HAL
open science

Evolving Fuzzy Classifiers: Application to Incremental Learning of Handwritten Gesture Recognition Systems

Abdullah Almaksour, Eric Anquetil, Solen Quiniou, Mohamed Cheriet

To cite this version:

Abdullah Almaksour, Eric Anquetil, Solen Quiniou, Mohamed Cheriet. Evolving Fuzzy Classifiers: Application to Incremental Learning of Handwritten Gesture Recognition Systems. International Conference on Pattern Recognition (ICPR), Aug 2010, Istanbul, Turkey. pp.4056-4059. hal-00582438

HAL Id: hal-00582438

https://hal.science/hal-00582438v1

Submitted on 1 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evolving Fuzzy Classifiers: Application to Incremental Learning of Handwritten Gesture Recognition Systems

Abdullah Almaksour Eric Anquetil
INSA de Rennes
Avenue des Buttes de Coesmes, F-35043 Rennes, France
{abdullah.almaksour,eric.anquetil}@irisa.fr

Solen Quiniou Mohamed Cheriet
Synchromedia-ETS
1100 rue Notre-Dame O., Montreal H3C1K3, Canada
solen.quiniou@synchromedia.ca, mohamed.cheriet@etsmtl.ca

***Abstract*—In this paper, we present a new method to design customizable self-evolving fuzzy rule-based classifiers. The presented approach combines an incremental clustering algorithm with a fuzzy adaptation method in order to learn and maintain the model. We use this method to build an evolving handwritten gesture recognition system. The self-adaptive nature of this system allows it to start its learning process with few learning data, to continuously adapt and evolve according to any new data, and to remain robust when introducing a new unseen class at any moment in the life-long learning process.**

***Keywords*-incremental learning; evolving; handwriting recognition; fuzzy classifier;**

I. Introduction and Background

Classification techniques appear frequently in many application areas, and become the basic tool for almost any pattern recognition task. The main problem in classification is to induce a classifier from a set of data samples. A large amount of samples is needed to set up and evaluate a classification system that can achieve a high accuracy, and it is practically very difficult to have such number of samples covering the expanse of the variability of classes. Therefore, life-long classifier adaptation becomes more and more an essential point. Moreover, in many application contexts, the classifier needs to take into account new unseen classes and to integrate them in the classification process, which increases the need for "evolving" classifiers. One good example of such applications is the use of online handwritten gesture classifiers which aim at facilitating interactions with computers using pen-based interfaces like whiteboards, tablet PCs, PDA...Etc. The main drawback in the current existing systems is that they are trained "offline" on a specific group of gestures and then implemented to operate without changing their structure during the use. This fixed structure does not allow the user to choose his own set of gestures or to add new ones according to his special needs, for example.

In our work, we aim at building an handwriting classifier, on-the-fly, from scratch and using only few data. Thus, the classifier will be incrementally adapted to achieve high recognition rates as soon as possible and to keep the system robust when introducing new unseen classes at any moment in the life-long learning process. An incremental learning algorithm is defined in [1] by the following criteria: it should be able to learn additional information from new data; it should not require access to the original data (*i.e.* data used to train the existing classifier); it should preserve previously acquired knowledge (it should not suffer from *catastrophic forgetting*, *i.e.* significant loss of original learned knowledge); and it should be able to accommodate new classes that may be introduced with new data. Many of the existing "incremental learning" algorithms are not truly incremental because at least one of the mentioned criteria is violated. These criteria can be briefly expressed by the so-called "plasticity-stability dilemma"[2]. It says that a system must be able to learn to adapt to a changing environment but that constant change can lead to an unstable system that can learn new information only by forgetting everything it has learned so far. We can distinguish two main types of incremental learning algorithms: algorithms for parameter learning and algorithms for structure learning. The incremental learning of parameters can be considered as an "adaptation" algorithm. The structure in such systems is fixed and initialized at the beginning of the learning process, and the system parameters are learned incrementally according to newly available data. Some examples of these systems are presented in [3], [4]. Most of the structure incremental learning algorithms are based on the principle of the ART clustering algorithm [5], such as [6], [7]. The main problem of these systems is that they are sensitive to the selection of the vigilance parameter, to the noise level in the training data and to the order in which the training data is presented. A promised incremental clustering approach had been presented in [8] based on the Mountain Clustering algorithm (originally introduced in [9]). The main idea of the proposed approach is that of a *potential* of a given point: it corresponds to a value representing the density in the data space at that point. The potential of a sample can be defined as the inverse of the sum of the distances between that data sample and all the other ones. Samples with high potential are then considered to be candidates to form a cluster.

