Applied Soft Computing 62 (2018) 315–327

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Evaluation of a novel fuzzy sequential pattern recognition tool (fuzzy elastic matching machine) and its applications in speech and handwriting recognition

Sina Shahmoradi *, Saeed Bagheri Shouraki

Artificial Creatures Lab (ACL), Department of Electrical Engineering, Sharif University of Technology, Azadi Ave, Tehran, Iran

ARTICLE INFO

Article history:
Received 23 November 2016
Received in revised form
26 September 2017
Accepted 21 October 2017
Available online 31 October 2017

Keywords:
Fuzzy sequential pattern recognition
Fuzzy elastic pattern
Speech recognition
Handwriting recognition
Hidden markov model

ABSTRACT

Sequential pattern recognition has long been an important topic of soft computing research with a wide area of applications including speech and handwriting recognition. In this paper, the performance of a novel fuzzy sequential pattern recognition tool named "Fuzzy Elastic Matching Machine" has been investigated. This tool overcomes the shortcomings of the HMM including its inflexible mathematical structure and inconsistent mathematical assumptions with imprecise input data. To do so, "Fuzzy Elastic Pattern" was introduced as the basic element of FEMM. It models the elasticity property of input data using fuzzy vectors. A sequential pattern such as a word in speech or a piece of writing is treated as a sequence of parts in which each part has an elastic nature (i.e. can skew or stretch depending on the speaker/writer's style). To present FEMM as a sequential pattern recognition tool, three basic problems, including evaluation, assignment, and training problems, were defined and their solutions were presented for FEMMs. Finally, we implemented FEMM for speech and handwriting recognition on some large databases including TIMIT database and Dr. Kabir's Persian handwriting database. In speech recognition, FEMM achieved 71% and 75.5% recognition rates in phone and word recognition, respectively. Also, 75.9% recognition accuracy was obtained in Persian handwriting recognition. The results indicated 18.2% higher recognition speed and 9–16% more immunity to noise in speech recognition in addition to 5% higher recognition rate in handwriting recognition compared to the HMM.

1. Introduction

Sequential data often emerge in the measurement of time series, e.g. the acoustic features at the time frames of speech signals [1] and kinematic and biological signals in human movements [2]. Hence, sequential pattern recognition is increasingly used in many data analysis fields such as handwriting recognition [3,4], speech recognition [5,6], activity recognition [7,8], and gesture recognition [2]. In this approach, sequences of information are utilized in decision-making [9]. Sequential pattern recognition in time-series information matches the input data with a pattern, both of which being sequences of information. However, inequality in the length of sequences between the input data and the pattern is a typical problem. Studies in 1960s showed that an appropriate pattern recognition tool should be developed to overcome this problem in speech recognition [10]. One of the main solutions to this problem is the "time-wrapping" idea, presented by Vintsyk [11]. He showed how dynamic programming can be used to find the best assignment between two sequences. In 1970s and 1980s, many researchers proposed various models that mostly included modeling of acoustic information in a network with some finite states, deriving the stochastic information of acoustic features and comparing them with the input signal using the dynamic time-wrapping method [12]. Some examples of these works can be seen in [13–16]. The most well-known method in this field is Hidden Markov Model (HMM) [17]. The capabilities of the HMM such as the ability to match with stochastic sequential time series and suitable training algorithms have made it a pervasive tool in speech recognition. Since mid-1990s, the HMM has been a dominant method in handwriting and speech recognition [18]. Besides, many researchers have demonstrated its superiority to short time classifiers such as MLP [19] and SVM [8].

Despite the powerful mathematical modeling of the HMMs, the main criticisms about them fall into two categories: 1) limitations due to the selected mathematical modeling assumptions. One of these assumptions is that the state duration in the HMM is

* Corresponding author.
E-mail address: sina.shah71@gmail.com (S. Shahmoradi).

https://doi.org/10.1016/j.asoc.2017.10.036
1568-4946/

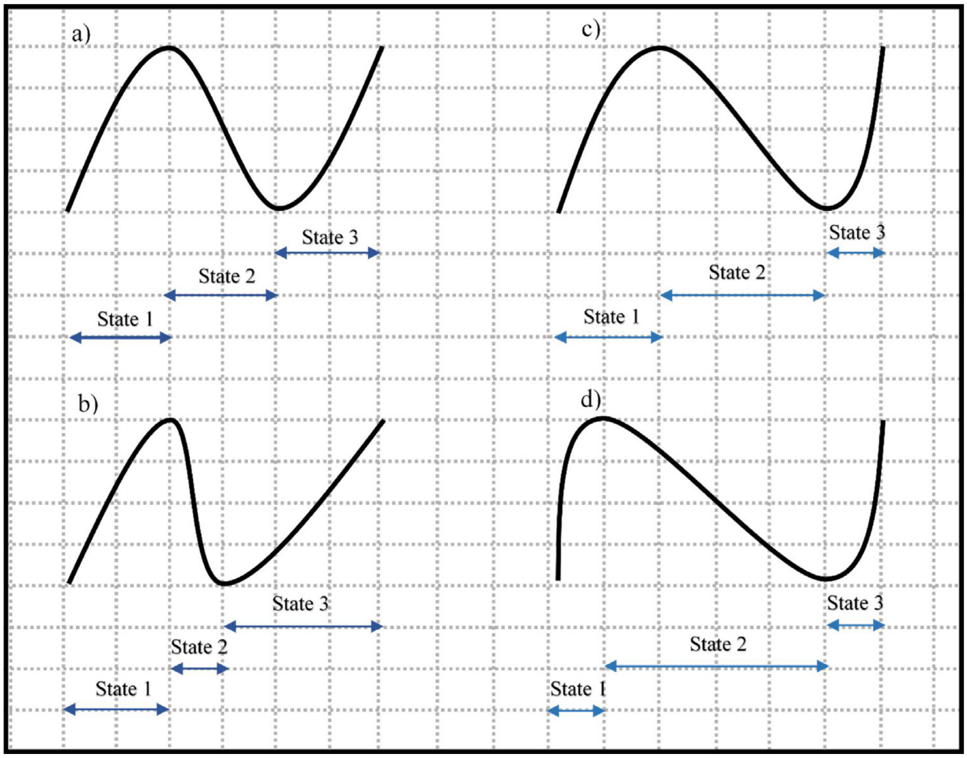

Fig. 1. Four samples that show a pattern. In the sequential approach, each sample is segmented into three segments. Each segment has specific features with an elastic nature. An appropriate sequential data model should describe the pattern according to these samples. Also, it should measure the similarity between the input data and this pattern.

implicitly a geometric distribution [20]. Consequently, the HMM is incapable of accommodating many important facets of the speech pattern structure [21]; 2) its inflexible mathematical structure, which is inconsistent with the imprecise nature of handwritten and speech data.

The main solution to the first problem is modeling the state duration explicitly using a state duration probability density. These works are referred to in the literature as the "Segmental Hidden Markov Model" [22] or "Hidden Semi-Markov Model" [20]. In addition, state duration modeling has been proven to be effective in the extensions of the HMMs such as Bayesian non-parametric HSMM [23] or explicit duration switch HMMs [24]. Although modeling the state duration has mostly yielded better recognition rates, more computational complexity is the main problem of standard HSMMs [25,26].

The major solution to the second problem is combining the HMM with fuzzy modeling in order to reduce its sensitivity to changes (e.g. different handwriting/speech styles) and to improve recognition performance in the presence of noise [27,28]. Tseng combined fuzzy vector quantization with HMMs (FVQ/HMM), thereby improving the recognition performance [29]. Besides, some researchers have attempted to express the elements of the HMM and extend Baum-Welch and Viterbi algorithms with Type1 [30] or Type2 [28] Fuzzy sets. In the same manner, Cheok proposed the generalized fuzzy HMM and showed that, compared to the HMM, this method can achieve the same recognition rate faster [31]. Due to the success of fuzzy HMM, it has been implemented in many applications [32–34]. Nevertheless, elasticity should be considered in fuzzy-based methods in order to deal with large and long datasets of imprecise speech and handwritten data. Elasticity is an inherent property of speech and handwritten data, which means that a part of the input data can stretch or skew based on the speaker/writer's style [35]. According to the authors' knowledge, no fuzzy sequential pattern recognition tool that can successfully handle the elasticity property in speech or handwritten data has been presented yet.

The goal of this paper was to present a fuzzy sequential pattern recognition tool which considers the elasticity property of input data. Unlike most studies that tried to improve the performance of the HMM by changing its structure or combining it with other methods [20–31], we focused on the definition of the basic element of a sequential pattern recognition tool using fuzzy elastic pattern. Since this tool is designed to compare and match input data with the fuzzy elastic pattern, it is called "Fuzzy Elastic Matching Machine" (FEMM).

