Stabilization by Conceptual Duplication in Adaptive Resonance Theory

Louis Massey

Royal Military College of Canada
Department of Mathematics and Computer Science
PO Box 17000 Station Forces
Kingston, Ontario, Canada, K7K 7B4
massey@rmc.ca

Abstract. Adaptive Resonance Theory (ART) neural networks are known for their plastic and stable learning of categories in data. Convergence to a stable representation is achieved through repetitive presentations of the data. It has been demonstrated previously that in a real-life setting, stability is not possible due to continuous novelty in data while the network forgets previous experiences. In this paper we show an improved version of ART that doesn't forget during stabilization. We present experimental results with this strategy in a task of text clustering.

Keywords: Adaptive Resonance Theory, neural networks, stable learning

1 Introduction

Adaptive Resonance Theory (ART) neural networks [1, 2] properties of stability and plasticity as well as their ability to process dynamic data efficiently make them attractive candidates for recognizing patterns in large, rapidly changing datasets generated in real-life environments. The applications of ART span many domains, including among others sonar signal recognition [3], parts management at Being [4] and text clustering [5].

Stability is an essential aspect of learning. Indeed, without stability a learning system becomes subject to catastrophic forgetting. Stability means that if an identical datum is presented several times to a learning system, it will be consistently recognized as being of the assigned category. It is trivial for a learning system to be stable: its merely has to stop learning on new data. This is what off-line learning systems do. ART on the other hand is an on-line learning system, allowing both continuous learning (plasticity) and guaranteeing a stable internal representation. ART converges to a stable representation after at most N-1 presentations of the N data items [6]. However, we have recently identified an important problem with ART stability while investigating its application to real-world problems [7]. In this paper, we present a new version of ART that resolves this difficulty. We test the new ART-based neural net in a topics recognition task using a benchmark text corpus.

2 Adaptive Resonance Theory

2.1 Description

In this paper we focus on the binary ART version known as ART1. The general architecture of an ART1 network is summarized on Fig. 1. The network is made of two interconnected layers of neurons and of an external control system (the box labeled C at the right of Fig. 1) that determines the operational mode of the layers. Weights w_{ij} exist on bottom-up connections going from input neuron i to output neuron j. There is one input neuron i for each component of an input vector $\mathbf{x}_k$ of dimension N. Weights t_{ji} are attributed to top-down connections, from output neuron j to input neuron i. Each output neuron j (j=1 to M) hence has an associated vector $\mathbf{t}_j$ constituted of the weights t_{ji} from the connections out of j. A vector $\mathbf{t}_j$ corresponds to the cluster *prototype*, that is the internal representation of the category learned by output neuron j. Similarly, there is an input activation vector $\mathbf{w}_j$ corresponding to the weights of connections going from the input layer to output neuron j.

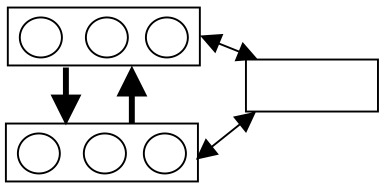

Fig. 1. The ART1 architecture showing the two interconnected layers of neurons and the external control system.

The input layer receives inputs and propagates them on the bottom-up connections, which causes activation of neurons on the output layer. The dot product between input x_k and bottom-up connections weight vectors determines the activation u_j of each output neuron j:

$$u_j = \mathbf{x}_k \cdot \mathbf{w}_j \ . \qquad (1)$$

Competitive selection takes place between output neurons. The winner selected is the neuron j* with maximum activation j*=arg max(u_j). The cluster represented by this output neuron is deemed to be the one with the greatest correlation with the input. The input is attributed to the winning output neuron and prototypes weights are updated as such:

$$\mathbf{t}_{j*}' = \mathbf{t}_{j*} \ \wedge \ \mathbf{x}_k \ . \qquad (2)$$

The prototype weight update with a logical AND guarantees a unidirectional movement of prototypes (monotonically decreasing magnitude) and thus also contributes to stability [8]. To stabilize, the network must iterate through the data until there is no further change of the input assignments to category nodes.

2.2 The Stability Problem

The ART stabilization process works as follows. First, assume that a datum x_k has just been processed by the neural network and is coded by prototype t_α. This is to say that x_k has been assigned to a cluster (or category node) of data represented by prototype t_α. Secondly, some of the further data processed by the neural net may also be assigned to this same cluster α and consequently prototype t_α will be updated to reflect the intersection of all assigned data as per formulae (2) above. Third, entering a stabilization iteration, datum x_k is presented again to the network and may not anymore be deemed similar enough to prototype t_α. This is possible because the prototype may have been changed by other data. The network must then find and reassign datum x_k to another prototype t_β. In a way, the ART network partially forgets previous experiences during stabilization to re-code assignments of data to new clusters. In other words, during stabilization, data is moved between concepts. When this movement stops, stabilization is achieved.

