

## Research Article

# Developments of Machine Learning Schemes for Dynamic Time-Wrapping-Based Speech Recognition

**Ing-Jr Ding, Chih-Ta Yen, and Yen-Ming Hsu**

*Department of Electrical Engineering, National Formosa University, No. 64, Wunhua Road, Huwei Township, Yunlin County 632, Taiwan*

Correspondence should be addressed to Chih-Ta Yen; [chihtayen@gmail.com](mailto:chihtayen@gmail.com)

Received 13 September 2013; Accepted 22 October 2013

Academic Editor: Teen-Hang Meen

Copyright © 2013 Ing-Jr Ding et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a machine learning scheme for dynamic time-wrapping-based (DTW) speech recognition. Two categories of learning strategies, supervised and unsupervised, were developed for DTW. Two supervised learning methods, incremental learning and priority-rejection learning, were proposed in this study. The incremental learning method is conceptually simple but still suffers from a large database of keywords for matching the testing template. The priority-rejection learning method can effectively reduce the matching time with a slight decrease in recognition accuracy. Regarding the unsupervised learning category, an automatic learning approach, called “most-matching learning,” which is based on priority-rejection learning, was developed in this study. Most-matching learning can be used to intelligently choose the appropriate utterances for system learning. The effectiveness and efficiency of all three proposed machine-learning approaches for DTW were demonstrated using keyword speech recognition experiments.

## 1. Introduction

Vocal and visual information that can be used as communication media to allow machines to interact with people has attracted considerable attention in the development of intelligent human-machine interaction devices [1]. Regarding the vocal aspect, the object of the information process is the voice uttered by speakers. The machines require equipment, such as a microphone (or an array of microphones that are widely used in mobile devices), to capture speech sequences that serve as the audio input from which the configuration of the surroundings, and even the status or situation reflected by the context, must be established such that the machine is claimed to be able to listen. A series of analyses must then be performed on the collected audio streams to enable speech recognition.

Automatic speech recognition (ASR) techniques have been widely used in numerous practical applications in recent years [2]. With the maturity and growth of handheld smartphone device applications, the ASR function is attracting much attention and becoming an essential application program in most mobile equipment. ASR also plays a key role in the field of speech processing. Considering ASR, proper uses

of speech pattern recognition techniques, such as speaker recognition, speaker verification, and audio event detection and classification, are being determined in the industry chain.

ASR techniques are classified into two categories: model-based and feature-based methods. Model-based speech recognition involves using a statistical model for recognizing the input testing utterance produced by a speaker. The hidden Markov model (HMM) [3], artificial neural network (ANN) [4], and support vector machine (SVM) [5] are frequently used computational models for performing speech recognition tasks. By contrast, feature-based speech recognition does not involve adopting a statistical model. Establishing (or training) a classification model in advance is not required for feature-based speech recognition, and therefore, this method is generally considered a conceptually simple and direct recognition technique. Dynamic time-wrapping (DTW), which belongs to the dynamic programming category, is a type of feature-based speech recognition [6]. DTW is essentially a type of optimal algorithm and has been widely used to solve numerous optimal problems, including speech recognition. Although DTW has only recently been used for speech recognition, DTW is still a prevalent and indispensable

technique because of its simplicity and inexpensive computation [7].

DTW is one of the mainstream techniques used in speech recognition, and related studies on improving DTW speech recognition have been conducted in recent years [8–11]. Most of these DTW-related studies have either developed improved template-matching algorithms [8, 9] or provided modified schemes for a DTW operation optimization framework [10, 11] for increasing the robustness of the recognition system. In [8], a partial sequence-matching method that involves using an unbounded DTW algorithm was proposed. In [9], the effectiveness of an improved end-point detection algorithm with reduced start- and end-points was validated using simulations. In the study conducted by [10], a feedback method for establishing a database of matching templates was presented. Chen et al. [11] systematically analyzed an optimal warping window size for DTW. Although several studies on improving the performance of DTW speech recognition have been conducted, developing DTW machine learning schemes by using utterances produced by a test speaker for tuning the recognition system is rare. Speaker learning functions for speech recognition, including DTW recognition, are crucial and necessary. Uttered voice data from a test speaker produce abundant information for adjusting the recognition system. By constantly tuning the DTW speech recognition system according to the utterances obtained from a test speaker, the system becomes speaker dependent and can maintain satisfactory recognition accuracy even when encountering unknown speakers. In general, speaker learning techniques for ASR are adopted in model-based speech recognition, particularly in HMM speech recognition. In HMM speech recognition, machine learning is also known as speaker adaptation [12–15]. However, these speaker learning methods are rarely observed in the field of feature-based speech recognition techniques. DTW, which is considered the representative of feature-based speech recognition techniques, displays an increase in recognition performance with well-designed speaker learning schemes.

