BRIEF ARTICLE

THE AUTHOR

1. Report

In the previous chapter, we have seen that using a MFCC representation of utterances with regions of silence removed leads to a large improvement in accuracy, time and computational complexity in the performance of DTW algorithm augmented with a euclidean metric..The main contributing factor behind the large time and computational complexity of the base line DTW is the **size** of the time series sequences. The computational cost of a DTW algorithm is (mn) where m and n denote the length of the two time series sequences currently compared. Using longer sequences increases the size of the DTW cost matrix hence resulting into a greater number of computations.

The DTW algorithm on its own is a domain independent algorithm that uses a similarity metric to compare any two sequences through comparison of their global trends. The algorithm employes dynamic programming to search a space of mapping between the time axis of the two respective sequences to determine the optimum alignment between them. The only difference between MFCC-augmented DTW and baseline DTW is the feature extraction stage. In machine learning, feature extraction refers to the pre-processing stage that involves the extraction of new features from a set of raw attributes through a suitable functional mapping. The extraction phase of MFCC features involves a segmentation of the time series followed by a functional mapping on the segmented windows. The resultant sequence of extracted feature vectors has a much smaller length compared to the length of the original sequence. Evident from the experiments done in the previous chapters, the use mel-cepstrum features extracted on 'cleaned' signals not only increases the accuracy of DTW but also reduces the time and computational cost through reduction of dimensionality of the original sequence.

Any time series sequence contains both local and global trends. In some time series datasets[], it has been observed that incorporating the information about these local trends and global shapes in the clustering /classification process does improve the performance of the DTW. The feature extraction methodology used in these works is domain and

application independent. The MFCC feature extraction on the other hand, is a domain and application dependent. This feature extraction process can only be applied to time series sequences corresponding to speech. From a scientific stand point, it will be interesting to compare and see how well/bad the domain independent methods are to domain dependent methods. In this chapter, I investigate an unsupervised methodology thats

- incorporates information about local and global trends in the feature extraction process
- employs an adaptive DTW that tackles the issue of time and computational complexity by moving from working on time series sequences to sequences of segmented time-slices. To achieve this, the algorithm uses a kernel function(self-proposed) that is designed to measure the similarity of sub-sequences more accurately than standard euclidean metric by being invariant toward time-dilation and scale.

1.1. **Feature extraction.** The fundamental problem of baseline DTW is that the numerical value of a data point in a time series sequence is not a complete picture of the data point in relation to the rest of the sequence.The euclidean metric computes distance based on numeral values. The context such as the position of the points in relation to their neighbours is ignored. To fix this issue, an alternative form of DTW know as Derivative DTW is proposed but the fundamental problem with this DTW is that it fails to detect significant common sub-patterns between two sequences(mainly global trends). Ideally we need to use features that contains information about the overall shapes of the sequences plus the local trend around the points. This allows the DTW to built a complete picture of the data point in relation to the rest of the sequence.

The methodology that I have used for my project is based on Xie and Wiltgen's paper[]. Each point in the time series sequence is replace by a 4 dimensional vector where the first two features corresponds to information regarding the local trends around the point and the last two features reflects the position of that point in the global shape of the sequence.

Definition of local feature:

$$f_{\text{local}}(r_i) = (r_i - r_{i-1}, r_i - r_{i+1})$$

Definition of global feature: Points to consider: must reflect information about the global trends and in order to be combined with local features, they must be of the same

scale.

$$f_{\text{global}}(r_i) = (r_i - \sum_{k=1}^{i-1} r_k, r_i - \sum_{k=i+1}^{M} \frac{r_k}{M+1})$$

The feature extraction stage is motivated from the extraction of MFCCs. The utterances are segmented into windows of width 10 ms and functional mapping is applied to each window. The mapping that I chose in this particular instance is as follows:

Kernel functions must be continuous, symmetric, and most preferably should have a positive (semi-) definite Gram matrix. Kernels which are said to satisfy the Mercer's theorem are positive semi-definite, meaning their kernel matrices have no non-negative Eigen values. The use of a positive definite kernel insures that the optimization problem will be convex and solution will be unique.

$$\begin{aligned} k(x,z) &= (x^T x')^2 \\ &= (x_1 z_1 + x_2 z_2)^2 \\ &= x_1^2 z_1^2 + 2x_1 z_1 x_2 z_2 + x_2^2 z_2^2 \\ &= (x_1^2, 2x_1 x_2, x_2^2)(z_1^2, 2z_1 z_2, z_2^2)^T \\ &= \phi(x)^T \phi(z) \end{aligned}$$

We saw that the simple polynomial kernel $k(x,z) = (x^T z)^2$ contains only terms of degree two. If we consider the slightly generalised kernel:

$$(x,z) = (x^T z + c)^2$$

with c >0, then the corresponding feature mapping $\phi(x)$ contains constant and linear terms as well as terms of order two. If we generalize this notion then $k(x,x') = (x^T z)^M$ contains all monomials of order M. For instance, if x and z are two images, then the kernel represents a particular weighted sum of all possible products of M pixels in the first image with M pixels in the second image.