Stabilization is similar to sleep in living organisms, a period during which experiences of the day are re-processed and properly coded and re-coded in memory. For an artificial learning system such as ART used in a real world, high-volume, 24/7 operation, stabilization would have to occur during system idle time. The various iterations may not occur immediately one after the other as there may be more urgent tasks required, such as processing newly arrived data and delivering it to a user.

For instance, suppose the system under consideration is one that routes, based on topics, intelligence and operational documents to various military analysts. This information is highly perishable and must be processed with high priority, before any further on-going stabilization iteration can continue.

One must question what happens in between stabilization iterations with data awaiting stabilization. During the first processing pass, data will be assigned to some clusters. Then during stabilization, data will be moved, defeating the whole purpose of providing a stable and consistent environment to users. That is, a datum x_k has been initially assigned to cluster α. x_k may be an important document attributed to topic α. The users expect to always find the document under that same topic folder once it has been saved there the first time. However, later on the document may be moved to another topic as stabilization continues. Users will not find the document under the same topic. This can happen several times and is clearly a problematic situation. In fact, the whole idea of stabilization rests on the premise that convergence to the so-called stable representation is achieved after the ART network has been able to iterate through the whole dataset several times. In an incremental setting as for most real-life streaming data problems, there is never a state of "complete" dataset: new data continues to be submitted and hence stabilization can never be truly achieved.

2.3 Conceptual Duplication

The solution we propose is to treat stabilization not as "conceptual shifts" (i.e. data moving between concept) but rather as conceptual duplications. This idea of *Conceptual Duplication* for ART1 neural networks modifies stabilization in such a way that all associations between data and categories are remembered by the network, even those that would be invalidated by traditional stabilization. In other words, once a datum has been attributed to a cluster, the network remembers this association.

The memory in which conceptual attributions are stored is not part of the ART1 neural network structure itself. We do not claim neurological plausibility; we take an engineering approach to solving a practical problem. Hence, an assignment table is used to accumulate the various categories attributed to data elements. Furthermore, it may not be desirable to remember every single attribution, so we also introduce a parameter called *evidence*. Evidence is a positive integer value that specifies how many times a category has to be attributed to a datum before it is deemed worthy to remember. If the threshold of evidence is not reached, only the last attributed category is retained.

3 Experimental Work

3.1 Methodology

The text classification benchmark dataset known as Reuter-21578 Distribution 1.0 ModApté split [9] is used for the experiment. Each document is transformed in the standard vector space model numerical representation [10]. A document d is translated into an N-dimensional binary vector. The vector's i^{th} component corresponds to the i^{th} word in the collection. A value of 1 indicates the presence of this word in d while a value of 0 signifies its absence. Since the resulting vectors are of very high dimensionality, a final preprocessing step is applied to reduce the number of features. To achieve this goal, words occurring in less than 77 documents were removed. The value of the vigilance parameter is incremented successively by fine steps and the quality of clustering is measured for each value of vigilance. This allows for obtaining an overall and complete view of clustering quality, rather than only measuring quality punctually at the pre-determined so-called natural number of clusters.

Quality evaluation is conducted by computing the F_1 external validity of the clustering solution at each level. Better quality is achieved with higher F_1 values, in the range [0,1]. Due to lack of space, we refer the reader to [11] for a detailed account of the definition and computing of F_1 for text clustering.

3.2 Results

Our experimental results (Fig. 2) show that the F_1 quality improves when the value of evidence is increased from 1 to 3 (in the legend, Ev=X means that a category had to be assigned to a document X times before it was deemed important enough to be

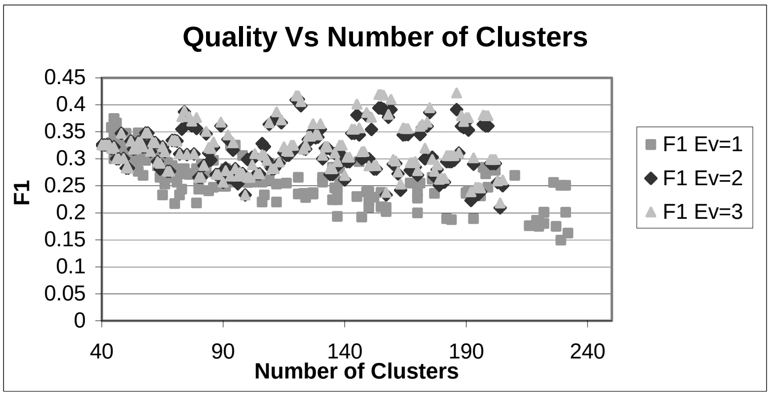

Fig. 2. Quality improves on all number of clusters found in the data as ART1 remembers previous concept attributions based on more evidence.

remembered during stabilization). The average quality for the three cases of evidence =1, 2 and 3 respectively is 0.28, 0.32 and 0.33.