Thus, machine learning schemes for DTW speech recognition were developed by using uttered data obtained from a test speaker. Supervised and unsupervised learning methodologies for DTW speech recognition are thoroughly explored in this paper. Regarding supervised learning, two learning methods for DTW were proposed in this study: incremental learning and priority-rejection learning. Regarding unsupervised learning, the most-matching learning method was developed, which extends the supervised priority-rejection learning to include a double-checking processing procedure of automatically verifying learning data. In summary, the three proposed machine learning methods for DTW speech recognition in this study have several advantages compared with those without the following

- (i) DTW speech recognition can be combined with system learning using data derived from a test speaker,
- (ii) speaker-dependent DTW behaves similarly to the HMM model-based technique, and

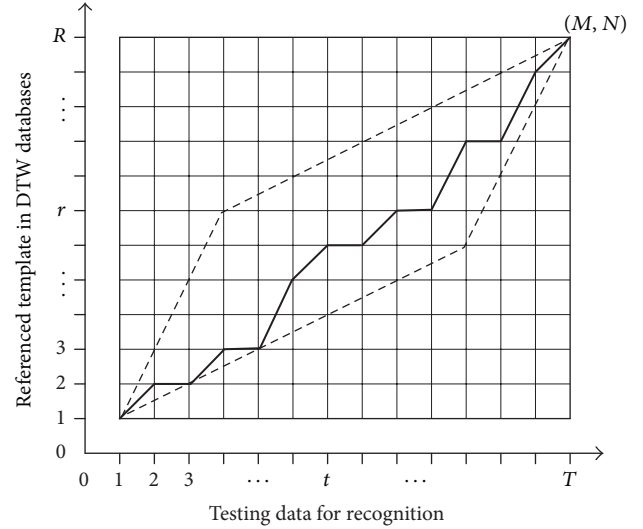


FIGURE 1: Template-matching operations on DTW speech recognition.

- (iii) additional robustness is provided to be adaptive in ordinary recognition environments, such as encountering an unknown test speaker.

## 2. DTW Speech Recognition

This section presents the conventional DTW speech recognition procedure without using any learning strategies. DTW, categorized into dynamic programming techniques, is a nonlinear warping algorithm that combines time-warping and appropriate template-matching calculations [6]. Figure 1 illustrates how the DTW algorithm was used to search for an optimal path between the testing data and the reference template. As illustrated in Figure 1, when computing the similarity degree between the testing data and the reference template, the low distortion between the two of them suggests a high similarity degree. The operation of DTW speech recognition is subsequently explained in this section. The testing utterance is composed of  $T$  frames and an arbitrary frame (a feature vector), denoted by  $t$ . The reference template consists of  $R$  frames and the arbitrary frame, indicated as  $r$ . The distortion between the  $T$  and  $R$  frames can be represented as  $d[T(t), R(r)]$ . The starting-point is  $(T(1), R(1)) = (1, 1)$  and the end-point is  $(T(M), R(M)) = (T, R)$ . Based on these DTW operational settings, the DTW distance from the optimal comparison path can be derived using (1). The arbitrary frame  $t$  in the testing data is generally not equal to the arbitrary frame  $r$  in the indices reference template

$$(\text{optimal}) D = \min \sum_{m=1}^M d(T(m), R(m)). \quad (1)$$

Assuming that the point  $(T(0), R(0)) = (0, 0)$  and  $d(0, 0) = 0$ , the accumulated distance that selects the optimal source path can be represented as follows:

$$\min D(t, r) = \min_{(t-1, r-1)} \{ \min D(t-1, r-1) + d(t, r) \}, \quad (2)$$

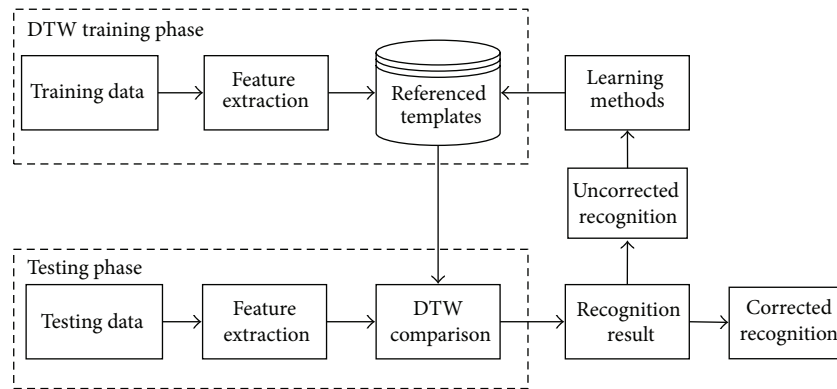


FIGURE 2: Frameworks of DTW speech recognition with machine learning.

where  $\min D(t, r)$  is the shortest distance from the starting position to position  $(t, r)$ . In Figure 1, the solid, black line represents the DTW optimal matching path with the distance derived using (2). The dotted line is the global path search constraint that was used to effectively reduce the searching time for acquiring an overall optimal path on DTW operations.

