Adaptive resonance theory based neural networks — the 'ART' of real-time pattern recognition in chemical process monitoring?

Dietrich Wienke *, Lutgarde Buydens
Nijmegen, The Netherlands

The family of artificial neural networks based on Adaptive Resonance Theory (ART) forms a collection of distinct mathematical pattern-recognition methods. The classification of sensor signals, process data analysis, spectral interpretation, and image analysis are discussed as applications of ART outside and within chemistry. The advantages of ART are considered. They include its use as a built-in detector for outliers, its rapid training speed, self-organizational behaviour, full chemical interpretability, and real-time and on-line applicability. A glossary of terms used in ART is given at the end of the article.

1. Introduction

Adaptive Resonance Theory (ART) based neural networks were introduced by Grossberg [1,2] as rather theoretical neural models, describing selected aspects of the classification behaviour of biological brains. However, ART models are also capable of solving technical and chemical pattern recognition tasks. The research on, and application of pattern recognition methods (PARC) in analytical chemistry, spectroscopy and chemical process monitoring was mainly initiated by the pioneering work of Jurs, Kowalski and Isenhour in 1969 on the classification of mass spectra using the Linear Learning Machine [3]. Implicitly, via a cross-reference (Ref. [2] in [3]) the authors had already made a link between their PARC method and the term 'artificial neural network'. Explicitly, this term was used in 1975 by Stonham et al. [4] within their method called the 'Adaptive Digital Learning Network', which also has been applied for classifying mass spectra. However, the PARC methods most used in chemistry in the seventies and the eighties were techniques based on statistical and graph theory, such as cluster analysis, discriminant analysis, principal component analysis and regression models [5]. The development and application of these classical pattern recognition methods dominated research in chemometrics. This found its expression in a number of contributions to textbooks for chemists [6–9]. It took another fourteen years before the term 'artificial neural network' was again used in chemometrics in connection with chemical pattern recognition [10]. Then, within a very short time, an enormous number of studies appeared on the application of artificial neural networks in chemistry, spectroscopy and process analysis. These studies dealt mainly with the 'Multilayer Feedforward Network' based on Rumelhard's back-propagation of error based learning rule (MLF-BP) and have also been inspired by Kohonen's self-organizing feature map, which was reflected in the book by Zupan and Gasteiger [11] and other reports [12,13]. Parallel with the interest of chemists in the theory of artificial neural networks PARC methods became increasingly applied in daily analytical practice in industry for chemical process monitoring, in routine quality control, in multisensor technology and in routine environmental analysis. The food and pharmaceutical industries are representative of this trend —by the use of remote rapid-scan near-infrared spectroscopy combined with a PARC method such as PCA, final and intermediate products can be checked continuously by their spectroscopic fingerprints to guarantee the company and its customers the consistently highest quality. However, this trend to the application of classical PARC methods in industry has also provided critical evaluation of the existing PARC methods. The present authors have often been told that there is a need for alternative PARC methods with the following properties:

- a higher speed of classification,
- a higher power of discrimination and prediction,
- ease of chemical interpretation, with no need of 'black boxes',

* Corresponding author. e-mail: wienke@sci.kun.nl.

0165-9936/95/$09.50

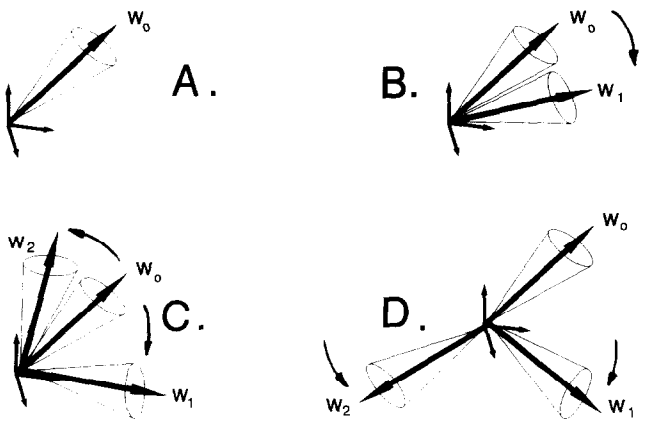

Fig. 1. Cluster formation in the features space during training an ART neural network.

- that they should be robust against outliers, and extra-and interpolations
- be simple to recalibrate and to configure for other applications, and
- be able to fulfil on-line and real-time classification tasks in production processes and in automated analytical systems.

The general theoretical interest in artificial neural networks and the critical evaluation of existing PARC methods by practical chemists, encouraged chemometricians to advance their research in PARC. In this, the study and application of artificial neural networks based on ART represent a trend of increasing importance.

