

Accelerating Competitive Learning Graph Quantization

Brijnesh J. Jain¹ and Klaus Obermayer¹

¹Berlin Institute of Technology, Germany
{bjb|oby}@cs.tu-berlin.de

Abstract. Vector quantization(VQ) is a lossy data compression technique from signal processing for which simple competitive learning is one standard method to quantize patterns from the input space. Extending competitive learning VQ to the domain of graphs results in competitive learning for quantizing input graphs. In this contribution, we propose an accelerated version of competitive learning graph quantization (GQ) without trading computational time against solution quality. For this, we lift graphs locally to vectors in order to avoid unnecessary calculations of intractable graph distances. In doing so, the accelerated version of competitive learning GQ gradually turns locally into a competitive learning VQ with increasing number of iterations. Empirical results show a significant speedup by maintaining a comparable solution quality.

1 Introduction

Vector quantization is a classical technique from signal processing suitable for lossy data compression, density estimation, and prototype-based clustering [10]. The aim of optimal vector quantizer design is to find a codebook of a given size k such that an expected distortion with respect to some (differentiable) distortion measure is minimized. Competitive learning VQ is one standard method for optimal vector quantizer design [5, 14, 30].

Graph quantization as formalized in [22] is a generalization of vector quantization to the quantization of combinatorial structures that can be represented by attributed graphs. The generalization from vectors to graphs opens applications to diverse domains like proteomics, chemoinformatics, and computer vision, where the patterns to be quantized are more naturally represented by trees, lattices, and graphs.

Designing optimal graph quantizers has been pursued as central clustering. Examples include competitive learning algorithms in the domain of graphs [12, 13, 18, 21] and k-means as well as k-medoids algorithms [7, 8, 18, 19, 22, 23, 28, 29]. A key problem of all these algorithms is that they are slow in practice for large datasets of graphs, because calculating the underlying graph distance (distortion) measure is a graph matching problem of exponential complexity. Given a training set of N graphs, competitive learning, k-means, and k-median calculate (or approximate) at least kN intractable graph distances during each cycle through the training set.

In this contribution, we propose an accelerated version of competitive learning GQ. We assume that the underlying distortion measure is a geometric graph metric, which arises in various different guises as a common choice of proximity measure in a number of applications [1–3, 11, 31, 33]. The proposed accelerated version of competitive learning GQ avoids unnecessary graph distance calculations by exploiting metric properties and by lifting the graphs isometrically to an Euclidean vector space whenever possible. Lifting graphs to vectors reduces competitive learning GQ to competitive learning VQ locally. If lifting becomes unfeasible, we switch back to competitive learning GQ. By switching back and forth between competitive learning GQ and competitive learning VQ locally, the accelerated version of competitive learning GQ reduces gradually to the efficient competitive learning algorithm for VQ.

The proposed accelerated version of competitive learning GQ has the following properties: First, it can be applied to finite combinatorial structures other than graphs like, for example, point patterns, sequences, trees, and hypergraphs. For the sake of concreteness, we restrict our attention exclusively to the domain of graphs. Second, any initialization method that can be used for competitive learning GQ can also be used for its accelerated version. Third, hierarchical methods for GQ benefit from the proposed accelerated scheme. Fourth, competitive learning GQ and its accelerated version perform comparable with respect to solution quality. Different solutions are caused by approximation errors of the graph matching algorithm and by multiple local minima but are not caused by the mechanisms to accelerate competitive learning.

The paper is organized as follows. Section 2 briefly describes competitive learning GQ. Section 3 proposes an accelerated version of competitive learning GQ. Experimental results are presented and discussed in Section 4. Finally, Section 5 concludes with a summary of the main results and future work.

2 Competitive Learning Graph Quantization

This section briefly introduces competitive learning as a design technique for graph quantization. For details on graph quantization and consistency results of competitive learning, we refer to [22].

2.1 Metric Graph Spaces

Let \mathbb{E} be a r -dimensional Euclidean vector space. An (*attributed*) *graph* is a triple $X = (V, E, \alpha)$ consisting of a finite nonempty set V of *vertices*, a set $E \subseteq V \times V$ of *edges*, and an *attribute function* $\alpha : V \times V \rightarrow \mathbb{E}$, such that $\alpha(i, j) = \mathbf{0}$ for each pair of distinct vertices i, j with $(i, j) \notin E$.

For simplifying the mathematical treatment, we assume that all graphs are of order n , where n is chosen to be sufficiently large. Graphs of order less than n , say $m < n$, can be extended to order n by including isolated vertices with attribute zero. For practical issues, it is important to note that limiting the maximum order to some arbitrarily large number n and extending smaller graphs to graphs of

order n are purely technical assumptions to simplify mathematics. For machine learning problems, these limitations should have no practical impact, because neither the bound n needs to be specified explicitly nor an extension of all graphs to an identical order needs to be performed. When applying the theory, all we actually require is that the graphs are finite.

A graph X is completely specified by its *matrix representation* $\mathbf{X} = (\mathbf{x}_{ij})$ with elements $\mathbf{x}_{ij} = \alpha(i, j)$ for all $1 \leq i, j \leq n$. By concatenating the columns of \mathbf{X} , we obtain a *vector representation* \mathbf{x} of X .

