Fixed Distortion Segmentation in Efficient Sound Segment Searching

Masahide Sugiyama
Graduate School of Computer Science and Engineering
The University of AIZU, Japan
sugiyama@u-aizu.ac.jp

Abstract

Searching query signal from stored signal is formulated as a segment searching problem where signal is converted into a sequence of feature vectors. As an efficient segment searching algorithm, a new pruning method in the segment sequence has been proposed and the effectiveness has been shown through experimental evaluation. The proposed searching algorithm is 20 - 30 times faster than the conventional Active Searching algorithm. As the first step of the proposed method distortion based segmentation is carried out. As searching criterion is based on $l_1$ norm, the segmentation is expected to be carried out based on $l_2$ criterion. This paper compares three segmentation methods; maximum $l_1$ distortion segmentation, average $l_2$ distortion segmentation and fixed length segmentation. The average $l_2$ distortion and fixed length segmentation methods are very efficient. On the other hand, the maximum $l_1$ distortion segmentation produces fixed distortion segmentation and does not require radius information. The experimental results show that first two methods have almost equal performance in segment searching when the number of segments are the same.

1. Introduction

As efficient segment searching techniques in time-space sequences various improvements of Active searching (AS) algorithms have been proposed and the effectiveness have been reported[1, 2]. Originally the algorithms have been formulated using the similarity measure. On the other hand, the mathematical distance based formulation of AS has been proposed and its geometrical interpretation has been described, and as a further generalization Fuzzy Active searching algorithm has been proposed[3]. Recently, in order to speed up AS algorithms, an efficient pruning technique using center vectors of local and global clusters for the histogram vector space has been proposed[4, 5]. Using segmentation and clustering for output probability vectors more efficient searching algorithm has been proposed[6, 7]. As the preprocessing of the algorithm the distortion based segmentation has been carried out. This paper compares there segmentation methods; maximum $l_1$ distortion segmentation, average $l_2$ distortion segmentation and fixed length segmentation methods. The first method does not require keeping radius information of each segment, however, the second and third methods have less segmentation computation. The experimental results show that with the same numbers of segments the first and second segmentation methods have the same performance, such that, the numbers of distance calculations in segment searching.

2. Efficient Segment Searching Method

In order to achieve more efficient searching algorithm than AS, [4, 5] proposed a local and global pruning technique of segment candidates. On the other hand, [6] proposed a new efficient searching algorithm which is based on segmentation and clustering based on $l_2$ norm. This is based on more geometrical viewpoint and can give geometrical insight and intuition rather than the similarity based formulation. The segmentation of the vector sequence is carried out prior to the clustering procedure. This gives continuous segments for searching, and using the mathematical triangle inequality all vectors in each segment is efficiently determined whether the vector satisfies the searching criterion.

For a series of output probabilities, $P = (p_t)$, and searching vector, $q$, the search problem is formulated as searching vectors which satisfy the following inequality:

$$d_i(p_t, q) < \theta$$  (1)

Here, $d_i$ is $l_i$ norm, and $\theta$ is positive and the threshold of searching. A proposed searching algorithm contains three procedures;

1. Determination using segment representatives: M1
   Sound sequence is initially segmented and the search problem is solved using each representative vector, $c_t$. This procedure is called as M1.

2. Determination using cluster representatives: M2
   A set of $c_t$ is clustered and more compressed representative vectors are generated. The search problem is solved using, $b_t$. This procedure is called as M2.

3. Distance pruning: M3
   The search problem is solved using pre-calculated distance matrix between cluster representatives, $b_t$.

The first procedure (M1: determination using segment representatives) in the above algorithm segments given a series of output probability vectors, $P = (p_t)$, into small segments (intervals) $[b_{i-1}, b_{i+1} - 1]$ and produces its representative vector $c_t$ which satisfies the following upper bound:

$$\forall p_t \ t \in [b_i, b_{i+1} - 1], \ d_i(p_t, c_t) \leq \delta_i \leq \max \delta_i$$  (2)

This procedure can be interpreted as geometrical expression in Fig.1. The given series of output probability vectors is covered by small balls where the center of $i$-th ball is $c_t$ and the radius is $\delta_i$. The determination (Eq.(1)) is done by the distance between the center of balls and $q$. The precise segmentation procedure (ball covering procedure) of the given series of output probability vectors is described in the next section. Table 1 shows relationship between search threshold, $\theta$, and number of distance calculations.
calculations. Here, the search threshold, \( \theta \), corresponds to the radius of an oblique-lined ball in Fig. 1. The average \( L_2 \) distortion segmentation method was applied and a segmentation threshold is set to \( \epsilon = 0.0001 \). A symbol "Full AS" means fully Active searching case, and "M1+M3+AS" means the combination of above M1 and M3 and AS is applied. In order to indicate the improvement two ratios of (Full AS)/(M1+M3+AS) are calculated; one is a ratio of number of distance calculations and the other is a ratio of processing times. When the searching threshold, \( \theta \), becomes smaller, the improvement ratio becomes larger.