### 3. Proposed Machine Learning Approaches for DTW Speech Recognition

The DTW technique cannot maintain satisfactory recognition performance levels in an ordinary recognition testing environment in which the uttered data from a test speaker is unmatched to the recognition system. Performing machine learning on DTW recognition may effectively resolve this phenomenon. Figure 2 illustrates the DTW recognition procedure combined with machine learning. As observed in Figure 2, the main contribution of this study is developing a method that can be used to continually adjust the DTW recognition system to become familiar with a speaker and then achieve outstanding recognition performance. The following subsection presents the proposed learning methods of DTW: incremental learning, priority-rejection learning, and most-matching learning.

**3.1. Incremental Learning.** The proposed incremental learning method for DTW is a supervised learning strategy. The supervisor (usually a system developer) monitors the overall speech recognition process. The system supervisor decides whether the test utterance should be returned to the DTW recognition system for learning according to the DTW recognition scores. The parameter  $D$ , which denotes the distortion of the entire comparison path in (1), is used to evaluate the DTW recognition score. If the system supervisor decides to perform the learning operation, the test utterance is considered to be a new template equipped with an appropriate label and added to the module of key word templates. This machine learning task should be conducted when the test utterance is incorrectly recognized. After learning, the updated DTW template set is closer to the uttered data derived from the test speaker and, therefore, the error recognition numbers are decreased. Figure 3 illustrates the processing flow of

the incremental learning scheme on DTW. To explain this learning scheme further, a pseudocode of the proposed DTW incremental learning method is presented in Pseudocode 1. As observed in Pseudocode 1, when performing incremental learning on DTW, the primary operations are

- (1) to label the learning data;
- (2) to add the learning data index into the referenced pattern index database; and
- (3) to add the feature of the learning data into the reference templates.

Thus, incremental learning provides a direct and conceptually simple learning technique. The primary disadvantage of incremental learning is that a large module of templates (reference template database in Figure 3) is presented for recognition comparisons because of numerous incorrectly recognized utterances found, which increases the number of computations on template matching and subsequently produces a heavy load of reference templates for real-time recognition responses.

**3.2. Priority-Rejection Learning.** To accelerate the computation of incremental learning using a large-scale reference template database and also maintain excellent recognition performance, an improved incremental learning scheme, priority-rejection learning, was further developed in this study and is presented in this section.

Priority-rejection learning offers the advantage of an invariant reference template database and can also immediately update content records in the template database when an utterance acquired from the test speaker is added into the template database for system learning. Figure 4 presents the processing procedure of the developed priority-rejection learning approach. As presented in Figure 4, after performing DTW recognition, two tasks were conducted. One task was to establish the recognition result among all of the possible template keyword candidates according to the DTW comparison scores, and the other was to record the value of the computed distance parameter, parameter  $D$  in (1), of each template keyword candidate. Priority-rejection learning, which involves processes that are similar to those used in incremental learning, is also a type of