Let $\mathcal{X} = \mathbb{E}^{n \times n}$ be the Euclidean space of all $(n \times n)$ -matrices and let \mathcal{T} denote a subgroup of the set \mathcal{P}^n of all $(n \times n)$ -permutation matrices. Two matrices $\mathbf{X} \in \mathcal{X}$ and $\mathbf{X}' \in \mathcal{X}$ are said to be equivalent, if there is a permutation matrix $P \in \mathcal{T}$ such that $P^\top \mathbf{X} P = \mathbf{X}'$. The quotient set

$$\mathcal{X}_{\mathcal{T}} = \mathcal{X} / \mathcal{T} = \{[\mathbf{X}] : \mathbf{X} \in \mathcal{X}\}$$

is a *graph space* of all *abstract graphs* $[\mathbf{X}]$ over the *representation space* \mathcal{X} induced by the transformation group \mathcal{T} . Note that the graph space $\mathcal{X}_{\mathcal{T}}$ is a Riemannian orbifold. The notion of orbifold is fundamental to extend analytical and geometrical properties of Euclidean spaces to graph spaces. For details, we refer to [22].

In the remainder of this contribution, we identify \mathcal{X} with \mathbb{E}^N ($N = n^2$) and consider vector- rather than matrix representations of abstract graphs. By abuse of notation, we sometimes identify X with $[\mathbf{x}]$ and write $\mathbf{x} \in X$ instead of $\mathbf{x} \in [\mathbf{x}]$. We say, X *lifts* to \mathbf{x} if we represent graph X by vector $\mathbf{x} \in X$.

Finally, we equip our graph space with a metric. Let $\|\cdot\|$ be a Euclidean norm on \mathcal{X} . Then the distance function

$$d(X, Y) = \min \{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x} \in X, \mathbf{y} \in Y\},$$

is a metric. It is well known that calculating the graph distance metric $d(X, Y)$ is a NP-complete graph matching problem [11].

Since \mathcal{T} is a group, we have

$$d_{\mathbf{y}}(X) = \min \{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x} \in X\} = d(X, Y),$$

where \mathbf{y} is an arbitrary vector representation of Y . By symmetry, we similarly have $d_{\mathbf{x}}(Y) = d(Y, X)$. Hence, the graph distance $d(X, Y)$ can be determined by fixing an arbitrary vector representation $\mathbf{y} \in Y$ and then finding a vector representation from X that minimizes $\|\mathbf{x} - \mathbf{y}\|$ over all vector representations \mathbf{x} from X .

A pair $(\mathbf{x}, \mathbf{y}) \in X \times Y$ of vector representations is called *optimal alignment* if $\|\mathbf{x} - \mathbf{y}\| = d(X, Y)$. Thus, we have $d(X, Y) \leq \|\mathbf{x} - \mathbf{y}\|$ for all vector representations $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, where equality holds if and only if \mathbf{x} and \mathbf{y} are optimally aligned.

Algorithm 1: Competitive Learning GQ

```
01  choose an initial codebook  $\mathcal{C} = \{Y_1, \dots, Y_k\} \subseteq \mathcal{X}_{\mathcal{T}}$ 
02  choose arbitrary vector representations  $\mathbf{y}_1 \in Y_1, \dots, \mathbf{y}_k \in Y_k$ 
03  repeat
04      randomly select an input graph  $X \in \mathcal{X}_{\mathcal{T}}$ 
05      let  $Y_X = \arg \min_{Y \in \mathcal{C}} d(X, Y)^2$ 
06      choose  $\mathbf{x} \in X$  optimally aligned with  $\mathbf{y} \in Y_X$ 
07      determine learning rate  $\eta_i > 0$ 
08      update  $\mathbf{y} = \mathbf{y} + \eta(\mathbf{x} - \mathbf{y})$ 
09  until some termination criterion is satisfied
```

2.2 Graph Quantization

Let $(\mathcal{X}_{\mathcal{T}}, d)$ be a metric graph space over the Euclidean space $(\mathcal{X}, \|\cdot\|)$. A *graph quantizer* of size k is a mapping of the form

$$Q : \mathcal{X}_{\mathcal{T}} \rightarrow \mathcal{Y}_{\mathcal{T}},$$

where $\mathcal{C} = \{Y_1, \dots, Y_k\} \subseteq \mathcal{X}_{\mathcal{T}}$ is a *codebook*. The elements $Y_j \in \mathcal{Y}_{\mathcal{T}}$ are the *code graphs*.

Adaptive graph quantization design aims at finding a codebook $\mathcal{C} = \{Y_1, \dots, Y_k\} \subseteq \mathcal{X}_{\mathcal{T}}$ such that the *expected distortion*

$$D(\mathcal{C}) = \sum_{j=1}^k \int_{\mathcal{X}_{\mathcal{T}}} \min_{1 \leq j \leq k} d(X, Y_j)^2 f(X) dX$$

is minimized, where $\mathcal{C} \in \mathcal{X}_{\mathcal{T}}^k$ and $f = f_{\mathcal{X}_{\mathcal{T}}}$ is a probability density defined on some measurable space $(\mathcal{X}_{\mathcal{T}}, \Sigma_{\mathcal{X}_{\mathcal{T}}})$.

