Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/28847

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Adaptive resonance theory based neural network for supervised chemical pattern recognition (FuzzyARTMAP)
Part 2: Classification of post-consumer plastics by remote NIR spectroscopy using an InGaAs diode array

D. Wienke [a,*], W. van den Broek [a], L. Buydens [a], T. Huth-Fehre [b], R. Feldhoff [b], T. Kantimm [b], K. Cammann [b]

[a] *Catholic University of Nijmegen, Laboratory for Analytical Chemistry, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands*
[b] *Institute for Chemical and Biochemical Sensor Research Münster, Mendel-Str. 11, 48149 Münster, Germany*

Abstract

The supervised working FuzzyARTMAP pattern recognition algorithm has been applied to automated identification of post-consumer plastics by near-infrared spectroscopy (NIRS). Experimentally, a remote operating parallel multisensor device, based on a rapid InGaAs diode array detector combined with new collimation optics, has been used. The laboratory setup allows on-line identification of more than 100 spectra per second. Internal parameter settings of FuzzyARTMAP were varied to explore their influence on the classifier's behavior. Discrimination results obtained were better than those from an optimized multilayer feedforward backpropagation artificial neural network (MLF-BP) and significantly better than those provided by the partial least squares method (PLS2). Additional advantages of FuzzyARTMAP compared to these two classifiers are a significantly higher speed of calibration, the chemical interpretability of network weight coefficients and a built-in detector against extrapolations.

Keywords: Artificial neural networks; Adaptive resonance theory (ART); Fuzzy set theory; Pattern recognition; Plastics recycling; Near-infrared spectroscopy; Multisensor array

1. Introduction

Adaptive resonance theory based artificial neural networks (ART) were recently found to be interesting for several chemical pattern recognition applications [1]. Whitley and Davis [2,3] proposed the use of ART-2 for chemical process monitoring and control by real-time classification of data taken from a running chemical reactor. Wienke and Kateman used ART-1 to classify UV/VIS and IR spectra and discussed the quantitative chemical interpretability of ART's weight coefficients [4]. Lin and Wang fitted industrial time series data and classified the parameter vectors by ART-2 [5]. Wienke et al. [6,7] applied ART-2a in on-line monitoring of the environment by

* Corresponding author. email: wienke@sci.kun.nl

a classification of airborne particles using scanning electron microscopy image data and X-ray fluorescence emission patterns. Another application of ART-2 to pattern recognition by image data was reported recently by Resch and Szabo [8]. Comparisons of ART-2a versus hierarchical cluster analysis, principal component analysis (PCA) and multidimensional scaling (MDS), the SIMCA classifier and backpropagation multilayer neural networks (MLF-BP) for rapid sorting of post-consumer plastics by remote NIR spectroscopy have recently been reported [9–11]. Numerous studies of the use of ART neural networks outside the field of chemistry were reviewed recently [1,14].

The reported general advantages of ART based classifiers compared to other types of multivariate and neural pattern recognition methods are short training times (usually less than 10 epochs), ability to model highly non-linear shaped clusters, robustness against outliers because of a built-in detector against extrapolations, the quantitative chemical interpretability of weight coefficients and the on-line applicability of ART in chemical apparatus.

Two supervised working ART algorithms (ARTMAP and FuzzyARTMAP) were recently introduced by Carpenter et al. [12,13]. The basic idea behind this has been to increase the discrimination power compared to former ART methods (ART-1, ART-2, ART-2a, ART-3, FuzzyART) that work unsupervised, in case of very closely located, non-linear shaped clusters in m-dimensional features space. ARTMAP is less interesting for chemical applications because it is restricted to the processing of binary coded input data. The present work, therefore, focusses on FuzzyARTMAP as the most advanced ART algorithm without any restriction on the type of input data. This work is presented in two parts because of the complexity of the FuzzyARTMAP algorithm, including aspects from adaptive resonance theory as well as from fuzzy set theory. By means of a moderately sized data set, the theory of FuzzyARTMAP and its properties were recently discussed in Part 1 [14]. Part 2 explores the on-line applicability of the FuzzyARTMAP network for a remote working near-infrared diode array detector for rapid identification of post-consumer plastic waste for automatic sorting. The present study reports new results from the running SIRIUS project (*Sensors and*

Artificial *I*ntelligence for *R*ecognition and *I*dentification of *U*sed Plastic*s*). Competitive studies in the field of automated plastic identification by NIR spectroscopy were recently published by Eisenreich et al. [16] and Ritzmann and Schudel [17]. In contrast to recent work of Alam and Stanton [15], it was found by the present authors [10] that MLF-BP neural networks, despite their excellent and widely accepted discrimination power between non-linear shaped data clouds, tend to produce uninterpretable nonsense responses if one tries to extrapolate them. Additionally to the normal classification error Type I (misclassification), this extrapolation deficiency of a MLF-BP network caused up to a 25% error of Type 0 (uninterpretable network answers) and of Type II (only seemingly correct answers for wrong samples (extrapolation case)) for NIR spectra of plastics [10]. Two further extrapolation experiments with ART networks were published recently (ART-2a [6], FuzzyARTMAP [14]). A missing built-in detector against outliers makes MLF-BP networks less robust with respect to undesired numerical extrapolations in an automated process-analytical environment. This is not acceptable and a serious disadvantage, in particular for automated control tasks. The theoretical reason for this deficiency is that MLF-BP networks do not perform their similarity analysis in the X-space of input data (NIR spectra) but only later in the Y-space (class membership). Compared to classical regression analysis, PCA or PLS, where high values of the X residuals simply help to signalize an extrapolation, the missing test strategy of a MLF-BP network is a step backward. In contrast to MLF-BP, the FuzzyARTMAP method performs its similarity check similarly to PLS in the spectral X-space, automatically providing a detection of outliers (and an extrapolation of the calibration subspace).

