

The Construction of Piano Teaching Innovation Model Based on Full-depth Learning

<https://doi.org/10.3991/ijet.v13i03.8369>

Anshi Wei

Baoji University of Arts and Sciences, Baoji, China

laowei135@163.com

Abstract—This paper presents a new method of building piano teaching innovation model based on full depth learning. The model includes the following main steps: (1) The normal behavior samples of piano teaching are obtained by the method of spectral clustering based on dynamic time homing (DTW), and the hidden Markov model; (2) to further train the hidden Markov model parameters in a large sample by means of iterative learning; (3) to use the maximum a posteriori (MAP) adaptive method to estimate the Hidden Markov Model (HMM) of the piano teaching behavior in a supervised manner; (4) The behavioral hidden Markov topology model is established for model estimation. The main features of this method are: it can automatically select the kinds and samples of the normal behavior patterns of piano teaching to establish an innovative model of piano teaching; the problem of under-learning of Hidden Markov Model (HMM) can be avoided in the case of fewer samples. The experimental results show that this model is more reliable than other methods.

Keywords—piano teaching innovation modeling; full depth learning; hidden Markov model

1 Introduction

Human behavior analysis has a wide application prospect and potential economic value in security monitoring, advanced human-computer interaction, video conference, behavior-based video retrieval and medical diagnosis. It is a hot research field in computer vision field. The ultimate goal of behavioral analysis is to obtain behavioral semantic description and understanding by analyzing the behavioral characteristics data, and behavior modeling and evaluation is a key step to achieve the above objectives. In many practical applications (such as behavior-based video We will focus on the piano teaching behavior evaluation. Therefore, this paper mainly discusses the construction of piano teaching innovation model, in which we will be a small amount of behavior in the scene is defined as piano teaching behavior, and a large number of recurring The general behavior of piano teaching is defined as normal behavior.

Although traditional modeling methods based on supervised learning [3] [7] can establish accurate behavior models in fixed scenarios, it is necessary to manually mark a large number of behavior sequences to obtain enough training samples, which

will result in a large number of human resources waste (8) .This method can automatically (semi-automatically) the establishment of behavioral models, can reduce the burden on people, while enhancing the applicability of the model. Zelnik-Manor and Irani [10] first established a behavioral event model automatically by multi-time-scale feature and spectral clustering method, which is simple and suitable for event evaluation, retrieval and time-series segmentation in large amounts of data. They also pointed out in [10] that the statistical distance discriminant model based on multi-time-scale features does not have good behavior recognition accuracy [11], and proposes an unsupervised model estimation method, It is not only to establish the behavior model directly, but to identify the piano teaching behavior by clustering analysis and simple similarity measure, so it is only suitable for the model estimation problem in the large amount of data in offline state. Xiang and Gong in [12] Firstly, the hidden Markov model (HMM) of normal behavior of piano teaching is established automatically by means of spectral clustering method. Then, it is judged by the given threshold value piano teaching behavior. This method not only solves the problem of time-varying behavior, but also real-time model estimation. The shortcoming of this method is that in the process of establishing the similarity matrix, the time series fragments of each behavior are directly Learning a HMM, this will be due to lack of learning HMM parameter estimation to bring the problem of unreliability [15]; In addition, in the model estimation using a fixed threshold will miss some of the normal behavior of piano teaching is similar to the piano teaching Behavior. One of the effective ways to solve the missed test is to establish a piano teaching innovation model while also establishing a variety of piano teaching behavior model. Piano teaching behavior of the sample is usually a small amount of, which brings direct to modeling a certain difficulty. [13] proposed a full-depth learning modeling method, that is, first through the method based on supervised learning to establish an innovative model of piano teaching, and then through the iterative adaptive method on this basis, the automatic establishment of piano teaching behavior Model. This method considers the existence of a piano teaching innovation model under the piano teaching behavior modeling, there is no discussion of the existence of a variety of piano teaching innovation model situation; In addition, also did not discuss the optimal number of iterations to determine the piano teaching behavior Model number of the problem.