A statistically consistent method to minimize the expected distortion using empirical data is competitive learning as outlined in Algorithm 1.

3 Accelerating Competitive Learning GQ

This section proposes an acceleration of competitive learning GQ that is based on a fixed training set $\mathcal{S} = \{X_1, \dots, X_N\}$ consisting of N independent graphs X_i drawn from $\mathcal{X}_{\mathcal{T}}$. At each cycle through the training set, competitive learning GQ as described in Algorithm 1 calculates kN graph distances, each of which is NP-hard. These distance calculations predominate the computational cost of Algorithm 1. Accelerating competitive learning GQ therefore aims at reducing the number of graph distance calculations.

Accelerating competitive learning GQ is based on the key idea that a small change of a vector representation $\mathbf{y} \in Y$ of a code graph $Y \in \mathcal{C}$ does not change

already optimally aligned vector representations $\mathbf{x} \in X$ of training graphs $X \in \mathcal{S}$ quantized by (closest to) Y . When keeping track of the most recent optimal alignment, calculating graph distances reduces to calculating Euclidean distances of vector representations, in particular in the final convergence phase. Thus, in order to avoid graph distance calculations, we lift graphs to their optimally aligned vector representations — if possible — and switch to Euclidean distance calculations.

To describe the accelerated version of Algorithm 1, we assume that $X \in \mathcal{S}$ is a training graph and $Y, Y_X \in \mathcal{C}$ are code graphs. By $Y_X = Q(X)$ we denote the *encoding* of X . Since the graph distance d is a metric, we have

$$u(X) \leq l(X, Y) \Rightarrow d(X, Y_X) \leq d(X, Y), \quad (1)$$

where

1. $u(X) \geq d(X, Y_X)$ denotes an upper bound of $d(X, Y_X)$ and
2. $l(X, Y) \leq d(X, Y)$ denotes a lower bound of $d(X, Y)$.

From Eqn. (1) follows that we can avoid calculating a graph distance $d(X, Y)$ if at least one of the following two conditions is satisfied

- (C_1) $Y = Y_X$
- (C_2) $u(X) \leq l(X, Y)$.

A code graph Y is a *candidate encoding* for X if both conditions (C_1) and (C_2) are violated. In this case, we apply the technique of *delayed distance evaluation*. For this, we proceed as follows:

1. We first test whether the upper bound $u(X)$ is out-of-date. An upper bound is out-of-date if $u(X) \geq d(X, Y_X)$. Note that we can check the state of $u(X)$ without calculating the distance $d(X, Y_X)$.
2. If $u(X)$ is out-of-date, we calculate the distance $d(X, Y_X)$ and set $u(X) = d(X, Y_X)$. Since improving the upper bound $u(X)$ might eliminate Y as being a candidate encoding for X , we recheck condition (C_2).
3. If condition (C_2) is still violated despite the updated upper bound $u(X)$, we have the following situation

$$u(X) = d(X, Y_X) > l(X, Y).$$

We update the lower bound by calculating $l(X, Y) = d(X, Y)$ and then re-examine condition (C_2).

4. If condition (C_2) remains violated, we have

$$u(X) = d(X, Y_X) > d(X, Y) = l(X, Y).$$

This implies that X is closer to code graph Y than to its current encoding Y_X . In this case Y becomes the new encoding of X .

Algorithm 2: Accelerated Competitive Learning GQ

```
01  choose an initial codebook  $\mathcal{C} = \{Y_1, \dots, Y_k\} \subseteq \mathcal{X}_{\mathcal{T}}$ 
02  choose arbitrary vector representations  $\mathbf{y}_1 \in Y_1, \dots, \mathbf{y}_k \in Y_k$ 
03  set  $u(X) = \infty$  and declare  $u(X)$  out-of-date for all  $X \in \mathcal{S}$ 
04  repeat
05      store  $Y$  in  $Y'$  for all  $Y \in \mathcal{C}$ 
06      repeat
07          randomly select  $X \in \mathcal{S}$ 
08           $Y_X = \text{CLASSIFY}(X)$ 
09          determine learning rate  $\eta > 0$ 
10          set  $\mathbf{y} = \mathbf{y} + \eta(\mathbf{x}_a - \mathbf{y})$ 
11           $\text{ESTIMATE\_BOUNDS}()$ 
12  until some termination criterion is satisfied
```

Algorithm 3: $\text{CLASSIFY}(X)$

```
01  for each  $Y \in \mathcal{C}$  do
02      if  $Y$  is a candidate encoding for  $X$ 
03          if  $u(X)$  is out-of-date
04               $\text{UPDATE\_BOUNDS}(X, Y)$ 
05          if  $Y$  is still a candidate encoding for  $X$ 
06              if  $d(X, Y) < u(X)$ 
07                   $\text{UPDATE\_BOUNDS}(X, Y)$ 
08                  set  $Y_X = Y$ 
09  return  $Y_X$ 
```

