DELTA-MSE DISSIMILARITY IN GLA BASED VECTOR QUANTIZATION

Mantao Xu

Department of Computer Science, University of Joensuu
Box 111, Fin-80101 Joensuu
FINLAND

ABSTRACT

The *generalized Lloyd algorithm* is one of popular partition-based algorithms to construct the codebook in vector quantization. We propose the Delta-MSE dissimilarity measurement between training vectors and code vectors based on the MSE distortion function. The Delta-MSE function is heuristically derived by calculating the difference of MSE distortion before and after moving a training vector from one cluster to another. We show that the Delta-MSE dissimilarity applies also to minimizing the F-ratio validity index of the vector quantizer. We incorporate the underlying dissimilarity into the generalized Lloyd algorithm in vector quantization with the initial codebook derived from the PCA-based k-d tree algorithm. Experimental results show that the proposed dissimilarity generally achieves better performance than the L2 distance in constructing the codebook of vector quantization.

1. INTRODUCTION

Vector quantization (VQ) s a method for data reduction that is widely used in low bit rate compression of image and audio data source [1, 2]. The objective of vector quantization is to search a *M* set of code vectors (codebook) with the minimum distortion between training vectors and their representative code vectors. One of most cited partition-based algorithms is the *generalized Lloyd algorithm* (GLA). It basically consists of two steps: the assignment of each training vector with a class label by finding its closest code vector and the computation of code vectors. There are many improved versions of the GLA algorithm such as the genetic GLA algorithms [3], the randomized local search algorithms [4, 5] and the fast implementations of GLA [6-7]. The standard GLA is applied as an integral part of the vector quantization algorithms above. Either the genetic algorithms or the randomized algorithms run the GLA algorithm many times during one run of the algorithms. The computation in the GLA mainly relies on the distance calculations between the training vectors and the code vectors. The fast implementations of GLA such as PDS [8] and MPS [9] reduce a number of distance calculations after several runs of partition in GLA. The GLA vector quantization algorithm can be also considered as a clustering algorithm on training sets. Hence its dissimilarity function or distance function can be reformulated to improve vector quantization performance.

The distortion function of VQ is always defined by the total dissimilarities between all training vectors and their code vectors. The definition of a new dissimilarity function often leads to the re-formalization of the distortion function, which also requires that the code vectors are re-computed consistently to minimize the distortion function. However, in this work, a heuristic and non-symmetric dissimilarity function is analytically induced from the predefined distortion function. The considered approach takes account into the dynamic nature of the GLA partition process, in which the cluster parameters (the cluster sizes) are subject to change all the time during the run of the algorithm. The above design paradigm can be applied to the *MSE* distortion function to derive a dissimilarity function between training vectors and code vectors.

The structure of this paper is organized as follows: We first describe the design paradigm of the Delta-MSE dissimilarity based on one partition of training vectors. In the following section, we show that the Delta-MSE dissimilarity is applicable also to the F-ratio clustering validity index. Then the algorithm is Delta-MSE dissimilarity incorporated into the GLA algorithm in next section. In experimental section, performance comparisons are reviewed between the Delta-MSE dissimilarity and the L_2 distance. Finally, the conclusions are drawn.

2. DELTA-MSE DISSIMILARITY

The aim of vector quantization is find the partition of the training set with the minimum distortion between all training vectors and their code vectors. The standard GLA vector quantization is an optimization problem specified by the minimization of the *MSE* function:

$$MSE(P,C) = \sum_{i=1}^{N} \| x_i - c_{p(i)} \|^2 \quad (1)$$

where

N is the number of data vectors;
k is the number of clusters (*NOC*);
$X = \{ x_1, x_2, \ldots\ldots x_N \}$ is a set of N training vectors;
$P = \{ p_i \mid i = 1, \ldots\ldots N \}$ is the set of class labels;
$C = \{ c_j \mid j = 1, \ldots k \}$ is the set of code vectors.

Assuming that a training vector x is moved from cluster i to cluster j, the change of the *MSE* function [10] caused by this move is:

$$v_{ij}(x) = \frac{n_j}{n_j + 1} \| x - c_j \|^2 - \frac{n_i}{n_i - 1} \| x - c_i \|^2 \quad (2)$$

where n_i and n_j are two cluster sizes respectively. The first part in the right hand side of equation (2), representing the increased value of the total variance of cluster j caused by this move, is denoted the *addition cost*. The second part, representing the decreased value of total variance of cluster i, is denoted the *removal cost*. The *addition cost* can be interpreted as the dissimilarity between training vector x and code vector c_j (x is outside cluster j). A smaller cluster size n_j obviously makes the *addition cost* more different from the L2 square distance.

