On Soft Learning Vector Quantization Based on Reformulation*

Jian Yu[1] and Pengwei Hao[2,3]

[1.] Dept. of Computer Science, Beijing Jiaotong University
Beijing, 100044, P.R .China
jianyu@center.njtu.edu.cn
[2.] Center of Information Science, Peking University
Beijing, 100871, P.R. China
phao@cis.pku.edu.cn
[3.] Dept. of Computer Science, Queen Mary, University of London,
London, E1 4NS, UK
phao@dcs.qmul.ac.uk

Abstract. The complex admissibility conditions for reformulated function in Karayiannis model is obtained based on the three axioms of radial basis function neural network. In this paper, we present an easier understandable assumption about vector quantization and radial basis function neural network. Under this assumption, we have obtained a simple but equivalent criterion for admissible reformulation function in Karayiannis model. We have also discovered that Karayiannis model for vector quantization has a trivial fixed point. Such results are useful for developing new vector quantization algorithms.

1 Introduction

It is well known that vector quantization is a data compression method, which encodes a set of data points into a reduced set of reference vectors. In the literature, many vector quantization strategies are proposed. One common way to design vector quantizer is based on cluster analysis. Therefore, many researchers studied cluster analysis and vector quantization together, e.g. in [1]. In [2], Karayiannis proposed an axiomatic approach to soft learning vector quantization and clustering based on reformulation. In this paper, this model is called Karayiannis model. As indicated in [2], Karayiannis model leads to abroad family of soft learning vector quantization and clustering algorithms, including FCM [3], fuzzy learning vector quantization (FLVQ) [4], entropy-constrained learning vector quantization (ECLVQ) [5], and so on.

Based on Karayiannis model, many variations for radial basis neural networks are proposed, see [6]. In the previous research, reformulated function plays a pivotal role

* This work was partially supported by the Key Scientific Research Project of MoE, China under Grant No.02031, the National Natural Science Foundation of China under Grant No.60303014, and the Foundation for the Authors of National Excellent Doctoral Dissertation, China, under Grant 200038.

F. Yin, J. Wang, and C. Guo (Eds.): ISNN 2004, LNCS 3173, pp. 168–173, 2004.

in constructing numerous reformulated radial basis function neural networks (RRBF). The admissibility condition for reformulated function is obtained based on the three axioms of radial basis function neural network (RBF). However, such three axiomatic requirements are not easily understood, especially in mathematical form. In this paper, we obtain an efficient criterion for admissible reformulation function in Karayiannis model based on a novel assumption of RBF and vector quantization, which is proved to be equivalent to the complex admissibility conditions for reformulated function in Karayiannis model.

The remainder of this paper is organized as follows: In Section 2, RBF, vector quantization and Karayiannis model are briefly related. In Section 3, a novel and intuitive assumption about RBF and vector quantization is proposed, which leads to a necessary condition for Karayiannis model to perform well. Moreover, we prove that such a condition is equivalent to the complex admissibility conditions for reformulated function in Karayiannis's model. In the final section, we draw the conclusion.

2 RBF, Vector Quantization, and Karayiannis Model

It is well known that the number and the centers of radial basis functions play an important role in the performance of RBF neural networks. Similarly, the size of codebook and the reference vectors are the key to vector quantization. Therefore, it is natural to establish a connection between vector quantization (in particular, LVQ) and RBF, see [2][6]. As noted in [6], a mapping can be the basis for constructing LVQ algorithms and RBF networks, more details can be found in [6]. In the following, we will describe this mapping.

Let $X=\{x_1, x_2, \ldots, x_n\}$ be a s-dimensional data set, $v=\{v_1,v_2,\ldots,v_c\}$ is the codebook, the centers (or prototype). Karayiannis defines a mapping $R^s \to R$ as below:

$$y = f\left(w_0 + \sum_{i=1}^{c} w_i g\left(\|x - v_i\|^2\right)\right) \tag{1}$$

where $f(x)$ and $g(x)$ are everywhere differentiable functions of $x \in (0,+\infty)$ and $\| \ \|$ denotes the Euclidean norm. Obviously, selecting appropriate $f(x)$ and $g(x)$ is very important for the performance of LVQ and RBF.

In order to tackle this issue, Karayiannis made the following assumption about RBF in [6]: an RBF is a composition of localized receptive fields, and the locations of these receptive fields are determined by the centers. And he also observed the three facts about RBF as following:

1. *The response of RBF's to any input is positive.*
2. *When the centers are considered as the centers of receptive fields, it is expected that the response of any RBF becomes stronger if an input approaches its corresponding center.*
3. *The response of any RBF is more sensitive to an input as this input approaches its corresponding center.*

Based on the above three observations, the three axioms of Karayiannis model are presented. In the following, we briefly introduce Karayiannis model.