In this paper, we extend the recursive mountain clustering by combining it with a robust fuzzy adaptation method,

and we use this hybrid algorithm to incrementally learn an evolving fuzzy rule-based classifier. Our system is used for the recognition of online handwritten gestures, and must be able to learn new classes of gestures and to evolve, sample after sample, without using all the old data.

A brief description of the system architecture is presented in Section II. Then, we explain the two main elements of the learning algorithm in Section III. Section IV studies the experimental evaluation results.

II. System architecture

Our system is based on a fuzzy rule-based classifier. The fuzzy rules make a link between intrinsic models (premises) and system outputs by consequent functions. For a K classes problem, a rule R_i is built for each fuzzy model P_i:

$$R_i : IF\ \vec{x}\ is\ P_i\ THEN\ y_i^1\ AND \ldots AND\ y_i^k \tag{1}$$

where $\vec{x}$ is the n-dimensional feature vector, P_i is a fuzzy model (prototype) defined by a center $\vec{\mu}_i$ and a covariance matrix Q_i. The degree of membership of $\vec{x}$ to P_i is given by the Mahalanobis distance:

$$\beta_i(\vec{x}) \quad = \quad 1/(1 + d_{Q_i}(\vec{x}, \vec{\mu}_i)) \tag{2}$$

For the consequent part, we can distinguish different structures [10]: (i) the zero order Takagi-Sugeno (TS) with binary consequents ($y_i^m = 1$ if P_i belongs to class m, 0 otherwise), (ii) the zero order TS with constant consequents ($y_i^m \in [0, 1]$ represents the participation of P_i in the description of class m), (iii) the first order TS, where the consequents are linear functions $y_i^m = \vec{a}_i^m \vec{x}$.

Finally, the sum-product inference is used to compute the system output for each class:

$$y^m(\vec{x}) = \sum_{i=1}^{R} \beta_i(\vec{x})\ y_i^m \tag{3}$$

III. Incremental learning algorithms

In order to incrementally learn a fuzzy rule-based classifier in an on-line manner, we need, on the one hand, to evolve its structure by adding (or deleting) rules, and, on the other hand, to adjust its parameters (prototypes' centers, covariance matrices and the consequent parameters). The emphasis in this paper is on the learning of the premise part of the system. We aim at extending the recursive mountain clustering by combining it with a robust adaptation method that can constantly re-center the fuzzy prototypes and re-shape their influence zones, according to each single data sample.

A. Recursive Mountain Clustering

As mentioned earlier in Section I, a recursive (on-line, one-pass, non-iterative) version of the mountain clustering method was introduced in [8]. The recursive formula avoids memorizing the whole previous data but keeps - using few variables - the density distribution in the feature space, based on the previous data:

$$P_k(x(k)) = \frac{k-1}{(k-1)\alpha(k) + \gamma(k) - 2\zeta(k) + k - 1} \tag{4}$$

where $P_k(x(k))$ denotes the potential of the k'th data sample and

$$\alpha(k) = \sum_{j=1}^{n} x_j^2(k) \tag{5}$$

$$\gamma(k) = \gamma(k-1) + \alpha(k-1), \quad \gamma(1) = 0 \tag{6}$$

$$\begin{aligned} \zeta(k) &= \sum_{j=1}^{n} x_j(k)\eta_j(k), \\ \eta_j(k) &= \eta_j(k-1) + x_j(k-1), \eta_j(1) = 0 \end{aligned} \tag{7}$$

