Towards Online Learning of a Fuzzy Classifier

Sofia Visa
Department of ECECS,
University of Cincinnati,
Cincinnati, OH 45221-0030, USA
svisa@ececs.uc.edu

Anca Ralescu
Department of ECECS,
University of Cincinnati,
Cincinnati, OH 45221-0030, USA
aralescu@ececs.uc.edu

***Abstract*— This study addresses issues related to the online applicability of a fuzzy classifier. In particular, it shows that a Fuzzy Classifier can be learned incrementally, and that in this process, imbalanced data sets, even when imbalance changes between classes can be used. Finally, it shows that for each class, examples and counter examples, can be effectively used. The most important aspect of the online fuzzy classifier is its perfect incremental aspect.**

Keywords: fuzzy classifier, online learning, adaptive learning systems, fuzzy modeling, pattern recognition.

I. Introduction

More and more computer applications require large volume of data, which of necessity may reside at different locations, or it is available at different times. On the other hand, pattern recognition algorithms must cope with dynamic patterns which change over time and for whose update all the data may be necessary. Therefore, in this context, online learning systems become more important, since they receive one or data items at a time and discharge it after learning. Such online learning systems are efficient under several aspects: are less memory consuming (by storing only rules or sufficient old information to enable correct updating), they can be used online even at intermediate stages (training and classification are not disjunct tasks), and they are faster since they involve only updating.

We present an architecture for online learning and classification using a fuzzy classifier(FC) first proposed in [7]. The main reason for which this classifier is suitable for online learning is that it is based only on data frequency. The online version successively updates this frequency distribution as new data arrives. This approach is also suitable for deriving a classifier in a distribute environment with *horizontally fragmented* data (i.e. when subsets of the training data reside at different locations).

In this study the online learning refers to the fact that data arrive in batches (or one at the time) and not the fact that there is no division between the train and test data [5].

II. The Fuzzy Classifier (FC)

Let C_i, $i = 1, \ldots, m$ denote classes of interest. To learn a multi-class classifier means to infer a decision rule (boundary) usually from examples of class membership (training data). To learn a fuzzy classifier means that the decision rule is expressed in terms of fuzzy sets.

For each class C_i, the training set T_i consists of:

- D_i: data for direct class membership - data points which belong to C_i;
- I_i: data for indirect class membership - data points that specifically do **not** belong to class C_i;

Therefore, if T denotes the training data for classes C_i, $i = 1, \ldots, m$, then the following hold:

$$T_i = D_i \cup I_i, \ i = 1, \ldots, m$$
$$T = \cup_{i=1}^{m} T_i$$

Note that any of these training sets may have points in common, that is, for any i and j, it is possible that: i and j, it is possible that:

$$D_i \cap D_j \neq \emptyset$$
$$I_i \cap I_j \neq \emptyset$$
$$D_i \cap I_j \neq \emptyset$$

Furthermore, for a given class C_i, either D_i, or I_i may be empty, but not both (because in that case the class cannot be learned).

The Fuzzy Classifier learns each class (directly or indirectly) independently, as a fuzzy set (or, as a pair of fuzzy sets, corresponding to the direct examples and indirect examples, respectively).

For a test data point, the membership degree (or degrees if the class is modeled as a pair of fuzzy sets) is used to decide class membership. In the simplest case, x is assigned to the class with largest membership value.

$$x \in C_{i_0}, \ \ where \ \ i_0 = arg \ maxFS_i(x)$$

It follows then from the previous discussion that the most important problem in designing a fuzzy classifier is to construct the membership functions for the fuzzy sets modeling each class. Often, fuzzy sets are constructed in an ad-hoc manner, by trial and error. Here, and for the problem at hand this aspect is very important, fuzzy sets are constructed in a systematic way, based on frequency of elements in a given class. The exact mechanism of deriving a fuzzy set membership function makes use of mass assignment theory (MAT) [3] and it is based on the relation between fuzzy sets, the probability distribution on their level sets (mass assignments) and the probability of selecting - *selection rule*- a particular value from a level set. When elements of each level set are equally likely to be selected this rule is called