In this paper, we propose a new method to build an innovative model of piano teaching based on full-depth learning. Unlike the above methods [10-12], this paper adopts dynamic time warping (DTW) we use the normalized distance to establish the similarity matrix, and use it for spectral clustering. DTW can solve the problem of time-varying behavior, is a dynamic programming process. In order to reduce the computational complexity, we randomly extract a large number of samples from the first part of the behavior the samples are classified by DTW-based spectral clustering, and then the HMM of the normal behavior of piano teaching can be further studied in a large sample by the iterative learning method. This method not only can automatically select the types and samples of the normal behavior patterns of piano teaching The HMM of the normal behavior of piano teaching is established to solve the problem of time-varying behavior, but also can effectively avoid the unreliability of HMM parameters estimation in Xiang and the literature [12] .At the same time, (HMM) of

the normal behavior of piano teaching, the HMM of all piano teaching behaviors is obtained by using the maximum a posteriori (MAP) adaptive method, which overcomes the limitation of the literature [13], which only applies to the existence of a piano teaching The normal behavior of the piano teaching model. Finally, through the normal behavior of piano teaching and piano teaching behavior model and the establishment of the piano teaching behavior discrimination model, to avoid the literature [12] the method of undetected problem, thereby reducing the error rate.

In the following sections, the key steps of the model are described in detail.

2 Time Series Segmentation and Representation

In the time-series segmentation, there are usually three methods: (1) according to the signal breakpoint segmentation; (2) according to the signal mutation point segmentation; (3) according to the overlap of the time window is divided according to different Application scenarios, you can choose a different segmentation method.

A continuous video sequence V is divided into N small segments, that is $V = \{v_1, v_2, \dots, v_n, \dots, v_N\}$. Ideally, each segment after segmentation represents a behavior $D_t = [d_{t1}, d_{t2}, \dots, d_{ti}, \dots, d_{tJ}]$. A feature vector is extracted for each data (where D_t is the eigenvector of the t -th data, d_{ti} ($1 \leq i \leq J$) is the eigenvector of the t -th data of the i -th Characteristics of the component, and J is the number of the eigen-components), each sequence can be expressed as $v_n = \{D_{n1}, D_{n2}, \dots, D_{nm}, \dots, D_{nM}\}$ (where: D_{nm} denotes the eigenvector of the n -th data of the m th video sequence; M denotes the length of the video sequence of the Nth segment.

In this paper, the image segmentation method is used to segment the sequence, and the feature extraction is performed by using the temporal and spatial filtering method. For details, see the experimental part.

3 Establishment of Innovative Model of Piano Teaching

DTW and HMM are often used in time series modeling [1, 16]. They can solve the problem of time-varying behavior well. However, DTW needs to establish a template for each kind of behavior, is a dynamic programming HMMM has more advantages than DTW in terms of learning ability and processing of undivided data streams [1, 2], but it needs to consider the initialization problem in use, because the HMM in a small number of samples, The direct learning HMM is a learning problem [15], and the obtained parameters are generally unreliable. The method of this paper makes full use of their respective characteristics, and puts forward the following modeling methods: Firstly, using spectral clustering based on DTW Then, the HMMs were selected by the threshold model [16], and then the HMMs were trained by the iterative learning method. Then the HMMs were trained by the iterative learning method.

In order to reduce the computational complexity, we randomly extract the H segment (usually $H = N / 10$) from the segment N video sequence for spectral clustering, the set of extracted data is denoted by $V' = (v'_1, v'_2, \dots, v'_H)$, and the remaining set of sample sequences is called A .

3.1 HMM Form of Spectral Clustering and Behavior Based on

Given the two time series signals Q and P (let Q be m and P be n), we use the normalized distance to evaluate the similarity between two time series signals

$$s = \exp[-\gamma(m, n) / G\sigma] \quad (1)$$

Here $\gamma(m, n)$ is the minimum cumulative distance of two time series signals Q and S obtained by DTW, G is the optimal path length obtained by DTW, and the range is $\max(m, n) \leq G \leq m + n - 1$, σ is the constant factor

For a randomly selected segment H , we can obtain a similarity symmetry matrix $S = [s_{mn}]_{H \times H}$ of $H \times H$ according to the similarity distance calculation formula (1).

Let S be the eigenvalue of $\lambda_i (i \in [1, H], \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_H)$, the variance contribution rate μ is defined as

$$\mu = \sum_{i=1}^L \lambda_i / \sum_{j=1}^H \lambda_j (L \leq H) \quad (2)$$

Select the variance contribution rate greater than T_μ (here in this passage, we use $T_\mu = 0.8$) and choose the L as the initial clustering number and use the method of spectral clustering (refer to [17]) to cluster the H segment into L class, and take the sampling The class $T (T = H / 10)$, which is greater than $C (C \leq L$ in this paper), $W_i (i = 1, 2, 3, \dots, C)$ is taken as the behavioral event (piano teaching normal behavior) in the time series.