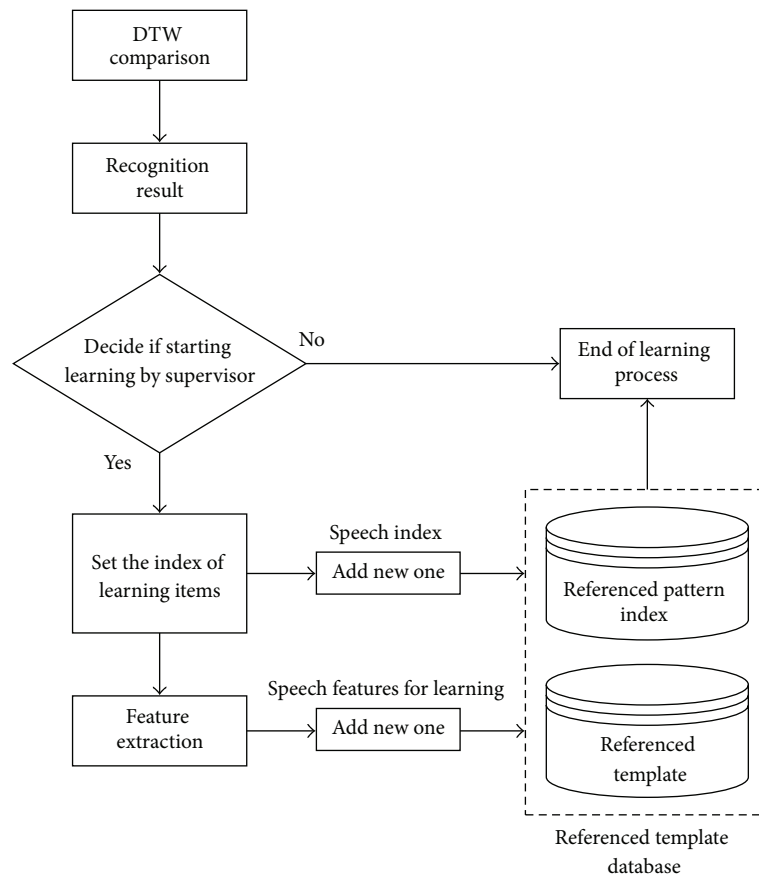


FIGURE 3: The proposed incremental learning method for DTW speech recognition.

```

Procedure DTW_Incremental_Learning ();
/*DTW recognition process*/
Perform DTW template matching;
Output recognition result before learning;
/* A decision of DTW system learning made by a supervisor */
If (Decision == "YES") then
  /* Correct recognition and then start the learning process */
  Label the recognition result and set the index  $i$ ;
  /*  $i$  is a relative index in referenced templates database */
  If ( $i$  setting == TRUE) then
    Convert the testing data to be the learning data;
    Feature extraction;
    Add  $i$  into the referenced pattern index;
  End If
  For each frame  $t$  ( $t = 1$  to total frames  $N$  of the learning data)
    Add features $_t$  into reference templates;
  End For
  /* End of learning process */
Else
  /* End of learning process (No learning) */
End If
  
```

PSEUDOCODE 1: The pseudocode of proposed incremental learning.

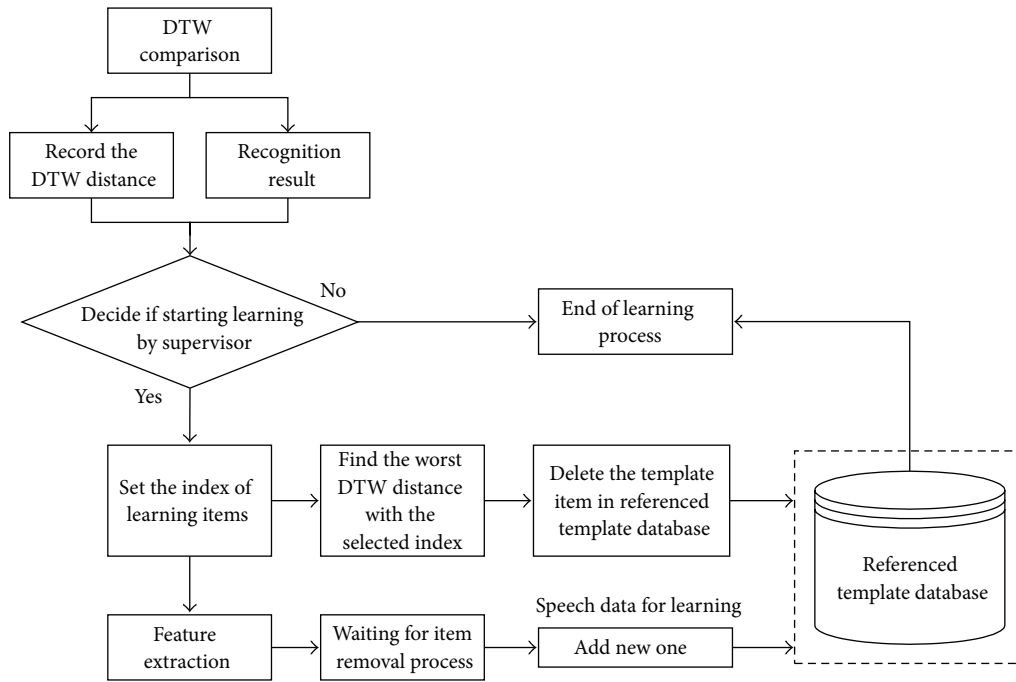


FIGURE 4: The developed priority-rejection learning method for DTW speech recognition.

