IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-4, Issue-10 E-ISSN: 2347-2693

Five Stage Dynamic Time Warping Algorithm for Speaker Dependent Isolated Word Recognition in Speech

Munshi Yadav[1*], Afshar Aalam[2]

[1]Dept. of IT , Guru Tegh Bahadur Institute of Technology, New Delhi. 110064
[2]Dept. of CS, Jamia Hamdard, Hamdard Nagar New Delhi, 110062

Available online at: www.ijcseonline.org

Received: 17/Sep/2016 Revised: 25/Sep/2016 Accepted: 20/Oct/2016 Published: 31/Oct/2016

Abstract- In speech recognition, a speaker dependent isolated word recognition system is used for small vocabulary in different applications for voice control systems. Dynamic Time Warping (DTW) algorithm is used for pattern matching when two sequences of unequal size are available. When test data and reference data or sequences are available of unequal in nature with time domain then existing DTW algorithm takes time more, while proposed solution will give the efficient algorithm which reduces the computation time without degradation of accuracy and efficiency.

***Keywords*:** Dynamic time warping, speech recognition, speaker dependent.

I. INTRODUCTION

Speech processing has two categories; one is to developing a new technique or method and second is to improve the existing method/algorithm with respect to different domain like time, space, efficiency, accuracy, delay, response time etc [1]. Fundamentally there are two types of speech, voiced and unvoiced speech [2]. In voiced speech there is a periodic signal where as unvoiced speech is non periodic or random signal somewhere others are mixture of both. In periodic signal pitch is varying with time [3]. Time domain analysis and frequency domain analysis in speech recognition is of two type's one is speaker dependent and other is speaker independent for spoken word recognition. Objective of the study is to develop the algorithm/technique which reduce the required time and space. There exists a space efficient approach called Sparse DTW algorithm which is efficient in space complexity [4]. Quantized DTW algorithm which increase the speed of pattern matching and takes less memory in comparison to existing DTW algorithm. In this technique it stores only one reference model for each word instead of multiple reference patterns. Where concept of vector quantization has been used based on classes which is represented by a centroid (codebook) [5].

Paper has organized in eight parts. Section (II) has discussed about acoustic feature extraction in speech recognition. In section (III) designing of the database. In section (IV) dynamic time warping algorithm. In section (V) five stage dynamic time warping algorithm. In section (VI) implementation of the proposed solution. In section (VII) results and in section (VIII) conclusion.

II. ACOUSTIC FEATURE EXTRACTION

Voice samples of spoken isolated words are captured through a closed talking boom mounted microphone. The detected analog signal is allowed to pass through the input signal conditioning circuitry (pre-amplifier, equalizer and low pass filter etc.), which emphasizes the higher frequencies as well as speaking level [6]. The pre-emphasized signal is passed through a spectrum analyzer based on the 16 contiguous filters, covering the frequency range from 300 Hz to 8000 Hz. The higher frequency range of 8000 Hz was chosen upon to cover most of the spectra of consonants [7] [8].

The output of each of the band pass filter is passed through a rectifier circuit followed by a low-pass filter with a cut-off frequency of 25 Hz. Thus, the output of individual channel is a signal which is proportional to the sound pressure level in that channel. Filter bank used for feature extraction in speech which is simple and inexpensive method for obtaining a spectral representation of speech signal. A filter bank consists of a number of band pass filters, covering adjacent frequency bands. The parameters available for experimentation are number of band pass filters, amplitude compression and centre frequency. It has been found that the more filters we have, the more accurate the spectral representation but it will take more memory space which increase the response time in pattern matching.

III. DESIGN OF SPEECH DATABASE

A speech database is a collection of recorded speech accessible on a computer. There are three categories of speech databases currently available. The first type, called analytic-diagnose, is used to improve our knowledge of the basic linguistic and phonetic elements of speech. A second type, defined as generic, includes non-specific vocabularies that are suitable for many applications. A third type, referred

to as specific, collects speech whose characteristics are related to the application target, such as an information request system. Data base has been develop on the basis of ten spoken words like zero, one, two, three, four and five, six, seven, eight, nine etc by a Male speaker and ten utterances of each word has been recorded.