2. The basic ideas of ART

The starting points in ART are n multivariate-described samples $\mathbf{X}$ (dimension $n \times m$), each given by measurements taken of m different features. A typical example is a set $\mathbf{X}$ of n spectra (or chromatograms or sensor data), whereby each $\mathbf{x}_i$ has a length m. The quantity n stands for different samples, mixtures or process stages, and m for different wavelengths, retention times, or sensor diodes, etc. Each multivariate measurement $\mathbf{x}_i$ is thus a directed vector in the m-dimensional space of m features. Some of the n vectors can be closer to each other in this space, forming groups if their m values are more similar to one another (Fig. 1). The aim of ART networks is to find groups with similar sample vectors $\mathbf{x}$ among the n vectors, whereby their number, c, is not a priori known. Classically expressed, ART is a multivariate data-clustering method. The clustering process, called 'network training', starts with a random selection of an arbitrary i-th sample, $\mathbf{x}_i$, out of the entire set of training samples $\mathbf{X}$ and copying it to a future 'long-term memory (LTM)' of 'weights'. In the easiest case, this will be simply a new vector $\mathbf{w}$ of length m representing a raw estimate of a first (new) cluster (Fig. 1). In other cases, several scaling operators follow this copy step. Anyway, the weights in ART are thus also vectors describing the direction of the clusters in the features space. After the initialization of a new (or first) cluster, another $\mathbf{x}_j$ is randomly selected and compared mathematically with all the hitherto existing k clusters, $\mathbf{W}$ ('network weights'). During such a comparison, a virtual image $\mathbf{x}_j^*$ of $\mathbf{x}_j$ is generated internally within the network. This virtual image either fits to the original or it significantly deviates. Both cases will be discussed later. Simultaneously, $\mathbf{x}_j^*$ also forms a kind of temporary short-term memory (STM) for $\mathbf{x}_j$. After comparison with all clusters ('weights'), a 'winner' among them can be found, providing a virtual image $\mathbf{x}_j^{*,\text{winner}}$ having the highest similarity (lowest dissimilarity) to $\mathbf{x}_j$. Dissimilarity is expressed in ART as ρ^{calc} via distinct formulae. One may think in the simplest version about a Euclidian angle between $\mathbf{x}_j$ and each $\mathbf{w}_j$. At this point, classical cluster analysis would stop with further similarity comparison. But in ART, the highest similarity between sample $\mathbf{x}_j$ and cluster $\mathbf{w}^{\text{winner}}$ is not the only satisfying criterion for their future fusion. In a second crucial step there is a check to find whether the virtual image $\mathbf{x}_j^{*,\text{winner}}$ generated by the network is similar enough to the original actual input, $\mathbf{x}_j$. In other words, the network asks itself (as we would do ourselves): "Did I ever see this input (e.g., this person), or do I not know it (e.g., him) yet ?". Technically, this is achieved by comparing the numerically calculated dissimilarity $\rho_{\text{winner}}^{\text{calc}}$ between $\mathbf{x}_j$ and $\mathbf{w}^{\text{winner}}$, using a constant, defined in advance, called a 'vigilance' parameter, $\rho^{\max}$. This $\rho^{\max}$ forms a fixed spatial limit around each cluster. Thus, if

$$\rho_{\text{winner}}^{\text{calc}} \leq \rho^{\max} \qquad (1)$$

then $\mathbf{x}_j^{*,\text{winner}}$ is similar enough to $\mathbf{x}_j$, or classically expressed, $\mathbf{x}_j$ falls inside the borders of the existing cluster $\mathbf{w}^{\text{winner}}$. For simplicity, $\rho_{\text{winner}}^{\text{calc}}$ and $\rho^{\max}$ are always given values 0–1 (achieved by suitable scaling procedures within ART). A value of $\rho^{\max}$ close to unity, means a high dissimilarity. Even a small deviation between $\mathbf{w}^{winner}$ and $\mathbf{x}_j$ is reason enough for opening a new separate cluster. High dissimilarity will thus provide many new clusters

with small diameter. In the reverse case, where ρ^{max} is chosen close to zero (low dissimilarity), only a few, but large, clusters will be generated. If Eq. 1 is fulfilled, one says that the network came into 'resonance' with this type of input, or it already 'knows' this phenotype.

In the other case, if

$$\rho_{winner}^{calc} > \rho^{max} \quad (2)$$

the network discovered a 'novelty'. Here, $\mathbf{x}_j^{*,winner}$ is not close enough to the closest $\mathbf{w}^{winner}$.

After the resonance check according to Eq. 1 and Eq. 2, the third crucial step follows —called 'adaptation'. In the case that Eq. 1 is fulfilled, the network changes the weights of the winning cluster by a small step, given by a step width called η (the 'learning rate'), towards the spatial position of the actual input vector $\mathbf{x}_j$. The entire calculation formula for adaptation of the weights containing η is called a 'learning rule'. In practice, η can be chosen between 0 and 1, whereby a value of η close to zero will provide only small changes in the weights. In this way, an ART network stores a weighted part of the present input vector in the LTM, just as any other artificial neural network does. If Eq. 1 is not true but Eq. 2 is fulfilled, the network does not adapt its weights but adapts its structure to the discovered novelty. 'Structure adaptation' means simply adding a new cluster to the existing ones. The novelty is immediately stored in the additional weights. This is another unique idea of Grossberg: ART neural networks not only use their weights but also their structure for information storage and for data fitting. After this step of adaptation, another input vector is randomly selected from training data, and the entire process of 'resonance' and 'adaptation' is repeated, whereby the former content of the STM is repeatedly overwritten by new virtual inputs, as happens in the biological brain. Random sampling of n times the training data matrix $\mathbf{X}$ is called one 'epoch' of training. Simultaneously, the contrast between clusters in the LTM increases ('generalization'). The process converges within a few epochs of training, with the formation of c clusters, whereby the previously chosen constants of learning rate, η, and vigilance parameter, ρ, determine how many clusters c are formed (Fig. 1). In other words, by a suitable choice of η and ρ the data cloud, $\mathbf{X}$, can be approximated by (or resolved to) a number, c, of clusters (Fig. 2). In this way, a variation of η and ρ can serve for active data exploration of $\mathbf{X}$. As in classical pattern recognition, it becomes clear also for