When \( \theta = 0.01 \), the ratio of the distance calculation and the processing times are 20.181 and 36.194, respectively.

3. Fixed Distortion Segmentation Method

This section describes fixed distortion segmentation methods which are applied in the first procedure in M1. "Fixed distortion" segmentation means that the radiiuses of all small balls are fixed or smaller than a specified threshold.

3.1. Maximum distortion segmentation method

From an arbitrary start position \( t_0 \) to an end position \( t_1 \) in \( X = (x_t) (t = 0, \ldots, T−1) \) \( L_1 \) distortion of subsequence, \( (x_{t}) (t = t_{0}, \ldots, t_{1}) \) is calculated and when the distortion is smaller than a segmentation threshold, \( \delta \), the end position is incremented. The maximum value of the end position \( t_1 \) and its center vector, \( c_i \), are determined. Here, this center vector, \( c_i \), corresponds to the center of the small ball and the distortion corresponds to \( \delta \) radius in Fig. 1. \[ t_1, c_i = \arg \max_{t_0 \leq t_1 < T} \left\{ \min_{t_0 \leq t \leq t_1} \| x_t - c_i \| \leq \delta \right\} \]

This procedure is called as "centroid method" because this requires a minimum distortion vector (centroid) based on the maximum \( L_1 \) distortion, however, it is not easy to solve this analytically. In spite of solving the centroid, searching the optimal vector, \( c_i \), in the subsequence is the finite combinatorial problem and can be solved using an exhaustive searching. This is called as "subsequence exhaustive search method".

\[ t_1, c_i = \arg \max_{t_0 \leq t \leq t_1} \left\{ \min_{t_0 \leq t \leq t_1} \| x_t - x_{m} \| \leq \delta \right\} \]

3.2. Average distortion segmentation method

The above method has the heavy computation amounts because of the exhaustive searching. The above problem is replaced with the problem where "maximum" is replaced by "average" and "\( \delta \)" is replaced by "\( \epsilon \)" norm. Therefore, the problem is replaced with that an average \( L_2 \) distortion is smaller than a threshold, \( \epsilon \). Here, the \( L_2 \) centroid is equal to the arithmetic mean vector of all vectors in the subsequence. Therefore, "average \( L_2 \) distortion" segmentation method can be solved analytically with small computation complexity.

\[ t_1 = \arg \min_{t_0 \leq t_1 < T} \frac{1}{t_1 - t_0 + 1} \sum_{t = t_0}^{t_1} \| x_t - x_{m} \|_2^2 \leq \epsilon \]

Initially \( t_0 = t_0 = 0 \), after determining \( t_1 \) using the above procedure, the next start position is set to \( t_1 = t_1 + 1 \), then iteratively \( t_1 \) is determined. This sequence corresponds to \( b_i \) in Eq.(2) and the average vector, \( x_{m} \), of the \( i \)-th segment corresponds to \( c_i \). To simplify the explanation, changing the position variable, \( t \), we can assume that \( t_0 = 1 \). Considering \( \| x_{m} \|^2 = \sum_{m} x_{m}^2 \) average \( L_2 \) distortion, \( D_n \), from the \( L_2 \) centroid is equal to the summation of \( L_2 \) distortion for all dimensions. Furthermore, considering each \( L_2 \) distortion is equal to the variance, the following Eq.(7) can be derived.

\[ x_{m} = \frac{1}{n} \sum_{t=1}^{n} x_{m,t} \]

\[ D_n = \frac{1}{n} \sum_{t=1}^{n} (x_{m} - x_{m,t})^2 = \frac{1}{n} \sum_{t=1}^{n} \sum_{m} (x_{m,t} - x_{m})^2 \]

\[ = \sum_{m} \left( \frac{1}{n} \sum_{t=1}^{n} x_{m,t}^2 - x_{m}^2 \right) = \alpha_n - \beta_n \]

\[ \alpha_n = \sum_{m} \frac{1}{n} \sum_{t=1}^{n} x_{m,t}^2, \quad \beta_n = \sum_{m} x_{m}^2 \]