supervised learning strategy. If the system supervisor decides to adjust the recognition system by using the test utterance, the learning target keyword template is first set. All of the keyword templates in the database with the same label as the target keyword template are subsequently removed. Following this step, the keyword template is removed from the system database to maintain the size of the reference template database after adding the test utterance to the database and subsequently establish a new reference template. In the reference template removal process, the determination policy of this study was to select the reference template with the lowest DTW comparison scores. The reference template that produces the highest value of parameter  $D$  is the least similar to the testing utterance, which causes correspondingly low DTW comparison scores to be produced. To explain the developed learning scheme further, a pseudocode of the proposed priority-rejection learning approach for DTW is presented in Pseudocode 2. As observed in Pseudocode 2, when performing priority-rejection learning on DTW, the main operations are

- (1) to search for the template item with the worst DTW distance;
- (2) to remove the identified item from the reference template database; and
- (3) to perform the primary incremental learning operations.

**3.3. Most-Matching Learning.** Unsupervised learning is an appropriate learning scheme for practical online speech recognition applications. This paper proposes an unsupervised learning method, namely, the most-matching learning

method, for DTW speech recognition. Most-matching learning is an extended version of the supervised priority-rejection learning method. The primary distinction between priority-rejection learning and most-matching learning is the decision-making scheme design of the recognition system adjustments for the test utterance. In contrast to the supervised learning scheme used in the priority-rejection learning method, the DTW speech recognition system involving most-matching learning uses an unsupervised learning scheme that determines whether the test utterance is appropriate for automatically performing system learning without any supervisors. The proposed most-matching learning method is illustrated in Figure 5. The continuous function blocks covered in the dashed line were integrated into a double-checking process, which verifies whether the test utterance was used in the learning process to update the DTW reference template database. In addition to the double-checking process, the operational functions in Figure 5 are almost similar to those of the priority-rejection learning method.

The double-checking process used in the most-matching learning method contains two fundamental steps to verify the test utterances produced by the speaker. The first step is to check if the calculated DTW score of Top-1 (indicating the reference template that is most similar to the test utterance) is greater than the predefined threshold  $T$ . If the score is lower than the value of  $T$ , the most-matching learning algorithm is immediately aborted because of the substandard test utterance. Otherwise, the most-matching learning process continues and then begins the second checking process. At the second checking step, reference templates with the same label as that of the Top-1 reference template are identified among the 10 highest DTW computational scores

```

Procedure DTW_Priority_Rejection_Learning ();
/* DTW recognition process */
Perform DTW template matching and store DTW distance;
Output recognition result before learning;
/* A decision of DTW system learning made by a supervisor */
If (Decision == "YES") then
  /* Correct recognition and then start the learning process */
  Label the recognition result and set the index  $i$ ;
  /*  $i$  is a relative index in referenced templates database */
  If ( $i$  setting == TRUE) then
    For each  $n$ 
      /*  $n$  is numbers of the estimated DTW-distance with  $i$  */
      Search the worst (the largest) DTW-distance;
    End For
    Return the template item with the worst DTW-distance;
    /* Removal process */
    Delete the found template item with the worst DTW- distance in the database;
    /* Process of learning data */
    Convert the testing data to be the learning data;
    Feature extraction;
    If (Removal process finished == TRUE) then
      Add  $i$  into the referenced pattern index;
      For each frame  $t$  ( $t = 1$  to total frames  $N$  of the learning data)
        Add features $_t$  into reference templates;
      End For
    End If
  End If
  /* End of learning process */
Else
  /* End of learning process (No learning) */
End If

```

PSEUDOCODE 2: The pseudocode of priority-rejection learning.

(Top-1 to Top-10) of the reference templates. If the number of searched reference templates (including the Top-1 reference template) is higher than a predefined value  $N$ , most-matching learning dictates that DTW recognition system learning be conducted. Otherwise, most-matching learning is aborted. The settings for the  $T$  and  $N$  values are established in an empirical procedure. The optimal values of thresholds  $T$  and  $N$  can be derived using a simple and direct trial-and-error testing procedure. To explain the unsupervised learning scheme further, a pseudocode of the most-matching learning approach for DTW is presented in Pseudocode 3. The primary functions performed using the most-matching learning method are summarized as follows:

- (1) to feed the learning data into an unsupervised double-checking process; and
- (2) to perform the primary operations of priority-rejection learning if the learning data are accepted.

#### 4. Experiments and Results

The experiments on DTW speech recognition involving the three proposed machine learning techniques were performed using a small vocabulary recognition application in which the test speaker was requested to utter a phrase for recognition

TABLE 1: Small-sized vocabulary keywords settings in DTW speech recognition.