2. Theory

In Part 1 [14], the theory and algorithm of the FuzzyARTMAP artificial neural network were discussed. The present theoretical part therefore concentrates on data coding for the application to automated identification of post-consumer plastics by remote NIR spectroscopy.

A FuzzyARTMAP neural network consists, in

principle, of three partial neural networks I, II and III, that are running simultaneously (Fig. 1 of Ref. [14]). During a so-called supervised calibration phase ('neural network training'), network No. I performs a cluster analysis in the m-dimensional space of n **X** data. In the present application, these are n NIR training spectra (given by reflectance values at m wavelengths) that will be clustered into c groups. For each training spectrum **x** its corresponding membership vector, **y**, to a particular type of plastic is also given. In the case of five distinct types of plastics (classes), **y** will form a binary coded vector of length $p = 5$, whereby the membership to one particular class is coded by a '1', otherwise a 0. In general, FuzzyARTMAP can also process real coded data in the **Y**-space. The second network (II) performs a cluster analysis in this p-dimensional **Y**-space of given class memberships **y**. For the chosen example of binary coded class memberships, network No. II will thus find exactly $b = 5$ clusters in the **Y**-space of training data. Note that network No. I in the **X**-space will find c clusters among the n NIR spectra, whereby c can be equal to or can be different from the number, b, of clusters that will be formed in the **Y**-space. Note further that the number of clusters c and b are unknown before the network training. The third simultaneously running neural network (the 'mapfield', No. III), monitors and controls the other two networks (Nos. I and II). During the training process, this third neural network links the c clusters of NIR spectra, formed in the **X**-space, to the b clusters of class membership, formed in the **Y**-space.

In the case of adaptive resonance theory, a cluster is described by a directed vector **w** ('network weight') that comes close to the centroid vector for this particular cluster. The cluster size is given, in principle, via a priory fixed constant, called the vigilance parameter ρ (usually between 0 and 1). Thus, by the choice of ρ the future number of formed clusters ('cluster resolution' of the experimental data cloud) and their size can be controlled to perform an explorative data analysis. Thus, in total three vigilance parameters have to be defined before training (an additional one for each partial network).

Each of the three partial networks within the FuzzyARTMAP algorithm makes use of three essential equations and one unequation during its training. First, as outlined in Part 1, each of the three net-

works determines the winning vector of network weights that is closest to a current input (competitive learning) by a fuzzy distance metric. Secondly, each network checks if the present input lies inside the winning cluster (resonance case). If not, the subsequent second place winner is checked and so on. If any resonating weight vector is found, a learning rule is applied as the third step to move the weights vector a little towards the current input. In the case that no resonance occurred with any existing weights vector (extrapolation case) an additional fresh cluster is linked to the network. Thus, in contrast to other neural network types, ART uses either weights adaptation or structure adaptation to model a data cloud during the network training process. This data driven adaptation can have two obvious advantages. First, it makes the choice of the network structure less subjective to a particular chemical user. Second, the training speed of the competitive learning process can become significantly faster than, for example, for a comparable Kohonen neural network or a linear vector quantizizer (LVQ) or for a classical cluster analysis method. The theoretical reason for this speed is that in ART the n pattern vectors have only to be compared with c weight vectors ('clusters' or 'neurons'), whereby c is, in the beginning, equal to 1. Even for increasingly large numbers n of spectra the computational expense remains small and only linearly increasing. In the case of a Kohonen network (or a LVQ), for example, the number c of winner competing weight vectors ('neurons') is fixed. This requires from the very beginning many more computational circles. For the other rather classical case of statistical clustering methods, half distance matrices of decreasing size, beginning with dimension $(n^2 - n)/2$, have to be calculated. As is very well known, the calculational expense increases more than quadratically (thus rapidly) for slowly (linear) increasing numbers n of training spectra. However, for highly scattered data sets, ART networks in general tend to an undesired proliferation of new clusters. If this happens, the high training speed from the beginning can slow down dramatically because c increases rapidly.

After completion of the network training, the network found a particular number c of clusters in **X**-space (or b in **Y**-space) with maximum contrast (dissimilarity) among them. The formed cluster struc-

tures in both variable spaces are the first result one can obtain from a FuzzyARTMAP network. In part 1 [14] and in Refs. [4,6,7,9] it has been shown how a variation of the vigilance parameter and learning rate can be used for an active elucidation of fine and raw cluster structures in the training data. As a second result, in the same references it has been illustrated that trained ART neural networks can be interpreted in chemical and spectroscopic terms after a suitable decoding of the scaled network weights. Third, the trained FuzzyARTMAP can be used for classification of unknown samples. In the present application, a NIR spectrum of an unknown type of plastic will be compared with the trained network in the X-space. The output of this network (No. I) forms the input to network No. III (mapfield) which will make a link to the partial network No. II in the Y-space. This latter network finally provides the membership to a particular type of trained plastic classes. However, if a NIR spectrum x that is offered to network No. I does not fall into any of the c clusters of the X-space, no network resonance will occur. The unknown is assigned as a novelty and this will be signalized as an extrapolation case. In this way the network self-protects against extrapolations and against a 'wash-out' of already learned information.