The HMM is constructed by using the HMM of each class in each class. Each HMM contains N hidden nodes. The output probability density function of each hidden node is Gaussian mixture function \square

$$p(y | \alpha, \theta) = \sum_{k=1}^K \alpha_k p_k(y | \mu_k, \Sigma_k) \quad (3)$$

Where: \mathcal{Y} is a D -dimensional random variable whose components are independent of each other; $p_k(\mathcal{Y} | \mu_k, \Sigma_k)$ is the Gaussian density function of the k component with the mean μ_k and the covariance Σ_k ; θ is the vector of the parameters of the component density function, including the mean μ of each Gaussian density function, the covariance Σ ; $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k) \left(\sum_{k=1}^k \alpha_k = 1 \right)$ is the mixture coefficient vector, α_k represents the probability that the k th component is selected.

We use BIC (Bayesian information criterion) [18] to automatically determine the number of components of mixed Gaussian:

$$\hat{K} = \arg \min_K \left\{ - \sum_{i=1}^T \log p(y_i | \alpha, \hat{\theta}(K)) + \frac{\xi(K)}{2} \log T \right\} \quad (4)$$

Where: A represents the length of the entire time series; B represents the number of parameters of the mixed model.

3.2 Threshold model

C threshold model as shown in Fig. 1 is constructed using the initialized HMMs in the previous section [14,16]. It is an HMM with each state traversal. The concrete construction process is as follows: All hidden nodes of C HMM are taken as the threshold model, the number of hidden nodes is $N = e * C$, the initial probability of each state is set to $1/N$, and the self-transition probability and the output probability of each hidden node are kept unchanged in the new model. The probability of transition between hidden states is given by:

$$a_{ij} = (1 - a_{ii}) / (N - 1) \quad (i \neq j) \quad (5)$$

Here a_{ii} is the self-transition probability of the hidden node i . The two black dots in Figure 1 denote the virtual start node and the end node, respectively, and they have no output and the transition probability is 1.

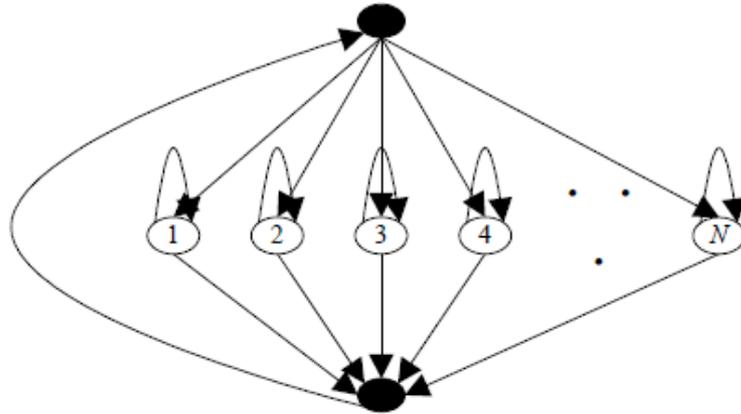


Fig. 1. Threshold model structure

Suppose that the current observation sequence is Y and the parameter set of the i HMM is $\lambda_i (i = 1, 2, 3, \dots, C)$ λ_i includes the initial state distribution vector, state transition probability matrix and output probability distribution matrix of the i th HMM), the parameter of the threshold model is λ_τ if Y belongs to Class i , then

$$i = \max_j \{P(Y | \lambda_j) | P(Y | \lambda_j) > P(Y | \lambda_\tau)\} \quad (j = 1, 2, 3, \dots, C) \quad (6)$$

Here $P(Y | \lambda_j)$ is the probability that the j -th HMM produces the observed sequence Y ; $P(Y | \lambda_\tau)$ is the probability that the threshold model generates the observed sequence Y .