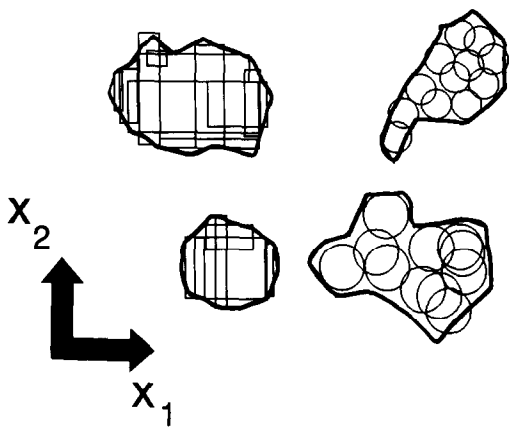

Fig. 2. The methods FuzzyART and FuzzyARTMAP approximate data clouds in the features space by distinctly numbers of distinct shaped hyperrectangulars. ART-1, ART-2, ART-2a, ARTMAP and ART-3 use distinct numbers of equally sized hyperregulars.

ART, that the more compact, point-shaped and well-separated the sub-clusters are, that are hidden in the n data vectors, the less important the ART user's choice of η and ρ will be. In this situation, ART will always find the true number, c, of hidden clusters. However, the more the data are scattered, or the more they form a continuous hypersurface of equidistant points, the higher the influence of η and ρ on the clustering result, and on the number c, will be.

3. Unsupervised working of ART for clustering

Historically, various unsupervised ART methods (ART-1 [1,2,14], ART-2 [15], ART-2a [16], ART-3 [17] and FuzzyART [18]) were developed before the supervised ART techniques (ARTMAP [19], FuzzyARTMAP [20]). 'Unsupervised' means that the dissimilarities among the n sample vectors are only considered in their measurement space $\mathbf{X}$ (X-space) for clustering. After clustering this X-space, the formed clusters can be assigned by the chemist to distinct process situations, product qualities, material types, etc. ART-1 only tolerates binary ('0' or '1' coded) numbers within an input vector $\mathbf{x}_i$. ART-2, ART-2a, ART-3 and FuzzyART can process any real number, scaled to the continuous range between 0 and 1.

In unsupervised ART, the n samples are not compared pairwise with each other with respect to their dissimilarity as they are in hierarchical cluster analysis or in the k-th nearest neighbor method, which

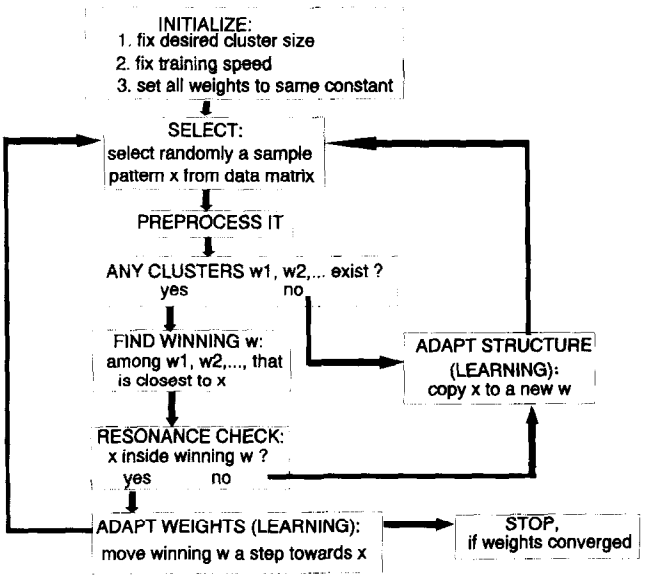

Fig. 3. Calculation steps during training of a FuzzyART (or an ART-2a) neural network.

require the calculation of a rapidly growing distance matrix of size $(n^2-n)/2$. In ART, the n samples have to be compared only with c clusters, keeping the number of computational steps growing only linearly for increasing n. Thus, unsupervised ART is attractive for rapid clustering of data sets with a large number, n, of multivariate samples.