The rest of this paper is organized as follows: In the second section, fuzzy elastic pattern is presented and its adequacy as an appropriate sequential data model is investigated. The third section is devoted to the introduction of FEMM and its basic elements. To make FEMMs applicable in pattern recognition problems, three basic problems with their solutions are presented for FEMMs in the fourth section. The fifth section discusses the results of implementing FEMM for speech and Persian handwriting recognition. The results and efficiency of FEMM are discussed in the sixth section. Finally, in the seventh section, we draw conclusions and future works are explained briefly.

2. Definition of fuzzy elastic pattern

Fig. 1 demonstrates four samples of a pattern that we modeled using the sequential approach. In this approach, the samples are segmented into three segments that come in a sequence, as shown in Fig. 1. With regard to these samples, the requirements of an appropriate sequential data model are enumerated below. Also, for recognition, the effective elements in measuring the degree of similarity between the input data and the pattern are indicated. First, the features of each segment should be described and the similarity

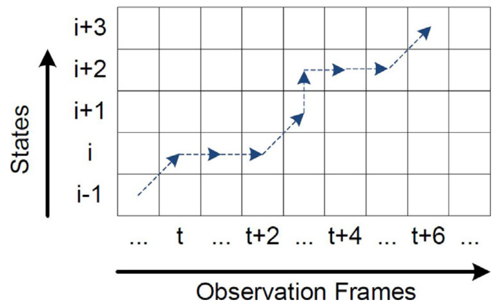

Fig. 2. An illustration of the state assignment of observation frames.

between the input data and the pattern descriptor of each segment should be measured.

Since writers have various writing styles, the pattern descriptor must deal with differences between a writer's style and the defined pattern [35]. Second, although all these samples show the same pattern, each segment has been written with a different length in each sample, indicating the elastic nature of the pattern. Thus, the duration range of each segment should be considered in the modeling of the pattern and similarity measurement. Thirdly, the sequence of segments ($1 \rightarrow 2 \rightarrow 3$) is the same in all the samples, which makes this pattern distinct from others. Therefore, the defined order of the segments should be involved in the procedure of similarity measurement which, in turn, requires an appropriate algorithm for the assignment of the sequence of input data to the segments.

To satisfy these requirements, the elastic pattern is defined as a sequential data model as follows: An elastic pattern is a sequence of several segments in which the length of each segment can skew or sketch in a specific region. The members of each segment have similar features. Each segment is described by a duplex with its first element describing its features, depending on the topic of the problem at hand, which satisfies the first requirement. The second element is used for describing the duration of the segment in the sequence. Thus, the second requirement would be satisfied. If these elements are described with the fuzzy set theory, the corresponding model is called the "Fuzzy Elastic Pattern" [36]. Consequently, the problem of differences between a writer's style and the pattern would be addressed due to the fuzziness of this model. For compatibility with the structure of the HMM, each segment is called a "state".

The satisfaction of the third requirement lies in the procedure of computing the degree of similarity between the input data and a fuzzy elastic pattern. In this procedure, at first the observation sequence of input data should be assigned to the states of the fuzzy elastic pattern, which is explained in detail in Section 4.2.2. Following this step, the degree of similarity is computed according to the membership values of the observations belonging to the fuzzy sets, describing the features and duration of each state. An example of the state assignment for six frames of an observation sequence is illustrated in Fig. 2. Based on this Figure, frames t, $t+1$ and $t+2$ of the observation sequence have been assigned to the ith state. Frame $t+3$ has been assigned to the $i+1$th and $i+2$th states and the observation frames $t+4$ and $t+5$ have been assigned to the $i+2$th state. In this assignment, the durations of the ith, $i+1$th, and $i+2$th states are three frames, half a frame, and two and a half frames, respectively.

To better understand this point, consider the following examples: An elastic pattern for describing a character in handwriting recognition consists of duplexes for each state in which the first element describes the angle of vectors constituting the character in that state, and the second element describes the ratio of the vectors in that state to the whole number of vectors. For instance, consider the "J" character as shown in Fig. 3. This character is segmented into three states according to the movement of hand while writing: top-to-bottom vectors (first state), right-to-left vectors (second state), and bottom-to-top vectors (third state). According to the definition of the elastic pattern, each state can skew or stretch depending on the writer's style, and members of each state have a similar angle. Similarly, the elastic pattern of a word in speech signal consists of duplexes for each state in which the first element describes the acoustic features of that state based on the MFCC features and the second element describes the relative length of that state in the word. Examples of such descriptions can be seen in [29,36] for speech recognition and in [35,36] for handwriting recognition.

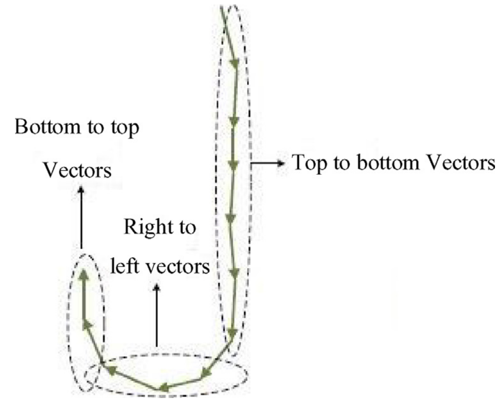

Fig. 3. An example of an elastic pattern for the 'J' character which is segmented into three states.

3. FEMM and its elements

According to the above definitions of fuzzy elastic pattern, FEMM is a Fuzzy sequential pattern recognition tool which compares and matches the input data with a fuzzy elastic pattern.

Basic elements of FEMM describe the pattern and duration of the states. A FEMM for matching the input data (d-dimensional observation vector) $\{O\}$ with a fuzzy elastic pattern with N states consists of the following two fuzzy vectors:

1. The fuzzy vector that describes the state duration $D = \left\{D_i(t)\right\} i = 1, 2, \ldots, N$, where $D_i(t)$ is a fuzzy set for describing the duration of the ith state with t being between 0 and 1.
2. The Fuzzy vector that describes the features of the pattern $P\left\{P_s^m(O_d)\right\}$, where $s = 1, 2, \ldots, N$, $m = 1, 2, \ldots, M$ with M being the number of the membership functions of each fuzzy set. This vector is constituted by several fuzzy sets that describe the observation vector for each state. Each fuzzy set, containing M membership functions, describes a dimension of the d-dimensional observation vector. For compatibility with the structure of the HMM [17], each membership function is called a "mixture". According to the above definition, $P_s^m(O_d)$ is the mth mixture describing the dth dimension of the observation vector O associated with the sth state. To fix the ideas, consider the "J" character in Fig. 3. An example of the Fuzzy Vector P for describing the angle of vectors is constituted by three fuzzy sets, each belonging to one state, as shown in Fig. 4. Each fuzzy set contains three trapezoidal membership functions or mixtures.

The compact notation $\lambda = (D, P)$ is used to show a FEMM. Fig. 5 illustrates the structure of a FEMM for matching an observation sequence with a fuzzy elastic pattern as explained above.

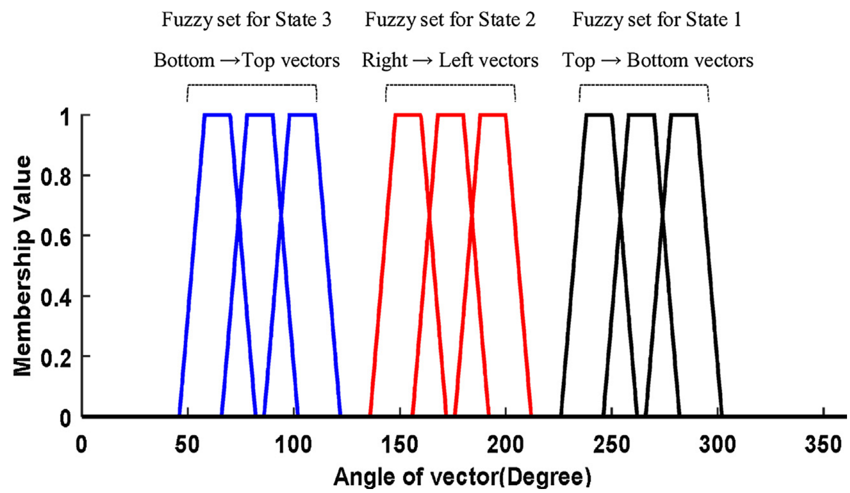

Fig. 4. An example of fuzzy sets describing the angle of vectors in the fuzzy elastic pattern of character "J" (Fig. 3). The Fuzzy set for each state contains three trapezoidal membership functions.