3.3 Iterative learning

By using the DTW-based clustering method, we obtain some sample sequences which reflect the class behavior of the long-term sequence information. We use these samples to initialize the HMM of the normal behavior of each class. As the number of these samples is far less In order to enhance the robustness of each HMM and make the HMM model more accurately reflect the main information of time series, we use the iterative learning method to further train each HMM, the main process is as follows:

1. Through the threshold model and the established HMM of the normal behavior of all kinds of piano teaching, judge the class of each sample sequence in A , and calculate the probability of generating the sample sequence through the corresponding HMM. Finally, the number of samples of normal behavior is $n_i (i = 1, 2, \dots, C)$

2. Take the maximum probability $n_i/5$ of each class from A (if $n_i < T'$, then stop extracting the sample, where T' is the given threshold), and put the sample sequence into the corresponding W_i respectively;
3. Using all the sample sequences in W_i , taking the existing parameters of each HMM as the initial values, and further training the HMM parameters through the EM model;
4. The parameters of the threshold model are updated according to the parameters obtained in (3) ;
5. Using the updated threshold model and various HMMs, we can judge whether there is sequence samples in A which meet the normal behavior of all kinds of piano teaching: if there is, go to (1); otherwise, stop.

3.4 A HMM Topological Structure Model of Piano Teaching Normal Behavior

After obtaining the C HMM model, we design a HMM topology model as shown in Figure 2. Figure 2 shows an HMM topology with $C + 2$ nodes, and the two black dots represent Virtual start node and end node, the transition probability between two nodes is 1. Each HMM of the normal behavior of the piano teaching as a state of the topology, each state has the same probability of transition $1/C$ between.

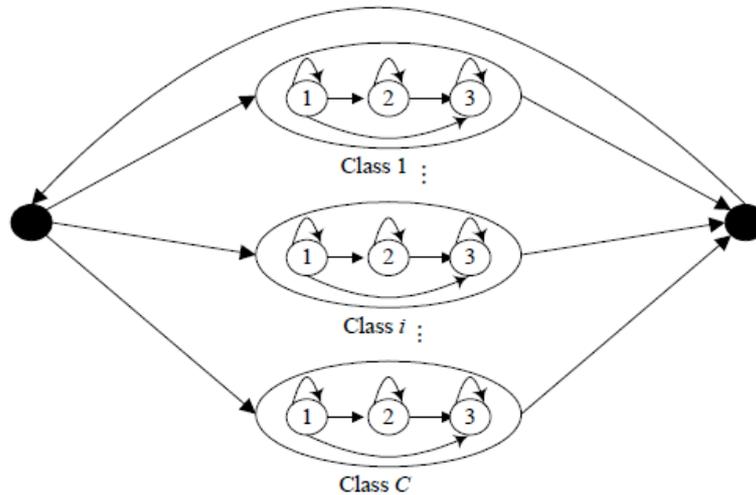


Fig. 2. piano teaching innovation model structures

4 Experimental results

In this paper, the average error rate (HTER) (half-total error rate) is used as the evaluation criterion [13]:

$$HTER = \frac{FAR + FRR}{2},$$

The false alarm rate (FAR) is the error rate of the piano teaching behavior. The false rejection rate (FRR) is the error rate of rejecting the piano teaching behavior. The calculation formula is:

$$FAR = \frac{\text{Number of false alarms}}{\text{Number of normal behavior samples}} \times 100\%,$$

$$FRR = \frac{\text{Number of rejects}}{\text{The number of samples for abnormal behavior}} \times 100\%.$$

4.1 Experimental data

The behavioral analysis models reported in the literature are usually evaluated on their own small-scale databases. We collected 21,327 data from 5 data collected from the collected data (of which: 7584 in the previous review, "Teaching", "teaching", "listening", "listening", "teaching", "listening", "listening", "listening", ") (1) First, the time series of each behavior is divided into two parts without overlapping: the time series of normal behavior of piano teaching is taken as the time series of each action. A random integer greater than 100 and less than 400 is used as the segment length value of each segmentation, and when the time series is not more than 100, the segmentation is stopped; for the time series of piano teaching behavior, As the segment length value of each segmentation, when the time series length is not more than 30, stop the segmentation; (2) all the behavioral fragments are randomly sorted to obtain the mixed behavior of the time series

The training sequence and the test sample sequence were obtained through the above synthetic process. In the experiment, eight different training sample sequences were used: keep the normal behavior of the piano teaching "review the front content", "sense teaching songs" 150, 350, 500, 700, 1000, 1300, and 1500, respectively, in the piano teaching practice "Understanding the Song", "Measuring Teaching Effect" and "Measuring and Evaluating Piano Teaching" The test samples consisted of 2584 normal behaviors, 3043 songs of perceptual teaching, and 400 of piano teaching behaviors, such as understanding songs, evaluating teaching effects and evaluating piano teaching .In the experiment, The time series data obtained by the above method were tested in different parameters and four kinds of correlation models, respectively.