IV. DYNAMIC TIME WARPING (DTW) ALGORITHM

DTW algorithm is based on dynamic programming approach of optimization based on bottom up fashion. It used for measuring the similarity between two sequences / pattern which may vary in time. DTW is very much used in different area like motion detector, image processing, speech processing signature verification curve verification face detection etc. DTW is a method that calculates an optimal match between two given sequences with certain restrictions of time duration [9] [10].

If two sequences A and B are strings of discrete symbols and having unequal dimensions then d(a, b) is defined as a distance between the symbols a and b [11][12].

d(a, b) = |a - b|,

DTW[i, j] := d(i, j) +

min {

DTW [i-1, j],

DTW [i, j-1],

DTW [i-1, j-1]

}

V. FIVE STAGE DYNAMIC TIME WARPING ALGORITHM

Algorithm_Five_Stage_DTW(X, Y)

{

1. X={ x_1, x_2, x_3 , ... x_m}
2. Y= {y_1,. Y_2, y_3, ... y_n}
3. n_1= |X|/5,
4. n_2= |Y|/5,
5. apply DTW algorithm for calculating the distance to corresponding filter values.
6. take the addition of calculated values for individual distance for every filter.
7. return (distance).

}

VI. IMPLEMENTATION IN MATLAB

```
Algorithm_DTW_Five_Stage( test, ref)

{

// test. is the test vector
// ref. is the reference vector
test= [test1,test4];
ref= [ref1,ref4];
for i=1:1;
 t=test(1:10,16*i+1:16*i+16);
 r=ref(1:10,16*i+1:16*i+16);
end
p=size(t);
q=size(r);
 count = 0;
for c=1:5;
h111=round((c-1)*p(1)/5);
h1=round(c*p(1)/5);
h211=round((c-1)*q(1)/5);
h2=round(c*q(1)/5);
 a=t(1+h111:h1,:);
b=r(1+h211:h2,:);
 m= h1-h111;
n= h2-h211;
 for k=1:16;
dtwi(1,1)= abs(a(1,k)-b(1,k));
    for j=2:n
       dtwi(1,j) = abs(a(1,k)-b(j,k))+ dtwi(1,j-1);
   end
        for i=2:m
       dtwi(i,1) = abs(a(i,k)-b(1,k)) +dtwi(i-1,1) ;
   end
 for i=2:m
   for j=2:n
```



```
            dtwi(i,j) =  abs(a(i,k)-b(j,k))+
          min( [dtwi(i-1,j),   dtwi(i,j-1) ,
                              dtwi(i-1,j-1) ] );
          d(k) = dtwi(i,j)/(m+n);
        count = count+1;
    end
end
 end
 total = sum(d);
 result(c) = total;
 end
 final = sum(result);
 }
```

VII. Results

Measured distance between spoken test words with reference word has been shown in the table. Minimum values have been shown in bold in the table given below.

Table: Comparison matrix for spoken word zero to ten and corresponding result of pattern matching.

TestRef.	zero	one	two	three	four	five	six	seven	Eight	nine
Zero	**420**	1694	926	850	1877	1753	1077	1657	1404	1493
One	844	**854**	748	1033	1238	1536	1267	1443	1538	1229
two	527	1668	**353**	1082	2072	1993	1089	2224	1750	1830
three	794	1345	1127	**396**	1725	1402	579	1643	1034	1105
four	458	1267	385	934	1446	1466	1159	1635	1362	1447
five	1145	1207	1298	1081	**1399**	**996**	1291	1297	1266	1007
Six	1048	1468	1273	542	1984	1473	**521**	1840	889	1066
Seven	742	1527	995	1046	1576	1411	1264	**1295**	1294	1269
Eight	1040	1474	1253	615	1638	1446	559	1727	**830**	1078
Nine	1203	1195	1409	868	1640	999	968	1386	1079	**707**