It should be noted that the change of variance by adding a training vector into one cluster is equivalent to the change of variance by removing the training vector from the new cluster. Hence the second part can be interpreted as the dissimilarity between training vector x and its former code vector c_i. Obviously, the training vectors in sparse clusters are moved frequently by the dissimilarity than those in dense clusters. The Delta-MSE dissimilarity between training vector x_i and code vector c_j is defined as:

$$D_{MSE}(x_i, c_j) = w_{ij} \cdot \| x_i - c_j \|^2 \quad (3)$$

where w_{ij} is defined as:

$$w_{ij} = \begin{cases} n_j / (n_j + 1) & p(i) \neq j \\ n_j / (n_j - 1) & p(i) = j \end{cases} \quad (4)$$

The square L2 distance in equation (3) can be replaced with the standardized L2 distance as:

$$D_{MSE}(x_i, c_j) = w_{ij} \cdot (x_i - c_j)^T D^{-1} (x_i - c_j) \quad (5)$$

where D is the diagonal matrix with diagonal elements given by v_j, which denotes the variance of the variable x^j over the N training vectors. The distribution of cluster sizes determines the clustering performance of the Delta-MSE dissimilarity. The sparser one cluster is, the more different the Delta-MSE dissimilarity can be in comparison to the L2 norm. When the codebook szie is increased, most of clusters become sparser. In this case, the proposed dissimilarity enables more reassignments of the training vectors in sparse clusters, consequently increasing the number of vector reassignments. The Delta-MSE dissimilarity therefore yields the better VQ distortion than the L2 distance.

3. F-RATIO VALIDITY INDEX

Many iterative clustering algorithms rely on F-ratio validity index in estimation of the codebook size. The F-ratio in is defined as the ratio of within-groups variance to between-groups variance. The total variance of the training set can be decomposed into the sum of within-groups variance and between-groups variance as:

$$\sigma(X) = \sum_{i=1}^{N} \| x_i - c_{p(i)} \|^2 + \sum_{j=1}^{k} n_j \| c_j - \bar{x} \|^2 \quad (6)$$

where $\bar{x}$ is the mean vector of training set. The F-ratio is an extension of Fisher's discriminant to measure the separability between clusters. The F-ratio clustering validity is calculated as the ratio of the total within-groups variance against the total between-groups variance as:

$$F = \frac{k \cdot \sum_{i=1}^{N} \| x_i - c_{p(i)} \|^2}{\sum_{j=1}^{k} n_j \| c_j - \bar{x} \|^2} = \frac{k \cdot MSE}{\sigma(X) - MSE} \quad (7)$$

The smaller the F-ratio is, the more separated the clusters are. The F-ratio validity index is useful in the estimation of codebook size, which also relies on the geometrical structure of training source.

Since the Delta-MSE dissimilarity is analytically derived from the MSE distortion, if the *removal cost* D_{MSE} (x, c_i) is greater than the *addition cost* D_{MSE} (x, c_j), the *MSE* distortion will decrease after this movement. In the partition phase, the training vector x is inclined to move into the cluster with the minimum addition cost, which brings the greatest decrease of *MSE* value. In the following, we will show that the property holds on to the F-ratio validity index as well.

Lemma: Given the partition of the training set that assigns training vector x into cluster i, if the *addition cost* $D_{MSE}(x, c_j)$ is greater than the *addition cost* $D_{MSE}(x, c_l)$, the F-ratio $F(x, c_j)$ after moving x to cluster j is greater than the F-ratio value $F(x, c_l)$ after moving x to cluster l.

Proof of Lemma: Suppose that training vector x is moved from cluster i to cluster j and cluster l respectively. From equation (3) and (7), the difference of $F(x, c_j)$ and F(x, c_l) is calculated as:

$$
\begin{aligned}
&F(x,c_j) - F(x,c_l) \\
&= \frac{k \cdot MSE(x,c_j)}{\sigma(X) - MSE(x,c_j)} - \frac{k \cdot MSE(x,c_l)}{\sigma(X) - MSE(x,c_l)} \\
&= \frac{k \cdot \sigma(X) \cdot (MSE(x,c_j) - MSE(x,c_l))}{(\sigma(X) - MSE(x,c_j)) \cdot (\sigma(X) - MSE(x,c_l))} \qquad (8) \\
&= \frac{k \cdot \sigma(X)(D_{MSE}(x,c_j) - D_{MSE}(x,c_l))}{(\sigma(X) - MSE(x,c_j)) \cdot (\sigma(X) - MSE(x,c_l))}
\end{aligned}
$$

The total variance $\sigma(X)$ is a positive constant; $\sigma(X)$-$MSE(x, j)$ and $\sigma(X)$-$MSE(x, l)$, representing the between-group variances after the two movements respectively, are also positive. Thus, the value of equation (10) is positive if and only if $D_{MSE}(x, c_j)$ is greater than $D_{MSE}(x, c_l)$, which proves the lemma.