4.2 Clustering and Parameter Selection

Figure 3 shows the clustering and parameter selection results.

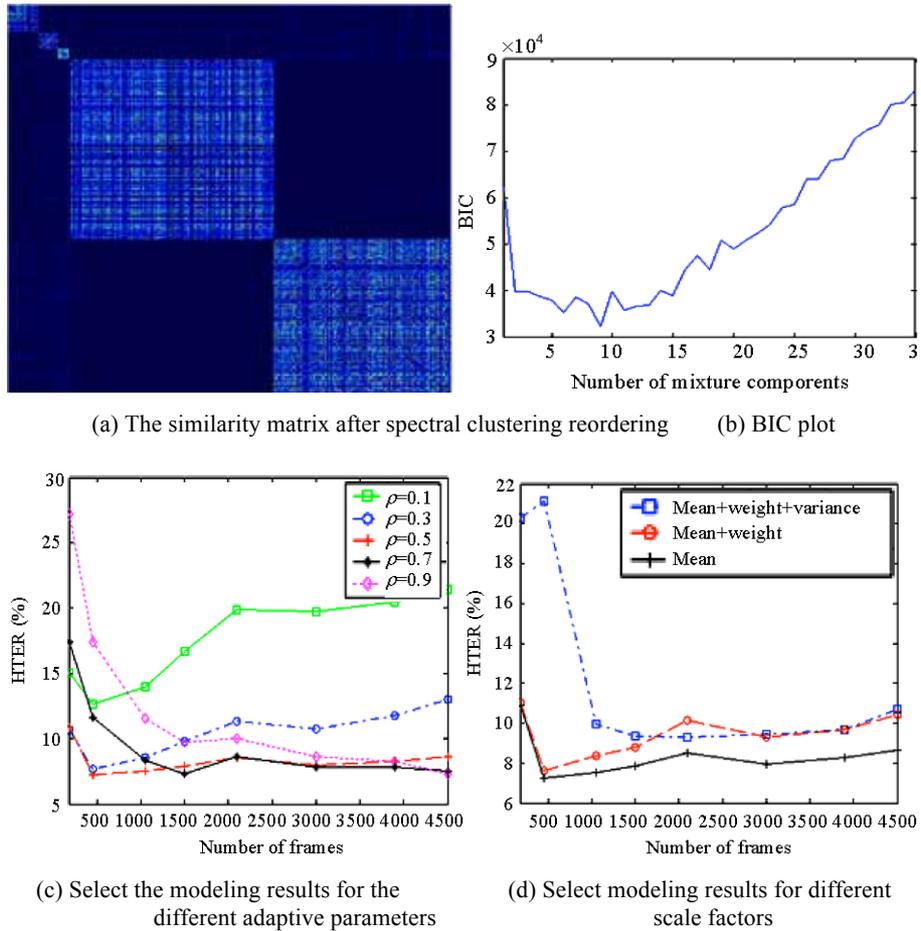


Fig. 3. Cluster and parameter selection results

Figure 3 (a) shows the training sample sequence synthesized by the method in Section 3.1 (including the piano teaching behavior sample "understanding song", "evaluation teaching effect", "evaluation of piano teaching" 350) by segmentation, sampling, spectrum Clustering and reordering. The main goal of this procedure is to get the right amount of normal behavior samples of piano teaching. It can be seen that the training set is clustered into five categories, of which: the largest sample of the two categories, respectively, that the two And the other three categories mainly include piano teaching behavior samples and some misclassified samples. Fig.3 (b) is the BIC curve obtained by using equation (6), and the other three types of piano teaching behavior samples Fig.3 (c) and Fig.3 (d) show the model selection of different adaptive param-

eters and scale factors, respectively, when the BIC value reaches the minimum value in $k = 9$. Therefore, the number of Gaussian mixture components selected in the experiment is 9. Fig.3 (c) Results the graphs and abscissa indicate the total number of samples of piano teaching activities (the sample sizes of all piano teaching activities are equal) in the training set, and the average error rate (HTER) in the ordinate. The three curves in Figure 3 based on $\rho = 0.5$ (c) (1) an adaptive HTER curve for the mean and the mixing coefficient (**mean + weight + variance**); (2) an adaptive HTER curve for the mean, the mixing coefficient and the variance (**mean + weight**); (3) The mean error rate, HTER, is only minimized when only the mean is adapted, since adaptation of too many parameters in a small number of samples degrades performance. 5 (d) show the average error rate (HTER) curves of the three scales, respectively, when the mean value is adaptive, and the scale factor ρ is 0.1,0.3,0.5,0.7,0.9 respectively. , ρ takes a smaller value, a better performance (generally $0.25 \leq \rho \leq 0.5$); Conversely, H take a larger value (generally take $0.7 \leq \rho \leq 0.9$), there is good performance. The reason may be: when the sample is small, the estimation parameters have a greater dependence on the current sample information. In the latter experiment, we choose the scale factor $\rho = 0.5$, only the mean of the adaptive model.