The unsupervised 'training' of ART-1, ART-2, ART-2a, ART-3 and FuzzyART finally provides three results that are interesting for the chemist. First, is the number, c, of clusters for a certain chosen size of ρ and η. This cluster structure says something about the heterogeneity of the data cloud $\mathbf{X}$. The number of clusters can often be related to different material types, process stages, product qualities, or environmental pollution situations, for example. Note that because of the unit-sized diameter of the c formed ART clusters, fixed via the previously defined vigilance constant, ρ^{max}, distinctly sized and shaped data clouds are approximated by distinct numbers of equally sized ART clusters. This is a very different approach from classical methods which try to model a single data cloud by a single, but distinctly shaped cluster. Second, the weight matrices $\mathbf{W}$ for most of the unsupervised ART methods can simply be rescaled and decoded again to the numerical level of the original input data, $\mathbf{X}$. The weight vectors can be considered as the centroid of a cluster reflecting in its m values the m main properties for this particular cluster. This opens the possibility of understanding the chemical nature of a cluster. Third, a trained ART-1, ART-2, ART-2a, ART-3 or FuzzyART neural network can always continue with the classification and learning of new samples. Thus, an unsupervised trained ART network can always be used as a supervised classifier, too. It requires that the nature of the formed network clusters can be assigned by the chemist to some distinct classes (material types, process stages, product qualities, etc.) after training and before network testing with unknown samples.

Table 1
Rescaled weights matrix from a trained ART-2a neural network

Body feature	Weight vector			
	$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{w}_3$	$\mathbf{w}_4$
Body length/cm	167.6	177.3	187.5	184.7
Body weight/kg	59.9	73.1	73.1	85.9
Footsize/a.u.	38.5	41.6	44.4	43.6
Stomach/cm	73.4	86.5	86.5	100.7
Neck/cm	33.9	37.6	37.6	41.5

The 47 male and female participants in courses of chemometrics were asked for their $m=5$ body measures 'weight (kg)', 'length (cm)', 'shoe size (a.u.)', 'stomach (cm)' and 'neck (cm)'. The obtained 47×5 data matrix has been clustered with ART-2a using the parameter settings $\rho^{max}=0.11$ and $\eta=0.1$. After ten training epochs, the network had stabilised dividing the 47 persons in $c=4$ clusters, given by their four corresponding weight vectors $\mathbf{w}_i$, each of length $m=5$.

Because of their high speed and their algorithmic simplicity, ART-2a and FuzzyART are the most interesting unsupervised ART methods for chemical applications (Fig. 3). After the training process, each of their c clusters is given by a directed weight vector in the m-dimensional X-space, thus in total providing a $c\cdot m$-dimensional weight matrix, $\mathbf{W}$, as LTM (Fig. 1). There are only two differences between ART-2a and FuzzyART. The first is in their measures of dissimilarity to estimate ρ^{calc} between $\mathbf{x}_j$ and $\mathbf{w}_I$ (Euclidian angle or intersection operator from fuzzy set theory). The second is in the way they preprocess their data (scaling of $\mathbf{X}$). The intersection operator based on fuzzy set theory in FuzzyART requires 'complement coded' input data between 0 and 1. Because 'complement coding' is also necessary for the supervised FuzzyARTMAP method, it will be separately explained later, in a special section.

An illustrative example for ART-2a, taken from Wienke [21], is given in Table 1. This weights matrix, $\mathbf{W}$, gave the result that the data cluster of male persons splits into three subclasses, $\mathbf{w}_4$ of

'large heavy men', $\mathbf{w}_3$ 'tall slender men' and $\mathbf{w}_2$ 'normal men', with significant distinct measurements of neck, shoe size and stomach: women form one homogeneous cluster $\mathbf{w}_1$. However, between $\mathbf{w}_1$ and $\mathbf{w}_2$ a particular overlap was observed, caused by a few mutual mis-classifications of men and women into both clusters. The reason for this was that some tall women from northern European countries were quite similar in the five body measurements taken to small male participants from southern Europe.

4. Supervised working of ART for classification

Recently, Carpenter et al. [19,20] introduced two supervised-working ART methods, called ARTMAP and FuzzyARTMAP. During their training phase these classifiers need for each sample two input vectors, $\mathbf{x}_i$ and $\mathbf{y}_i$, where $\mathbf{x}_i$ again contains measurement values (a spectrum, a chromatogram, or multivariate process data). The additional vector, $\mathbf{y}_i$, contains p 'desired outputs' or p 'given class memberships'. An example is the membership of a spectrum, $\mathbf{x}_i$, in a particular molecular structure class coded by $\mathbf{y}_i$. Another example is a vector $\mathbf{x}_i$ of measured material properties (hardness, color, etc.), obtained by setting p process parameters $\mathbf{y}_i$ (e.g., pressure and temperature) for the chemical reactor in which the product is synthesized. The ARTMAP (or FuzzyARTMAP) now consists of three partial ART networks (Fig. 4). Each of these three partial networks (including the 'mapfield') needs its individual, previously fixed learning rate, η, and vigilance parameter, ρ^{max}, giving in total $3 \times 2 = 6$ parameters. Changing the size of these parameters allows a further data exploration with distinct (fine or raw) cluster resolutions, but now of the two coupled variable-spaces $\mathbf{X}$ and $\mathbf{Y}$ simultaneously. For example, by choice of suitable parameter sizes many small clusters in the X-space can be linked via the mapfield to only a few large clusters in the Y-space, or vice versa. For chemical applications FuzzyARTMAP is more interesting than ARTMAP, because it tolerates not only binary coded input data, as ARTMAP does, but also any other real coded data in $\mathbf{X}$ and $\mathbf{Y}$.