4. IMPLEMENTATION OF GLA ALGORITHM

The Delta-MSE dissimilarity is incorporated into the generalized Lloyd algorithm in this work. The incorporated GLA algorithm can also be accelerated by the triangular inequality elimination technique (TIE) by Chen and Hsieh [11]. The values of all weight numbers $\{w_{ij} \mid i = 1, \ldots N, j = 1, \ldots k\}$ are reserved in two k-dimensional arrays in each partition phase. The partition of training vectors by the Delta-MSE dissimilarity can also be exactly accelerated by application of two triangular inequalities. The number of Delta-MSE calculations can be reduced by the following two inequalities:

$$
\begin{aligned}
&\| c_a - c_j \| > \left(1 + \sqrt{\frac{w_{ia}}{w_{ij}}}\right) \cdot \| c_a - x \| \\
&\text{abs}(\| c_b - x \| - \| c_b - c_j \|) > \sqrt{\frac{w_{ia}}{w_{ij}}} \cdot \| c_a - x \|
\end{aligned} \qquad (9)
$$

where x is training vector, c_a and c_b are its nearest code vector and farthermost code vector found so far; c_j is the code vector to be detected. If one of the above equations holds, the calculation of $D_{MSE}(x, c_j)$ can be avoided. A practical implementation of the acceleration utilizes the $k \times k$ matrix of the L2 distances between code vectors. The calculation of the matrix usually takes $O(k^2 d)$ time. Assuming that $k << N$, the accelerated partition with the Delta-MSE dissimilarity takes $O(d(k\text{-}s)N)$ time where s is the average number of avoided Delta-MSE calculations in reassignments of all training vectors.

The initial code vectors here are chosen by a *k-d tree* algorithm based on the nested principal component analysis, which is proposed in [12-13]. The code vectors are selected as k number of *k-d tree* bucket centers. The *k-d tree* algorithm ensures that its bucket centers can be as appropriate candidate code vectors as training vectors. The time complexity of selecting the initial code vectors from *k-d tree* buckets is $O(dkN)$.

Table 1: Comparison between the L2 distance and the Delta-MSE dissimilarity

Dataset		*MSE*	*F-ratio*
Air5	*L2*	8.381	1.908
	Delta-MSE	8.298	1.888
Bridge	*L2*	269.2	6.240
	Delta-MSE	267.0	6.185
Bridge2	*L2*	5744	57.72
	Delta-MSE	5654	55.88
Camera	*L2*	202.4	3.177
	Delta-MSE	190.5	2.974
Housec5	*L2*	5.202	1.683
	Delta-MSE	5.119	1.656

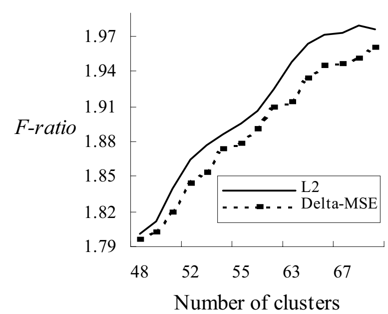

Fig. 1: F-ratios of the vector quantizers for *Air5*

5. EXPERIMENTAL RESULTS

We first study the training sets generated from four standard images: *Air5* and *Housec5* are the training sets with the RGB-values from image *Airplane* and *House* - quantized to 5 bits per color; *Bridge* and *Camera* are the training sets with 4×4-blocks from image *Bridge* and *Cameraman*; *Bridge2* is the training set with 4×4 binarized blocks from image *Bridge*. Both L2 norm and Delta-MSE dissimilarity are tested in the standard GLA algorithm. The initial code vectors are selected from the PCA-based *k-d tree* bucket centers. The average *MSE* and F-ratio values over the codebook size from 48 to 70 are displayed in table 1. The F-ratios of *Air5* are plotted against the number of clusters (codebook size) in figure 1.

It turns out in figure 1 that the Delta-MSE dissimilarity achieves significantly smaller *F-ratio* distortions than the L2 distance with the increase of codebook size. The clusters become sparser with the increased codebook size, which makes the Delta-MSE dissimilarity more different from the L2 square distance and consequently enables more heuristicity for minimizing the distortions. Table 1 shows the proposed dissimilarity generally performs better than the L2 norm in the GLA based vector quantization.

Table 2: Comparison between the standardized L2 distance and its corresponding Delta-MSE dissimilarity.