3.3. Fixed length segmentation method

Fixed length segmentation method simply segments in a given number of frames based on Eq.(5). The probability distance value between a middle point \( m = (t_0 + t_1)/2 \) of the given segment \( [t_0, t_1] \) and an arbitrary point \( t \) is bounded as follows;

\[ \| x_{t} - x_{m} \|^2 \leq \max_{t_0 \leq t \leq t_1} \| x_{t} - x_{m} \| \]

\[ \leq \frac{\max_{t_0 \leq t \leq t_1} 2|t - m|}{L} = \max_{t = t_0} \frac{2|t - 2m|}{L} \]

\[ = \left\{ \begin{array}{ll} \frac{t_1 - t_0}{L} & (t_0 + t_1 = 0 \mod 2) \\ \frac{t_1 - t_0 + 1}{L} & (t_0 + t_1 = 1 \mod 2) \end{array} \right. \]

Here, \( t_0 + t_1 = t_1 - t_0 \mod 2 \). For example, segmentation which satisfies \( t_1 = t_0 \leq \delta L, \delta L, t_0 = 0 \mod 2 \) gives \( t_1 \).
3.4. Relationship between average $l_1$ and $l_2$ distortion

The relationship between the average $l_1$ distortion and average $l_2$ distortion can be derived by using the following inequality between $l_1$ and $l_2$ norms:

$$\frac{1}{n} \sum_{i=1}^{n} ||x_i - x_0||_1 \leq \sqrt{M} \frac{1}{n} \sum_{i=1}^{n} ||x_i - x_0||_2 \leq \sqrt{M}$$

Furthermore, a same inequality for the number of addition in the sum the following inequality can be derived:

$$\frac{1}{n} \sum_{i=1}^{n} ||x_i - x_0||_2 \leq \sqrt{M} \frac{1}{n} \sum_{i=1}^{n} ||x_i - x_0||_1 \leq \sqrt{M}$$

4. Experimental Evaluation

4.1. Experimental condition

The first data of CampusWave sound database[8] is used for the evaluation experiment. This database is a series of music request programs which local FM radio station broadcasted, and each 1 hour length sound contains that dialogue voice intervals by two female personalities, requested musics (mainly pops musics), commercial sounds, and so on. The experimental setup is shown in Table 2. Sound signal is transformed to a series of LPC analysis. The searching query segments is set to 10 second length. In this experiment each searching segment is a part of commercial sound interval and appears only once. The number of searching segments is 33. The VQ codebook with 32 codes is generated from the same sound data using LBG algorithm. The computer specification is Pentium4(2.66GHz), 512MB memory size, OS Vine Linux 3.1, gcc 3.3.2 with O3 optimization option.

4.2. Experimental Results

Table 3 shows the relationship between segmentation threshold and number of segments (maximum, minimum and average lengths of the segments). Here maximum $l_1$ distortion method is applied. Table 4 shows the relationship between the segmentation threshold and the number of segments in average $l_2$ distortion method. As the minimum length satisfies the inequality $\geq \delta L_c$ the minimum length is relatively larger than that given by average $l_2$ distortion method.

Fig.2 shows the relationship between number of segments and number of distance calculation in three methods. This shows that searching times for the first and the second methods are almost same when the numbers of the segments are same. The third method is slightly larger than the others. Fig.3 shows the relationship between the segmentation thresholds $(\delta \sqrt{M})$ and the number of segments. The bold line represents the result by the maximum $l_1$ segmentation method and the dashed line represents the result by the average $l_2$ segmentation method. This figure shows two methods give almost same number of segments when $\sqrt{M} - \delta = c > 0 (c = 0.015)$. Eq.(8) estimates the value of threshold, $\epsilon$, which produces the same number of segments as $\delta$.

$$\epsilon = \frac{1}{M} (\delta + c)^2$$

Combining with Fig.2, average $l_2$ segmentation method can give the same searching performance as maximum $l_1$ segmentation method with the specified threshold, $\delta$.

Table 5 compares the processing times for three segmentation methods. The parameters in each method are set to give the almost equal number of segments. The fixed length segmentation is fastest and the maximum $l_1$ segmentation method is very slow. Including the processing time of distance matrix calculation in M3, the processing times using fixed length and average $l_2$ are 17.040 and 17.310, respectively, and the processing times of those two methods are not so significantly different.
5. Conclusions

As the fixed distortion segmentation method, this paper compared three methods; the maximum $l_1$ distortion segmentation, average $l_2$ distortion segmentation and fixed length segmentation method. Average $l_2$ distortion and fixed length segmentation are very efficient. On the other hand, $l_1$ distortion segmentation does not require radius information. The experimental results show the first two methods have almost equal performance in the segment searching when the number of segments are the same. As future plans efficient radius information keeping in average $l_2$ distortion method and the improvement of the maximum $l_1$ distortion method will be investigated.

6. References