By observing that the FCM and ECFC algorithms correspond to reformulation functions of the same basic form, Karayiannis [2] proposed a family of functions in the general form as follows:

$$R = \frac{1}{n}\sum_{k=1}^{n} f(S_k),\ S_k = \frac{1}{c}\sum\nolimits_{i=1}^{c} g(d_{ik}), d_{ik} = \|x_k - v_i\|^2,\ f(g(x)) = x \tag{2}$$

Obviously, (1) is a basis to develop (2). As Karayiannis noted, minimization of admissible reformulation functions (2) using gradient descent can produce a variety of batch LVQ algorithms. The gradient of R with respect to v_i is:

$$\frac{\partial R}{\partial v_i} = -\frac{2}{nc}\sum\nolimits_{k=1}^{n} f'(S_k)g'(\|x_k - v_i\|^2)(x_k - v_i) \tag{3}$$

Hence, the update equation for the centers can be obtained as:

$\Delta v_i = -\eta_i \frac{\partial R}{\partial v_i} = \eta_i \sum\nolimits_{k=1}^{n} \alpha_{ik}(x_k - v_i)$, where η_i is the learning rate for the prototype v_i, and $\alpha_{ik} = f'(S_k)g'(\|x_k - v_i\|^2)$ is the competition function.

Then the LVQ algorithms can be implemented as follows:

Set the generator $f(x)$*, the initial centers* $\{v_{i,0} \mid 1 \le i \le c\}$*, the termination limit* ε*, the maximum number of iteration* L*, and t=1*

Step 1. $v_{i,t} = v_{i,t-1} + \eta_{i,t}\sum\nolimits_{k=1}^{n} \alpha_{ik,t}(x_k - v_{i,t-1}),\ 1 \le i \le c$ *where* $\{v_{i,t-1} \mid 1 \le i \le c\}$ *is the set centers obtained after the* $(t-1)$*-th iteration,* $\eta_{i,t}$ *is the learning rate at iteration t and* $\alpha_{ik,t} = f'(S_{k,t-1})g'(\|x_k - v_{i,t-1}\|^2)$ *with* $S_{k,t-1} = c^{-1}\sum\nolimits_{i=1}^{c} g(\|x_k - v_{i,t-1}\|^2)$.

Step 2. If $\max_i \|v_{i,t} - v_{i,t-1}\| < \varepsilon$*, or l>* L*, then stop; else l=l+1 and go to step 1. where* $\eta_{i,t-1}$*, are learning rates between 0 and 1.*

According to the above three axioms, Karayiannis found that the properties of admissible competition functions should satisfy the following three axioms:

Axiom 1: if c=1, then $\alpha_{1k} = 1,\ 1 \le k \le n$;

Axiom 2: $\alpha_{ik} \ge 0,\ 1 \le i \le c,\ 1 \le k \le n$;

Axiom 3: If $\|x_k - v_p\| > \|x_k - v_q\| > 0$, then $\alpha_{pk} < \alpha_{qk},\ \forall p \ne q$.

According to Axiom 1-3, Karayiannis proved the following theorem.

Theorem 1 (Karayiannis, [2]): Let $X=\{x_1, x_2, \ldots, x_n\} \subset R^s$ be a finite set of feature vectors which are represented by the set of $c<n$ prototypes $V=\{v_1, v_2, \ldots, v_c\} \subset R^s$. Then the function R defined by (2) is admissible reformulation function of the first (second) kind in accordance with the axiomatic requirements 1-3 if $f(x)$ and $g(x)$ are differentiable everywhere functions of $x \in (0,+\infty)$ satisfying $f(g(x)) = x$, $f(x)$ and $g(x)$ are both monotonically decreasing (increasing) functions of x, and $g'(x)$ is a monotonically increasing (decreasing) function of $x \in (0,+\infty)$.

Obviously, the condition of admissible functions for Karayiannis model is too complex. In other words, it is not easy to judge whether or not a function is admissible for Karayiannis model, and Karayiannis himself also made many efforts to study more specific admissible reformulation function for Karayiannis model, see [2], [7], [8].

In Section 3, we present a simple assumption about vector quantization and RBF, which leads to a more efficient criterion for selecting function $f(x)$.

3 A Simple Assumption About RBF and Vector Quantization

When encoding or vector quantization, it is expected that all reference vectors are different. For example, Buhmann & Kuhnel thought that configurations with degenerate reference vectors are inadmissible in [9]. Similarly, the centers of receptive fields of RBF networks are not expected to be coincidental. However, many soft learning vector quantization algorithms (like ECLVQ) can output degenerate reference vectors, i.e. coincidental reference vectors. Therefore, it is naturally supposed that degenerate reference vectors are not the stable solution of vector quantization algorithms. The similar assumption has been used for clustering algorithms, see [10], [11], and [12].