Introducing a new sample affects the potential values of the centers of the existing clusters, which can be recursively updated by:

$$P_k(\mu_i) = \frac{(k-1)P_{k-1}(\mu_i)}{k - 2 + P_{k-1}(\mu_i) + P_{k-1}(\mu_i)\sum_{j=1}^{n} \|\mu_i - x\|_j^2} \tag{8}$$

If the potential of the new sample is higher than the potential of the existing centers, then this sample will be a center of a new cluster (and a new fuzzy rule will be formed in the case of fuzzy rule-based classifier). If the high potential sample is close to an existing center $\vec{\mu}_i$, then this sample will replace $\vec{\mu}_i$ and no new cluster will be created.

B. Fuzzy Vector Quantization

As can be noted in section III-A, the condition to have a high potential is a very hard one, and it is inversely proportional to the growing number of data. In this way, we can imagine a cluster center $\vec{\mu_i}$ which is not really in the optimal center position (according to the data history), but that remains the center because it still has the highest potential value. Therefore, the incremental clustering process of the premise part of the fuzzy classifier will not be able to take advantage of the data points that do not have a very high potential to move (or reshape) the existing clusters. We enhance the incremental clustering process (described in section III-A) by an adaptation algorithm that allows the modification of all the fuzzy prototypes by re-centering and re-shaping them for each new data point. For this purpose, we use a fuzzy version of the Vector Quantization algorithm [11]. In this method, the farther the normalized activation β_i of the premise of the rule i is away from its objective score β_i^*, the more it must be moved:

$$\vec{\Delta\mu}_i = \lambda * (\beta_i^* - \beta_i(\vec{x})) * (\vec{x} - \vec{\mu}_i) \tag{9}$$

where the adaptation parameter λ lies between 0 and 1. The objective score β_i^* is 1 if the prototype P_r and $\vec{x}$ belong to the same class and 0 otherwise. In the same way, a fuzzy

recursive formula is given in [12] to update the inverse of the covariance matrix as follows:

$$Q_i^{-1} \Leftarrow \frac{Q_i^{-1}}{1-\alpha\delta_i} - \frac{\alpha\delta_i}{1-\alpha\delta_i} \cdot \frac{(Q_i^{-1}\vec{d}) \cdot (Q_i^{-1}\vec{d})^T}{1+\alpha\delta_i(\vec{d}^T Q_i^{-1}\vec{d})} \quad (10)$$

$$\delta_i = \beta_i^* - \beta_i(\vec{x}) \quad (11)$$

where $\vec{d} = \vec{x} - \vec{\mu_r}$ and α lies between 0 and 1.

C. Learning algorithm

The incremental learning algorithm of the consequent parameters depends on the type of the used fuzzy system. If we take the three structures mentioned in Section II: (i) no consequent learning is needed for the simple structure with binary consequents, (ii) for the second structure, an online estimation of the constant consequents is presented in [13] and (iii) the linear consequents learning problem in a first-order TS can be solved by the weighted Recursive Least Square method (wRLS)[8]. The complete learning algorithm can be summarized by Algorithm 1.

Algorithm 1 Online incremental learning algorithm

```
for all new sample x⃗ do
  if x⃗ is the first sample of a new class then
    add a new fuzzy prototype centered on x⃗ to the
    system; let its potential be 1
  else
    calculate the potential of x⃗ by [4]
    update the potentials of the existing prototypes cen-
    ters using [8]
    if P(x⃗) > P_k(μ⃗_i)  ∀i ∈ [1, R]  then
      if x⃗ is close to a center μ⃗_i then
        let x⃗ be the center of the prototype P_i
      else
        add a new fuzzy prototype centered on x⃗ to the
        system; let its potential be 1
      end if
    else
      apply premise adaptation according to x⃗ by [9]
      and [10]
    end if
  end if
  update the consequent parameters;
end for
```