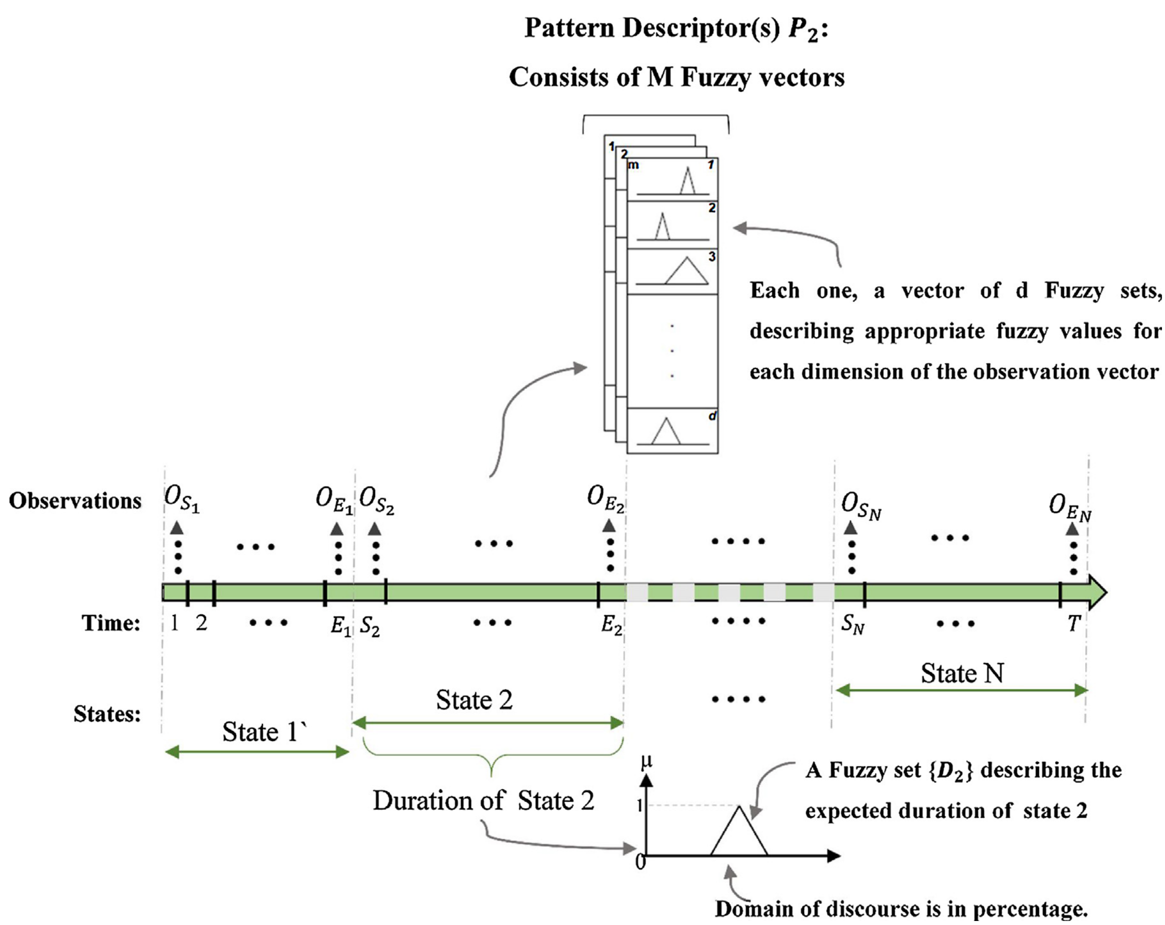

Fig. 5. An illustration of FEMM structure which matches the observation sequence with the fuzzy elastic pattern containing N states: The observations are assigned to the states, e.g. the segment $\{O_{S_2}...O_{E_2}\}$ of the observations has been assigned to the second state. The features of this segment are described by the fuzzy vector P_2. Also, the expected duration of this state is described by the fuzzy set D_2.

4. Three basic problems for FEMMs and their solutions

Since FEMMs are designed for sequential pattern recognition problems, the three basic problems of interest for a sequential pattern recognition such as the HMM [17] should be defined and solved for FEMMs as follows:

4.1. The three basic problems of FEMMs

Problem 1. Given the observation sequence $O = \{O_1O_2..., O_T\}$, the FEMM $\lambda = (D, P)$ and the assignment $\langle S, E\rangle$ between the observations (Fig. 5), how do we compute $\lambda(O, \langle S, E\rangle)$, the degree of membership of the observation sequence to the given FEMM? The solution to this problem allows us to choose the fuzzy elastic pattern which best matches the observation among several competing fuzzy elastic patterns.

Problem 2. Given the observation sequence O= $\{O_1 O_2 \ldots, O_T\}$, and the FEMM $= (D, P)$, how do we choose an optimal state sequence $s_1 s_2 \ldots s_T$ (i.e. maximizes $\lambda(O)$)? The solution to this problem helps us find the optimal state sequence, especially in connected word recognition problems [17].

Problem 3. Given the observation sequence $O = \{O_1 O_2 \ldots, O_T\}$, labels set: $C = \{C_1, C_2, \ldots, C_n\}$, and the assignment function A from the observations to the $[1\ldots n]$ set in which each member of the observations is assigned to one and only one member from the labels set and the FEMM set $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ in which $\lambda_i = (D_i, P_i)$, how do we adjust the parameters of λ_i to maximize the following function?

$$\sum_{i=1\ldots T} \left[A(O_i) = argmax_j \left(\lambda_j(O_i)\right)\right] \tag{1}$$

where [] is an Iversonian bracket which returns 1 if the content is true and returns 0 if the content is false [37]. The solution to this problem leads to optimizing the parameters of FEMMs.

4.2. Solutions to the three basic problems of FEMMs

4.2.1. Solution to the first problem

The degree of membership of the observation sequence $O = O_1 O_2 \ldots O_T$, given the FEMM $\lambda(D.P)$, is computed as follows:

At first, the degree of membership of an observation at the lth frame to the ith state of a FEMM should be defined. The degree of similarity between a d-dimensional observation vector at the lth frame $\{O_l^1 O_l^2 \ldots O_l^d\}$ and the describing pattern of the ith state is defined based on the max-product composition with the aid of T-Norm as the multiplication operator and S-Norm as the adding operator [38] as follows:

$$P_i(O_l) = max_{j=1\ldots M}\left(\Pi_{k=1\ldots d} P_i^j\left({O_l}^k\right)\right) \tag{2}$$

Similar to the HMMs, the degree of membership of the segment of an observation sequence $\{O_{S_i} \ldots O_{E_i}\}$, which has been assigned to the ith state, is proportional to the multiplication of the degree of membership of each observation frame to the pattern descriptor of that state. The state duration should also be considered in this definition. Thus, the degree of membership is defined as follows:

$$P_i\left(O_{S_i} \ldots O_{E_i}\right) = \left(\prod_{t=S_i \ldots E_i} P_i(O_t)\right) * D_i\left({}^{\Delta t}/_{L_o}\right) \tag{3}$$

in which L_o is the length of the whole observation sequence and Δ_t is the length of the observation segment which has been assigned to the ith state.

Finally, the degree of membership of the whole observation sequence to the FEMM $\lambda(D, P)$ is computed as follows:

$$\lambda(O, \langle S_i, E_i \rangle) = \prod_{i=1\ldots n} P_i\left(O_{S_i} \ldots O_{E_i}\right) \tag{4}$$

where O_{S_i} is the first observation frame that is assigned to the ith state and O_{E_i} is the last observation frame which is assigned to the ith state.

4.2.2. Solution to the second problem

Finding the "optimal" assignment of an observation sequence to a FEMM is similar to the procedure used in the HMMs, utilizing the Viterbi algorithm. However, in simple HMMs, the state duration is not considered in the definition of the probability of an observation sequence to an HMM. Although the state duration has been modeled in the HSMMs, it led to complex relations [20]. In FEMMs, finding the optimal state assignment of a given observation sequence for each state involves the following two concepts simultaneously:

1) Degree of similarity between an observation sequence and the features of that state which is measured by the corresponding fuzzy sets (P Fuzzy vector).
2) State duration: the ratio of the length of an observation sequence assigned to a state to the length of the whole observation vector. The state duration is only considered at transition to the next state, which is measured by the corresponding fuzzy set (D Fuzzy vector).

To consider both concepts, instead of using the two-dimensional (2D) matrix in the Viterbi algorithm [17], a three-dimensional matrix (3D) is defined as follows: A 3D matrix V is defined to find the best state sequence, such that $V(s, f, t)$ is the maximum membership value of an observation sequence $\{O_1 O_2 \ldots O_f\}$ to a single path from the first state to the sth state, such that t % of the observation frames belong to the sth state up to the fth frame.

To retrieve the state sequence, a 3D matrix PS is defined, such that $PS(s, f, t)$ is the previous state in which $V(s, f, t)$ has been maximized for each f and t.