Algorithm 4: $\text{UPDATE_BOUNDS}(X, Y)$

```
01  set  $u(X) = d(X, Y_X)$ 
02  set  $l(X, Y_X) = d(X, Y_X)$ 
03  declare  $u(X)$  as up-to-date
04  store  $\mathbf{x}_a \in X$  with  $d(X, Y_X) = \|\mathbf{x}_a - \mathbf{y}\|$ 
```

Algorithm 5: $\text{ESTIMATE_BOUNDS}()$

```
01  compute  $\delta(Y) = d(Y, Y')$  for all  $Y \in \mathcal{C}$ 
02  set  $u(X) = \min \{u(X) + \delta(Y_X), \|\mathbf{x}_a - \mathbf{y}\|\}$  for all  $X \in \mathcal{X}_{\mathcal{T}}$ 
03  declare  $u(X)$  as out-of-date for all  $X \in \mathcal{X}_{\mathcal{T}}$ 
04  set  $l(X, Y) = \max \{l(X, Y') - \delta(Y), 0\}$  for all  $X \in \mathcal{X}_{\mathcal{T}}$  and for all  $Y \in \mathcal{C}$ 
```

Otherwise, if at least one of both conditions (C_1) and (C_2) is satisfied or becomes satisfied during examination, Y_X remains the encoding of X .

Crucial for avoiding NP-hard graph distance calculations are good estimates of the lower and upper bounds $l(X, Y)$ and $u(X)$ after each cycle through the training set \mathcal{S} . For this, we keep record of the most recent optimal alignment of an input graph X and its current code graph Y_X . Suppose that \mathbf{y} is a vector representation of Y_X during a cycle through the training set. Each time we calculate a distance $d(X, Y_X)$, we obtain a vector representation $\mathbf{x} \in X$ such that (\mathbf{x}, \mathbf{y}) is an optimal alignment of $X \times Y_X$. By $\mathbf{x}_a \in X$ we denote the vector representation of X of an optimal alignment obtained by the most recent distance calculation $d(X, Y_X)$. Updating the bounds is then carried out as follows: After each cycle through the training set \mathcal{S} , we compute the change $\delta(Y)$ of each centroid Y by the distance

$$\delta(Y) = d(Y, Y'),$$

where Y' is the code graph before the t -th cycle through the training set \mathcal{S} and Y is the current code graph after the t -th cycle through \mathcal{S} . Based on the triangle inequality of d , we set the bounds according to the following rules:

$$l(X, Y) = \max \{l(X, Y') - \delta(Y), 0\} \quad (2)$$

$$u(X) = \min \{u(X) + \delta(Y_X), \|\mathbf{x}_a - \mathbf{y}\|\}, \quad (3)$$

Both rules guarantee that $l(X, Y)$ is always a lower bound of $d(X, Y)$ and $u(X)$ is always an upper bound of $d(X, Y_X)$. Note that calculating an Euclidean distance and the computational overhead of storing aligned vector representations \mathbf{x}_a of each training graph X is computationally negligible compared to calculating the NP-hard graph distance d .

Finally, we declare upper bounds as out-of-date if

$$\delta(Y) = d(Y, Y') > \theta$$

where $\theta \geq 0$ is some prespecified control parameter that trades accuracy against speed. This is motivated by the following considerations. For sake of simplicity, we assume that $\theta = 0$. From

$$\delta(Y) = d(Y, Y') = 0$$

follows that the current and the previous code graph are identical. This implies that an optimal alignment $(\mathbf{x}, \mathbf{y}')$ of $X \times Y_X'$ is also an optimal alignment (\mathbf{x}, \mathbf{y}) of $X \times Y_X$ for all training graphs X encoded by Y_X . According to eqn. (3) the upper bounds are then of the form

$$u(X) = \|\mathbf{x} - \mathbf{y}\| = d(X, Y_X).$$

As a consequence, we may declare $u(X)$ as up-to-date rather than out-of-date. This makes a delayed distance evaluation of $d(X, Y_X)$ unnecessary in the next

iteration. To further accelerate competitive learning GQ, we can generalize this idea for small changes

$$\delta(Y) = d(Y, Y') \leq \theta.$$

The underlying assumption of this heuristic is that the more similar the previous and the recomputed code graphs are, the more likely is an optimal alignment $(\mathbf{x}, \mathbf{y}')$ of $X \times Y'_X$ also an optimal alignment (\mathbf{x}, \mathbf{y}) of $X \times Y_X$. In fact, we have

$$d(X, Y_X) \leq \|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}'\| + \theta = d(X, Y'_X) + \theta$$

showing that the upper bound $\|\mathbf{x} - \mathbf{y}\|$ deviates at most by θ from $d(X, Y_X)$.

Algorithm 2 describes the accelerated version of competitive learning GQ. The accelerated version calls the subroutines CLASSIFY, UPDATE_BOUNDS, and ESTIMATE_BOUNDS described in Algorithm 3, 4, and 5 respectively. The subroutine CLASSIFY determines the encoding of the current training graph by applying the principle of delayed distance evaluation. The subroutine UPDATE_BOUNDS is an auxiliary subroutine for updating the bounds and keeping book of the most recent optimal alignment of the current training graph and its encoding. Finally, ESTIMATE_BOUNDS is a subroutine that re-estimates the lower and upper bounds after each cycle through the training set.