IV. Evaluation

We will particularly focus in our experiments on the rapidity of the performance improvement in the beginning of the incremental learning process, and on the stability and the recovery speed of the performance when introducing new unseen classes. We led the experiments on the "SIGN" database, which is a database of on-line handwritten gestures. It is composed of 17 different gestures drawn by 17 different writers on Tablet PCs. Each writer has drawn 100 samples of each gesture, *i.e.* 1,700 gestures in each writer-specific dataset. The dataset (and additional information on the data collection protocol) can be found in [14]. Each gesture is described by a set of 21 features. The presented results are the average of results of 17 different tests for the 17 writers. In order to get the results unbiased by the data order effect, we repeat the experiment for each writer 40 times with different random data orders and the mean results are considered. We used about half of the database for the incremental learning process and the rest is used to estimate the evolution of the performance during the learning process. Two fuzzy incremental learning models are compared in these experiments: (a) ETS: Evolving zero-order TS classifier with binary consequents and recursive mountain clustering learning, (b) ETS+: our extended version of ETS in which we integrate the fuzzy vector quantization algorithm in the incremental learning process. The parameters λ and α are set to 0.005 and 0.001 respectively. In the first experimental protocol, we introduce the 17 gestures in the beginning of the learning process. A new sample from each class is presented to the system between each two consecutive evaluation points. In order to have referential values in the evaluation of the recognition rate on the used dataset, we trained two well-known non-incremental classifiers using the whole training set in batch mode, and we measured their performance on the test dataset. We choose a Multi-Layer Perceptron (MLP) classifier and a K-Nearest Neighbor (K-NN, with K=5) classifier. We note from Figure 1 that

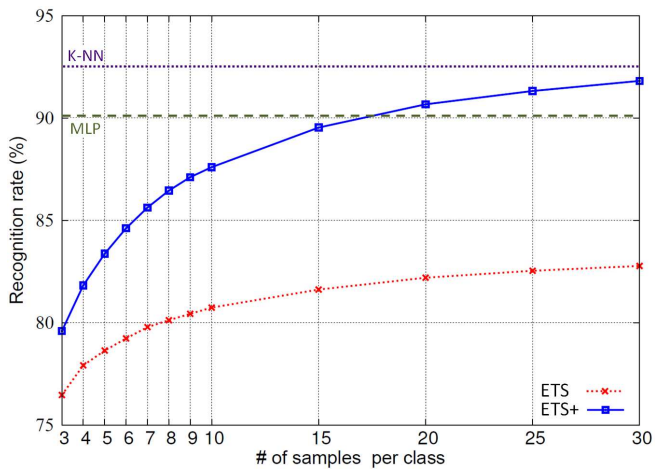

Figure 1: Evolution of the recognition rate in the beginning of the incremental learning process.

integrating the premise adaptation in the model significantly boosts the incremental learning process. By comparing the performance of ETS+ with that of ETS, we see that the recognition error rate decreases by about 40% thanks to

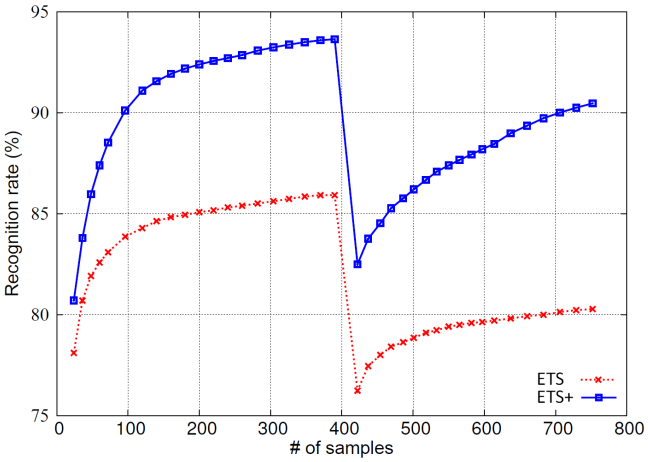

Figure 2: Performance recovery when introducing new unseen classes.