4.3 Model comparison and analysis

We compare this method with the following three methods: (1) Supervised method. All samples are manually labeled, and then each HMM is learned by EM model. (2) Based on full depth learning (HMM), which is used to establish the piano teaching behavior model automatically, and then to mark the piano teaching behavior sample by hand, and to learn the HMM of the piano teaching behavior through EM model learning; (3) (Xiang Tao and MAP). Using the method of Xiang and Gong [12], the HMM of normal behavior of piano teaching is established automatically, then the sample of piano teaching behavior is marked by hand, and the piano teaching behavior is obtained by MAP adaptive method. HMM.

Figure 4 shows the comparison curves of the various models, the horizontal and vertical coordinates of the meaning of Figure 3 (c) the same.

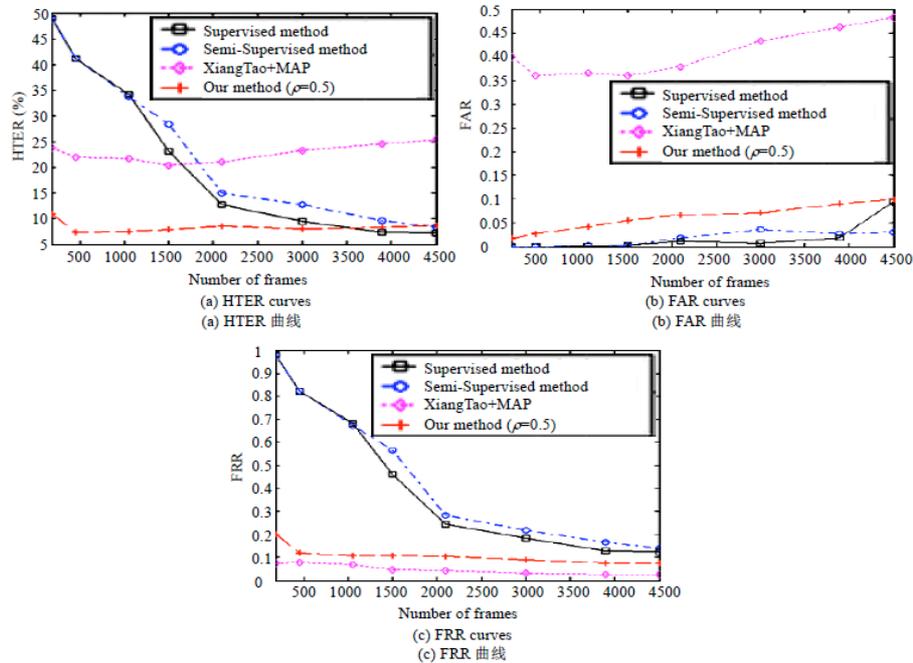


Fig. 4. Comparison of the results of related models

It can be seen from the figure: Method (1), method (2) in a small number of cases of piano teaching behavior of the sample has a higher false rejection rate and average error rate, but with the increase of samples, false rejection rate and average error rate decrease. The reason is that, in the case of a small number of samples, learning EMM directly from the EM model is a problem of under-learning, and the parameters obtained are very unreliable. When there are a large number of samples, the parameters obtained are reliable (3) has high false alarm rate and average error rate, mainly because of the direct use of the HMM distance [12] to build similarity matrix, HMM does not consider a small sample initialization problem. In this paper, the establishment of HMM the paper chooses the appropriate initial sample by DTW-based spectral clustering method, which avoids this problem effectively. At the same time, in the case of a small number of piano teaching behavior samples, the MAP adaptive method is used to establish the piano teaching behavior Of the HMM, so, in the case of a small number of samples can get a good evaluation results.