5. Complement coding as data preprocessing

Before complement coding a vector, its elements have to be scaled to the range between 0 and 1.

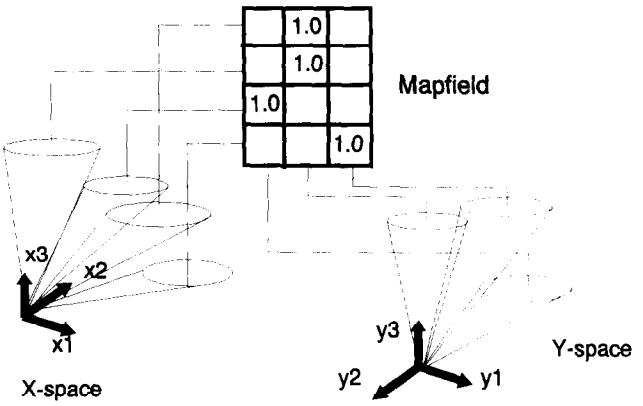

Fig. 4. One of the three partial neural networks in the FuzzyARTMAP algorithm clusters the n vectors $\mathbf{X}_{n,m}$ in their m-dimensional X-space. The second one clusters the n $\mathbf{Y}_{n,p}$ vectors in their p-dimensional Y-space. Both processes run parallel and are watched, monitored and controlled ('supervised') by a third neural network having the special name, 'mapfield'. After training, this mapfield is finally nothing else then a data array, $\mathbf{W}_{c,b}$ of size $c \times b$. It contains numbers, called 'inter-networks weights', that describe the strength of the link between the $k = 1 \ldots c$ clusters in the X-space to the $l = 1 \ldots b$ clusters in the Y-space.

'Complement coding' then means that each input vector, $\mathbf{x}_i$, is doubled in length, m, by appending its complement vector, $\mathbf{1} - \mathbf{x}_i$, before offering it to the ART neural network. We illustrate this by the following example. Assume an arbitrarily chosen five-element input vector

$$\mathbf{x}_i = [0.1, 0.7, 0.2, 0.1, 0.8]$$

After complement coding, the ten-element input vector

$$\mathbf{x}_i^{cc} = [0.1, 0.7, 0.2, 0.1, 0.8, 0.9, 0.3, 0.8, 0.9, 0.2]$$

is obtained. In the case of the supervised-working FuzzyARTMAP, the desired output, $\mathbf{y}_i$, is complement coded, too, by its corresponding vector $\mathbf{1} - \mathbf{y}_i$. For example, assume a two-class problem, coded by a binary one-element output vector $\mathbf{y}_1$, whereby $\mathbf{y}_1 = 1$ is defined for a sample from the desired output class 1 and $\mathbf{y}_1 = 0$ as the desired output for a sample of class 2. After complement coding, the following binary coded two-element vectors $\mathbf{y}_1^{cc}$ are obtained

$$[\mathbf{y}_1, \mathbf{1} - \mathbf{y}_1] = \mathbf{y}_1^{cc} = [1, 0] \text{ for class 1}$$

and

$$[\mathbf{y}_1, \mathbf{1} - \mathbf{y}_1] = \mathbf{y}_1^{cc} = [0, 1] \text{ for class 2}$$

are obtained.

Complement coding gives 'absent features', having low values of their data, high contributions in the fuzzy-set-theory-based measures for dissimilarity just as 'present features' that usually have higher values. Complement coding can also be considered as a technique that substitutes a single sharply defined feature by a 'feature range' with an upper and a lower limit. In this way, for two or more features hyperrectangular-shaped boxes are obtained for each cluster instead of single directed vectors. The boxes grow during training, from a small point vector to large hyperrectangular ones. Thus, an experimental data cluster will be approximated by FuzzyART (or FuzzyARTMAP) by a sequence of overlapping hyperrectangulars. The outer limits of such a hyperrectangular are called in the present work 'upper fuzzy bounds' and 'lower fuzzy bounds'. After rescaling, a cluster position, obtained from a FuzzyART network run (or FuzzyARTMAP), is given by 'fuzzy ranges' for all m features instead of single sharp values as given by ART-1, ART-2, ART-2a and ART-3 and ARTMAP.

6. Qualification versus quantification

Summarizing the former theoretical discussion, it became obvious that ART networks are more applicable for pattern recognition (qualification) than for function fitting (quantification). This is because each ART cluster (or class) describes a limited local area in the features space. In this way, complicated irregular shaped data clouds can be approximated, in principle, by a sequence of overlapping ART hyperclusters (radial or rectangular). It has been shown [19,20,22,23] in theoretical studies that FuzzyARTMAP can approximate data clouds that are located within each other or are formed from twisted spirals. MLF-BP networks, for example, were less able to model such data. However, the more a data set forms a continuous function in the variables space, the more new single-weight vectors are required. Function fitting is thus always possible with ART, but the cost one has to pay for this is a rapid proliferation of an enormous number of new single clusters.