The elements of the V matrix are computed using dynamic programming with an algorithm similar to the Viterbi algorithm which is used in the HMMs [17]. Consider the d-dimensional observation sequence with the length L as $\{O_1 O_2 \ldots O_L\}$ $whereO_i = \{O_i^1 O_i^2 \ldots O_i^d\}$. The elements of the V and PS matrices are initialized for the first frame and all the states as follows:

$$V(s, 1, {}^1\!/_L \times s) = (\prod_{i=1\ldots s} P_i(O_1)) \times (\prod_{i=1\ldots s-1} D_i({}^1\!/_L \times s)), s > 1 \tag{5}$$

The elements of the V and PS matrices for other frames are computed by relations (6–8) in a recursive manner, depending on their corresponding assumed condition. Each of these relations computes new values for the elements of the V and PS matrices for each triple (s, f, t), where $1 < f < L-1$, $1 < s < N-1$, and the derived value is replaced if it is greater than the current $V(s, f, t)$ value. These relations are as follows:

$$\begin{aligned} &Foreach(s, f, t) suchthat V(s, f, t) > 0, s < N, f < L \\ &V(s, f+1, t+{}^1\!/_L) = V(s, f, t) \times P_s(O_{f+1}) \\ &PS(s, f+1, t+{}^1\!/_L) = s \end{aligned} \tag{6}$$

$$\begin{aligned} &\text{For each } (s, f, t) \text{ such that } V(s, f, t) > 0,\ s < N,\ f < L, \\ &V\left(s+i, f+1, {}^1\!/_L \times i\right) = V(s, f, t) \times D_s(t) \times \\ &\left(\prod\nolimits_{j=1\ldots i-1} P_{s+j}\left(O_{f+1}\right) \times D_{s+j}\left({}^1\!/_L \times i\right)\right) \times P_{s+i}\left(O_{f+1}\right) \\ &PS\left(s+i, f+1, {}^1\!/_L \times i\right) = s,\ i \in [1, \cdots, N-s] \end{aligned} \tag{7}$$

$$\begin{aligned} &\text{For each } (s, f, t) \text{ such that } V(s, f, t) > 0, s < N, f < L \\ &V\left(s+i, f+1, {}^1\!/_L \times (i+1)\right) = V(s, f, t) \times D_s\left(t + {}^1\!/_L \times (i+1)\right) \times P_s\left(O_{f+1}\right) \\ &\times \left(\prod\nolimits_{j=1\ldots i-1} P_{s+j}\left(O_{f+1}\right) \times D_{s+j}\left({}^1\!/_L \times (i+1)\right)\right) \times P_{s+i}\left(O_{f+1}\right) \\ &PS\left(s+i, f+1, {}^1\!/_L \times (i+1)\right) = s, i \in [1, \cdots, N-s] \end{aligned} \tag{8}$$

Relation (6) is used when it is assumed that a new frame is assigned to the current state; Relation (7) is employed when it is assumed that a new frame is assigned to the $s+1$th state up to the $s+i$th state; and Relation (8) is suitable for the situation when a new frame is assigned to the sth state up to the $s+ith$ state.

Moreover, the PS matrix is utilized to keep track of the path, such that $PS(s, f, t)$ records the state that has maximized $V(s, f, t)$.

Finally, Relation (9) gives the degree of membership of the whole observation sequence to a FEMM as follows:

$$P(O) = \max(V(N, L, t) \times D_N(t) \quad (9)$$

The algorithm of finding the best state assignment of an observation sequence to a FEMM based on the above relations is demonstrated below:

Algorithm: Finding the best state assignment between an observation sequence and a FEMM

Initialization:

1) Compute the elements of the V and PS matrices for all states and all the time slots at the first frame, using Relation (5).

Recursion:

2) For each frame, compute the elements of V and PS matrices for the next frame and all the states and all the time slots, with corresponding $V(s, f, t)$ being nonzero as follows:
Compute the elements of $V(s, f, t)$ and $PS(s, f, t)$ matrices individually for Relations (6-8) and update $V(s, f, t)$ and $PS(s, f, t)$ if the derived value is greater than current $V(s, f, t)$ value (i.e. we have a higher degree of membership in that situation).

Termination:

3) For each frame f and each state s, beginning from the last frame and the last state, find the maximum degree of membership among all the time slots associated with the (s, f) pair.
Find a path over all the frames using the fact that $PS\{s, f, argmax_t(V(s, f, t))\}$ is the number of the state at the previous frame.

Compute the degree of membership of the whole observation sequence to FEMM using Relation (9).

The pseudo-code of this algorithm is presented in Appendix A.

Remark 1. In computing $V(s, f, t)$, $t \in [0.1]$ for each pair (s, f), t or time quantization slot can be any decimal number between [0,1]. Although more time quantization slots let us enhance the modeling of the elastic nature of input data, it demands a large memory size and increases the computational burden, lowering the time and space efficacy of the algorithm. To solve this problem, the maximum number of time quantization slots is limited for each (s, f) pair according to the "principle of incompatibility" (Zadeh, 1973) that states: "The ability to make the precise and significant statements about the system behavior diminishes until a specific threshold and beyond that, precision and significance are mutually exclusive" [38]. To do so, a new value for $V(s, f, t)$ is computed with relations (6–8) every time and there exist at least Q non-zero elements in V matrix. We select $V(s, f, x)$ and$V(s, f, y)$ such that $|x - y|$ is minimum (i.e. x and y are the nearest time slots) and among them only the element corresponding to the time slot with the higher degree of membership is kept. Therefore, the number of available time slots for each pair (s, f) is kept at Q slots. The optimum value of Q can be empirically adjusted depending on the available memory size, computational power, and required recognition performance.

4.2.3. Solution to the third problem

As a solution to the third problem, we should find the "optimal" parameters of FEMM. The difficulty lies in the definition of the optimal parameters of FEMM. If we define "optimal" parameters to maximize the degree of membership of the observation sequence to FEMMs, similar to the one that has been defined in the HMMs [17], one can simply adjust the parameters of the fuzzy sets, such that the domain of discourse has the maximum membership value (1) to the fuzzy sets. This means that all the training samples, including the negative training samples (samples that do not belong to that FEMM) would have the maximum degree of similarity to that FEMM and, thus, classification error may be even 100%. Therefore, FEMMs are only trainable by minimum classification error methods [39]. As a result, training a FEMM should be performed in the presence of the training samples of other FEMMs. Thus, the training method for FEMMs is proposed based on the segmental K-means (SKM) [40] and First Choice Hill Climbing (FCHC) [41] methods as follows:

First, the center points of all the fuzzy sets of each FEMM are separately adjusted by the Segmental K-Means method, such that the degree of membership of the training samples to each FEMM reaches an appropriate limiting point using the procedure described in Chart 1. Then, other parameters of the fuzzy sets (e.g. for Gaussian membership functions, the width of the membership function) are adjusted in a Hill Climbing procedure for all FEMMs as shown in Chart 2. Termination criteria in the training procedure can be set as achieving an acceptable recognition rate or a predefined number of iterations.

5. Experimental results

FEMM was tested on a large speech dataset which consisted of two smaller datasets. The first dataset consisted of 39 phones from all the DR1, DR2, and DR3 branches of TIMIT dataset [42] for all the speakers. Training data included 60979 samples and test data included 20324 samples. This dataset is called "P39". The second dataset included the 100 most repetitive words of TIMIT dataset. For each word, all the repetitions of the training and test datasets were used, including 21747 samples for training and 7998 samples for test. This dataset is called "W100".

In the acoustic feature extraction level, a speech signal was converted into the MFCCs of 39 dimensions with the first-order Pre-Emphasis filter and 30 ms Hamming windowing. Tests were performed on P39 and W100 datasets in the presence of additive white and babble noises with different signal-to-noise ratios (SNRs). Five state HMMs with three Gaussian mixtures and three state FEMMs with three mixtures for P39 dataset were employed. For the W100 dataset, HMMs with seven states and three Gaussian mixtures and FEMMs with five states and three mixtures were utilized.