5 Conclusion and Outlook

This paper presents a method to build an innovative model of piano teaching based on full-depth learning, which can automatically select the normal behavior pattern of piano teaching under the condition of large sample, which can be used to construct the piano teaching innovation model and avoid the lack of learning problem in the case of

a small number of samples and establish a reliable piano teaching behavior model. The experiment also validates the model. The model is suitable for the training set including all kinds of behaviors the future work will focus on the acquisition of more data, the performance of the model testing and research cannot be obtained in advance for each type of piano teaching behavior model of the case of piano teaching innovation model to build the problem.

6 References

- [1] Wang L, Hu WM, Tan TN. Recent developments in human motion analysis. *Pattern Recognition*, 2003,36(3):585–601. [https://doi.org/10.1016/S0031-3203\(02\)00100-0](https://doi.org/10.1016/S0031-3203(02)00100-0)
- [2] Hu WM, Tan TN, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. *IEEE Trans. on Systems, Man, and Cybernetics—Part C*, 2004,34(3):334–352. <https://doi.org/10.1109/TSMCC.2004.829274>
- [3] Haritaoglu I, Harwood D, Davis LS. W4: Real-Time surveillance of people and their activities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000,22(8):809–830. <https://doi.org/10.1109/34.868683>
- [4] Bobick AF, Davis JW. The recognition of human movement using temporal templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001,23(3):257–267. <https://doi.org/10.1109/34.910878>
- [5] Gong SG, Xiang T. Recognition of group activities using dynamic probabilistic networks. In: Werner B, ed. *Proc. of the Int’l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2003. 742–749.
- [6] Li H, Greenspan M. Multi-Scale gesture recognition from time-varying contours. In: Werner B, ed. *Proc. of the Int’l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2005. 236–243.
- [7] Laptev I, Linderberg T. Space-Time interest points. In: Werner B, ed. *Proc. of the Int’l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2003. 432–439. <https://doi.org/10.1109/ICCV.2003.1238378>
- [8] Hu WM, Xie D, Tan TN, Shen J. A hierarchical self-organizing approach for learning the patterns of motion trajectories. *Chinese Journal of Computers*, 2003,26(4):417–426 (in Chinese with English abstract).
- [9] Shechtman E, Irani M. Space-Time behavior based correlation. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Press, 2005. 405–412. <https://doi.org/10.1109/CVPR.2005.328>
- [10] Zelnik-Manor L, Irani M. Event-Based analysis of video. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2001. 123–130. <https://doi.org/10.1109/CVPR.2001.990935>
- [11] Zhong H, Shi JB, Visontai M. Detecting unusual activity in video. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2004. 819–826. <https://doi.org/10.1109/CVPR.2004.1315249>
- [12] Xiang T, Gong SG. Video behaviour profiling and abnormality detection without manual labeling. In: Werner B, ed. *Proc. of the Int’l Conf. on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2005. 1238–1245.
- [13] Zhang D, Gatica-Perez D, Bengio S, McCowan I. Semi-Supervised adapted HMMs for unusual event detection. In: Jacobs A, Baldwin T, eds. *Proc. of the IEEE Conf. on Com-*

- puter Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2005. 611–618. <https://doi.org/10.1109/CVPR.2005.316>
- [14] Ma GY. Temporal sequence clustering and recognition [Ph.D. Thesis]. Beijing: Tsinghua University, 2004 (in Chinese with English abstract).
- [15] Oliver N, Garg A, Horvitz E. Layered representations for learning and inferring office activity from multiple sensory channels. *Computer Vision and Image Understanding*, 2004,96(2):163–180. <https://doi.org/10.1016/j.cviu.2004.02.004>
- [16] Lee HK, Kim JH. An HMM-based threshold model approach for gesture recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1999,21(10):961–973. <https://doi.org/10.1109/34.799904>
- [17] Weiss Y. Segmentation using eigenvectors: A unifying view. In: Werner B, ed. Proc. of the Int'l Conf. on Computer Vision. Los Alamitos: IEEE Computer Society Press, 1999. 975–982. <https://doi.org/10.1109/ICCV.1999.790354>
- [18] Xiang T, Gong SG. Visual learning given sparse data of unknown complexity. In: Werner B, ed. Proc. of the Int'l Conf. on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2005. 701–708. <https://doi.org/10.1109/ICCV.2005.250>

7 Author

Anshi Wei is with the Music College, Baoji University of Arts and Sciences, Baoji, China (laowei135@163.com).

Article submitted 07 February 2018. Final acceptance 23 February 2018. Final version published as submitted by the author.