4 Experiments

This section reports the results of running standard competitive learning GQ and its accelerated version.

4.1 Data.

We selected four data sets described in [27]. The data sets are publicly available at [16]. Each data set is divided into a training, validation, and a test set. In all four cases, we considered data from the test set only. The description of the data sets are mainly excerpts from [27]. Table 1 provides a summary of the main characteristics of the data sets.

data set	#(graphs)	#(classes)	avg(nodes)	max(nodes)	avg(edges)	max(edges)
letter	750	15	4.7	8	3.1	6
grec	528	22	11.5	24	11.9	29
fingerprint	900	3	8.3	26	14.1	48
molecules	100	2	24.6	40	25.2	44

Table 1. Summary of main characteristics of the data sets.



Fig. 1. Example of letter drawings: Prototype of letter A and distorted copies generated by imposing low, medium, and high distortion (from left to right) on prototype A.

Letter Graphs. We consider all 750 graphs from the test data set representing distorted letter drawings from the Roman alphabet that consist of straight lines only (A, E, F, H, I, K, L, M, N, T, V, W, X, Y, Z). The graphs are uniformly distributed over the 15 classes (letters). The letter drawings are obtained by distorting prototype letters at low distortion level. Lines of a letter are represented by edges and ending points of lines by vertices. Each vertex is labeled with a two-dimensional vector giving the position of its end point relative to a reference coordinate system. Edges are labeled with weight 1. Figure 1 shows a prototype letter and distorted version at various distortion levels.

GREC Graphs. The GREC data set [4] consists of graphs representing symbols from architectural and electronic drawings. We use all 528 graphs from the test data set uniformly distributed over 22 classes. The images occur at five different distortion levels. In Figure 2 for each distortion level one example of a drawing is given. Depending on the distortion level, either erosion, dilation, or other morphological operations are applied. The result is thinned to obtain lines of one pixel width. Finally, graphs are extracted from the resulting denoised images by tracing the lines from end to end and detecting intersections as well as corners. Ending points, corners, intersections and circles are represented by vertices and labeled with a two-dimensional attribute giving their position. The vertices are connected by undirected edges which are labeled as line or arc. An additional attribute specifies the angle with respect to the horizontal direction or the diameter in case of arcs.

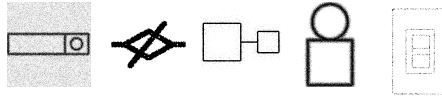


Fig. 2. GREC symbols: A sample image of each distortion level

Fingerprint Graphs. We consider a subset of 900 graphs from the test data set representing fingerprint images of the NIST-4 database [32]. The graphs are uniformly distributed over three classes *left*, *right*, and *whorl*. A fourth class (*arch*) is excluded in order to keep the data set balanced. Fingerprint images are converted into graphs by filtering the images and extracting regions that are relevant



Fig. 3. Fingerprints: (a) Left (b) Right (c) Arch (d) Whorl. Fingerprints of class arch are not considered.

[26]. Relevant regions are binarized and a noise removal and thinning procedure is applied. This results in a skeletonized representation of the extracted regions. Ending points and bifurcation points of the skeletonized regions are represented by vertices. Additional vertices are inserted in regular intervals between ending points and bifurcation points. Finally, undirected edges are inserted to link vertices that are directly connected through a ridge in the skeleton. Each vertex is labeled with a two-dimensional attribute giving its position. Edges are attributed with an angle denoting the orientation of the edge with respect to the horizontal direction. Figure 3 shows fingerprints of each class.

Molecules. The mutagenicity data set consists of chemical molecules from two classes (mutagen, non-mutagen). The data set was originally compiled by [24] and reprocessed by [27]. We consider a subset of 100 molecules from the test data set uniformly distributed over both classes. We describe molecules by graphs in the usual way: atoms are represented by vertices labeled with the atom type of the corresponding atom and bonds between atoms are represented by edges labeled with the valence of the corresponding bonds. We used a 1-to- k binary encoding for representing atom types and valence of bonds, respectively.

4.2 General Experimental Setup

In all experiments, we applied standard competitive learning GQ (std) and its accelerated version (acc) to the aforementioned data sets by using the following experimental setup:

Setting of competitive learning GQ algorithms. To initialize the standard and accelerated competitive learning GQ algorithm, we used a modified version of the "furthest first" heuristic [15]. For each training set \mathcal{S} , the first code graph Y_1 is initialized to be a graph closest to the sample mean of \mathcal{S} (see [19, 20] for details on computing the sample mean of graphs). Subsequent code graphs are initialized according to

$$Y_{i+1} = \arg \max_{X \in \mathcal{S}} \min_{Y \in \mathcal{C}_i} d(X, Y),$$

where \mathcal{C}_i is the set of the first i centroids chosen so far. We terminated both competitive learning GQ algorithms after 150 cycles through the training set.

Graph distance calculations and optimal alignment. For graph distance calculations and finding optimal alignments, we applied a depth first search algorithm on the letter data set and the graduated assignment [11] on the grec, fingerprint, and molecule data set. The depth first search method guarantees to return optimal solutions and therefore can be applied to small graphs only. Graduated assignment returns approximate solutions.