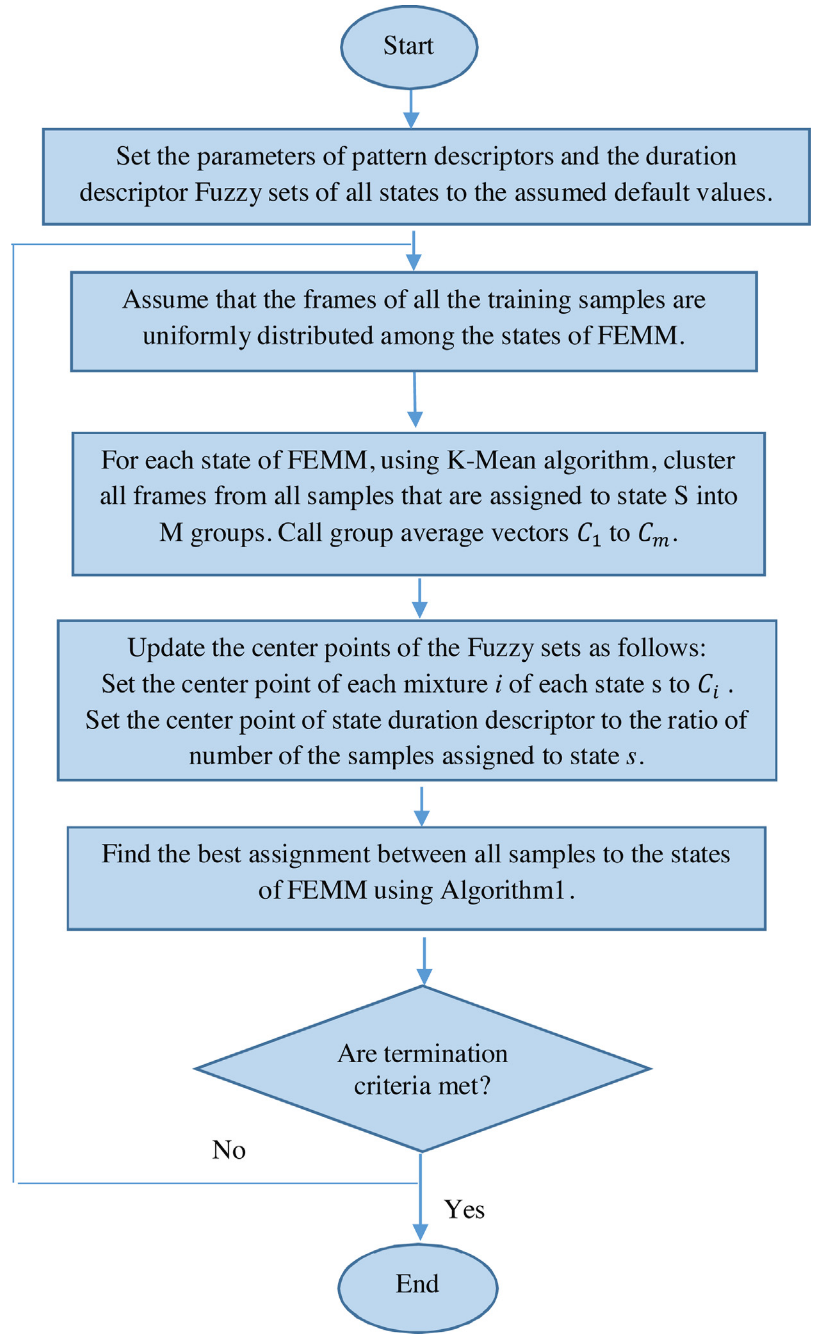

Chart 1. Training the center points of the Fuzzy sets, describing the pattern and duration of each state in a FEMM with the SKM method.

Table 1
Comparison of the results of speech recognition performance between FEMM and the HMM on the P39 dataset.

Method	Phone Recognition Rate (clean data)	Average Phone Recognition Rate for Noisy Data (SNR (0–20db))	Recognition Time (s)
HMM	70.14%	24.3%	62
FEMM trained with SKM method	69.74%	35.6%	41
FEMM trained with SKM + FCHC method	71.05%	37.1%	41

Table 2
Comparison of the results of speech recognition performance between FEMM and the HMM on the W100 dataset.

Method	Word Recognition Rate (clean data)	Average Word recognition Rate for Noisy Data (SNR (0–20db))	Recognition Time (s)
HMM	76.6%	33.8%	454
FEMM trained with SKM method	70.32%	42.5%	381
FEMM trained with SKM + FCHC method	75.51%	43.1%	381

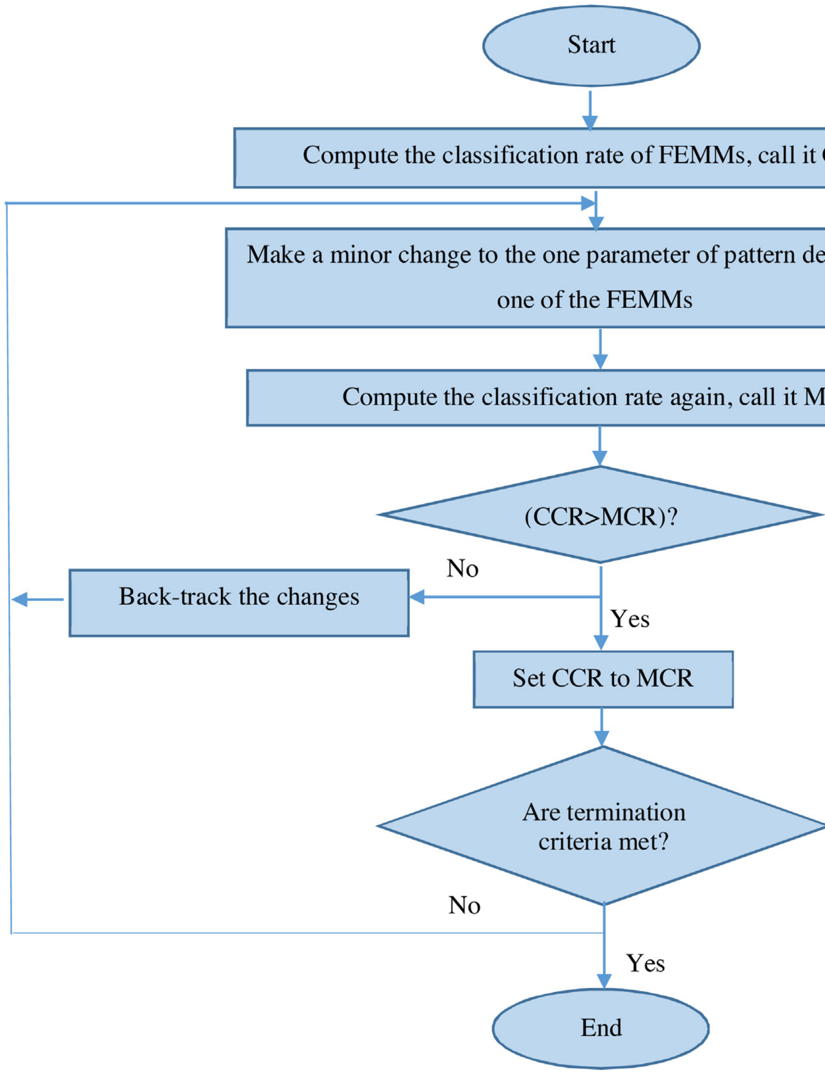

Chart 2. Training the parameters of the fuzzy sets, describing the pattern of each state in FEMMs with the FCHC method.

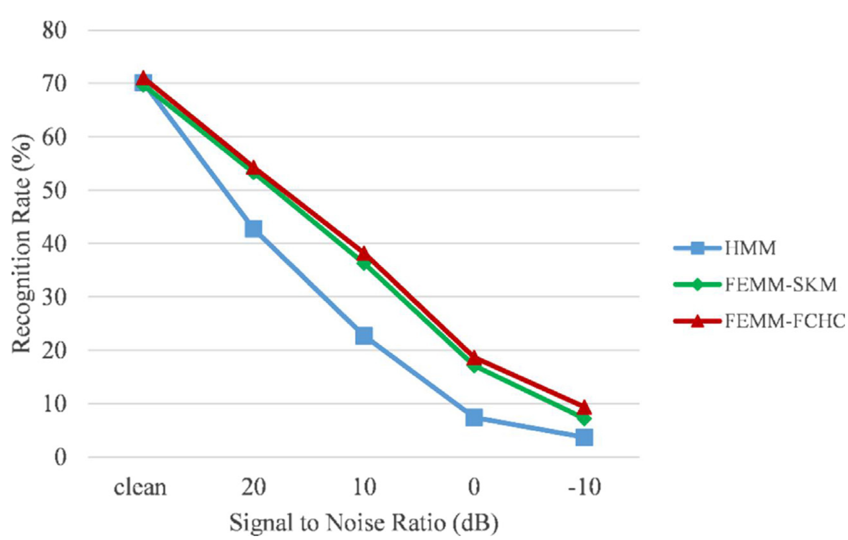

Fig. 6. Comparison of the recognition rates between the HMM, FEMMs trained with SKM method and FEMMs trained with SKM + FCHC method on P39 test dataset versus different SNR ratios of additive white noise.