3. Experimental

FuzzyARTMAP has been implemented in an experimental setup for rapid identification of post-consumer plastics (Fig. 1). The hardware setup has recently been described in detail by Huth-Fehre et al. [18]. A new collimation optics, developed by Huth-Fehre et al. [19], enables acquisition of NIR transflectance spectra from remote plastic objects at a conveyor. The construction of the optics allows a variation of location, shape and distance of the object to the optics within a range up to 50 cm, always providing measurable NIR spectra. The optics integrates over a fixed part of the conveyor. The spectrum of the empty conveyor serves as the reference spectrum. The transflected radiation is spectroscopically resolved using a new generation NIR diode array spectrometer with a 256 pixel InGaAs diode array detector (PolyTec XDAP, IKS Optoelectronics, Duisburg, Germany [10,18]). The spectrometer provides up to 158 complete NIR spectra per second in the optical range 800–1700 nm (shortwave range near-infrared). The low absorption coefficients in this spectral region enable a sample penetration depth of several millimeters. In this way, the deeper located sample bulk is also analysed instead of the sample surface
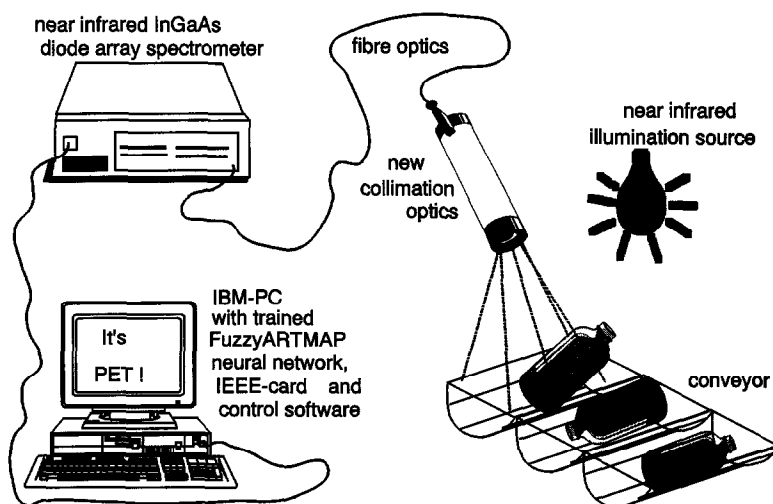

Fig. 1. Scheme of the experimental setup for rapid automated on-line identification of post-consumer plastic waste by remote near-infrared sensing combined with real-time pattern recognition by a FuzzyARTMAP artificial neural network.
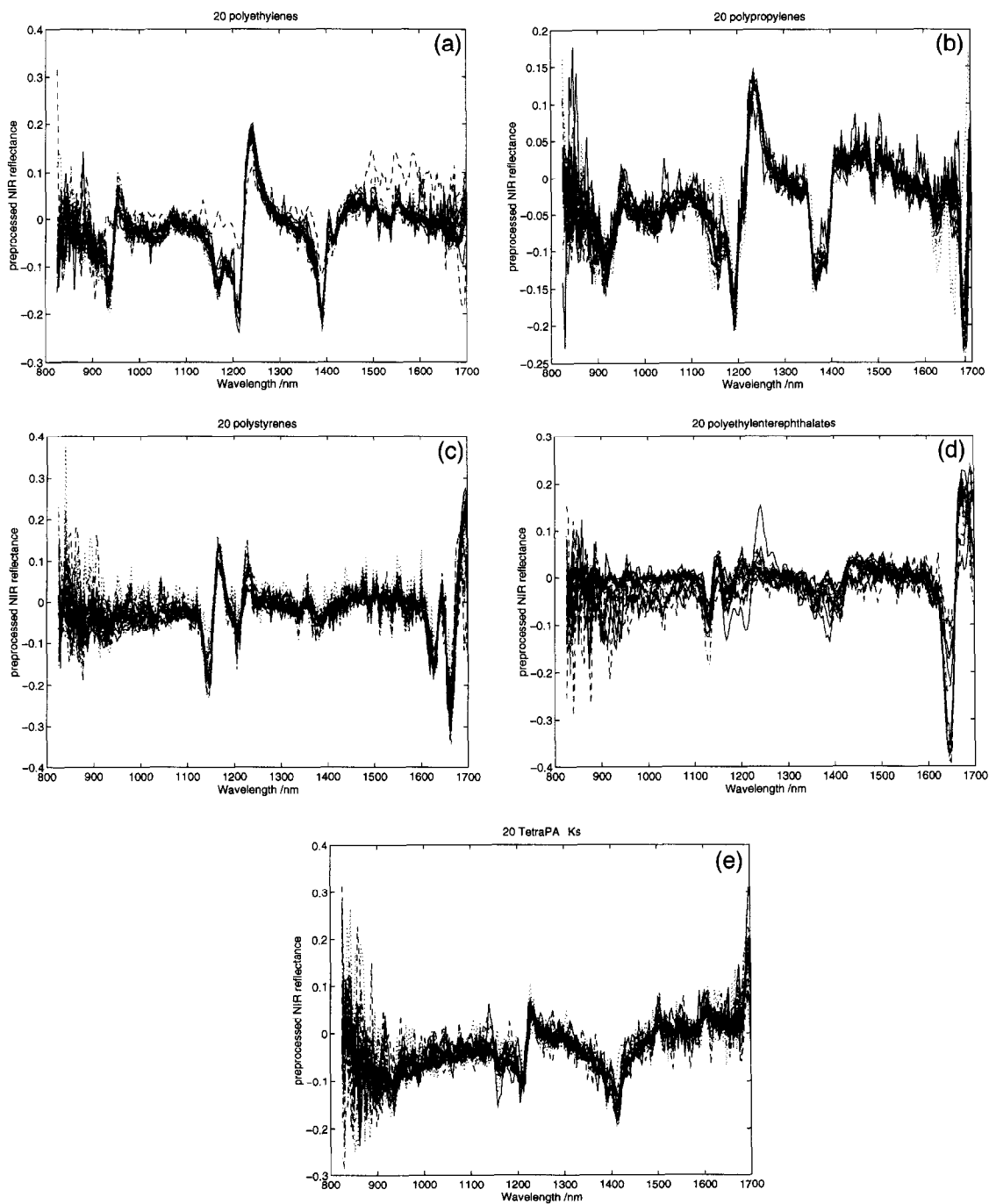
Fig. 2. Selected near-infrared transflectance spectra of distinct grades of post-consumer packaging materials. The spectra were recorded by the setup shown in Fig. 1 and shown (after pre-processing, see text) as they are offered to the FuzzyARTMAP neural network.