For brevity, let Ω denote all fixed points of Karayiannis model, then it includes the saddle points and attractive points. A saddle point is unstable, i.e. not able to stand sufficiently small disturbing. All attractive points are stable. Therefore, the degenerate reference vectors are stable if they are the attractive points of Karayiannis model. It is better if Karayiannis model has no degenerate reference vectors, but unfortunately, it is easy to prove that $\forall 1 \le i \le c, v_i = \bar{x}$ is a fixed point of Karayiannis model, where $\bar{x} = \sum_{k=1}^{n} x_k \big/ n$. Therefore, the output of Karayiannis model will be $\bar{x}$ with a great probability if $\forall 1 \le i \le c, v_i = \bar{x}$ is stable. Obviously, it is not the case we hope to face. What measure can we take to avoid such happening? First, we need to do is to find a criterion to judge whether a fixed point is stable for Karayiannis model. Theorem 2 offers such a criterion.

Theorem 2: If $\forall x, f''(x) < 0$ then $\forall 1 \le i \le c, v_i = \bar{x}$ is a stable fixed point of Karayiannis model, i.e. $\forall 1 \le i \le c, v_i = \bar{x}$ is a strict local minimum of *R*.

Proof: See Appendix A.

As analyzed above, it is unacceptable for $\forall 1 \le i \le c, v_i = \bar{x}$ is a stable fixed point of Karayiannis model. Therefore, it is a natural requirement for Karayiannis model that $\forall x, f''(x) \ge 0$. It seems inconsistent with Theorem 1, but we indeed have Theorem 3.

Theorem 3: $f(x)$ and $g(x)$ are differentiable everywhere functions of $x \in (0, +\infty)$ satisfying $f(g(x)) = x$, and both monotonically decreasing (increasing) functions of *x*, and $g'(x)$ is a monotonically increasing (decreasing) function of $x \in (0, +\infty)$ if and only if *f*(*x*) and *g*(*x*) are differentiable everywhere functions of $x \in (0, +\infty)$ satisfying $f(g(x)) = x$ and $f''(x) > 0$.

Proof: See Appendix B.

Apparently, Theorem 3 verifies that the new criterion is simple but equivalent to the complex condition in Theorem 1. Moreover, Theorem 2 tells us that our criterion has its own meaning totally different from three observations on radial basis functions. It also opens a new door to understand the properties of radial basis function networks. As a matter of fact, the optimality test for Karayiannis model is obtained, too.

4 Conclusions and Discussions

In [2], Karayiannis proposed a soft learning vector quantization based on reformulation. The complex admissibility conditions for reformulated function in Karayiannis model is obtained based on the three axioms of radial basis function neural network. Based on an easier understood assumption about vector quantization and radial basis function neural network, we obtained a simple but equivalent criterion about admissible reformulation function in Karayiannis's model. Moreover, we have found the optimality test for Karayiannis model, and discovered that Karayiannis model for vector quantization has a trivial fixed point. Such results are also useful for constructing new vector quantization algorithms.

References

1. Equitz, W.H.: A New Vector Quantization Clustering Algorithm. IEEE Transactions on Acoustics, Speech, and Signal Processing, 37(10) (1989) 1568-1575
2. Karayiannis, N.B.: An Axiomatic Approach to Soft Learning Vector Approach Quantization and Clustering. IEEE Transactions on Neural Networks, 10(5) (1999) 1153-1165
3. Bezdek, J.C.: Cluster Validity with Fuzzy Sets. J.Cybernt., 3(3) (1974) 58-72
4. Karayiannis, N.B., Bezdek, J.C.: An Integrated Approach to Fuzzy Learning Vector Quantization and Fuzzy C-means Clustering. IEEE Transactions on Fuzzy Systems, vol.5 (1997) 622-628
5. Karayiannis, N.B.: Entropy-Constrained Learning Vector Quantization Algorithms and Their Application in Image Compression. SPIE Proc., vol.3030 (1997) 2-13
6. Karayiannis, N.B.: Reformulated Radial Basis Neural Networks Trained by Gradient Descent. IEEE Transactions on Neural Networks, 10(3) (1999) 657 - 671
7. Karayiannis, N.B., Randolph-Gips, M.M.: Soft Learning Vector Quantization and Clustering Algorithms Based on Non-Euclidean Norms: Multinorm Algorithms. IEEE Transactions on Neural Networks, 14 (1) (2003) 89-102
8. Karayiannis, N.B.: Soft Learning Vector Quantization and Clustering Algorithms Based on Ordered Weighted Aggregation Operators. IEEE Transactions on Neural Networks, 11(5) (2003) 1093 - 1105
9. Buhmann, J., Kuhnel, H.: Vector Quantization with Complexity Costs. IEEE Trans. On Information Theory, 39(4) (1993) 1133-1145
10. Yu, J.: General c-means Clustering Model and Its Applications. CVPR2003, v.2 (2003) 122-127
11. Yu, J., Yang, M.S.: A Study on Generalized FCM. RSFDGrC'2003, Lecture Notes in Computer Science, no.2639, Springer-Verlag Heidelberg (2003) 390-393
12. Yu, J., Cheng, Q.S., Huang, H.K.: Analysis of the Weighting Exponent in the FCM. IEEE Transactions on Systems, Man and Cybernetics-part B: Cybernetics, 34(1) (2004) 634-639