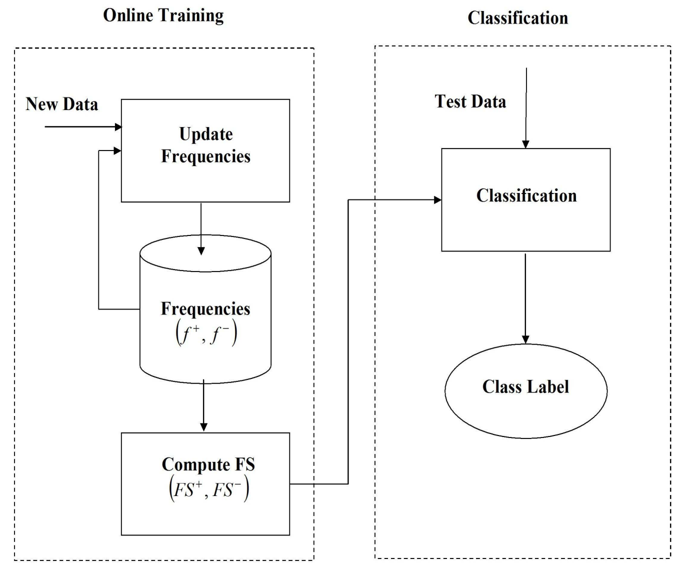

Fig. 1. General layout for online learning and classification using fuzzy classifiers.

least prejudiced rule. In this case, the fuzzy set, constructed from the relative frequency distribution of the class elements, is obtained as follows: Given the class relative frequency distribution, $\{f_{(k)}; k = 1, \ldots, n; 1 \geq f_{(1)} \geq f_{(2)} \geq \cdots \geq f_{(n)} \geq 0, \sum_{i=1}^{n} f_{(i)} = 1\}$, the least prejudiced fuzzy set is obtained from the Equation 1:

$$\mu_{(k)}^{lpd} = kf_{(k)} + f_{(k+1)} + f_{(k+2)} + ... + f_{(n)} \tag{1}$$

where $\mu_{(k)}^{lpd}$ denotes the kth largest value of the membership function. For an extensive treatment of this subject see [3] and [2].

III. General Framework for the Online Fuzzy Classifier(OFC)

The problem of online updating of the membership function is considered next, in the framework of a two-class classifier (that is $n = 1$). For historical reasons, we refer to these classes as the *positive* and *negative* class. Figure 1 shows the general layout of the online learning of the 2-class Fuzzy Classifier, henceforth to be referred to as the OFC.

Current data information is stored in the (cumulative) class frequencies (f^+, f^-). This information will be used to: (1) derive current fuzzy sets, and (2) for updating fuzzy sets when new data is presented. A simple update operation of the frequencies from which the fuzzy sets are recomputed. This way, the system learns continuously and at the same time, it can be used for classification. Note that class cumulative frequency can be updated using direct data only or, alternatively using indirect data, or again, in the latter case the indirect data can be used to update the complement of a class.

This paper illustrates the online learning of the fuzzy classifier on a one-dimensional artificial data set. In addition, application for multi-class classification and for multi-dimensional data sets is also explored.

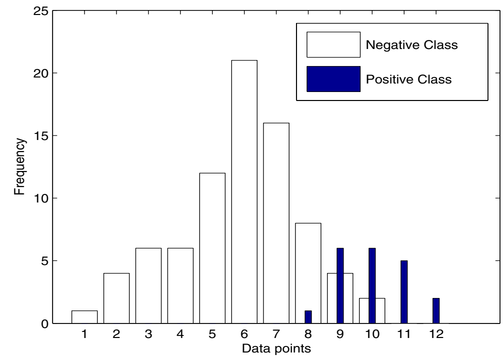

Fig. 2. The data set histogram.

IV. Experiments using the Online Fuzzy Classifier

Three experiments are carried out as follows:

1) **Experiment 1** shows that online learning of the classifier from direct training data is perfectly incremental, in the sense that the resulting fuzzy sets is identical to the fuzzy set which would have been obtained had the entire data been seen;
2) **Experiment 2** shows that the OFC can be applied to imbalanced data sets even when, as new data is presented for learning, the imbalance shifts between classes;
3) **Experiment 3** addresses the use of online learning of counter examples, that is, for a class C_i, the training point $x \in I_i$.