only. This causes significantly less disturbance from a polluted sample surface (stickers, dirt, humidity). On the other hand, much less spectroscopic information is coded in this part of the spectrum, requiring especially powerful pattern recognition methods. The raw NIR spectrum (vector **x** of $m = 256$ reflectance numbers, each of 14 bit intensity resolution) is then transferred on-line via an IEEE card (and corresponding control software) into the RAM of a personal computer (Intel 80486 processor). Recent examples of raw spectra of post-consumer plastics, recorded by this diode array spectrometer, can be found in Refs. [10,18]. The raw spectrum is then preprocessed by a moving average filter followed by calculation of the first derivative and scaling to spectrum unit length. This combination of preprocessing methods provides compact, well separable clusters in the spectral space [10,18]. Examples of preprocessed spectra, now called feature vectors of length m, **x**, are shown in Fig. 2 for distinct grades, colors, sizes and sample placement of pieces of post-consumer plastics. The used samples belong to a set of dirty bottles and packing collected via the public waste collection system of the german town of Cologne. The feature vector **x** is further preprocessed by FuzzyARTMAP using complement coding [12–14] and scaling of all m features to the range $\langle 0...1 \rangle$. They were subsequently classified. An identification result is immediately presented in real time on the computer screen as a large text signal (Fig. 1). In the present study, five post-consumer packaging materials were classified: PET (polyethyleneterephthalate), PE (polyethylene), PP (polypropylene), PS (polystyrene) and TetraPAK™ (a packaging material consisting of paper and PE layers). Fig. 2 shows that the spectral similarity between PE, PP and TetraPAK™ is greater than that between PET and PS. The available training and test spectra used in the present study (Table 1) originate from samples of different size, color and distinct sample placement on the conveyor.

4. Software and computations

The program 'FuzzyArtMap.c' was written by the author in ANSII-C computer language for UNIX computers (SUN) for the gcc-compiler. The alternative DOS version, now implemented on the master PC

Table 1
Number of NIR spectra of distinct samples from five different packaging materials used for training and testing the FuzzyARTMAP artificial neural network

Packaging material	Samples in test set	Samples in training set
PET	106	107
PE	113	120
PP	127	108
PS	101	106
TetraPAK™	104	104
Total	545	551

of the fine-sorter (Fig. 1) for plastic waste, has been translated into Borland-C. The control software for the IEEE data acquisition card has also been written in Borland-C, forming together with FuzzyARTMAP and the data preprocessing algorithms one large software package for on-line measurements and real-time pattern recognition. A Microsoft Windows based version, written in Visual Basic and DLL codes, has recently been completed and will be described separately [21]. The training of the neural network is done externally at a SUN Spark 10 Unix workstation. With the given experimental data (Table 1) the FuzzyARTMAP algorithm has been studied with respect to the influence of the size of the vigilance parameters $\rho^{x,max}$ and to the size of the learning rate η^x. The number of clusters formed and the error of classification were used as criteria to find optimal initial network parameter settings. The separation power of FuzzyARTMAP was finally compared with the separation power of the partial least squares algorithm (PLS) and an optimized multilayer feedforward backpropagation artificial neural network (MLF-BP). The MLF-BP network (from MATLAB neural network toolbox) made use of the Ngyuen–Widrow initialization, adaptive size of learning rate, momentum term, a sigmoidal transfer function (hidden layer) and a log-linear transfer function (outputs 0–1). The hidden layer included six hidden neurons. The following section summarizes and discusses the obtained results.