7. Applications of ART outside chemistry

Grossberg, Carpenter and their co-workers tested their ART algorithms mostly with simulated data sets of their own or from other authors. However, in their paper introducing ARTMAP [19] they used a large database of mushrooms for a classification. A first application of ART-2 to automated identification of written characters has been reported by Gan and Lua [24]. Burke [25] compared ART with the k-means clustering method. Caudell presented the use of ART-1 in combination with parallel operating opto-electronic multichannel detector arrays for commercial and military remote control tasks [26]. This study is interesting for developers and researchers dealing with rapid, parallel data processing for chemical multisensor arrays. Real-time processing of sensor data becomes an increasingly serious problem in process control and environmental monitoring. Optical, opto-electronic and electronic hardware implementations of ART-1 and ART-2 were reported by Kane and Paquin [27], Wunsch et al. [28,29] and by Ho et al. [30]. These hardware implementations by optical lenses make ART extremely fast and very attractive for high-speed pattern recognition tasks. In fact, an optical realization of a neural network is mostly faster than an implementation via a microelectronic neural chip.

An application of ART-2 to pattern recognition with image data was reported recently by Resch and Szabo [31]. Benjamin et al. [32] found that MLF-BP neural networks have a better discrimination power than ART-2, in a study using economic data on the optimal location of future industrial facilities in the USA. Hohenstein [22] studied the applicability of FuzzyARTMAP in medical data analysis. He performed a supervised off-line classification of electromagnetic field patterns taken experimentally from the brains of individual patients, and found that FuzzyARTMAP discriminates at least as well as the MLF-BP neural network. Willems [23] has given a theoretical overview of all ART methods, from the point of view of information science.

8. Applications of ART in chemistry

During last two years, chemists have started to study and apply a variety of ART methods. Lin and Wang [33] fitted process-analytical time series data by an autoregressive model. The characteristic pattern vectors of autoregressive function parameters which were obtained were successfully classified using ART-2. Whitley and Davis [34,35] proposed to use a trained ART-2 network

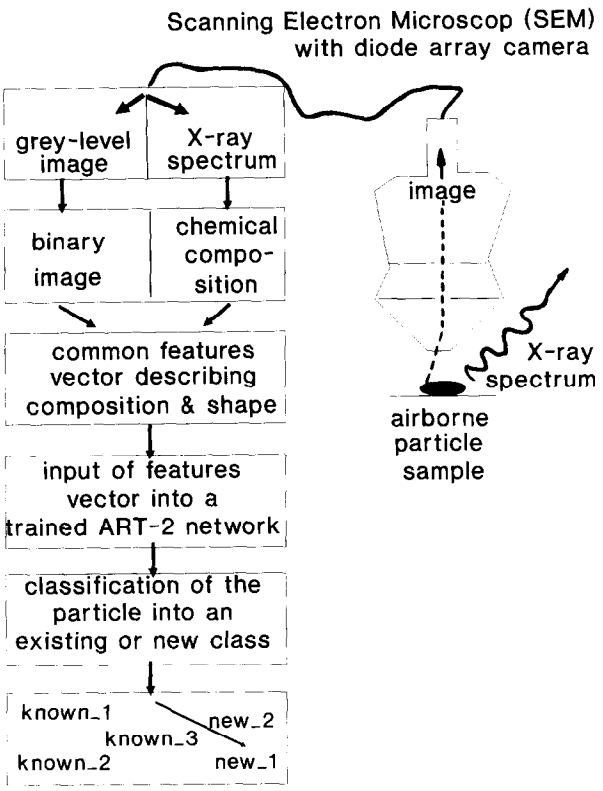

Fig. 5. Classification of airborne particles with respect to their origin using an unsupervised trained ART-2a neural network as a future supervised classifier. The scanning electron microscopy image and the spectral X-ray fluorescence emission of the particle are combined to a common features vector as input to the ART-2a network. In this particular application, ART is very useful because it can give a signal if an 'unknown particle', originating from a 'new, unexpected' air pollution source, is detected.

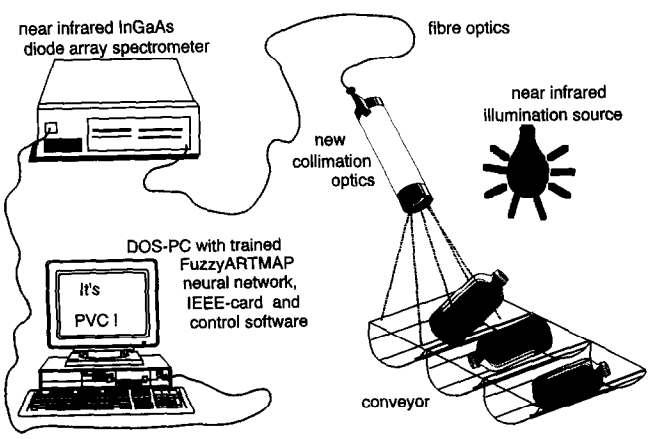

Fig. 6. A trained FuzzyARTMAP neural network as a real-time classifier has been implemented in a remote optical sensor device for rapid sorting of post-consumer plastic waste.

for monitoring and control of chemical reactors by real-time interpretation of sensor data taken from a reactor. They demonstrated that the ART network correctly predicted 'normal' and 'not normal' situations of the reactor during the running chemical reaction.