Performance measures. We used the following measures to assess the performance of an algorithm on a dataset: (1) empirical distortion, (2) classification accuracy, (3) silhouette index, and (4) number of graph distance calculations.

The *empirical distortion* is here given by

$$\hat{D}(\mathcal{C}) = \frac{1}{2} \sum_{i=1}^N \min_{1 \leq j \leq k} d(X_i, Y_j).$$

The *silhouette index* is a cluster validation index taking values from $[-1, 1]$. Higher values indicate a more compact and well separated cluster structure. For more details we refer to Appendix A and [30].

Accelerated competitive learning GQ incurs computational overhead to create and update auxiliary data structures and to compute Euclidean distances. This overhead is negligible compared to the time spent on graph distance calculations. Therefore, we report number of graph distance calculations rather than clock times as a performance measure for speed.

4.3 Performance Comparison

We applied both competitive learning GQ algorithms to all four data sets in order to assess and compare their performance. The control parameter θ of the accelerated version is set to zero. For each data set 10 runs of each algorithm were performed and the best result with respect to the error value (empirical distortion) is selected. For taking the average, 10 runs are too less in order to provide statistically significant results. For conducting a higher number of runs in order to obtain statistical significant results, computational resources were not sufficient.

Table 2 summarizes the results. The first observation to be made is that the solution quality of standard and accelerated competitive learning GQ is comparable with respect to error, classification accuracy, and silhouette index. Deviations are due to the non-uniqueness of the solutions and the approximation errors of the graduated assignment algorithm. The second observation to be made from Table 2 is that the accelerated version outperforms standard competitive learning GQ with respect to computation time on all data set.

The results show that the speedup factor increases with increasing number k of codebook graphs but obviously is independent of the average cluster size N/k . Contrasting the silhouette index and the dimensionality of the data to the speedup factor gained by accelerated competitive learning GQ, we make the following observation: First, the silhouette index for the letter, grec, and

data set	measure						
letter		$k = 15$		$k = 30$		$k = 45$	
		$N/k = 50.0$		$N/k = 25.0$		$N/k = 16.7$	
		std	acc	std	acc	std	acc
	error	19.2	19.3	11.1	11.1	8.0	8.1
	accuracy (in %)	72	72	90	90	94	94
	silhouette	0.43	0.42	0.40	0.40	0.34	0.34
	matchings ($\times 10^5$)	16.9	2.2	33.8	1.7	50.6	1.5
grec	speedup	1.0	7.5	1.0	19.6	1.0	34.7
		$k = 22$		$k = 33$		$k = 44$	
		$N/k = 24.0$		$N/k = 16.0$		$N/k = 12.0$	
		std	acc	std	acc	std	acc
	error	37.5	38.9	27.6	29.4	25.5	23.5
	accuracy (in %)	74	75	87	87	88	91
	silhouette	0.40	0.40	0.42	0.44	0.45	0.53
fingerprint	matchings ($\times 10^5$)	17.4	3.7	26.1	3.0	34.8	2.8
	speedup	1.0	4.7	1.0	8.7	1.0	12.2
		$k = 15$		$k = 30$		$k = 60$	
		$N/k = 60.0$		$N/k = 30.0$		$N/k = 15.0$	
		std	acc	std	acc	std	acc
	error	76.7	85.6	17.2	18.2	1.3	1.3
	accuracy (in %)	64	63	67	65	79	79
molecules	silhouette	0.29	0.29	0.34	0.33	0.34	0.36
	matchings ($\times 10^5$)	20.3	4.5	40.5	2.0	81.0	1.2
	speedup	1.0	4.5	1.0	19.9	1.0	66.3
		$k = 2$		$k = 6$		$k = 10$	
		$N/k = 50.0$		$N/k = 16.7$		$N/k = 10.0$	
		std	acc	std	acc	std	acc
	error	61.9	62.1	55.7	54.8	52.6	52.9
molecules	accuracy (in %)	64	62	66	66	66	68
	silhouette	0.06	0.06	0.03	0.03	0.01	0.00
	matchings ($\times 10^4$)	6.0	5.7	18.0	11.2	30.0	12.3
	speedup	1.0	1.1	1.0	1.6	1.0	2.4

Table 2. Results of standard competitive learning GQ (std) and accelerated competitive learning GQ (acc) on four data sets. Rows labeled *matchings* give the number of graph distance calculations, and rows labeled *speedup* show how many times an algorithm is faster than standard competitive learning GQ for graphs.

fingerprint data set are roughly comparable and indicate a cluster structure in the data, whereas the silhouette index for the molecule data set indicates almost no compact and homogeneous cluster structure. Second, the dimensionality of the vector representations is largest for molecule graphs, moderate for grec graphs, and relatively low for letter and fingerprint graphs. Thus, the speedup factor of accelerated competitive learning GQ apparently decreases with increasing dimensionality and decreasing cluster structure. This behavior is in line with Elkan’s k-means for vectors [6] and Elkan’s k-means for graphs [23] as well as with findings in high-dimensional vector spaces. According to [25], there will be little or no acceleration in high dimensions if there is no underlying structure in the data. This view is also supported by theoretical results from computational geometry [17].