5. Results and discussion

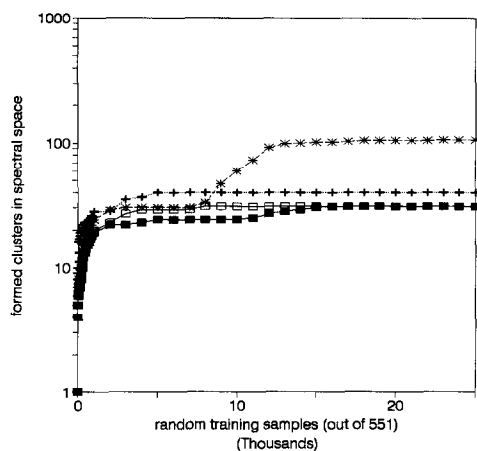In former studies of ART-1 and ART-2a neural networks [4,6,7,9,10] it has been found that, for in-

Fig. 3. Presentation of the growing resolution of a data set into finer subclusters by an increase of the vigilance parameter (here $\rho^{x,max}$ in the X-space) as a function of the number of training epochs for four distinct sizes of the vigilance parameter ($\rho^{x,max} = 0.4$ (■), 0.5 (□), 0.6 (+), 0.7 (∗). In general, fast convergence of the training process can be observed for a FuzzyARTMAP network within a few epochs. The higher the desired cluster resolution is, the more time the convergence takes (see also Fig. 5).
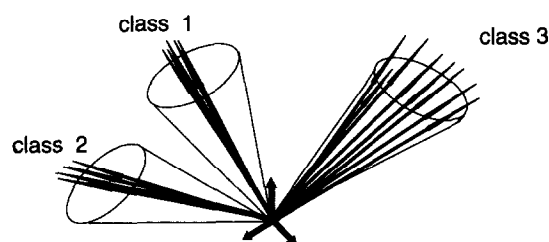

Fig. 4. The more compact the data are (left, classes 1 and 2), the less they scatter. In such a case a change of the cluster box size via the vigilance parameter $\rho^{x,max}$ has almost no effect. The more the data scatter (class 3, right) or if they form a continuous function in place of a cluster, the more sensitive the cluster resolution is to small changes in $\rho^{x,max}$. Note that the concept of vigilance ρ follows a distance concept and not a similarity concept.

creasing vigilance parameter $\rho^{x,max}$, the size of the formed clusters decreases and simultaneously their number increases. (Note the inverse relation, because ρ corresponds to a measure of distance and not to a similarity measure.) A similar trend has been observed for FuzzyARTMAP (Fig. 3), despite the fact that $\rho^{x,max}$ is now calculated by the fuzzy set theory based intersection operator that differs from the Euclidian-like distance metrics in ART-1 and ART-2a. For $\rho^{x,max} = 0.7$ and larger, a proliferation of new clusters is observed. For $\rho^{x,max} = 0.4$ and smaller, the network quickly stabilizes at a fixed number of clusters within a few training epochs. A further decrease of $\rho^{x,max}$ to values smaller than 0.4 did not decrease the number of clusters formed significantly. This can easily be understood by the scheme in Fig. 4.

The remarkable speed of convergence of FuzzyARTMAP's training process has already been mentioned in Part 1. Figs. 3 and 5 illustrate that only 3–5 training epochs (thus between 3×545 and 5×545 randomly selected training samples) are required for these NIR data sets. The MLF-BP network that is described in the last part of this section needed 100 or more epochs for the same data to converge. Looking additionally at the necessary internal algorithmic cal-

culations (see Part 1), FuzzyARTMAP requires fewer computing steps than the MLF-BP network per training spectrum, if the number of formed classes in the X-space stays at the moderate magnitude of the desired number of output classes. In general, a higher training speed allows an evaluation of more distinct network parameter combinations within the same short time. This allows the user a comfortable and rapid optimization of the settings of the classifier and
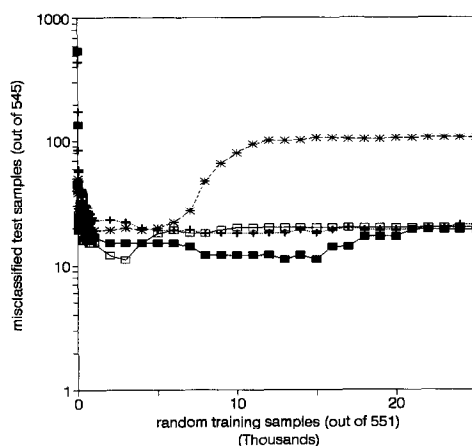

Fig. 5. Convergence of FuzzyARTMAP's training process shown by a presentation of the prediction error (test data set, Table 1) for four distinct sizes of the vigilance parameter ($\rho^{x,max} = 0.4$ (■), 0.5 (□), 0.6 (+), 0.7 (∗)) as a function of the number of training epochs. In general, the reclassification error rapidly decreases within the first few epochs. With increased vigilance parameter $\rho^{x,max}$ the number of formed clusters rapidly grows (see also Fig. 3) providing a growth in the classification error for the test set. FuzzyARTMAP network overtraining becomes visible.

it requires a less powerful computer. Additionally to training speed, an additional feature has been found in Fig. 3. Clusters concerning many training patterns are always formed within the first few training epochs. This is caused by a higher statistical chance of the contributing training patterns becoming randomly selected. Thus, the raw cluster structure is always formed in the very beginning of the training, small separate clusters and single pattern clusters are split up later on. Therefore, the influence of $\rho^{x,max}$ on the number of clusters within the first epochs is rather small (Fig. 3). This causes the classification results to be rather independent in the beginning on a particular choice of $\rho^{x,max}$ (Fig. 5).

Another important observation can be made concerning Figs. 3 and 5. A FuzzyARTMAP network can be overtrained if $\rho^{x,max}$ is chosen so large that uncontrolled cluster proliferation occurs (here $\rho^{x,max} = 0.7$ and larger). After stabilization at a relatively low level, the classification error can increase again (Fig. 5) if one tries to form too many single clusters that do not reflect the general cluster structure of the data. Obviously, the separation power decreases again in such a case because too many single clusters are separated in the area between two larger main clusters. In such a case, very similar clusters with very similar weight vectors belonging to distinct Y-classes are produced. However, such overtraining happens later after a relatively large number of epochs when the general cluster structure starts to split into many small subclusters.

