# Adaptive Scaling of Cluster Boundaries for Large-Scale Social Media Data Clustering

**Shrinidhi M 192IT024**

**V Venkatesh Kumar 192IT025**

## 1.Abstract

In this Project, we are dealing with three algorithms based on the fuzzy adaptive resonance theory (Fuzzy ART). It use a single parameter, i.e. the vigilance parameter to identify data clusters, and are robust to modest parameter settings. First discover the vigilance region (VR) that essentially determines how a cluster in the Fuzzy ART system recognizes similar patterns in the feature space. Second, we introduce the idea of allowing different clusters in the Fuzzy ART system to have different vigilance levels in order to meet the diverse nature of the pattern distribution of social media data. We propose three vigilance adaptation methods, namely, the activation maximization (AM) rule, the confliction minimization (CM) rule, and the hybrid integration (HI) rule. With an initial vigilance value, the resulting clustering algorithms, namely, the AM-ART, CM-ART, and HI-ART, can automatically adapt the vigilance values of all clusters during the learning epochs in order to produce better cluster boundaries.

## 2. Introduction

Adaptive Resonance Theory (ART) is a neural theory of cognitive information processing, which states that fast learning is a resonant phenomenon in neural circuits. This has led to the development of the ART 1 neural network model for unsupervised learning. ART 1 performs unsupervised learning by modeling clusters as memory prototypes and encoding binary input patterns incrementally through a two-way similarity measure for searching suitable clusters, which simulates how a human brain captures, recognizes and memorizes information regarding objects and events. As long as the difference between the input pattern and the selected winner cluster from the category field does not exceed a certain threshold, called the vigilance parameter, the input pattern is considered a member of the winner cluster. Fuzzy ART replaces the intersection operator (∩) in ART 1 by the min operator (∧) of fuzzy set theory so as to learn both binary and analog patterns. Fuzzy ART inherits the advantages of ART 1 including fast and stable learning and the incremental clustering. Although Fuzzy ART and its variants are useful for unsupervised learning in many areas, they require manual value selection for the vigilance parameter. Specifically, Fuzzy ART still relies on an empirically fixed vigilance value to scale the cluster size, which makes its performance highly dependent on the vigilance parameter value. For example, with a small vigilance value, Fuzzy ART permits high generalization, which may lead to the generation of several big clusters mixed with patterns from multiple classes. On the other hand, a large vigilance value may incur the over-generation of clusters such that one class may be represented by multiple small clusters. Therefore, similar to the selection of the number of clusters in the K-means clustering algorithm, selecting a suitable vigilance value for Fuzzy ART poses a great challenge.

## 3. Literature Survey

Despite the advantages of fast and stable learning, Adaptive Resonance Theory (ART)[1] still relies on an empirically fixed vigilance parameter value to determine the vigilance regions of all of the clusters in the category field (F2), causing its performance to depend on the vigilance value. It would be desirable to use different values of vigilance for different category field nodes, in order to fit the data with a smaller number of categories. Two methods are being introduced namely Activation Maximization Rule (AMR) and the Confliction Minimization Rule (CMR). Despite their differences, both ART with AMR (AM-ART) and with CMR (CM-ART) allow different vigilance levels for different clusters, which are incrementally adapted during the clustering process. process. Specifically, AMR works by increasing the vigilance value of the winner cluster when a resonance occurs and decreasing it when a reset occurs, which aims to maximize the participation of clusters for activation. On the other hand, after receiving an input pattern, CMR first identifies all of the winner candidates that satisfy the vigilance criteria and then tunes their vigilance values to minimize conflicts in the vigilance regions.

ART 2-A[2], an efficient algorithm that emulates the self-organizing pattern recognition and hypothesis testing properties of the ART 2 neural network architecture, but at a speed two to three orders of magnitude faster. Analysis and simulations show how the ART 2-A systems correspond to ART 2 dynamics at both the fast-learn limit and at intermediate learning rates. Intermediate learning rates permit fast commitment of category nodes but slow recording, analogous to properties of word frequency effects, encoding specificity effects, and episodic memory. Better noise tolerance is hereby achieved without a loss of learning stability. The speed of ART 2-A makes practical use of ART 2 modules in large-scale neural computation.

Adaptive resonance theory (ART)[3] network for classification tasks which does not use the vigilance parameter. This feature is due to the geometry of categories in PTAM, which are irregular polytopes whose borders approximate the borders among the output predictions. During training, the categories expand only towards the input pattern without category overlap. The category expansion in PTAM is naturally limited by the other categories, and not by the category size, so vigilance is not necessary. PTAM works in a fully automatic way for pattern classification tasks, without any parameter tuning, so it is easier to employ for non expert users than other classifiers. PTAM achieves lower error than the leading ART networks on a complete collection of benchmark data sets, except for noisy data, without any parameter optimization.

The increasing popularity of social media is shortening the distance between people. Social activities, e.g., tagging in Flickr, bookmarking in Delicious, twittering in Twitter, etc. are reshaping people's social life and redefining their social roles. People with shared interests tend to form their groups in social media, and users within the same community likely exhibit similar social behavior (e.g., going for the same movies, having similar political viewpoints), which in turn reinforces the community structure. The multiple interactions in social activities entail that the community structures are often overlapping, i.e., one person is involved in several communities. We propose a novel co-clustering framework[4], which takes advantage of networking information between users and tags in social media, to discover these overlapping communities. In our method, users are connected via tags and tags are connected to users. This explicit representation of users and tags is useful for understanding group evolution by looking at who is interested in what. The efficacy of our method is supported by empirical evaluation in both synthetic and online social networking data

## 4. Methodology

The architecture of Fuzzy ART (Fig. 1) consists of the input field $F1$ for receiving input patterns and the category field $F2$ for clusters. Fuzzy ART described as

1) *Input Vectors:* Let I = x denotes the input pattern in the input field $F1$. Note that for x = $(x1, \ldots, xm)$, $xi \in [0, 1]$ $(i = 1, \ldots, m)$. With complement coding [2], x is further concatenated with its complement vector x such that I = $($x, .x$)$, where .x = 1 − x.