Index of keywords	Keywords (in Chinese)
1	回家
2	請開燈
3	請關燈
4	開門
5	關門

testing. All of the uttered data were recorded in an office using a close-talking microphone. The speech signal was sampled at 44.1 kHz and recorded with monochannel settings and 16 bit resolution. The analysis frames were 20 ms wide with a 10 ms overlap. For each frame, a 10-dimensional cepstral vector was extracted. Table 1 presents the small vocabulary database that was composed of five keyword patterns. Each of the test speakers was asked to provide utterances that served as training data for establishing the DTW reference template database. Each of the five reference templates in Table 1 acquired 10 copies from the test speakers. Fifty reference templates were present in the database for DTW speech recognition. The recognition testing experiments for



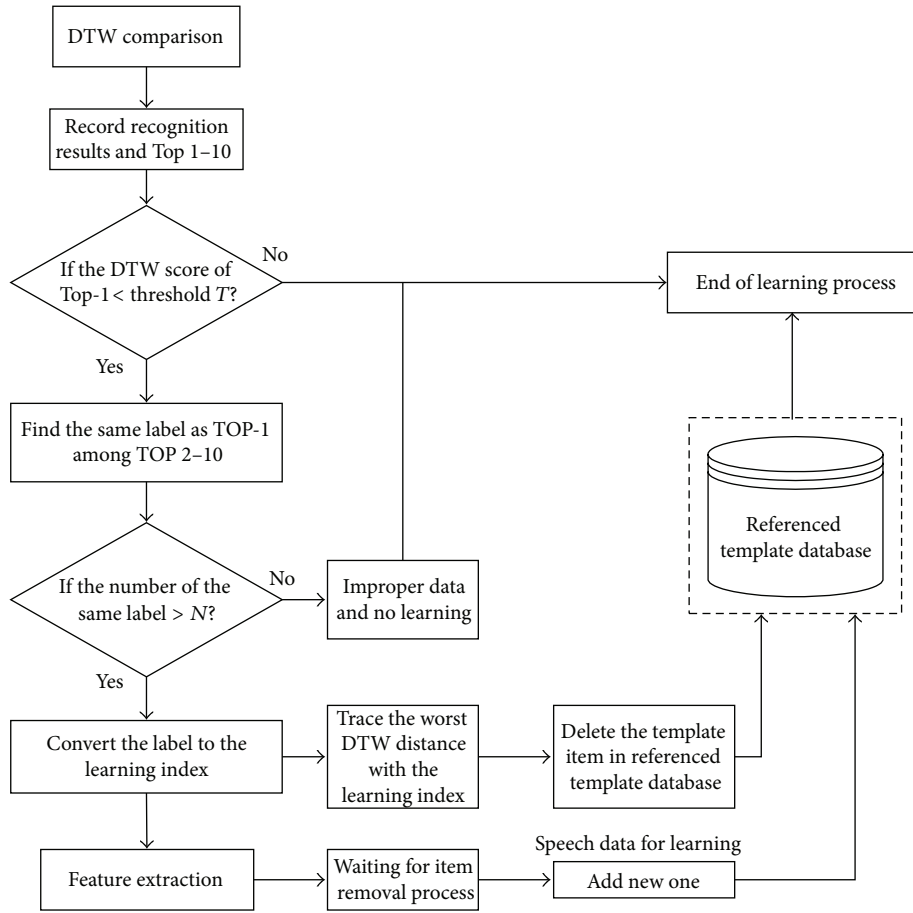


FIGURE 5: The presented most-matching learning method for DTW speech recognition.

TABLE 2: Recognition performances of inside testing DTW speech recognition.

Index of keywords	Recognition rates
Index 1	100%
Index 2	100%
Index 3	100%
Index 4	100%
Index 5	100%

evaluating the three proposed learning methods for DTW comprised two parts: inside testing and outside testing.

Table 2 presents the experimental results of performing inside testing using DTW speech recognition on the keyword patterns listed in Table 1. As shown in Table 2, the recognition accuracy of each template item reached 100% and, therefore, no learning was required.

In the outside testing recognition experiments, the utterances used to test the DTW speech recognition system completely differed from those used for establishing the DTW reference template database. Unique utterances were acquired from the test speakers. In addition, the utterances used as the learning data for the proposed learning methods were also obtained from the same test speakers. The baseline

TABLE 3: Recognition rates of outside testing DTW speech recognition (baselines without any learning).