For the Y data and for the mapfield, the choice of $\rho^{y,max}$ does not play that important a role for the present application, because these data are binary coded, giving point-like ideal clusters without any variance. In general it is true that the more point-like (or compact) a training data set is, the less important are the choice and influence of the size of $\rho^{x,max}$ for the number of clusters found by FuzzyARTMAP (Fig. 4). On the other hand, a data set having the highest scattering will be most sensitive to a small increase of $\rho^{x,max}$. Data sets with a high variance will thus tend to high cluster proliferation. From Fig. 4 it becomes clear that FuzzyARTMAP is more suitable for pattern recognition applications (qualification) than for function fitting (quantification). Part 1 has already outlined this.
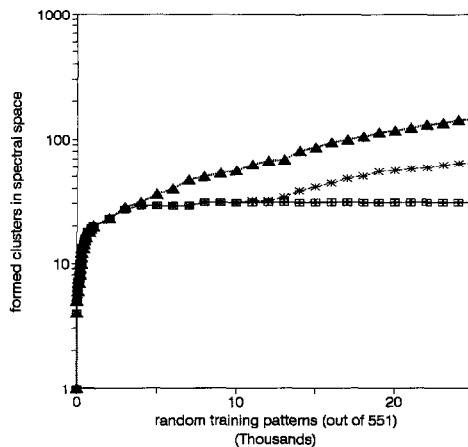
With respect to the influence of the chosen learn-

Fig. 6. With increasing learning rate η^x (($\square$) $\eta^x = 0.001$, ($*$) $\eta^x = 0.005$, ($\blacktriangle$) $\eta^x = 0.05$) a FuzzyARTMAP neural network tends to uncontrolled proliferation of new clusters caused by rapid oscillations of existing clusters in the variable space (see also Fig. 7).

ing rate η on the number of clusters it has recently been found for ART-1 [4] that a large η value (> 0.5) causes an undesirable proliferation of clusters. The same was also observed, in principle, for FuzzyARTMAP (Figs. 6 and 7). The reason for this is that a training pattern that is situated inside a cluster, but close to its border, falls quickly outside the cluster if the actual weight vector of this cluster moves with a too large step away from its actual position (fast
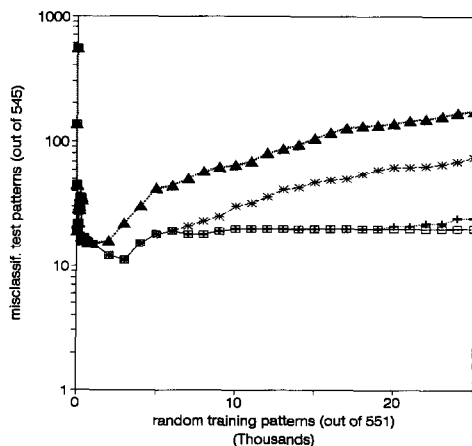

Fig. 7. Rapid oscillations of existing clusters caused by a large learning rate η^x increase the prediction error of a FuzzyARTMAP neural network (($\square$) $\eta^x = 0.001$, ($*$) $\eta^x = 0.005$, ($\blacktriangle$) $\eta^x = 0.05$). A moderate learning rate of $\eta^x = 0.005$–0.01 seems a practical suitable choice.

Table 2
Mutual classification of NIR spectra of 545 test samples, belonging to five packaging materials, by a FuzzyARTMAP neural network after almost one training epoch (500 randomly selected training samples out of 551). 525 out of 545 test spectra were correctly classified (96.3%), whereby the network generated 15 clusters in X-space

Packaging material	Outside all classes	PET	PE	PP	PS	TetraPAK™
PET	0	94	10	1	0	1
PE	0	0	101	0	0	3
PP	0	0	2	118	0	0
PS	0	1	1	0	106	0
TetraPAK™	0	0	1	0	0	106

learning). For simplicity, we consider the two extremes $\eta = 0.0$ (no learning) and $\eta = 1.0$ (fast learning). In the first case, the first randomly selected training pattern will completely dominate from the very beginning the entire future cluster structure without any modification during training. The initial cluster structure will propagate throughout the entire training phase without change. The short-term memory of the network will be identical to the future long-term memory. In the second case, with $\eta = 1.0$, the network will have no long-term memory. Any existing cluster will be continuously overwritten and corrected by any newly presented training pattern. The cluster will hardly oscillate in several spatial directions causing in this way an uncontrolled proliferation of new clusters. By comparing Table 2 (no overtraining) with Table 3 (overtraining) it can be seen that most of the misclassified samples are placed in new classes. This proves the hypothesis of rapid oscillations of class boxes in the case of the chosen size of η being too large.

The final result of this study concerns a comparison of an optimized FuzzyARTMAP neural network with PLS and with an optimized MLF-BP network. All three methods were calibrated ('trained') with the training data sets of NIR spectra given in Table 1. In a subsequent step, the three classifiers were validated with the test data sets (Table 1). For the PLS runs, a classical PLS2 model for the combined NIR training data sets (Table 1) for PE, PP, PS, PET and Tetra-PAK™ was fitted against the binary coded class membership Y. The network settings for the opti-

mized MLF-BP neural network have already been given in Section 4.