VIII. CONCLUSIONS AND FUTURE WORK

An experiment has been done over English digit for isolated spoken word from zero to nine numbers. Ten utterances of each spoken word have been taken by a Male speaker. The time taken by proposed solution in speech recognition using five stage dynamic time warping algorithm is very very less in comparison to speech recognition using single stage dynamic time warping algorithm. This study can be improve using other existing feature extraction technique like LPC and MFCC for five stage dynamic time warping algorithm is very very less in comparison to speech recognition using single stage dynamic time warping algorithm. This study can be improve using other existing feature extraction technique like LPC and MFCC for feature extraction and HMM technique for pattern matching.

IX. ACKNOWLEGEMENT

We would like to thanks Dr. Abdul Mobin (Chief Scientist Retd. from National Physical Laboratory, Delhi) for his valuable suggestions, without which it could not be possible to complete this work.

X. REFERENCES

[1] Titus Felix FURTUNA, Dynamic Programming Algorithms in Speech Recognition, Revista Informatica Economica nr. 2(46), 2008, pp 94- 99.

[2] B. H. Jaung and L.R. Rabiner , Automatic Speech Recognition – A Brief History of The Technology, Elsevier Encyclopedia of Language and Linguistics, Second Edition, 2005.

[3] Rubita Sudirman, Sh.-hussain Salleh, Ting Chee Ming, Local DTW Coefficients and pitch feature for back-

propagation NN digit recognition. Proceedings of the IASTED International Conference on Networks and Communication Systems 2006. pp-201-206.

[4] Ghazi Al- Naymat, Sanjay Chawala, Javid Taheri, Sparse DTW A Novel Approach to Speed up Dynamic Time Warping, Proc. of 8th Australasian Data Mining Conference (AusDM'09), pp 117- 127.

[5] Tiberius Zaharia, Svetlana Segarceanu, Marius Cotescu, Alexandru Spataru, " Quantized Dynamic Time Warping (DTW) Algorithm" , IEEE, 2010, pp 91-94.

[6] L.R. Rabiner and R. W. Schafer, Thery and Applications of Digital Speech Processing, Prentice Hall Inc., 2011.

[7] L.R. Rabiner and R. W. Schafer, Introduction to Digital Speech Processing, Foundations and Trends in Signal Processing, Vol. 1, Nos. 1-2, NOW Publishers, Boston, pp. 1-200, 2007.

[8] S. Dusan and L.R. Rabiner. Can Automatic Speech Recognition Learn More From Human Speech Perception, Trens in Speech Technology, Proc. of the third conference on Speech Technology and Human Computer Dialogue. pp. 21-36, May 2005.

[9] Eiji Mizutani, The Dynamic Time Warping Algorithms. Lecture Note for Mechanical Engineering Seminar, Tokyo Metropolitan University. 2006. pp 1-11.

[10] Rubita Sudirman, Sh.-hussain Salleh. Ting Chee Ming, NN Speech Recognition Utilizing Aligned DTW Local Distance Scores, ICMT-192. pp 1-5.

[11] Sakoe, H. & S. Chiba.(1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE, Trans. Acoustics, Speech and Signal Proc., Vol. ASSP-26. pp 43-49.

[12] Somya Adwan, Hamzah Arof, A Novel Double Stage Dynamic Time Warping Algorithm for Image Template Matching, Proceeding of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications (ICMSA 2010), University Tunku Abdul Rahman , Kaula Lumpur Malaysia. Pp. 667-676.

About the Author

Munshi Yadav: Author is graduated in Electronics and Communication Engineering in 1998 from G. B. Pant Engineering College, Pauri Garhwal, Uttranchal and did M.Tech in Information Technology from University School of Information Technology, Guru Govind Singh Indraprastha University, Delhi in 2006 and pursuing PhD in computer Science from Jamia Hamdard Delhi. Currently working as Associate Professor in the department of Information Technology, Guru Tegh Bahadur Institute of Technology, Delhi.

