Dynamic Programming Algorithms in Speech Recognition

Kayte C.N[1]., Kayte J.N. [2], Kayte S.N. [3], R.R. Manza[4]
[1]Indraraj Arts ,Commerec and Science College Sillod,Dist Aurangabad(mh)-431112
[2]Arts ,Commerec and Science College Badnapur,Dist Jalna(mh)
[3]MGM Dr.P.Y.Pathrikar College Aurangabad
[4]Dept. of Computer Science ,Dr.B.A.M.University Aurangabad

Abstract

In a system of speech recognition containing words, the recognition requires the comparison between the entry signal of the word and the various words of the dictionary. The problem can be solved efficiently by a dynamic comparison algorithm whose goal is to put in optimal correspondence the temporal scales of the two words. An algorithm of this type is Dynamic Time Warping. This paper presents two alternatives for implementation of the algorithm designed for recognition of the Marathi isolated words.

Keywords: dynamic programming, speech recognition, word detection.

Introduction

India is multilingual country of more than 1 billion population with 18 constitutional languages and 10 different scripts. Devnagari, an alphabetic script, is used by a number of Indian languages. A numeral word analysis is done after taking an input through microphone from a user. The aim of this paper is to build numeral word recognition tool for Marathi language. This is a isolated word speech recognition tool.

Today's vocal recognition systems are based on the general principles of forms' recognition [6][7][10]. The methods and algorithms that have been used so far can be divided into four large classes:

1. Discriminant Analysis Methods based on Bayesian discrimination
2. Hidden Markov Models
3. Dynamic Programming -Dynamic Time Warping algorithm (DTW) [9];
4. Neuronal Networks.

Dynamic Time Warping Algorithm (DTW)

DTW algorithm is based on Dynamic Programming. This algorithm is used for measuring similarity between two time series which may vary in time or speed. This technique also used to find the optimal alignment between two times series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis. This warping between two time series can then be used to find corresponding regions between the two time series or to determine the similarity between the two time series. The Fig.1. Shows the example of how one times series is „warped" to another.

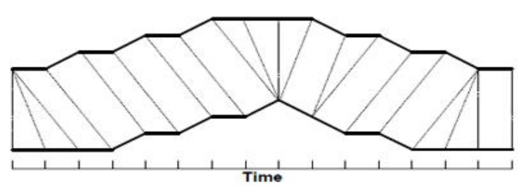


Fig.1: Warping between two time series.

To align two sequences using DTW, an *n*- by- *m* matrix where the (*i*th, jth) element of the matrix contains the distance *d* (*q*i, *cj*) between the two points *q*i and *c*j is constructed. Then, the absolute distance between the values of two sequences is calculated using the Euclidean distance computation as shown in "(2)".

$$d\,(q_i,c_j) = (q_i - c_j)^2 \qquad (2)$$

Each matrix element (*i*, *j*) corresponds to the alignment between the points *q*i and *c*j. Then, accumulated distance is measured by "(3)".

$$D\,(i, j) = \min[D(i-1, j-1), D(i-1, j), D(i, j-1)] + d(i, j) \quad \ldots (3)$$

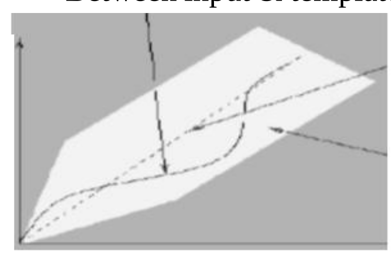
This is shown in Fig.2, where the horizontal axis represents the time of test input signal, and the vertical axis represents the time sequence of the reference template. The path shown results in the minimum distance between the input and template signal. The shaded area represents the search space for the input time to template time mapping function. Any monotonically non decreasing path within the space is an alternative to be considered. Using dynamic programming techniques, the search for the minimum distance path can be done in polynomial time P (t), using equation below:-

$$p\ t = o[N^2\ V] \qquad (4)$$

Where, N is the length of the sequence, and V is the number of templates to be considered[4].

Linear time warp. Minimum distance mapping
Between input & template.

Template Signal time



DTW Search Space

Fig.2: Dynamic time warping (DTW).

Theoretically, the major optimizations to the DTW algorithm arise from observations on the nature of good paths through the grid[7].

The global warp cost of the two sequences is defined as shown below:

$$GC = \frac{1}{p}\sum_{i=1}^{p} w_i\ ,$$

where wi are those elements that belong to warping path, and p is the number of them. The calculations made for two short sequences are shown in figure 3 including the highlight of the warping path.

	-2	10	-10	15	-13	20	-5	14	2
3	5	12	25	37	53	70	78	89	90
-13	16	28	15	43	37	70	78	105	104
14	32	20	39	16	43	43	62	62	74
-7	37	37	23	38	22	49	45	66	71
9	48	38	42	29	44	33	47	50	57
-2	48	50	46	46	40	55	36	52	54

Fig.3 warping path

There are three conditions imposed on DTW algorithm that ensure them a quick convergence:

1. **monotony** – the path never returns, that means that both indices i and j used for crossing through sequences never decrease.
2. **continuity**–the path advances gradually,step by step; indices i and j increase by maximum 1 unit on a step.

3. **boundary** –the path starts in left-down corner and ends in right-up corner.