premise adaptation. We note that using ETS+, a recognition rate of 90% is reached after only 15 samples per class. It is important to mention as well that the classifier contains, in average, less than 20 fuzzy rules for 17 classes of gesture, which is computationally reasonable. Furthermore, we can note from the same figure that the presented incremental one-pass learning model (ETS+) can achieve or exceed the performance of some well-known iterative (or batch) classification methods. In the second experiment, we emulate the real application context in which the classifier starts with few classes of gestures, and then the user adds a set of new gestures according to his needs. We aim at studying the ability of the classifier to learn new classes of data without fully destroying the knowledge learned from the old ones. Thus, the learning process starts with 10 classes of gestures, and then, few samples of 7 new unseen gestures and some samples of the already learned ones are introduced between each two consecutive evaluation points. Figure 2 shows how ETS+ resists better when introducing new classes and it is able to re-estimate rapidly all the covariance matrices and to improve rapidly the recognition performance for the old and the new gestures.

V. Conclusion

In the context of handwritten gesture recognition systems, we present in this paper an online incremental learning algorithm for fuzzy rule-based classifiers. Using this algorithm, the recognition system can start to learn from scratch and with few learning data. Moreover, the dynamic nature of the presented classifiers allows as well adding new unseen classes without destroying the already learned ones.

References

[1] R. Polikar et al., "Learn++: An incremental learning algorithm for supervised neural networks," *IEEE Tr. on Systems, Man, and Cybernetics*, vol. 31, no. 4, pp. 497–508, 2001.

[2] J. Zwickel and A. J. Wills, *New Directions in Human Associative Learning*. Psychology Press, 2005.

[3] J. LaViola et al, "A practical approach for writer-dependent symbol recognition using a writer-independent symbol recognizer," *PAMI*, vol. 29, no. 11, pp. 1917–1926, 2007.

[4] J.-S. Jang, "Anfis: adaptive-network-based fuzzy inference system," *IEEE Tr. on Systems, Man and Cybernetics (Part B)*, vol. 23, no. 3, pp. 665–685, 1993.

[5] G. A. Carpenter and S. Grossberg, "The art of adaptive pattern recognition by a self-organizing neural network," *Computer*, vol. 21, no. 3, pp. 77–88, 1988.

[6] J. Sadri et al, "A new clustering method for improving plasticity and stability in handwritten character recognition systems," in *ICPR*, 2006, pp. 1130–1133.

[7] E. Lughofer, "Flexfis: A robust incremental learning approach for evolving takagi-sugeno fuzzy models," *IEEE Tr. on Fuzzy Systems*, vol. 16, no. 6, pp. 1393–1410, 2008.

[8] P. Angelov and D. Filev, "An approach to online identification of takagi-sugeno fuzzy models," *IEEE Tr. Systems, Man, and Cybernetics*, vol. 34, no. 1, pp. 484–498, 2004.

[9] R. R. Yager and D. P. Fileu, "Learning of fuzzy rules by mountain clustering," in *SPIE*, B. Bosacchi and J. C. Bezdek, Eds., vol. 2061, 1993, pp. 246–254. [Online]. Available: http://link.aip.org/link/?PSI/2061/246/1

[10] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE TSMC*, vol. 15, no. 1, pp. 116–132, 1985.

[11] F. Chung and T. Lee, "Fuzzy competitive learning," *Neural Networks*, vol. 7, no. 3, pp. 539–551, 1994.

[12] S. De Backer et al, "Texture segmentation by frequency-sensitive elliptical competitive learning," *Image and Vision Computing*, vol. 19, no. 9–10, pp. 639–648, 2001.

[13] J.-C. de Barros and A. L. Dexter, "On-line identification of computationally undemanding evolving fuzzy models," *Fuzzy Sets and Systems*, vol. 158, no. 18, pp. 1997–2012, 2007.

[14] "http://www.synchromedia.ca/web/ets/gesturedataset."