $\tilde{\xi}_n(i,j)$ using the generalized division of TFN and $\tilde{\zeta}_n(i)$ can be calculated using $\tilde{\zeta}_n(i) = \max_{1 \le j \le s} \tilde{\xi}_n(i,j)$.

References

1. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *IEEE Proceedings* **77**(2) (1989) 257–286.
2. J. M. Koo and C. K. Un, Fuzzy smoothing of HMM parateters in sppech recognition, *Electronic Letters* **26**(11) (1990) 7443–7447.
3. J. Dai, J. E. M. Tyler and I. G. Mac Kenzie, Fuzzy smoothing of HMM parameters using the Parzen window, *Electronic Letters* **29**(22) (1993) 1911–1913.
4. D. Tran and M. Wagner, Fuzzy hidden Markov models for speech and speaker recognition, *18th Int. Conf. North American Fuzzy Information Processing Society — NAFIPS* (1999) pp. 426–430.
5. M. R. Tarihi, A. Taheri and H. Bababeyk, A new method for fuzzy hidden Markov models in speech recognition, *Int. Conf. Emerging Technologies, IEEE*, September 2005, pp. 36–40.
6. L. Li, J. Hu, Q. Huang and J. Song, A fuzzy hidden Markov model for traffic status classification based on video features, *Multi conf. Computational Engineering in systems Applications*, October 2006, pp. 2050–2055.
7. Y. Li, R. Wang and J. Xu (eds.), Intrusion detection method based on fuzzy hidden Markov model, *Sixth Int. Conf. Fuzzy Systems and Knowledge Discovery*, August 2009, pp. 470–474.
8. X. Zhang and F. Naghdy, Human motion recognition through fuzzy hidden Markov model, *Int. Conf. Computational Intelligence for Modelling, Control and Automation, and Int. Conf on Intelligent Agents, Web Technologies and Internet Commerce*, November 2005, pp. 450–456.
9. H. Uguz, A. Ozturk, R. Saracoglu and A. Arslan, A biomedical system based on fuzzy discrete hidden Markov model for the diagnosis of the brain diseases, *Expert Systems with Applications* **35** (2008) 1104–1114.
10. H. Uguz, A. Arslan, R. Saracoglu and I. Turkoglu, Detection of heart valve diseases by using fuzzy discrete hidden Markov model, *Expert Systems with Applications* **34** (2008) 2799–2811.
11. F. Salzenstein and W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, *Graphical Models and Image Processing* **59**(4) (1997) 205–220.
12. W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, *International Journal of Approximate Reasoning* **45** (2007) 1–16.
13. F. Salzenstein, C. Collet, S. Lecam and M. Hatt, Non-stationary fuzzy Markov chain, *Pattern Recognition Letters* **28** (2007) 2201–2208.
14. K. S. Trivedi, *Probability & Statistics with Reliability, Queuning and Computer Science Applications*, 2nd edn. (John Wiley & Sons, 2002).
15. D. Ergu and G. Kou, Questionnaire design improvement and missing item scores estimation for rapid and efficient decision making, *Annals of Operation Research* **197** (2012) 5–23.
16. G. Kou and C. Lou, Multiple factor hierarchical clustering algorithm for large scale web page and search engine clickstream data, *Annals of Operation Research* **197** (2012) 123–134.

17. A. G. L. Herrera, E. H. Viedma, M. J. Cobo, M. A. Martinez, G. Kou and Y. Shi, A conceptual snapshot of the first decade (2002−2011) of the international journal of information technology decision making, *International Journal of Information Technology & Decision Making* **11**(2) (2012) 247−270.
18. M. Mizumoto and K. Tanaka, Some properties of fuzzy numbers, *Advances in Fuzzy set Theory and Applications*, eds. M. M. Gupta *et al.* (North Holland, Amsterdam, 1979), pp. 153−164.
19. J. J. Buckley and E. Eslami, Fuzzy Markov chains: Uncertain probabilities, *Math Ware and Soft Computing* **9** (2002) 33−41.
20. L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* **1** (1978) 3−28.
21. K. Y. Cai, *Introduction to Fuzzy Reliaiblity* (Kluwer Academic Publishers, 1996).
22. B. Praba and R. Sujatha, A fuzzy Markov model for web testing, *Ganita Sandesh* **21**(2) (2007) 111−120.
23. L. Stefanini, A generalization of Hukuhara difference and division for interval and fuzzy arithmetic, *Fuzzy Sets and Systems* **161** (2010) 1564−1584.
24. D. Dubois, Possibility theory and statistical reasoning, *Computational Statistics & Data Analysis* **51**(1) (2006) 47−69.
25. M. G. Iskander, Comparison of fuzzy numbers using possibility programming: Comments and new concepts, *Computers and Mathematics with Applications* **43** (2002) 833−840.
26. Argmax, http://en.wikipedia.org/wiki/Arg_max.