Further increases in evidence result in little further gain. In fact, as one becomes more demanding with evidence, there is less and less opportunities for conceptual duplication to actually occur. Consequently, the solution tends to turn into the usual hard clustering of ART1. Although not shown on Fig. 2, this is exactly how the ART1 quality Vs number of clusters evolves *without* conceptual duplication, that is with the usual ART stabilization. In this latter case, average F1 quality is 0.34.

Hence, quality is not improved by conceptual duplication but the decrease in quality is minimal, particularly in the case of an evidence of 3 for which $F_1 = 0.33$ compared to $F_1 = 0.34$ without conceptual stabilization. Furthermore, the advantages of solving the stated problem and offering a soft-clustering algorithm still make the approach worthy. Soft-clustering is for example very useful in text clustering since multiple topics can be assigned to a document, making them more easily accessible for users. It is actually a more natural way to organize documents than hard clustering since documents are rarely of a single topic according to human classifiers, a phenomenon known as the *inter-indexer inconsistency* [12].

4 Conclusion

In this paper, we discuss the problem with data elements moving between categories in a dynamic environment when using ART1. We have proposed conceptual duplication to resolve this issue. In this case, the nature of stability has changed radically; from stability that iteratively fixes a single concept attribution onto each datum of a static dataset to a stabilization that accumulates evidence to determine appropriate, possibly multiple conceptual attributions. With conceptual duplication, past concept attributions that meet the evidence criterion are remembered rather than being forgotten as in the existing ART1 algorithm.

Experimental results in a text clustering task have shown that as greater evidence is demanded, average F_1 quality increases but rapidly tapers off since higher evidence forces back the network into its usual behavior. Stabilization with conceptual duplication does not result in text clustering that exceeds the quality of the usual ART1 stabilization. However, it offers two major advantages: First, it resolves the important problem about data moving between categories during stabilization in a dynamic data environment; and second, it provides a soft-clustering solution.

References

1. Grossberg, S.: Adaptive pattern classification and universal recording : I. Parallel development and coding of neural feature detectors", Biological Cybernetics, Vol 23 (1976) 121-134.
2. Carpenter, G.A. and Grossberg, S.: Adaptive Resonance Theory (ART). In: Handbook of Brain Theory and Neural Networks, Ed.: Arbib M.A., MIT Press (1995).
3. Carpenter, G.A. and Streilein, W.W.: ARTMAP-FTR: A neural network for fusion target recognition, with application to sonar classification. AeroSense. In: Proceedings of SPIE's 12th Annual Symposium on Aerospace/Defense Sensing, Simulation, and Control. Orlando, April 13-17, 1998, Bellingham, WA: Society of Photo-Optical Instrumentation Engineers.
4. Caudell, T., Smith, S.D.G., Johnson, C., Wunsch, D., and Escobedo, R.: An industrial application of neural networks to reusable design. Adaptive Neural Systems, Technical Report BCS-CS-ACS-91-001, Seattle, WA: The Boeing Company (1991) 185-190.
5. Massey, L.: On the Quality of ART Text Clustering , Neural Networks, 16(5–6): 771–778 (2003).
6. Georgiopoulos,M., Heileman, G.L. and Huang, J.: Convergence properties of learning in ART1. Neural Computation, 2(4) 502-509, 1990.
7. Massey L.: Real-World Text Clustering with Adaptive Resonance Theory Neural Networks, In: Proceedings of 2005 International Joint Conference on Neural Networks, Montreal, Canada, July 31- August 4, 2005.
8. Moore B.: ART and Pattern Clustering, Proceedings of the 1988 Connectionist Models Summer School, p. 174-183, 1988.
9. Apte, C. Damerau, F. and Weiss, S.M.: Automated learning of decision rules for text categorization, ACM Transactions on Information Systems, 12(2):233-251, 1994.
10. Salton, G. and Lesk, M.E.: Computer evaluation of indexing and text processing. Journal of the ACM, Vol 15, no. 1, pp 8-36, January 1968.
11. Massey L.: An Experimental Methodology for Text Clustering, In: Proceedings of 2005 IASTED International Conference on Computational Intelligence (CI 2005), Calgary, Canada, July 4-6, 2005.
12. Cleverdon, C.: Optimizing convenient online access to bibliographic databases. Information Services and Use 4 (1). (1984). pp.37-47.