A. The Data Set

The one-dimensional artificial data set consists of two classes. The positive class has 20 data obtained from the normal distribution with $\mu = 9$ and $\sigma = 1$; the negative class consist of 80 points obtained from the normal distribution with $\mu = 5, \sigma = 2$. The range of the whole data set spans the interval $[1, 12]$. Fig. 2 shows the histogram for the whole data set. As it can be observed, the two classes overlap in three data points: $x = 8$, $x = 9$ and $x = 10$ are valid examples for both classes. However, fuzzy classifiers such as the one presented here can deal with overlap: data in overlapping are classified according to the membership degrees to each of the classes.

B. Experiments

For the experiments presented in this section, the following steps are used in learning the online data:

Input: *k batches of data arriving online;*

Output: *The fuzzy sets corresponding to each class FS^+ and FS^-;*

For each batch $T_i = (T_i^+; T_i^-)$, $i = 1 \ldots k$ do:

Step 1: *From T_i compute New* $\begin{pmatrix} x^+; x^- \\ f^+; f^- \end{pmatrix}$;

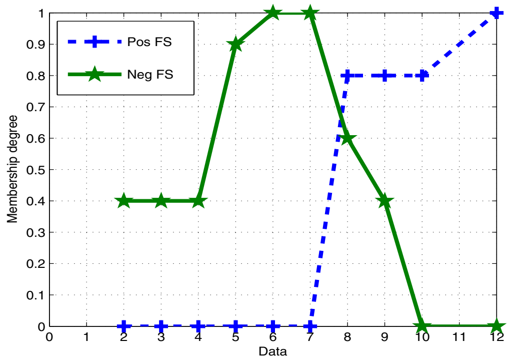

Fig. 3. Experiment 1: The fuzzy sets after the first batch.

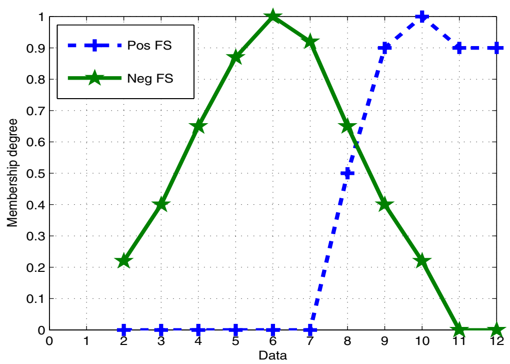

Fig. 4. Experiment 1: The fuzzy sets after the second batch.

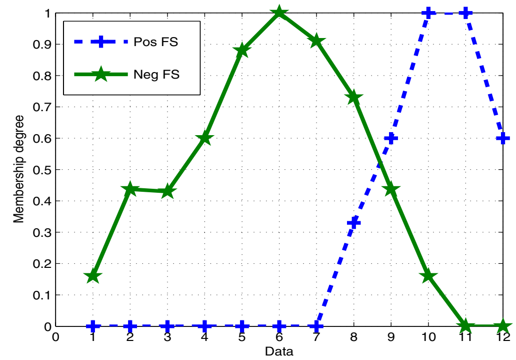

Fig. 5. Experiment 1: The fuzzy sets after the third batch.

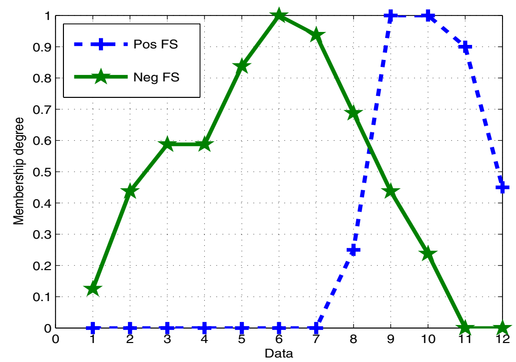

Fig. 6. Experiment 1: The fuzzy sets after the forth batch.

Step 2:

$$Current(x^+; x^-) = Old(x^+; x^-) \bigcup New(x^+; x^-);$$

$$Current(f^+; f^-) = 0 \vee Old(f^+; f^-) \pm New(f^+; f^-);$$

Step 4: *Store*

$$Current \begin{pmatrix} x^+; x^- \\ f^+; f^- \end{pmatrix};$$

Step 5: *Compute the FS using the updated frequencies from Step* 4.

