1993-01

On the Relationship Between the Generalized Equality Classifier and ART 2 Neural Networks

https://hdl.handle.net/2144/1989

ON THE RELATIONSHIP BETWEEN THE GENERALIZED EQUALITY CLASSIFIER AND ART 2 NEURAL NETWORKS

Alper K. Caglayan and Magnus Snorrason

January 1993

Technical Report CAS/CNS-93-012

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215

On the Relationship between the Generalized Equality Classifier and ART 2 Neural Networks

Alper K. Caglayan & Magnús Snorrason
Charles River Analytics Inc.
55 Wheeler Street
Cambridge, MA 02138
akc@crasun.cra.com

Abstract

In this paper, we introduce the Generalized Equality Classifier (GEC) for use as an unsupervised clustering algorithm in categorizing analog data. GEC is based on a formal definition of inexact equality originally developed for voting in fault tolerant software applications. GEC is defined using a metric space framework. The only parameter in GEC is a scalar threshold which defines the approximate equality of two patterns. Here, we compare the characteristics of GEC to the ART2-A algorithm (Carpenter, Grossberg, and Rosen, 1991). In particular, we show that GEC with the Hamming distance performs the same optimization as ART2. Moreover, GEC has lower computational requirements than ART2 on serial machines.

1. Introduction

In section 2 of this paper, we introduce the Generalized Equality Classifier (GEC) for use as an unsupervised clustering algorithm in categorizing analog data. Then, after a brief review of ART2-A in section 3, we compare the characteristics of GEC to the ART2-A algorithm in section 4. In particular, we show that GEC with the Hamming distance performs the same optimization as ART2. We then present empirical results showing the relationship between the GEC threshold and ART vigilance parameters, and a comparison of the computational complexity of the two algorithms in digital simulations.

2. Generalized Equality Classifier

Generalized Equality Classifier is based on the generalized equality concept introduced in (Lorczak, Caglayan and Eckhardt, 1989) in the context of voting algorithms for fault tolerant software. Here, we briefly describe the approach using the framework of metric spaces. Suppose there are n patterns in a vector space X which we would like to organize into clusters. Suppose that there is a metric d defined on X, that is, (X, d) is a metric space. Recall that this means that d is a real-valued function defined on the cartesian product X x X with the following properties:

(i) $d(x, y) \geq 0$
(ii) $d(x, y) = 0$ implies $x = y$
(iii) $d(x, y) = d(y, x)$
(iv) $d(x, z) \leq d(x, y) + d(y, z)$

for all elements x, y and z in X.

Generalized Equality

In categorizing data into clusters, we would like to assign any two patterns into the same cluster whenever the distance between the two patterns is less than a selected threshold, ε. In order to formalize this notion of equality, let us examine this notion of "*same*"patterns.

This definition lacks important properties associated with any mathematical definition of equality, the most important property being transitivity. That is, if x_1 is the "same" pattern as x_2 and x_2 is the "same" pattern as x_3, then x_1 is not necessarily the "same" pattern as x_3. For example, consider the case where X is a subset of real numbers and d is the usual distance on the real line. For an arbitrary real number x, the patterns $x_1 = x-\varepsilon$, $x_2 = x$, and $x_3 = x+\varepsilon$ satisfy

$$d(x_1, x_2) = \varepsilon$$
$$d(x_2, x_3) = \varepsilon$$

while

$$d(x_1, x_3) = 2\varepsilon > \varepsilon$$

If transitivity is, in some sense, forced (e.g. x_1 is taken to be identical to x_3 in the example above by virtue of the fact that both are "equal" to x_2) problems can arise. This definition of equality which now includes the transivity property described above would declare the sequence of patterns

$$x, x + \varepsilon, x + 2\varepsilon, x + 3\varepsilon, \ldots, x + k\varepsilon$$

to be identical since each term is within ε of its predecessor for arbitrarily large values of k.

The problem described above can be avoided by insisting that *any* pair of patterns from a cluster agree to within ε by the chosen metric. This provides the following formal algorithm for the Generalized Equality Classifier. Let $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ denote the set of n patterns for classification. The Generalized Equality Classifier constructs k clusters $V_1, V_2, \ldots, V_k$ such that

$V_1, \ldots, V_k$ is maximal with respect to the property that for *any* x, y in V_i, $d(x, y) \leq \varepsilon$

where k is not predetermined. Let $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ be a metric space with the distance metric d. The GEC algorithm can be implemented as follows:

1) select a distance metric and a threshold e
2) let the set $A = \mathbf{X}$
3) select any element x from A
4) set $S = \{x\}$
5) select y from A-S
6) let $S = S \cup \{y\}$ if $d(y, x) \leq \varepsilon$ for all x in S
7) repeat steps five and six until no new elements are added to S
8) replace A with A-S, clear S
9) if A is the empty set then stop, else return to step 3

Classical clustering methods assign a feature vector **x** into one of a predefined number of categories $V_1, \ldots V_k$. In contrast, the number of categories is not preset in the Generalized Equality Classifier. There are several definitions for a cluster (Jain, 1986). It is generally accepted that no single definition of a cluster is adequate. In applying GEC, we recommend the use of the most appropriate distance metric for a given application domain as discussed in (Lorczak, et al., 1989).