Dataset		*MSE*	*F-ratio*	*Test Error*
Speaker1	*L2*	6712	22.57	5.888
	Delta-MSE	4850	21.83	5.851
Speaker2	*L2*	6379	22.82	5.535
	Delta-MSE	6289	22.31	5.529
Speaker3	*L2*	4596	22.89	5.118
	Delta-MSE	4474	21.95	5.102
Speaker4	*L2*	4614	21.84	5.546
	Delta-MSE	4534	21.11	5.552
Speaker5	*L2*	3208	19.82	4.941
	Delta-MSE	3158	19.34	4.904

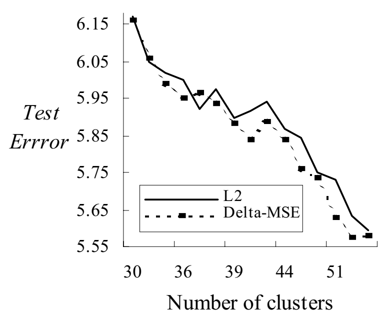

Fig. 2: Test errors of the vector quantizers for *Speaker1*.

We secondly study the five real speaker datasets from TIMIT speech corpus by using the stepwise GLA algorithm. The Delta-MSE dissimilarity in equation (5) and the standardized L2 distance are investigated in the GLA algorithm. Each dataset is separated into a training set and a test set (about 25%: 75%). The two dissimilarities are incorporated into the stepwise GLA algorithm. Then the vector quatnizers are tested by their test sets. The average test errors are shown in table 2. The average *MSE* and F-ratio values displayed in the table are calculated over the codebook size from 30 to 55. Figure 2 plots the test error of the vector quantizers for *Speaker1* against their codebook size. It turns out that the proposed dissimilarity generally has better performance than the standardized L2 distance in the GLA based vector quantization. With the increase of the codebook size, the performance gains of the proposed dissimilarity are increased.

6. CONCLUSIONS

We have proposed Delta-MSE dissimilarity function between training vectors and code vectors. The dissimilarity function is calculated as the change of within-group variance before and after moving a given training vector from one class to another. The dissimilarity function provides the heuristic direction of training vector movements, in which the VQ distortion function will possibly decrease most. Although derived from the MSE distortion, the dissimilarity function applies also to the minimization of the F-ratio validity index.

The experimental results show that the proposed dissimilarity undermines more reassignments of training vectors than the L2 norm. With the increase of codebook size, the performance gains of the GLA algorithm is increased as well.

The weakness of the Delta-MSE dissimilarity function lies in its non-symmetric formalization. The cluster size can be one of the dominant factors only if the cluster is sparse enough.

7. REFERENCES

[1] N. M. Nasrabadi and R. A. King, “Image coding using vector quantization: A review,” *IEEE Trans. Commun*, 36(8): pp. 957--971, August 1988.

[2] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*, KLUWER, Boston, MA., 1992.

[3] P. Fränti, “Genetic algorithm with deterministic crossover for vector quantization,” *Pattern Recognition Letters*, Elsevier, 21 (1), pp. 61-68, 2000.

[4] P. Fränti and J. Kivijärvi, “Randomized local search algorithm for the clustering problem,” *Pattern Analysis and Applications*, 3 (4), 358-369, 2000.

[5]. P. Fränti, J. Kivijärvi and O. Nevalainen: “Tabu search algorithm for codebook generation in VQ,” *Pattern Recognition*, 31 (8), 1139-1148, August 1998.

[6] T. Kaukoranta, P. Fränti and O. Nevalainen, “A fast exact GLA based on code vector activity detection,” *IEEE Trans. on Image Processing*, 9 (8), 1337-1342, August 2000

[7] J. Z. C. Lai and C. C. Lue, “Fast Search Algorithms for VQ Codebook Generation,” *Journal of Visual Communication and Image Representation*, (7) 2, pp. 163-168, June 1996.

[8] C. D. Bei and R. M. Gray, “An improvement of the minimum distortion encoding algorithm for vector quantization,” *IEEE Trans. Commun.*, vol. 33, pp. 1132–1133, Oct. 1985.

[9] S. W. Ra and J. K. Kim, “A fast mean-distance-ordered partial codebook search algorithm for image vector quantization,” *IEEE Trans. Circuits Syst.*, vol. 40, pp. 576–579, Sept. 1993.

[10] H. Späth, *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*, Ellis Horwood Publ., Chichester, U.K., 1980.

[11] S.-H. Chen and W. M. Hsieh, “Fast algorithm for VQ codebook design,” *Proc. Inst. Elect. Eng.*, vol. 138, pp. 357–362, Oct. 1991.

[12] R. F. Sproull. “Refinement to nearest-neighbour searching in k-d trees,” *Algorithmica*, No.6, pp. 579--589, 1991.

[13] A. Likas, N. Vlassis and J. J. Verbeek, “The global k-means clustering algorithm,” *Pattern Recognition*, 36 (2), pp. 451-461, 2003.