Figure 4 shows how the number of graph distance calculations of accelerated competitive learning GQ decreases with increasing number of cycles through the training set. The standard version of competitive learning GQ requires kN graph distance calculations at each cycle corresponding to 100%. For the letter, grec, and fingerprint data set, the plot shows that after a few cycles more than 80% out of kN graph distance calculations can be avoided and replaced by calculations of simple Euclidean distances. Accelerated competitive learning GQ operates in both spaces, the Euclidean graph space and its underlying Euclidean representation space. The graph space is used for choosing vector representations optimally aligned to their code graphs. Given optimal aligned vector representations, competitive learning GQ reduces to competitive learning VQ in the representation space. Switching back to the graph space serves as correcting the optimal alignments between the vector representations of the input and their closest code graphs. As the plot shows, the corrective function of the graph space is less required with increasing number of cycles through the training set.

5 Conclusion

Accelerated competitive learning GQ avoids graph distance calculations by exploiting the metric properties of the graph distance and by lifting input graphs from the training set to vector representations that are optimally aligned to their encodings. In doing so, accelerated competitive learning GQ switches between two spaces, the graph space and its underlying Euclidean representation space. The competitive learning in the graph space serves as a corrective of the fast but erroneous counterpart in the representation space. With increasing time competitive learning for GQ gradually turns into competitive learning for VQ without loss of solution accuracy. In particular, during the long final convergence phase most graph distance calculations can be avoided and replaced by Euclidean distances, provided the data possesses a cluster structure. We conclude that accelerated competitive learning GQ is a first step to avoid intractable graph distance calculations without loss of structural information. We believe that the proposed acceleration can also be applied to GQ using more general graph edit distance metrics as underlying distortion measures. As a next step, future work

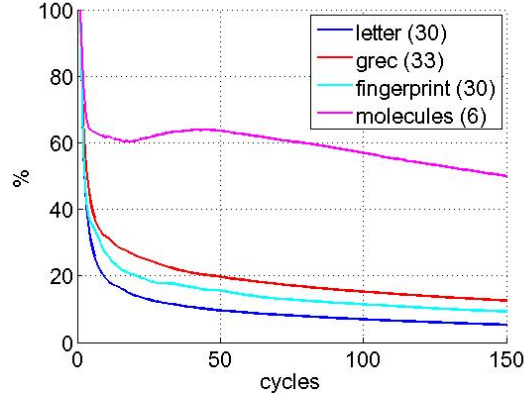


Fig. 4. Percentage of graph distances calculated by accelerated competitive learning GQ at each cycle through the training set. The standard version of competitive learning GQ requires 100% (corresponding to kN) graph distance calculations at each cycle. The number k for each training set is shown in parentheses next to the identifiers of the data sets in the legend box.

is concerned with an empirical investigation of the lifting threshold θ to trade solution quality against speed.

A The Silhouette Index

Suppose that $\mathcal{S} = \{X_1, \dots, X_m\}$ is a sample of m patterns. Let $\mathcal{C} = \{\mathcal{C}_1, \dots, \mathcal{C}_k\}$ be a partition of \mathcal{S} consisting of k disjoint clusters with

$$\mathcal{S} = \bigcup_{i=1}^k \mathcal{C}_i.$$

We assume that D is the underlying distance function defined on \mathcal{S} . The distance between two subsets $\mathcal{U}, \mathcal{U}' \subseteq \mathcal{S}$ is defined by

$$D(\mathcal{U}, \mathcal{U}') = \min \{D(X, X') : X \in \mathcal{U}, X' \in \mathcal{U}'\}.$$

If $\mathcal{U} = \{X\}$ consists of a singleton, we simply write $D(X, \mathcal{U}')$ instead of $D(\{X\}, \mathcal{U}')$.

Let

$$D_{\text{avg}}(X, \mathcal{U})$$

denote the average distance between pattern $X \in \mathcal{S}$ and subset $\mathcal{U} \subseteq \mathcal{S}$. Suppose that pattern $X_i \in \mathcal{S}$ is a member of cluster $\mathcal{C}_{m(i)} \in \mathcal{C}$. By $\mathcal{C}'_{m(i)}$ we denote the set $\mathcal{C}_{m(i)} \setminus \{X_i\}$. For each pattern $X_i \in \mathcal{S}$ let

$$a_i = D_{\text{avg}}(X_i, \mathcal{C}'_{m(i)})$$

be the average distance between pattern X_i and subset $\mathcal{C}'_{m(i)}$. By

$$b_i = \min_{j \neq m(i)} D_{\text{avg}}(X_i, \mathcal{C}_j)$$

we denote the minimum average distance between pattern X_i and all clusters from \mathcal{C} not containing X_i . The *silhouette width* of X_i is defined as