Using all $m = 239$ features in the pre-processed NIR spectra, the PLS model wrongly classified 33 out of 545 test spectra. This corresponds to 94% correct predictions. In contrast, the best MLF-BP model only had a prediction error of 11 samples (97.9% correctly classified test spectra). The best FuzzyARTMAP model ($\rho^{x,max} = 0.4$, $\eta^x = 0.005$) provided 12–18 wrongly classified test spectra (96.7–97.8% correctly classified test spectra) (Figs. 6 and 8). For $m = 239$ features, FuzzyARTMAP always formed between 20 and 40 clusters from the 551 training spectra in the X-space, depending on the random order of NIR spectra during training (Figs. 3 and 6). This first result indicates that the PLS classifier is less powerful for this particular application compared to both neural networks. It also shows that FuzzyARTMAP is no better than the MLF-BP network with respect to its discrimination power.

However, in a former study the present authors found [4] that ART neural networks are very sensitive to wrong and noisy features. It was shown that a suitable feature selection can significantly improve the classification results of ART-1. In the present application of FuzzyARTMAP to NIR spectra, at least two data pre-processing techniques were required that commonly tend to increase the relative contribution of noise to the data (first derivative per spectrum, range scaling per wavelength). Therefore, it was decided to

Table 3
Overtraining of a FuzzyARTMAP neural network shown by a detailed list of mutual classification and misclassification of NIR spectra of 545 test samples, belonging to five packaging materials, after 37 training epochs (about 20000 randomly selected training samples out of 551). Only 487 of 545 test spectra were correctly classified (89.3%), whereby the relatively high number of 104 X-clusters were generated by the network. Compared to Table 2 this is a significant increase of misclassification

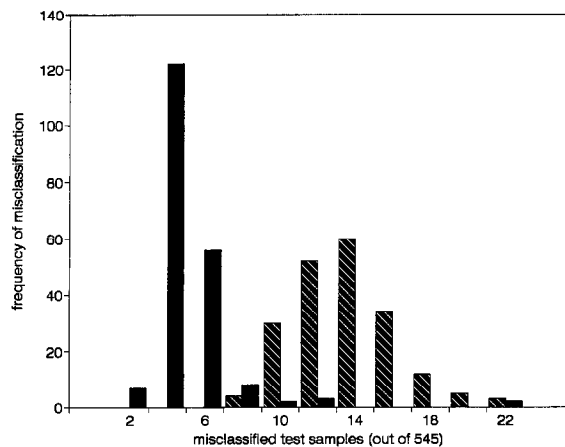| Packaging material	Outside all classes	PET	PE	PP	PS	TetraPAK™
PET	10	95	0	1	0	0
PE	40	0	64	0	0	0
PP	2	0	0	116	2	0
PS	2	0	0	0	106	0
TetraPAK™	1	0	0	0	0	106

Fig. 8. Frequency of particular observed classification errors for the same test set (545 NIR spectra of plastics) for 200 runs with varied random generator seed, for an optimized MLF-BP neural network (white-shadowed bars) and an optimized FuzzyARTMAP network (black bars) (for the parameters of MLF-BP, see Section 4; parameters for optimized FuzzyARTMAP were $\rho^{x,max} = 0.5$, $\rho^{y,max} = 0.95$, $\rho^{xy,max} = 0.95$, $\eta^{xy} = 1.0$, $\eta^{y} = 1.0$, $\eta^{x} = 0.005$, number of training epochs = 10, 491 training samples, 551 test samples).

repeat the runs with PLS2, MLF-BP and FuzzyARTMAP with an optimal subset of $m = 50$ selected features out of the 239 features. The subset selection was reached by a genetic algorithm or Fisher weighing, described in another study concerning optimal wavelength selection for NIR spectra [20]. For PLS, again only 94% correct classifications were found. The optimum number of latent PLS variables was chosen by cross validation. The MLF-BP runs and the FuzzyARTMAP runs were repeated 200 times with distinct random generator seeds. For MLF-BP this always causes slightly distinct initializations of the network weight matrices, giving finally a slight variation in the classifications results. For FuzzyARTMAP, a different random generator seed only influences the random order of selection of the training pattern (but not the initialization of weights). The final result for MLF-BP and FuzzyARTMAP (Fig. 8) shows that an optimized FuzzyARTMAP network with an optimally selected subset of $m = 50$ features reduces the error further to only three misclassified test samples (out of 545), which corresponds to 99.5% correct classifications. On the one hand, the

distribution for FuzzyARTMAP's classification error shows a large tail, giving a rather log-normal shape (see the single run with 22 wrong classifications). On the other hand, for MLF-BP in no case could a calibration model be obtained for these NIR spectra with the excellent low prediction errors of FuzzyARTMAP. Note further that, despite the slight tailing, the total statistical chance for FuzzyARTMAP to produce a calibration model with lower prediction errors is obviously much higher for this particular application than for MLF-BP. FuzzyARTMAP provided much more frequent stable and low classification errors closer to the origin ('zero classification error') than MLF-BP. More than 120 runs (out of 200) of FuzzyARTMAP provided a prediction error smaller than five samples (out of 545); and more than 180 (out of 200) runs still provided an error smaller than 10. This is a significantly high probability to reproduce a very predictive calibration model. In contrast, the low chance for MLF-BP in this particular case was only around 30 runs out of 200 to obtain an error smaller than 10 samples. The chance of reproducing a MLF-BP model with low classification error is, for this application, statistically significantly lower.