Because optimal principle in dynamic programming is applied using "backward" technique, identifying the warp path uses a certain type of dynamic structure called "stack". Like any dynamic programming algorithm, the DTW one has a polynomial complexity. When sequences have a very large number of elements, at least two inconveniences show up:

- memorizing large matrices of numbers;
- performing large numbers of distances calculations.

There is an improvement in standard DTW algorithm that sorts out the two problems named above: Fast DTW (Fast Dynamic Time Warping) [Stan Salvador, Philip Chan - 6]. The proposed solution consists of dividing distances matrices into 2,4,8.16, etc. matrices of smaller dimensions through a repeatedly process of splitting in two the input sequences. This way, the distance calculations are performed only on these smaller matrices and the warp path is then put together by merging the warp paths calculated for smaller matrices. From algorithmic point of view, the proposed solution is based on "Divide et Impera" method.

Words Identification Using DTW Algorithm

Words identification can be performed by straight comparison of the numeric forms of the signals or by signals spectrogram comparison. The comparison process in both cases must compensate for both the different length of the sequences and non-linear nature of the sound. The DTW Algorithm succeeds in sorting out these problems by finding the warp path corresponding to the optimal distances between two series of different lengths. There are some particularities when the algorithm is applied to the two cases:

The numeric sequence can have some thousand numeric values, while a subsequence can have some hundred. Decreasing the number of numeric values can be performed by removing those ones between the extreme

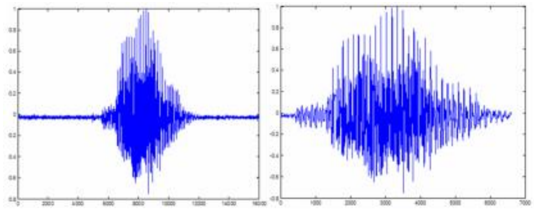


Fig.4. Vocal Signal for the word "nau"

Points. This process of reducing the length of the numeric sequence must not alter its form. Apparently, the process leads to a decrease in recognizing precision. However, taking into account an increase in speed, the precision is, in fact, increased by enlarging the number of words in the dictionary.

1. Signals spectrogram representations and applying the DTW algorithm for comparison of two spectrograms. The method consists of splitting the numeric signal in a number of "windows" (intervals) which will overlap. For each window, real number intervals, (sound frequencies) the Quick Fourier's transform… will be calculated and it will be stored in a matrix: the sound spectrogram. The parameters will be the same for all calculation operations of the: the window length, Fourier's transforms length, the overlap length for two successive windows. The Fourier's transform is symmetrical related to the center and the complex numbers from the second half are the conjugated complex number of the symmetrical numbers from the first half. Due to this fact, only values from the first half can be retain, so that the spectrogram will be a complex numbers matrix, its number of lines equaling half of Fourier's transform length and its number of columns

depending on the sound length. The DTW will be applied on a real number matrix resulted from conjugating the spectrogram values, matrix called energies matrix.

CONCLUSIONS

DTW Algorithm is very useful for isolated words recognition in a limited dictionary. For a fluent speech recognition, Hidden Markov Chains are used. Using dynamic programming ensures a polynomial complexity to the algorithm: $O(n^2v)$, where n is sequences' lengths and v is the number of words in the dictionary. There are some weaknesses of the DTW. First, the $O(n^2v)$ complexity may not be satisfactory for a larger dictionary which could ensure an increase in the success rate of the recognition process. Secondly, it is difficult to evaluate two elements from two different sequences, taking into account that there are many channels having distinct features. However, DTW remains an easy-to implement algorithm, open to improvements, very appropriate for applications that need simple words recognition: telephones, car computers, security systems, etc.

REFERENCES

[1] Kayte C.N. *Isolated Word Recognition for Marathi Language using VQ and HMM*., Science Research Reporter 2(2):161-165, April 2012

[2] Kayte C.N. , *A Multi-HMM Marathi Isolated Word Recognizer*, Science Research Reporter 2(2):175-177, April 2012

[3]Benoit Legrand, Nallasivam Palanisamy, *Chromosome classification Using Dynamic Time Warping*, ScienceDirect Pattern Recognition Letters 29(2008) 215– 222

[4] Cory Myers, Lawrence R. Rabiner, Aaron E. Rosenberg, *Performance Tradeoffs in Dynamic TimeWarping Algorithms for Isolated Word Recognition*, Ieee Transactions On Acoustics, Speech, And Signal Processing, Vol. Assp-28,No. 6, December 1980

[5] F. Jelinek. "*Continuous Speech Recognition by Statisical Methods*." IEEE Proceedings 64:4(1976): 532-556

[6] Siddheshwar S, *Speech Processinf for Marathi Numeral Recognition using MFCC and DWT features*, IJERA,Mar 2012,p219

[7] Rabiner, L. R., *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proc. of IEEE, Feb. 1989

[8] Rabiner, L. R., Schafer, R.W., *Digital Processing of Speech Signals, Prentice Hall*, 1978.

[9] Stan Salvador, Chan, FastDTW: *Toward Accurate Dynamic Time Warping in Linear Time and Space,IEEE Transactions on Biomedical*.Engineering, vol. 43, no. 4

[10] Young, S., *A Review of Large-Vocabulary Continuous Speech Recognition*, IEEE Signal Processing Magazine, pp. 45-57, Sep. 1996

[11] Sakoe, H. & S. Chiba.*Dynamic programming algorithm optimization for spokenword recognition*. IEEE, Trans. Acoustics,Speech, and Signal Proc., Vol. ASSP-26.