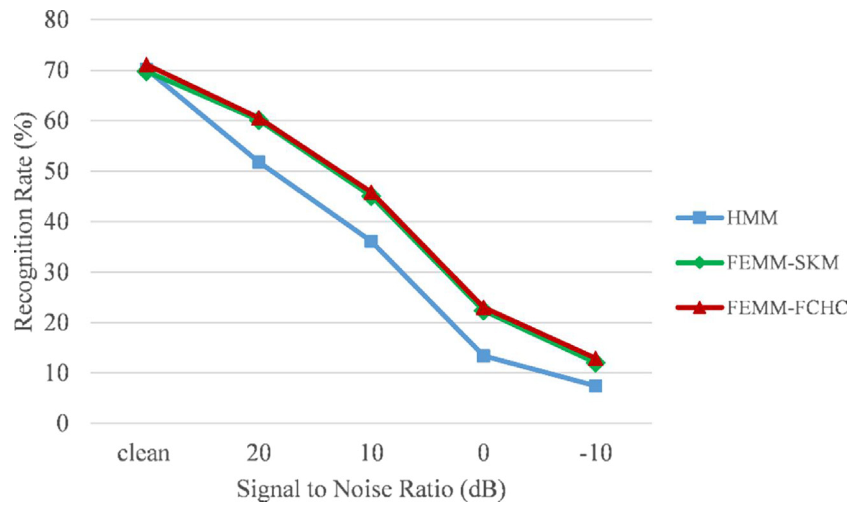

Fig. 7. Comparison of recognition rates between the HMM, FEMMs trained with SKM method and FEMMs trained with SKM + FCHC method on W100 test dataset versus different SNR ratios of additive babble noise.

Table 3
Comparison of handwriting recognition performance between FEMM and the HMM on the N32 dataset.

Method	Recognition Rate	Recognition Time
HMM	70.78%	100s
FEMM trained with SKM method	67.35%	95s
FEMM trained with SKM + FCHC method	75.92%	95s

Clean data were used for training and tests were performed on the noisy data. Tables 1 and 2 compare the recognition rates for clean data, average recognition rates for noisy data, and recognition times of FEMM with the HMM for P39 and W100 datasets, respectively. In addition, Figs. 6–7 compare the recognition rates of FEMM with the HMM for P39 and W100 datasets versus different signal-to-noise ratios of additive white and babble noises, respectively.

In handwriting recognition, Dr. Kabir's discrete Persian words dataset [43] was incorporated. This dataset comprises 5843 samples, including the 18 basic Persian characters, 0–9 digits, comma, dot, and question mark,[1] written by 124 persons and presented in 32 classes. This dataset is called N32.

In feature extraction, the length and the angle of vectors were extracted from the input handwritten samples, with the angles of vectors being in the range of $[-\pi, \pi]$. Without any preprocessing, the length of each vector was normalized with respect to the whole character length, and the angles of the vectors were normalized with respect to the angle between the first and the last points of the character. All the tests were performed ten times and the reported recognition rates are the average of all tests. The N32 dataset was randomly segmented into four groups. Each time we used three groups for training, and the test was performed on the fourth group.[2] The tests were conducted using 10 state FEMMs with 3 mixtures and 27 state HMMs with 3 Gaussian mixtures. Table 3 compares the recognition rates and recognition times of FEMM with the HMM.

6. Discussion

6.1. Effect of FEMM structure on recognition performance

To investigate the effect of FEMM structure on recognition performance, the effect of the number of states on recognition performance was examined. FEMM was tested on P39 dataset with different numbers of states, and the results are presented in Fig. 8.

The results showed that increasing the number of states up to three improved recognition rate, but further increment led to additive learning, even reducing the recognition rate. Thus, the number of states for each problem should be determined empirically.

6.2. Effect of elasticity modeling on recognition performance

To explore the effect of elasticity modeling on recognition performance, FEMM was tested on the W100 dataset with different numbers of maximum time quantization slots Q. Recognition rates and recognition times are presented in Fig. 9. It seemed that the lack of elasticity modeling (Q = 1) caused a 2.5% reduction in recognition rate.

Increasing Q up to 5 time slots improved recognition rate, but recognition rate remained constant with further increment. On the other hand, further increment in Q led to excessive computation time (Fig. 9). As expected, modeling the elastic nature of speech data with FEMM improved the recognition performance, but the maximum number of time quantization slots has to be adjusted, depending on the recognition time which, in turn, gives a flexible structure for choosing the optimal Q or an appropriate number of time slots for modeling the elasticity.

6.3. Comparing the performances of FEMM and the HMM

As depicted in Tables 1 and 2, FEMM has shown approximately the same speech recognition rate with a 15–30% higher recognition speed comparing with the HMM, approving the lower computational cost of FEMM. In handwriting recognition, although Persian has a fully cursive handwriting [35], FEMM which is trained with the SKM-FCHC method has shown a 5% higher recognition rate with approximately the same speed, compared with the HMM. This result indicates that the FCHC training method enhances the recog-

[1] ا، ب، ح، د، ر، س، ص، ط، ع، ف، ق، ک، ل، م، ن، و، ه، ی، ء، ۰. ۱. ۲. ۳. ۴. ۵. ۶. ۷. ۸. ۹،"،"، "."، "؟"

[2] Four-Fold Test

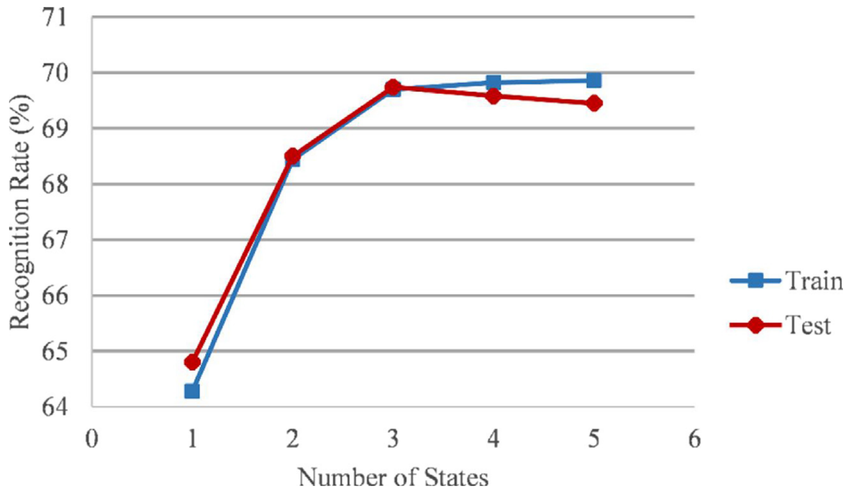

Fig. 8. Speech recognition rate changes versus different number of states for P39 dataset.

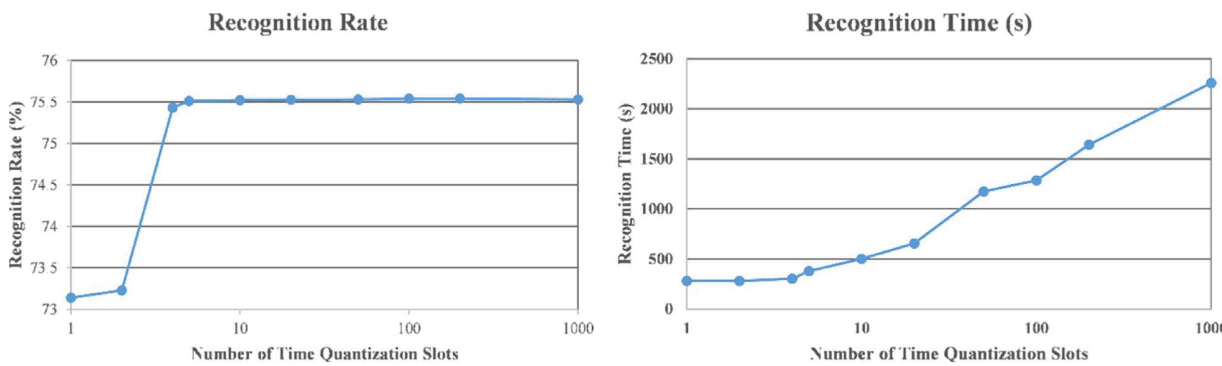

Fig. 9. The recognition rate and time changes versus the different number of time quantization slots for three states FEMMs with three mixtures on the W100 dataset.

nition performance by adjusting the parameters of FEMMs. The speech recognition results in Tables 1 and 2 also prove this issue.

Moreover, as depicted in Figs. 6–7, due to the fuzziness of fuzzy elastic patterns, FEMM has demonstrated 9–16% more immunity against noise in the normal noisy environment (SNR: 20–0 dB). This immunity still remains in noise levels which make the input not identifiable for human beings.