Wienke and Kateman [36] used ART-1 to classify UV/VIS-and IR-spectra for spectral interpretation. They showed the quantitative chemical interpretability of the ART weights after a suitable back-transformation in terms of the spectral peaks. Wienke, Xie and Hopke [37,38] applied ART-2a in environmental monitoring (Fig. 5).

A comparison of ART-2a versus methods based on principal component analysis, the SIMCA classifier and MLF-BP neural networks, has been given by Wienke et al. [39]. They applied ART-2a to the rapid sorting of post-consumer plastics by remote NIR spectroscopy. By a large cross validation study with distinct training and test data sets it was shown quantitatively that ART-2a classifies better than SIMCA but less well than MLF-BP neural networks. However, the advantages of ART-2a against MLF-BP networks are the much higher training speed, built-in detector against outliers, and full chemical interpretability. A comparison of FuzzyARTMAP with MLF-BP networks and the partial least squares method for rapid sorting of post-consumer plastic waste (Fig. 6) has been reported by Wienke et al. [40,41]. After careful data preprocessing by scaling and feature-selection it was found that FuzzyARTMAP was able to train much faster and to classify significantly better. Additionally, the trained neural network could be fully interpreted in spectroscopic terms.

9. Software and computers

ART neural networks can be trained using personal computers if the training data set, n, and the number of features, m, per sample are of moderate size. In the present study, data sets of $n = 500$ and $m = 256$ and more caused no trouble. For significantly larger data sets, a Unix workstation is recommended. The software for ART-1 has been written in MATLAB. ART-2a was developed in TurboPascal and FuzzyARTMAP in C-language (Borland-C, C for Unix-gcc-compiler).

10. ART Glossary

Artificial Neural Networks: empirical algorithms that try to model the way in which biological neural networks perform pattern recognition.

Network training: stepwise, very often repeated, fitting of a set of parameters (see 'weights') to experimental data.

Stability-Plasticity-dilemma: a fundamental question in machine-assisted learning is, "how can a network have enough plasticity and simultaneously enough stability to be able to adapt to 'new knowledge' and also to 'distinct, very new knowledge' "?

ART: Adaptive Resonance Theory, solves the stability-plasticity dilemma by offering two alternative ways of network learning (see 'adaptation').

Weights in ART: a set of numerical fitting parameters which model the locations of clusters in the data space.

Adaptation: concerns two distinct types of ART network learning, by changing either the ART weights alone, or by changing the ART network structure towards the offered training data.

Resonance: a trained ART network identifies certain data offered to it (new knowledge) as belonging to an already 'known' phenotype. It learns these data by weights adaptation.

No resonance: implies strongly deviating data (very new knowledge) that do not fit into the existing network structure. These identified novelties require from the network an adaptation of its structure.

Vigilance parameter: models the outer spatial border line of an ART cluster. It serves as a decision limit defining whether additional training data represent novelties or not.

STM: Short Term Memory means, the way ART reflects in its 'mind' a presented input pattern (as temporary, abstract or virtual mirror image of the original).

LTM: Long Term Memory, is the stored, weighted contribution of the present input data in the form of weights in an ART network

Fast ART learning: if weights are changed with a large step size towards the present input data (reached by a high learning rate, η). The content of the LTM will closely follow the usual rapid oscillations of the content of the STM (almost no LTM will be formed by training).

Slow ART learning: after their initialization, weights are almost unchanged in subsequent training steps (reached by a learning rate, η, close to zero). The contribution of subsequent training data to the content of the LTM will thus be almost negligible.

Acknowledgements

The authors are grateful to Prof. E. Pretsch (ETH Zürich) for his kind invitation to write the present article. Thanks are extended to the reviewer for his valuable comments.

References

[1] S. Grossberg, Biol. Cybern., 23 (1976) 121–134.
[2] S. Grossberg, Biol. Cybern., 23 (1976) 187–203.
[3] P.C. Jurs, B.R. Kowalski and T.L. Isenhour, Anal. Chem., 41 (1969) 21–26.
[4] T.J. Stonham, I. Aleksander, M. Camp, W.T. Pike and M.A. Shaw, Anal. Chem., 47 (1975) 1817–1822.
[5] C. Albano, W. Dunn III, U. Edlund, E. Johansson, B. Nordén, M. Sjöström and S. Wold, Anal. Chimica Acta, 103 (1978) 429.
[6] K. Varmuza, Pattern Recognition in Chemistry, Lecture Notes in Chemistry, Springer-Verlag, Berlin, Heidelberg, New York, 1978.
[7] P.J. Lewi, Multivariate Data Analysis in Industrial Practice, Research Studies Press, Chichester 1982.
[8] M.A. Sharaf, D.L. Illman and B.R. Kowalski, Chemometrics, Wiley, New York, 1986.
[9] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufman, Chemometrics —A Textbook (Data Handling in Science And Technology), Elsevier, Amsterdam, 1988.
[10] J.U. Thomson and B. Meyer, J. Magn. Reson., 84 (1989) 84.
[11] J. Zupan and J. Gasteiger, Neural Networks for Chemists, VCH, Weinheim, 1993.
[12] J.R.M. Smits, W.J. Melssen, L.M.C. Buydens and G. Kateman, Chemom. Intell. Lab. Syst., 22 (1994) 165–189.
[13] J.R.M. Smits, W.J. Melssen, L.M.C. Buydens and G. Kateman, Chemom. Intell. Lab. Syst., 23 (1994) 267–291.
[14] S. Grossberg, Studies of Mind and Brain, Reidel, Dordrecht, 1982.
[15] G.A. Carpenter and S. Grossberg, Appl. Opt., 12 (1987) 4919–4930.
[16] G.A. Carpenter and S. Grossberg, Neural Networks, 3 (1990) 129–152.
[17] G.A. Carpenter, S. Grossberg and D.B. Rosen, Neural Networks, 4 (1991) 493–504.