The error improvement obtainable by feature reduction is thus for FuzzyARTMAP much greater than for the MLF-BP network. Feature reduction also provided smaller models with respect to the size of the calibration (or weights) matrices for PLS, MLF-BP and FuzzyARTMAP. In the case of FuzzyARTMAP, the number of clusters that were formed in the X-space for the training data now decreased from 20–40 down to 10–15. This finally provided a more than five times higher classification speed (m reduced from 256 down to 50 and c reduced from 40–20 down to 10–15), which becomes important in practice for rapid automated sorting of post-consumer plastics.

6. Conclusions

FuzzyARTMAP is a new supervised classifier based on adaptive resonance theory that can be used to solve non-linear chemical pattern recognition tasks.

FuzzyARTMAP's training phase converges remarkably rapidly. The few calculation steps in the

subsequent classification phase also allow on-line application of FuzzyARTMAP in an process-analytical environment. This has been demonstrated by a practical implementation in an identification device for automatic recognition of post-consumer plastics by remote near-infrared spectroscopy. The discrimination power of FuzzyARTMAP for this particular application was highly comparable to an optimized multilayer feedforward neural network and significantly better than that of the partial least squares method. Together with its rapid convergence in training and with the good discrimination power, three additional useful properties are provided by the FuzzyARTMAP algorithm. First, the weights of the trained network can be interpreted in terms of original variables (thus no black box) after suitable decoding and rescaling. In an example (see Part 1) it has been demonstrated that by means of such an interpretation important feature combinations can be extracted that contribute to a particular cluster. This helps to understand why certain clusters were formed automatically by FuzzyARTMAP. Second, FuzzyARTMAP showed in this particular application significantly less dependence on initialization settings than, for example, MLF-BP networks. This has been demonstrated by more reproducible working trained networks with more stable and lower classification errors. Third, the double phase calculation concept of competitive learning and the possibility for network structure extension works as a built-in detector for novel patterns. This helps to prevent undesired extrapolations of the calibration subspace and it helps to avoid any senseless classification of outliers (for an example, see Part 1).

However, there are also two serious disadvantages with the FuzzyARTMAP algorithm. For highly scattered data clusters or for function fitting FuzzyARTMAP is less suitable because it is a typical local modelling technique. It would form in such cases many small subclusters to approximate the experimental data cloud (cluster proliferation). Additionally, the strong dependence of FuzzyARTMAP results on suitable data pre-processing is a serious deficiency that requires significant additional research. The more compact the data clusters can be made beforehand by choosing a suitable data pre-processing scheme, the less the user's future choice of training parameter settings (initialisation, sizes of learning

rates and vigilance parameters) will be subjective to the result.

References

[1] D. Wienke and L. Buydens, Tr. Anal. Chem., 99(9) (1995) 1–18.

[2] J.R. Whitley and J.F. Davis, Comput. Chem. Eng., 18(7) (1994) 637–661.

[3] J.R. Whitley and J.F. Davis, Pattern, IEEE Expert, 4 (1993) 54–63.

[4] D. Wienke and G. Kateman, Chemom. Intell. Lab. Syst., 23(2) (1994) 309–329.

[5] C.C. Lin and H.P. Wang, Comput. Ind., 22(2) (1993) 143–152.

[6] D. Wienke, Y. Xie and P.K. Hopke, Chemom. Intell. Lab. Syst., 25 (1994) 367–387.

[7] Y. Xie, P.K. Hopke and D. Wienke, Environ. Sci. Technol., 28 (1994) 1921–1928.

[8] C. Resch and Z. Szabo, J. Nucl. Med., 5(35) (1994) 182 (abstract).

[9] D. Wienke, in L. Buydens and W. Melssen (Eds.), Chemometrics: Exploring and Exploiting Chemical Information, University Press, University of Nijmegen (NL), 1994, p. 197.

[10] 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, Anal. Chim. Acta, 1995, in press.

[11] D. Wienke, L. Buydens and K. Cammann, Presented at the International Chemometrics Meeting, Veldhoven (NL), 3–5 July 1994.

[12] G.A. Carpenter, S. Grossberg and J.H. Reynolds, Neural Networks 4 (1991) 565–588.

[13] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds and D.B. Rosen, IEEE Trans. Neural Networks, 3(5) (1992) 698–713.

[14] D. Wienke and L. Buydens, Chemom. Intell. Lab. Syst., 32 (1996).

[15] M.K. Alam and S.L. Stanton, Process Control Qual., 4 (1993) 245–252.

[16] N. Eisenreich, H. Knull and E. Theines, 23rd Annual Conference of the Institute of Chemistry, Tech Karlsruhe (Germany), 1992, pp. 59/1–59/12.

[17] H.P. Ritzmann and D. Schudel, Kunststoff 84 (1994) 582–584.

[18] T. Huth-Fehre, R. Feldhoff, Th. Kantimm, L. Quick, F. Winter, K. Cammann, W. van den Broek, D. Wienke, W. Melssen and L. Buydens, J. Molec. Struct., 348 (1995) 143–146.

[19] T. Huth-Fehre, T. Kantimm, R. Feldhoff and L. Quick, German patent application P 4423 770.7 (1994).

[20] W. van den Broek, D. Wienke, W. Melssen and L. Buydens, Chemom. Intell. Lab. Syst., in preparation.

[21] D. Wienke, J.-W. Schoenmakers, W. van den Broek, R. Feldhoff and L. Buydens, Chemom. Intell. Lab. Syst., in preparation.