3. Adaptive Resonance Theory (ART2)

ART2-A is an algorithmic version of ART2 (Carpenter and Grossberg, 1987) which is a dynamic system that can perform unsupervised classification of an arbitrary number of analog spatial patterns. ART2 is a three layer network, where layer F0 performs preprocessing, F1 is a feature representation field, and F2 is a category representation field with competitive learning. In ART2-A, the problem of implementing ART2 on a sequential digital computer is approached by capturing the essential computational steps in an algorithm, rather than in a layered neural network.

After preprocessing (normalization and thresholding), the *best match* category node is chosen by taking the inner product of the preprocessed input vector **I** with each of the weight vectors $\mathbf{Z}_j$ which gate the signal to the category nodes:

$$T_J = \max_j \{\sum_i I_i Z_{ji} : j = 1 \ldots N\}$$

where J is the index of the best *match* node. Note that because $\|\mathbf{I}\| = \|\mathbf{Z}_j\| = 1.0$, this equation. simplifies to:

$$T_J = \max_j \{\cos(\mathbf{I}, \mathbf{Z}_j) : j = 1 \ldots N\}$$

Since T_J is used in determining whether the match is *good enough*, the measure of *goodness* is simply the angle between the preprocessed input vector and the weight vector of the best match category node. Then to determine whether this *best match* is in fact *good enough*, T_J is compared to the *vigilance* parameter ρ. If $T_J \geq \rho$, then node J is considered *committed* and it learns the given input vector. However, if $T_J < \rho$, then the value of J is reset to the index of an arbitrary *uncommitted* node which then learns the input vector. Learning is done in a single iteration: if J is an uncommitted node, then the weights on pathways connecting to that node are set equal to the input vector, such that $Z_J = \mathbf{I}$. If J is committed, a convex combination of previous learning and the preprocessed input is learned.

4. Analysis

Here, we show that the Generalized Equality Classifier with the Hamming distance metric essentially performs the same optimization problem as in ART2, discuss the relationship between the GEC threshold and the ART vigilance parameters, and compare the computational complexity of the two algorithms on serial machines.

In ART2, an input vector is assigned to a matching node if the projection of the input vector onto that node is greater than the vigilance parameter, ρ. Here, we show that this maximization problem is equivalent to assigning the input vector onto a matching node if the corresponding Hamming distance is less than a give threshold value.

Let S be the unit cube in R^n defined by the Hamming distance, i.e.

$$S = \left\{ x \in \mathbf{R}^n : \sum_{i=1}^{n} |x_i| = 1 \right\}$$

Let **X** be a subset of **S** and **s** be an arbitrary element of **S** and **x** be an arbitrary element of **X**. The vector **x*** in **X** that is closest to s according to the Hamming distance is the vector in **X** with the maximum projection on s. Mathematically, this can be stated as follows:

Let **x*** be defined by

$$\| \mathbf{x}^* \| \leq \| \mathbf{x} - \mathbf{s} \| \text{ for all } \mathbf{x} \text{ in } \mathbf{X}$$

where $\| \bullet \|$ is the Hamming distance norm. Let $\overline{\mathbf{X}}$ be defined by

$$\cos(\overline{\mathbf{X}}, \mathbf{s}) \geq \cos(\mathbf{x}, \mathbf{s}) \text{ for all } \mathbf{x} \text{ in } \mathbf{X}$$

then

$$\mathbf{x}^* = \overline{\mathbf{X}}$$

In order to see this, let s be (1,0,..., 0) without loss of generality. Then

$$\|\mathbf{x}-\mathbf{s}\| = |x_1 - 1| + |x_2| + \ldots\ldots + |x_n|$$

$$= |x_1 - 1| + 1 - |x_1|$$

$$|\cos(\mathbf{x},\mathbf{s})| = |x_1|$$

$$\arg\max_{x_1} |\mathbf{x}_1| = \arg\min_{x_1} |x_1 - 1| + 1 - |x_1|$$

Therefore, the GEC and ART classification optimizations are equivalent. The Generalized Equality Classifier algorithm requires the specification of a threshold parameter, ε. This threshold is used within the algorithm as the maximum distance that patterns can be apart while still being put into the same cluster. That is, larger values of ε result in smaller number of clusters with more elements each, where as smaller values for ε result in larger number of clusters with fewer elements each. The ART2-A paradigm includes the specification of a vigilance parameter, ρ. The vigilance parameter controls the granularity of clustering: 1.0 means that a perfect match is required and therefore only identical patterns end in the same cluster; lower values produce fewer clusters with more elements each.