2) *Weight Vectors:* Let w$j$ denote the weight vector associated with the $j$ th cluster $c\,j$ ($j = 1, \ldots, J$) in the category field $F2$.

3)*Parameters:* The Fuzzy ART dynamics are determined by choice parameter $\alpha$ >0, learning parameter $\beta \in [0, 1]$, and vigilance parameter $\rho \in [0, 1]$.

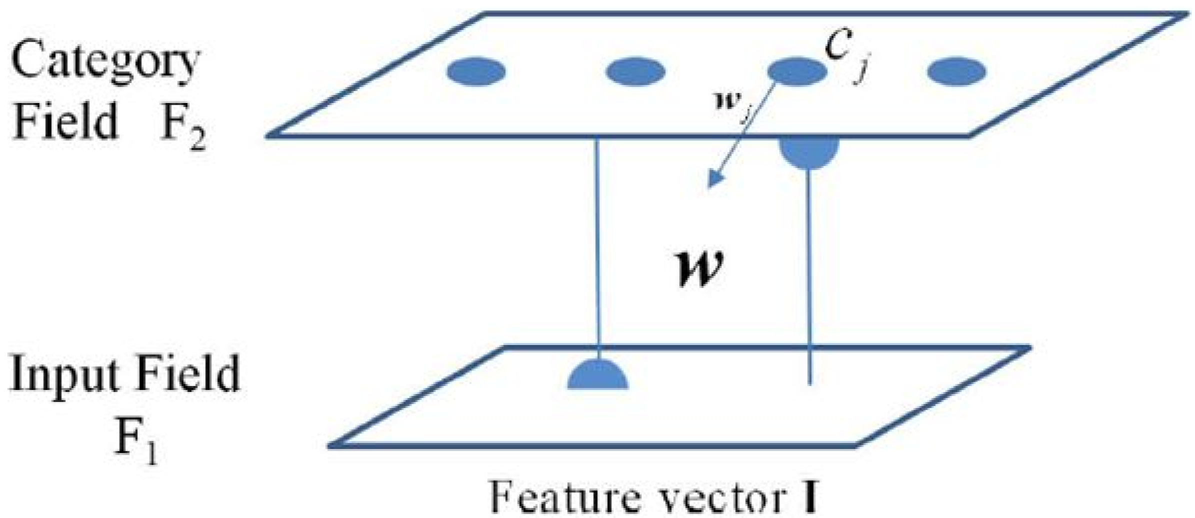


Fig1. Fuzzy Architecture.

The clustering process of Fuzzy ART has three key steps.

*1. Category Choice: For each input pattern, Fuzzy ART calculates a choice value according to the choice function for each of the clusters in the category field.*

*2. Template Matching: The similarity between the input pattern I and the winner cj is evaluated using a match function Mj.*

*3. Prototype Learning: If cj satisfies the vigilance criteria, its corresponding weight vector wj will be upload through a learning function.*

*Rules for adapting VP in Fuzzy ART are*

*1. Activation Maximization Rule: Input patterns are likely to incur resonances for the same clusters with a small vigilance value, while a large vigilance value may lead to the reset.*

*2. Confliction minimization Rule: The CMR minimizes the overlap between VRs of close cluster to produce better cluster boundaries. Incorrect recognition of* *patterns usually is caused by a small vigilance value. Well partitioned boundaries between clusters can minimize the risk of miscategorization.*

*3. Hybrid Integration of AMR and CMR: Candidate selection: Select the input.*

*a. Winner Identification: Select the cluster using match function.*

*b. Conflict Minimization: Update the vigilance parameters of all winner candidates*

*c. Activation Maximization: Search in the remaining clusters to identify the set of clusters.*

## 5.Conclusion

We investigated making the vigilance parameter in Fuzzy ART self-adaptable so as to permit Fuzzy ART to consistently produce high-quality clusters under modest parameter settings for large-scale social media data sets. The contributions of this paper are twofold. We proposed three adaptation rules, namely, the AMR, the CMR, and the HIR, for the vigilance parameter so that the clusters in the Fuzzy ART system would have individual vigilance levels and be able to adaptively tune their VR boundaries during the clustering process.

## 6.References

[1]L. Meng, A. Tan and D. C. Wunsch, "Vigilance adaptation in adaptive resonance theory," The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, 2013, pp. 1-7.

[2]G. A. Carpenter, S. Grossberg, and D. B. Rosen, “ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition,” NeuralNetw., vol. 4, pp. 493–504, Jul. 1987.

[3] D. G. Amorim, M. F. Delgado, and S. B. Ameneiro, “Polytope ARTMAP: Pattern classification without vigilance based on general geometry categories,” IEEE Trans. Neural Netw., vol. 18, no. 5,pp. 1306–1325, Sep. 2007.

[4]X. Wang, L. Tang, H. Gao, and H. Liu, “Discovering overlapping groups in social media,” in Proc. IEEE Int. Conf. Data Mining, Dec. 2010, pp. 569–578.