Index of keywords	Baselines (no learning) Recognition rates
Index 1	96%
Index 2	80%
Index 3	70%
Index 4	62%
Index 5	66%
Average	74.8%

recognition rates of each reference template in the database are presented in Table 3. The data in Table 3 indicate a distribution from the highest recognition rate, 96%, to the lowest recognition rate, 62%, which suggests that the overall recognition performance was not ideal. The average recognition rate produced using conventional DTW speech recognition without any learning was 74.8%. The performance evaluations of DTW speech recognition combined with the proposed incremental learning method are presented in Table 4. As shown in Table 4, after performing five learning iterations, a recognition rate improvement was apparent. The

```

Procedure DTW_Most_Matching_Learning ();
Initialize the values of thresholds  $T$  and  $N$  to be constants;
/* DTW recognition process */
Perform DTW template matching;
Record DTW-distances of recognition results of Top-1 to Top-10;
/* Decide if starting learning by the unsupervised method */
If (DTW score of Top-1 < Threshold  $T$ ) then
/* Start learning */
  For each result label of Top- $t$  ( $t = 2$  to 10)
    Search the same label as Top-1 among Top 2-10;
  End For
  If (Numbers of the same labels as that of Top-1 >  $N$ ) then
    Convert the label of Top-1 to the learning index  $i$ ;
    If ( $i$  setting == TRUE) then
      For each  $n$ 
        /*  $n$  is the number of the DTW-distance with  $i^*$  */
        Search the worst (the largest) DTW-distance;
      End For
      Return the template item with the worst DTW-distance;
      /* Removal Process */
      Remove the template item with the worst DTW-score in the database;
      /* Process of learning data */
      Feature extraction;
      If (Removal process finished == TRUE) then
        Add  $i$  into the referenced pattern index;
        For each frame  $t$  ( $t = 1$  to total frames  $N$  of the learning data)
          Add features $_t$  into reference templates;
        End For
      End If
    End If
  Else
    /* Improper data and no learning */
  End If
Else
  /* End of learning process (No learning) */
End If

```

PSEUDOCODE 3: The pseudocode of most-matching learning.

TABLE 4: Recognition results of DTW speech recognition with incremental learning.

Index of keywords	Recognition rates after incremental learning (supervised learning)		
	One time	Two times	Five times
Index 1	96%	98%	100%
Index 2	84%	88%	94%
Index 3	80%	92%	96%
Index 4	64%	66%	76%
Index 5	66%	68%	72%
Average	78%	82.4%	87.6%

TABLE 5: Recognition results of DTW speech recognition with priority-rejection learning.

Index of keywords	Recognition rates after priority-rejection learning (supervised learning)		
	One time	Two times	Five times
Index 1	96%	96%	98%
Index 2	90%	90%	92%
Index 3	82%	86%	94%
Index 4	68%	68%	72%
Index 5	68%	66%	70%
Average	80.8%	81.2%	85.2%

third template item exhibited the greatest recognition rate improvement, which was 26% (from 70% to 96%). The recognition rate of the first item reached 100% after completing the fifth incremental learning. Table 5 shows the performance of the priority-rejection learning method when applied to DTW

speech recognition. As shown in Table 5, after completing five learning iterations, the first template item exhibited the highest recognition rate, 98%, which is a nearly 100%, of incremental learning. Regarding the increase in the recognition rate after performing priority-rejection learning, as



TABLE 6: Comparisons of incremental learning and priority-rejection learning methods on DTW computation time.

Learning times	Incremental learning method			
	No learning	One time	Two times	Five times
DTW computation time (sec.)	0.594	0.609	0.718	0.828
Learning times	Priority-rejection method			
	No learning	One time	Two times	Five times
DTW computation time (sec.)	0.594	0.609	0.609	0.609

TABLE 7: Recognition results of DTW speech recognition with most-matching learning (settings of thresholds:  $T = 300$  and  $N = 4$ ).

Index of keywords	Recognition rates after most-matching learning (unsupervised learning)		
	One time	Two times	Five times
Index 1	100%	100%	98%
Index 2	86%	86%	90%
Index 3	66%	64%	66%
Index 4	62%	62%	60%
Index 5	66%	68%	58%
Average	76%	76%	74.4%

TABLE 8: Recognition results of DTW speech recognition with most-matching learning (settings of thresholds:  $T = 400$  and  $N = 4$ ).