In figure 1, we show the number of clusters generated from a set consisting of 50 patterns with 3 features each, as a function of the ε and ρ parameters. The patterns are from a binary classification problem but the features are continous valued. The results for the GEC are generated with a max norm distance metric. As ε increases, the number of clusters decrease until ε ≥ 0.8 when all training patterns fall into the same cluster. As ρ approaches 1.0, the number of clusters increase until the total number of patterns, 50, is reached.

Figure 2 shows computational time in real-time seconds for both methods as a function of number of patterns. The interesting thing to note is the large disparity in computational time between the GEC and ART2-A as the number of patterns increases. This is due to the fact that ART2-A continues categorization until steady state is achieved, whereas the Generalized Equality Classifier only performs categorization once. Moreover, there are no multiplications involved in the distance metric used in this version of GEC (also true of the Hamming distance version), compared to the projection computations in ART which require multiplications. We stress that these results are only valid for a serial machine. The implementation architecture would obviously change the results substantially.

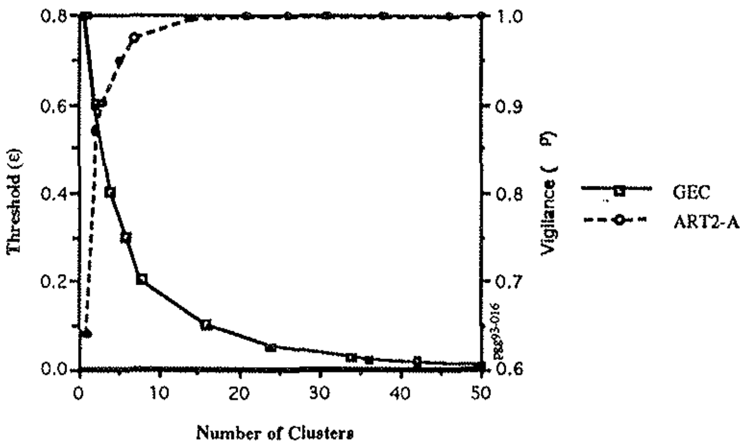

Figure 1: Number of Clusters vs. ε and ρ

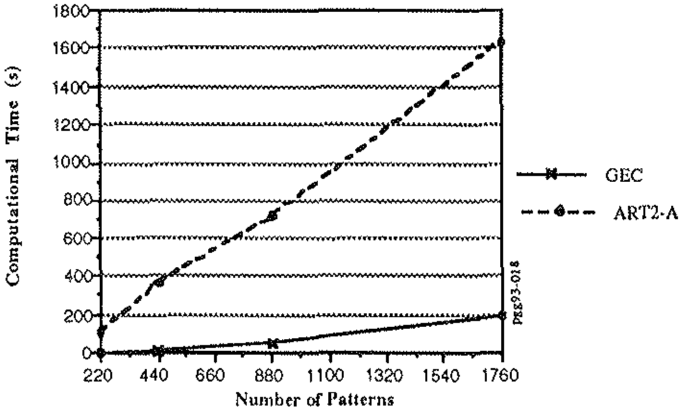

Figure 2: Computational Time Complexity: GEC vs. ART2-A

5. Conclusions

In this paper, we have introduced the Generalized Equality Classifier which is based on a formal definition of equality using a metric space framework. We have compared the GEC algorithm to the ART2 algorithm both analytically and empirically. In particular, we have shown that GEC with the Hamming distance performs the same optimization as ART2. Moreover, GEC has lower computational requirements than ART2 on serial machines.

Acknowledgments

We acknowlege the help of our colleages, Mr. Paul G. Gonsalves, and Dr. Adam X. Miao, in obtaining the results in section 4.

References

Carpenter, G. and Grossberg, S. 1987. "ART2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns." *Applied Optics*, Vol. 26: pp. 4919-4930.

Carpenter, G.A., Grossberg, S. and Rosen, D.B. 1991. "ART2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition." *Neural Networks*, Vol. 4.

Jain, A.L. 1986. "Cluster Analysis." *Handbook of Pattern Recognition and Image Processing*. Young and Fu, ed. Academic Press. pp. 33-57.

Lorczak, P.R., Caglayan, A.K. and Eckhardt, D.E. 1989. "A Theoretical Investigation of Generalized Voters for Redundant Systems." *FTCS 19*, Chicago, IL.