$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)}.$$

The *silhouette of cluster* $\mathcal{C}_j \in \mathcal{C}$ is given by

$$S_j = \frac{1}{|\mathcal{C}_j|} \sum_{i: X_i \in \mathcal{C}_j} s_i.$$

The *silhouette index* is then defined as the average of all cluster silhouettes

$$\mathfrak{S} = \frac{1}{k} \sum_{j=1}^k S_j.$$

References

1. H. Almohamad and S. Duffuaa, "A linear programming approach for the weighted graph matching problem", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5)522–525, 1993.
2. T.S. Caetano, L. Cheng, Q.V. Le, and A.J. Smola, "Learning graph matching" *International Conference on Computer Vision*, p. 1–8, 2007.
3. T. Cour, P. Srinivasan, and J. Shi, "Balanced graph matching", *NIPS 2006 Conference Proceedings*, 2006.
4. P. Dosch and E. Valveny, "Report on the second symbol recognition contest", *GREC 2005 Conference Proceedings*, p. 381–397, 2006.
5. R.O. Duda, P.E. Hart, and D.G. Stork *Pattern Classification*, Wiley & Sons, 2000.
6. C. Elkan, "Using the triangle inequality to accelerate k-means", *ICML 2003 Conference Proceedings*, p. 147–153, 2003.
7. M. Ferrer, *Theory and algorithms on the median graph. application to graph-based classification and clustering*, PhD Thesis, Univ. Autònoma de Barcelona, 2007.
8. M. Ferrer, E. Valveny, F. Serratosa, I. Bardají, and H. Bunke, "Graph-Based k-Means Clustering: A Comparison of the Set Median versus the Generalized Median Graph" *CAIP 2009 Conference Proceedings*, 2009.
9. M. Garey, D. Johnson, and H. Witsenhausen, "The complexity of the generalized Lloyd-Max problem", *IEEE Trans. Inform. Theory*, 28, p. 255–256, 1982.
10. A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.
11. S. Gold and A. Rangarajan, "Graduated Assignment Algorithm for Graph Matching", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18:377–388, 1996.

12. S. Gold, A. Rangarajan, and E. Mjolsness, "Learning with preknowledge: clustering with point and graph matching distance measures" *Neural Computation*, 8(4):787–804, 1996.
13. S. Günter and H. Bunke, "Self-organizing map for clustering in the graph domain", *Pattern Recognition Letters*, 23(4):405–417, 2002.
14. J.A. Hertz, R.G. Palmer, and A. Krogh, *Introduction to the Theory of Neural Computation*, Perseus Books, 1991.
15. D. Hochbaum and D. Shmoys, "A best possible heuristic for the k-center problem", *Mathematics of Operations Research*, 10(2):180–184, 1985.
16. IAM Graph Database Repository of the Computer Vision and Artificial Intelligence Research Group, University Bern, Switzerland (accessed 2009).
URL: <http://www.iam.unibe.ch/fki/databases/iam-graph-database>
17. P. Indyk, A. Amir, A. Efrat, and H. Samet, "Efficient algorithms and regular data structures for dilation, location and proximity problems", *Proceedings of the Annual Symposium on Foundations of Computer Science*, p. 160–170, 1999.
18. B. Jain and F. Wysotzki, "Central Clustering of Attributed Graphs", *Machine Learning*, 56, 169–207, 2004.
19. B. Jain and K. Obermayer, "On the sample mean of graphs", *IJCNN 2008 Conference Proceedings*, p. 993–1000, 2008.
20. B. Jain and K. Obermayer, "Algorithms for the Sample Mean of Graphs", *CAIP 2009 Conference Proceedings*, 2009.
21. B. Jain and K. Obermayer, "Structure Spaces", *Journal of Machine Learning Research*, 10:2667–2714, 2009.
22. B. Jain and K. Obermayer, "Graph Quantization", [arXiv:1001.0921v1](https://arxiv.org/abs/1001.0921v1) [cs.AI], 2010.
23. B. Jain and K. Obermayer, "Elkan's k-Means for Graphs", [arXiv:0912.4598v1](https://arxiv.org/abs/0912.4598v1) [cs.AI], 2009.
24. J. Kazius, R. McGuire, and R. Bursi, "Derivation and validation of toxicophores for mutagenicity prediction", *Journal of Medicinal Chemistry* 48(1):312–320, 2005.
25. A. Moore, "The anchors hierarchy: Using the triangle inequality to survive high dimensional data", *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, p. 397–405, 2000.
26. M. Neuhaus and H. Bunke, "A graph matching based approach to fingerprint classification using directional variance", *AVBPA 2005 Conference Proceedings*, pp. 191–200, 2005.
27. K. Riesen and H. Bunke, "IAM Graph Database Repository for Graph Based Pattern Recognition and Machine Learning", *SSPR 2008 Conference Proceedings*, pp. 287–297, 2008.
28. A. Schenker, M. Last, H. Bunke, and A. Kandel, "Clustering of web documents using a graph model", *Web Document Analysis: Challenges and Opportunities*, pp. 1–16, 2003.
29. A. Schenker, M. Last, H. Bunke, and A. Kandel, *Graph-Theoretic Techniques for Web Content Mining*, World Scientific Publishing, 2005.
30. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Elsevier, 2009.
31. S. Umeyama, "An eigendecomposition approach to weighted graph matching problems", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):695–703, 1988.
32. C. Watson and C. Wilson, *NIST Special Database 4, Fingerprint Database*, National Institute of Standards and Technology, 1992.

33. M. Van Wyk, M. Durrani, and B. Van Wyk, "A RKHS interpolator-based graph matching algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):988–995, 2002.