6.4. Comparing the efficiencies of FEMM and the HMM

The efficiency of FEMM can be described and compared with that of the HMM based on the time and space complexity of the proposed algorithms, presented as solutions to the three basic problems in the fourth section. The basic algorithm, which is used in all the subsequent algorithms, is the definition of the similarity between the input data and the fuzzy elastic pattern (see Eq. (2)). For the sake of comparison, in continuous HMMs, this definition is equivalent to emission probabilities which are modeled with Gaussian mixture models (GMMs) [17]. The proposed definition is more computationally effective due to the following reasons: (1) granularity of the fuzzy elastic pattern, which is represented in Eq. (2) by the max operator, similar to the generalized fuzzy HMM [31]; (2) for a sequential pattern with N states and a D-dimensional observation vector with length L, the time complexity of GMM algorithm with M mixtures costs $O\left(N*L*M*D^3\right)$ [44], while the proposed definition (Eq. (2)) costs only $O(N*L*M*D)$. Regarding space complexity, FEMM requires less variable memory, since two matrices (P, D) with the space complexity of $O(N*M*D*W)$ should be stored, where W is the number of necessary parameters for defining the membership functions (e.g. for Gaussian membership functions, W is two (center and width of the membership function)). However, in the continuous HMM, more matrices (mean, covariance, and coefficients of GMMs) with the space complexity of $O\left(N*M*D^2\right)$ should be stored.

Since finding the "optimal" assignment of an observation sequence to FEMM and the HMM is based on the Viterbi algorithm, the time complexity of these methods in the assignment algorithm should be in the same order. Nevertheless, modeling the state duration in FEMM increases time and space complexity from $O(N^2*T)$ and $O(N*T)$ in the HMMs [45] to $O(N^2*T*Q)$ and $O(N*T*Q)$ in FEMMs, respectively, where Q is the maximum time slots for modeling the state duration. Thus, increasing Q lowers the efficiency of the assignment algorithm, especially for large numbers of time slots ($Q > 10$). On the other hand, according to the results in Section 6.2, small numbers of Q ($\sim$3-5) are usually enough to achieve a good recognition performance. In summary, there is a trade-off between less complexity in computing the degree of similarity to FEMM and more complexity in the assignment algorithm of FEMMs. Experimental results in Tables 1–3 proved that, between the two noted factors, the factor of "less time complexity in computing the degree of similarity" is the dominant one, especially when the number of dimensions of input data grows. In this regard, the superiority of the recognition time of FEMM compared to the HMM in speech recognition is more substantial than the one in handwriting recognition (see Tables 1–3), since the number of feature dimensions of a speech signal is much higher than those of a handwritten one.

FEMM was trained using segmental k means and First Hill Choice Climbing (FCHC) algorithms (Section 4.2.3), both based on the Viterbi algorithm. The segmental k means method is more efficient than the Baum-Welch algorithm in training [46]. On the other hand, the time complexity of the FCHC algorithm is dependent on the number of iterations in the algorithm, which can potentially lower the time efficiency of training (if any shortcoming regarding time efficiency is observed, depending on the problem at hand, the FCHC method could be replaced with a more efficient optimization algorithm without changing the structure of FEMM).

7. Conclusion and future works

The HMM has been a dominant tool in sequential pattern recognition problems such as speech and handwriting recognition. Despite the powerful modeling structure of the HMM, its low consistency with the imprecise nature of speech and handwritten data has limited its performance. Although several studies have attempted to surmount these limitations, the problem still persists as no sequential pattern recognition tool is appropriately fast, robust to noisy data, and accurate enough to be an ultimate answer. To have a fast sequential pattern recognizer with a consistent structure, when imprecise input data are given, a fuzzy sequential pattern recognition tool was presented in this paper. We focused on the basic definition of the elements of a pattern recognition tool by introducing the fuzzy elastic pattern. Modeling speech and handwriting with fuzzy elastic patterns gives us the ability to model the skewing or stretching of a part in the input data, which is an inherent property in speech and handwritten data. Since this tool matches the input data with the fuzzy elastic pattern and benefits from the fuzzy linguistic description, it is called the Fuzzy Elastic Matching Machine (FEMM).

In comparison with previous fuzzy logic-based approaches, the elastic matching concept was included in FEMM, leading to a more consistent model with speech and handwritten data. The major difference between FEMM and previous methods which have attempted to consider the elasticity property in the HMM by including the state duration, is a structure devoid of too complex relations with the aid of fuzzy set theory. Consequently, FEMM is expected to have a higher recognition speed and more immunity to noise, compared with the HMMs.

To investigate the performance of FEMM, it was tested in speech and Persian handwriting recognition on large word and phone datasets selected from TMIT dataset for speech recognition and Dr. Kabir's Persian character dataset for handwriting recognition. Results revealed that, in speech recognition, FEMM showed the same recognition accuracy with a 15–30% higher recognition speed compared with the HMM. Furthermore, FEMM demonstrated a 9–16% higher immunity to noise in the presence of different signal-to-noise ratios of additive white and babble noises. In Persian handwriting recognition, FEMM indicated a 5% higher recognition rate compared with the HMM. These results prove that FEMM could be an alternative tool for sequential pattern recognition problems.

As a next step in improving the performance of FEMM, we are working on implementing it in online speech and handwriting recognition. Since sequential pattern recognition is used in many applications such as bio-signal processing, gait recognition, and gesture recognition, using FEMMs in these applications will make the capabilities and potentials of FEMMs more obvious.

Acknowledgements

The authors would like to thank Dr. Ramin Halavati for his helpful suggestions. We also thank Mr. Ashkan Ghanbarzadeh and Ms. Nasim Bagheri Shouraki for their useful comments.

Appendix A.

Inputs: Observation vector $O\{O_1 \ldots O_L\}$ *where* $O_i = \{O_i^1 \ldots O_i^d\}$

Fuzzy vector $P\{P_n^m(O^d)\}$, which describes feature of states

Fuzzy vector $D\{D_i(t)\}$ *where* $i = 1 \ldots N$, which describes duration of states

Outputs: path [1…L], assigned state for each of the observation

Degree of membership of the observation sequence to FEMM

Pseudo Code

1. Create the three dimensional matrices V and PS, set both to zero matrices
2. For each s from 1 to N
   2.1 $V(s, 1, {}^1/_{L \times s}) = (\prod_{i=1\ldots s} P_i(O_1)) \times (\prod_{i=1\ldots s-1} D_i({}^1/_{L \times s}), s > 1)$
   2.2 $PS(s, 1, {}^1/_{L \times s}) = 0$
3. For each f from 1 to L-1
   For each s from 1 to N
   For each t, such that $V(s, f, t) \neq 0$
   3.1. $NewVal = V(s, f, t) \times P_s(O_{f+1})$
   3.2. $NewTime = t + {}^1/_L$
   3.3. $SelfMore(s, f+1, NewTime, NewVal, s)$
   3.4. $For\ each\ i\ from\ 1\ to\ N - s$
   3.4.1. $NewTime = {}^1/_{L \times i}$
   3.4.2. $NewVal = V(s, f, t) \times D_s(t) \times (\prod_{j=1\ldots i-1} P_{s+j}(O_{f+1}) \times D_{s+j}(NewTime)) \times P_{s+i}(O_{f+1})$
   3.4.3. $SelfMore(s+i, f+1, NewTime, NewVal, s)$
   3.4.4. $NewTime = {}^1/_{L} \times (i+1)$
   3.4.5. $NewVal = V(s, f,) \times D_s(t + NewTime) \times P_s(O_{f+1}) \times (\prod_{j=1\ldots i-1} P_{s+j}(O_{f+1}) \times D_{s+j}(NewTime)) \times P_{s+i}(O_{f+1})$
   3.4.6. $SelfMore(s+i, f+1, NewTime, NewVal, s)$
4. $For\ each\ t\ such\ that\ V(N.L-1.t) \neq 0$
   4.1. $V(N, L-1, t) = V(N, L-1, t) \times D_N(t)$
5. $CurrentState = N$
6. $For\ each\ f\ from\ L\ to\ 1$
   6.1. $BestTime = argmax_t V(CurrentState, f, t)$
   6.2. $PreviousState = PS(CurrentState, CurrentFrame, BestTime)$
   6.3. $Path(f) = CurrentState$
   6.4. $CurrentState = PreviousState$
7. $degree\ of\ membership\ of\ observation\ to\ FEMM = max_t(V(N, L, t))$

The pseudo code of state assignment of observation in FEMM.

Sub function: $SelfMore(s, f, t, Value, PreviousState)$

1. Set T to the set of times as t', such that $V(s, f, t')$ has a non-zero value.
2. If $(t \in T \; and \; Value > V(s, f, t))or(|T| < Q)$
   - 2.1. $V(s, f, t) \leftarrow Value$
   - 2.2. $PS(s, f, t) \leftarrow PreviousState$
   - 2.3. Exit
1. Put t in T
2. Store $V(s, f, t) \leftarrow Value$
3. Store $PS(s, f, t) \leftarrow PreviousState$
4. Find x and y in T such that $|x - y|$ is minimum
5. $if \; V(s, f, x) < V(s, f, y) \; Then \; swap \; x \; and \; y.$
6. $Remove \; V(s, f, y) \; and \; PS(s, f, y)$

The SelfMore sub function used to find the best state assignment pseudo code, which updates values of the V and PS matrices.