Index of keywords	Recognition rates after most-matching learning (unsupervised learning)		
	One time	Two times	Five times
Index 1	98%	96%	98%
Index 2	84%	88%	90%
Index 3	72%	68%	58%
Index 4	62%	64%	54%
Index 5	66%	66%	58%
Average	76.4%	76.4%	71.6%

the incremental learning, the third template item achieved the most improvement, 24% (from 70% to 94%), which was still slightly lower than 26% of incremental learning. The finding that the priority-rejection learning performance is inferior compared with that of incremental learning is completely reasonable and correct because priority-rejection learning maintains a fixed reference template database (50 templates in this scenario), and the reference template database of incremental learning gradually increases after learning. Table 6 presents a comparison of incremental learning and priority-rejection learning computational speed. As observed in Table 6, priority-rejection learning was superior to incremental learning because priority-rejection learning requires fewer template-matching comparison operations. The performance of unsupervised most-matching learning is shown in Tables 7 and 8 with various threshold settings for parameters  $T$  and  $N$ . Using unsupervised most-matching learning is evidently less favorable than using incremental learning or priority-rejection learning. Although most-matching learning is operated without a supervisor, the recognition performance still improved after learning in most situations. However, under circumstances in which DTW speech recognition encountered substandard test utterances for most-matching learning, the recognition performance was dissatisfactory and the recognition rate was substantially lower than the baseline (e.g., the recognition rates of the third template item listed in Tables 7 and 8 were lower than the baseline after unsupervised learning operations were conducted). DTW speech recognition achieved the highest average recognition rate of 76.4%, which was higher than the baseline of 74.8%, when the unsupervised learning method was used.

## 5. Conclusion

This study focused on DTW-based speech recognition for developing machine learning schemes in recognition systems. Two categories of learning mechanisms, supervised and unsupervised learning, used for DTW speech recognition were thoroughly explored in this paper. Regarding supervised learning, this study proposed two methods, incremental learning and priority rejection learning, for performing DTW. Both incremental learning and priority-rejection learning are conceptually simple and improve the recognition accuracy of conventional DTW. Regarding unsupervised learning, the most-matching approach was developed for DTW in this study. The most-matching approach was based on the concept that priority-rejection learning can automatically perform DTW system learning without any human supervisors. DTW that applies any of the three proposed learning methods uses processes that are similar to those used in model-based speech recognition and can adjust the recognition system properly by using the utterances produced by the speaker.

## References

- [1] F. Harashima and S. Suzuki, "State-of-the-art intelligent mechatronics in human-machine interaction," *IEEE Industrial Electronics Magazine*, vol. 4, no. 2, pp. 9–13, 2010.
- [2] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, "A review on speech recognition technique," *International Journal of Computer Applications*, vol. 10, no. 3, pp. 16–24, 2010.
- [3] L. R. Rabiner, "A Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

- [4] A. K. Paul, D. Das, and M. M. Kamal, "Bangla speech recognition system using LPC and ANN," in *Proceedings of the IEEE 7th International Conference on Advances in Pattern Recognition (ICAPR '09)*, pp. 171–174, Kolkata, India, February 2009.
- [5] A. Ganapathiraju, J. E. Hamaker, and J. Picone, "Applications of support vector machines to speech recognition," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2348–2355, 2004.
- [6] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [7] P. G. N. Priyadarshani, N. G. J. Dias, and A. Punchihewa, "Dynamic time warping based speech recognition for isolated Sinhala words," in *Proceedings of IEEE International Midwest Symposium on Circuits and Systems*, 2012.
- [8] X. Anguera, R. Macrae, and N. Oliver, "Partial sequence matching using an unbounded dynamic timewarping algorithm," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 3582–3585, Dallas, Tex, USA, March 2010.
- [9] Y.-S. Lin and C.-P. Ji, "Research on improved algorithm of DTW in speech recognition," in *Proceedings of the International Conference on Computer Application and System Modeling (ICCASM '10)*, vol. 9, pp. V418–V421, Taiyuan, China, October 2010.
- [10] X. Chen, J. Huang, Y. Wang, and C. Tao, "Incremental feedback learning methods for voice recognition based on DTW," in *Proceedings of International Conference on Modeling, Identification and Control*, 2012.
- [11] Q. Chen, G. Hu, F. Gu, and P. Xiang, "Learning optimal warping window size of DTW for time series classification," in *Proceedings of International Conference on Information Sciences, Signal Processing and their Applications*, 2012.
- [12] C.-H. Lee, C.-H. Lin, and B.-H. Juang, "A study on speaker adaptation of the parameters of continuous density hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 806–814, 1991.
- [13] C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech and Language*, vol. 9, no. 2, pp. 171–185, 1995.
- [14] R. Kuhn, J.-C. Junqua, P. Nguyen, and N. Niedzielski, "Rapid speaker adaptation in eigenvoice space," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 695–706, 2000.
- [15] B. K.-W. Mak and R. W.-H. Hsiao, "Kernel eigenspace-based MLLR adaptation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 784–795, 2007.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