[18] G.A. Carpenter, S. Grossberg and D.B. Rosen, Neural Networks, 4 (1991) 759–771.
[19] G.A. Carpenter, S. Grossberg and J.H. Reynolds, Neural Networks, 4 (1991) 565–588.
[20] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, D.B. Rosen, Neural Networks, 3 (1992) 698–713.
[21] D. Wienke, in L. Buydens and W. Melssen (Editors), Chemometrics: Exploring and Exploiting Chemical Information, University Press, University of Nijmegen, 1994, p. 197.
[22] R. Hohenstein, Classification of Neuro-Magnetic Field Patterns Using a Fuzzy ARTMAP Network, Master Student Research Report, Wilhelms-University Münster, 1994.
[23] P. Willems, The ART Neural Network Models Enlighted: Implementation on Sequential and Parallel Computer Systems, Master Student Research Report, Catholic University of Nijmegen, 1994.
[24] K.W. Gan and K.T. Lua, Pattern Recogn., 25 (1992) 877–882.
[25] L.I. Burke, Neural Networks, 4 (1991) 485–491.
[26] T.P. Caudell, Appl. Opt., 31 (1992) 6220–6229.
[27] J.S. Kane and M.J. Paquin, IEEE Trans. Neural Networks, 4 (1993) 695–702.
[28] D.C. Wunsch, D.J. Morris, R.L. McGann and T.P. Caudell, Appl. Opt., 32 (1993) 1399–1407.
[29] D.C. Wunsch, T.P. Caudell, C.D. Capps, R.J. Marks and R.A. Falk, IEEE Trans. Neural Networks, 4 (1993) 673–684.
[30] C.S. Ho, J.J. Liou, M. Georgiopoulos, G.L. Heileman and C. Christodoulou, Int. J. Electron., 76 (1994) 271–291.
[31] C. Resch and Z. Szabo, J. Nucl. Med., 5 (1994) 182–182 (abstract).
[32] C.O. Benjamin, S.-C. Chi, T. Gaber and C.A. Riordan, Comput. Ind. Eng., 28 (1995) 43–50.
[33] C.C. Lin and H.P. Wang, Comput. Ind., 22 (1993) 143–152.
[34] J.R. Whitley and J.F. Davis, Comput. Chem. Eng., 18 (1994) 637–661.
[35] J.R. Whitley and J.F. Davis, IEEE Expert, 4 (1993) 54–63.
[36] D. Wienke and G. Kateman, Chemom. Intell. Lab. Syst., 23 (1994) 309–329.
[37] D. Wienke, Y. Xie and P.K. Hopke, Intell. Lab. Syst., 26 (1994) 367–387.
[38] Y. Xie, P.K. Hopke and D. Wienke, Environ. Sci. Technol., 28 (1994) 1921–1928.
[39] D. Wienke, W. van den Broek, W. Melssen, L. Buydens, R. Feldhoff, T. Huth-Fehre Kantimm, L. Quick, F. Winter and K. Cammann, Anal. Chim. Acta, (1994) submitted.
[40] D. Wienke and L. Buydens, Chemom. Intell. Lab. Syst., (1995) in press.
[41] D. Wienke, W. van den Broek, R. Feldhoff, T. Huth-Fehre, T. Kantimm, T. Huth-Fehre, L. Quick, W. Melssen, F. Winter, K. Cammann and L. Buydens, Chemom. Intell. Lab. Syst., (1995) submitted for publication.

Dr. Dietrich Wienke is at the Laboratory of Analytical Chemistry, Catholic University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, Netherlands. He gained his Ph.D. in Analytical Chemistry at the University of Jena (Germany) with a process-analytical multivariate-statistical study. His interests concern evolution methods such as the Kalman Filter, the Genetic Algorithm and ART neural networks for their potential in analytical chemistry and neural network methodology for processing of environmental data and for analytical imaging. He leads the laboratory for remote infrared imaging spectroscopy (NIRIS/SIRIUS) doing research for the rapid automatic sorting of post-consumer materials.
Professor Dr. Lutgarde Buydens, Head of the Analytical Chemistry Laboratory, studied pharmacy and informatics at the Free University of Brussels (Belgium), where she also obtained her Ph.D. in the field of Quantitative Structure Activity Relationships (QSAR). She studied the capabilities of expert systems for analytical applications and artificial intelligence. She is interested in research and applications of multivariate statistics, genetic algorithms, and artificial neural network technology to chemical-analytical problems and QSAR.