References

[1] Christopher M. Bishop, Pattern recognition, Mach. Learn. 128 (2006).

[2] M. Georgi, C. Amma, T. Schultz, Recognizing hand and finger gestures with IMU based motion and EMG based muscle activity sensing, Biosignals (2015) 99–108.

[3] S. Bhoi, D.P. Dogra, P.P. Roy, Handwritten text recognition in odia script using hidden markov model, in: 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), IEEE, December, 2015, pp. 1–3.

[4] O. Samanta, U. Bhattacharya, S.K. Parui, Smoothing of HMM parameters for efficient recognition of online handwriting, Pattern Recogn. 47 (11) (2014) 3614–3629.

[5] A. Kuamr, M. Dua, T. Choudhary, Continuous hindi speech recognition using gaussian mixture HMM, in: 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), IEEE, March, 2014, pp. 1–5.

[6] M. Debyeche, J.P. Haton, A. Houacine, A new vector quantization approach for discrete HMM speech recognition system, Int. J. Comput. 5 (1) (2014) 72–78.

[7] N. Raman, S.J. Maybank, Activity recognition using a supervised non-parametric hierarchical HMM, Neurocomputing 199 (2016) 163–177.

[8] H. Hoettinger, F. Mally, A. Sabo, Activity recognition in surfing-a comparative study between hidden markov model and support vector machine, Procedia Eng. 147 (2016) 912–917.

[9] A.J. Young, A.M. Simon, N.P. Fey, L.J. Hargrove, Intent recognition in a powered lower limb prosthesis using time history information, Ann. Biomed. Eng. 42 (3) (2014) 631–641.

[10] L. Bahl, and Frederick Jelinek. Decoding for channels with insertions, deletions, and substitutions with applications to speech recognition, IEEE Trans. Inf. Theory 21 (4) (1975) 404–411.

[11] T.K. Vintsyuk, Element-wise recognition of continuous speech composed of words from a specified dictionary, Cybern. Syst. Anal. 7 (2) (1971) 361–372.

[12] Richard Bellman, "1Dynamic Programming, 1 Princeton, New Jersey.", 1957.

[13] L. Bahl, et al., Automatic Recognition of Continuously Spoken Sentences from a finite state grammer, IEEE, in: IEEE International Conference on ICASSP'78 Acoustics, Speech, and Signal Processing, vol. 3, 1978.

[14] James K. Baker, Stochastic modeling for automatic speech understanding, in: Readings in Speech Recognition, Morgan Kaufmann Publishers Inc., 1990.

[15] Bruce Lowerre, The HARPY speech understanding system, in: Readings in Speech Recognition, Morgan Kaufmann Publishers Inc., 1990.

[16] S. Glinski, et al., The graph search machine (GSM): A programmable processor for connected word speech recognition and other applications, IEEE, in: IEEE International Conference on ICASSP'87 Acoustics, Speech, and Signal Processing, vol. 12, 1987.

[17] Lawrence R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.

[18] Alain Biem, Minimum classification error training for online handwriting recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (7) (2006) 1041–1051.

[19] A.D. Chan, K.B. Englehart, Continuous myoelectric control for powered prostheses using hidden Markov models, IEEE Trans. Biomed. Eng. 52 (1) (2005) 121–124.

[20] Shun-Zheng. Yu, Hidden semi-Markov models, Artif. Intell. 174 (2) (2010) 215–243.

[21] M. Russell, A. Cook, Experimental evaluation of duration modelling techniques for automatic speech recognition, IEEE, in: IEEE International Conference on ICASSP'87 Acoustics, Speech, and Signal Processing, Vol. 12, 1987.

[22] Martin Russell, A segmental HMM for speech pattern modelling, IEEE, in: ICASSP-93., 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993, Vol. 2, 1993.

[23] M.J. Johnson, A.S. Willsky, Bayesian nonparametric hidden semi-Markov models, J. Mach. Learn. Res. 14 (February) (2013) 673–701.

[24] S. Chiappa, Explicit-duration markov switching models, Found. Trends® Mach. Learn. 7 (6) (2014) 803–886.
[25] S.Z. Yu, H. Kobayashi, An efficient forward-backward algorithm for an explicit-duration hidden Markov model, IEEE Signal Process. Lett. 10 (1) (2003) 11–14.
[26] S.M. Bhatti, M.S. Khan, J. Wuth, F. Huenupan, M. Curi2lem, L. Franco, N.B. Yoma, Automatic detection of volcano-seismic events by modeling state and event duration in hidden Markov models, J. Volcanol. Geotherm. Res. 324 (2016) 134–143.
[27] R. Halavati, S.B. Shouraki, S.H. Zadeh, Recognition of human speech phonemes using a novel Fuzzy approach, Appl. Soft Comput. 7 (3) (2007) 828–839.
[28] J. Zeng, Z.Q. Liu, Type-2 fuzzy hidden moarkov models, in: Type-2 Fuzzy Graphical Models for Pattern Recognition, Springer, Berlin Heidelberg, 2015, pp. 57–83.
[29] H.-P. Tseng, M.J. Sabin, E.A. Lee, Fuzzy vector quantization applied to hidden markov modeling, Proceedings of ICASSP'87 (1987) 641–644.
[30] D. Tran, M. Wagner, Fuzzy hidden Markov models for speech and speaker recognition, in: Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American, IEEE, 1999.
[31] A.D. Cheok, S. Chevalier, M. Kaynak, K. Sengupta, K.C. Chung, Use of a novel generalized fuzzy hidden Markov model for speech recognition, IEEE, in: The 10th IEEE International Conference on Fuzzy Systems, 2001, Vol. 3, 2001, pp. 1207–1210.
[32] S. Sun, J. Yun, H. Lin, N. Zhang, A. Abraham, H. Liu, Granular transfer learning using type-2 fuzzy HMM for text sequence recognition, Neurocomputing 214 (2016) 126–133.
[33] S.M. Darwish, M.M. Madbouly, M.B. Khorsheed, Hand gesture recognition for sign language: a new higher order fuzzy HMM approach, Int. J. Eng. Technol. 8 (3) (2016) 157.
[34] H. Lyu, J. Suraj Nandiganahalli, I. Hwang, Human automation interaction issue detection using a generalized fuzzy hidden markov model, AIAA Information Systems-AIAA Infotech@ Aerospace (2017) 0344.
[35] R. Halavati, S.B. Shouraki, Recognition of Persian online handwriting using elastic Fuzzy pattern recognition, Int. J. Pattern Recognit Artif Intell. 21 (03) (2007) 491–513.
[36] R. Halavati, Automatic Concept Extraction to Improve the Recognition Performance for Sequential Patterns, Sharif University of Technology, Tehran, Iran, 2009 (PHD thesis).
[37] R.L. Graham, D.E. Knuth, O. Patashnik, S. Liu, Concrete mathematics: a foundation for computer science, Comput. Phys. 3 (5) (1989) 106–107.
[38] H.J. Zimmermann, Fuzzy control, in: Fuzzy Set Theory—and Its Applications, Springer, Netherlands, 1996, pp. 203–240.
[39] S. Katagiri, C.H. Lee, B.H. Juang, New discriminative training algorithms based on the generalized probabilistic descent method, in: Proceedings of the 1991 IEEE Workshop Neural Networks for Signal Processing [1991], IEEE, September, 1991, pp. 299–308.
[40] L.R. Rabiner, J.G. Wilpon, B.H. Juang, A segmental k-means training procedure for connected word recognition, AT&T Techn. J. 65 (3) (1986) 21–31.
[41] S.J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, 2nd edition, Prentice-Hall, 2002, pp. 111–118.
[42] W.M. Fisher, G.R. Doddington, K.M. Goudie-Marshall, The DARPA speech recognition research database: specifications and status, in: Proc. DARPA Workshop on Speech Recognition, February, 1986, pp. 93–99.
[43] S.M. Razavi, E. Kabir, A data base for online persian handwritten recognition, in: 6th Conference on Intelligent Systems, Kerman, Iran, 2004.
[44] R.C. Pinto, P.M. Engel, A fast incremental gaussian mixture model, PLoS One 10 (10) (2015) e0139931.
[45] R. Datta, J. Hu, B. Ray, On efficient Viterbi decoding for hidden semi-Markov models, in: 19th International Conference on Pattern Recognition, 2008. ICPR 2008, IEEE, December, 2008, pp. 1–4.
[46] T.K. Bhowmik, J.P. van Oosten, L. Schomaker, Segmental K-means learning with mixture distribution for HMM based handwriting recognition, in: International Conference on Pattern Recognition and Machine Intelligence, Springer Berlin Heidelberg, June, 2011, pp. 432–439.