Appendix A: Proof of Theorem 2

$$1 \le i, j \le c\ , \frac{\partial R^2}{\partial v_i \partial v_j} = \frac{4}{c^2 n}\sum_{k=1}^{n} f''(S_k) g'(d_{jk}) g'(d_{ik})(x_k - v_i)(x_k - v_j)^T$$
$$+\frac{4}{nc}\delta_{ij}\sum_{k=1}^{n} f'(S_k) g''(d_{ik})(x_k - v_i)(x_k - v_i)^T + \frac{2}{nc}\delta_{ij}\sum_{k=1}^{n} f'(S_k) g'(d_{ik}) I_s \tag{4}$$

Set $\vartheta = (v_1, v_2, \cdots, v_c) = (\bar{x}, \bar{x}, \cdots, \bar{x}) \in \Omega$, and $d_{k\bar{x}} = (x_k - \bar{x})^T (x_k - \bar{x})$. Noticing $f(g(x)) = x$, we have $f'(g(x))g'(x) = 1$. Then, $\forall i$, $d_{ik} = d_{ck}$, $f'(S_k) g'(d_{ik}) = 1$. By straight calculation, the second term of the Taylor series expansion of R on the point $\vartheta = (v_1, v_2, \cdots, v_c) \in \Omega$ is as follows:

$$\varphi_v^T \left(\frac{\partial^2 R}{\partial v_i \partial v_j} \right) \varphi_v = \frac{4}{n}\sum_{k=1}^{n} f''(g(d_{k\bar{x}}))(g'(d_{k\bar{x}}))^2 \left(\sum_{i=1}^{c} c^{-1} \varphi_{v_i}, (x_k - \bar{x}) \right)^2$$
$$+ 2\sum_{i=1}^{c} c^{-1} \varphi_{v_i}^T \left(I_{s\times s} + \sum_{k=1}^{n} \frac{2g''(d_{k\bar{x}})}{n \times g'(d_{k\bar{x}})} (x_k - \bar{x})(x_k - \bar{x})^T \right) \varphi_{v_i} \tag{5}$$

Noticing that $\left(\sum_{i=1}^{c} c^{-1} \varphi_{v_i}, (x_k - \bar{x})\right)^2 \le \sum_{i=1}^{c} c^{-1} \left(\varphi_{v_i}, (x_k - \bar{x})\right)^2$, and $f(g(x)) = x$ we have $f''(g(x))(g'(x))^2 = -\frac{g''(x)}{g'(x)}$, then we can obtain (6) from (5) if $f''(x) < 0$:

$$\varphi_v^T \left(\frac{\partial^2 R}{\partial v_i \partial v_j} \right) \varphi_v \ge -\frac{4}{n}\sum_{k=1}^{n} \frac{g''(d_{k\bar{x}})}{g'(d_{k\bar{x}})} \sum_{i=1}^{c} c^{-1} \left(\varphi_{v_i}, (x_k - \bar{x})\right)^2 + 2\sum_{i=1}^{c} c^{-1} \varphi_{v_i}^T \varphi_{v_i}$$
$$+ \sum_{k=1}^{n} \frac{4g''(d_{k\bar{x}})}{n \times g'(d_{k\bar{x}})} \sum_{i=1}^{c} c^{-1} \left(\varphi_{v_i}, (x_k - \bar{x})\right)^2 = 2\sum_{i=1}^{c} c^{-1} \varphi_{v_i}^T \varphi_{v_i} \ge 0 \tag{6}$$

■

Appendix B: Proof of Theorem 3

As $f(g(x)) = x$, $f'(g(x))g'(x) = 1$, obviously, we have $\forall x \ge 0\ f'(x) \ne 0$, and $g'(x) \ne 0$. According to Darboux's Theorem, we know that $\forall x,\ f'(x) > 0$, otherwise $\forall x > 0,\ f'(x) < 0$. It is easy to prove that $f'(x)$<0 if and only if $g'(x)$<0 or that $f'(x)$>0 if and only if $g'(x)$>0. Noticing that $f''(g(x))(g'(x))^2 + f'(g(x))g''(x) = 0$, we have $f''(x) > 0$ if and only if $\frac{g''(x)}{g'(x)} < 0$. ■