In the above algorithm k denotes the number of batches (each batch may consist of zero, one or more data for each class), T_k are the actual data for a given batch and $x^{+,-}$ are the data in the domain, which are stored in a pair with their corresponding frequencies $f^{+,-}$. The frequencies can be updated in both ways (see the above Step 3): the addition operation updates the frequencies when examples of a class arrive, whereas the subtraction version is used for the counterexamples of a given class (this case is discussed in Example 3).

Experiment 1: In the first experiment, the online arrival of data (for $k = 4$ batches) is simulated. Each of the four batches consist of five positive examples and twenty negative examples. After receiving each batch the frequencies are computed/updated and the data are discarded. At each such intermediate step the fuzzy sets can be compute and used for classification purpose (see Figures 3-6). OFC produces the same FS as the offline FC when applying to the same data. This is the major importance of the online version of the FC. To guarantee such an equivalence between on- and off-line learning, other Machine Learning algorithms must store the data and recompute (e.g. the decision trees must restructure the tree as new examples arrive [4], neural network must retrain).

Figure 3 shows the FS for the two classes after receiving the first batch of data (5 positives and 20 negatives). The data are dropped and only the frequencies corresponding to each data already in the domain are stored. For the next batch of data the frequencies are updated and the obtained FS are displayed in Figure 4. Figure 5 shows the FS obtained after receiving the third batch of data. After receiving the last batch of data the fuzzy sets (Figure 6) computed from the newly updated frequencies are exactly the same as the ones obtained offline using all data (Figure 7).

Experiment 2: The second experiment shows that the OFC applies for imbalanced data, even when the imbalanced class shifts over time. For this experiment $k = 3$ and the distribution of the three batches of data are shown in Table I. After learning the first batch, the majority class is the negative class, but the arrival of the second batch shifts the majority to the positive class.

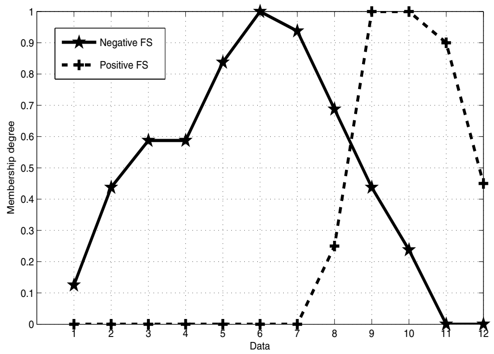

Fig. 7. The fuzzy sets obtained offline using all data.

TABLE I

EXPERIMENT 2: THE DATA DISTRIBUTION OF THE ONLINE BATCHES.

Batch	Card|Pos|	Card|Neg|
1	5	10
2	15	5
3	0	65

(B_1) shows the state after the first batch of examples: for each data point (first row) the corresponding frequencies (second row) and the membership degree of the fuzzy sets (third row).

$$\begin{pmatrix} x^- \\ f^- \\ FS^- \end{pmatrix} = \begin{pmatrix} 5 & 6 & 7 & 8 & 9 \\ 4 & 3 & 1 & 1 & 1 \\ 1 & 0.9 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$
$$\begin{pmatrix} x^+ \\ f^+ \\ FS^+ \end{pmatrix} = \begin{pmatrix} 8 & 9 & 10 & 12 \\ 1 & 1 & 1 & 2 \\ 0.8 & 0.8 & 0.8 & 1 \end{pmatrix} \qquad (B_1)$$

Obviously, the majority class is the negative class. Further, the data points $x = 8$ and $x = 9$ are valid examples (with frequency one) for both the positive and negative class. But their membership values to the positive and negative fuzzy sets are different, since they are computed relatively to the class size. Thus, when testing, this classifier assigns the data points $x = 8$ and $x = 9$ to the positive class ($FS^+ = 0.8 > FS^- = 0.5$), even though the negative class is the majority class. This example shows that the fuzzy classifier can be applied for overlapping and imbalance data where the class of interest is the minority class. In a way, the fuzzy classifier captures here the fact that the examples $x = 8$ and $x = 9$ are more representative for the minority class than for the majority class. This is a particular feature of the FC that gives chances to the minority class as well, by considering the class size in deciding the class label.

After receiving the second batch of data, the imbalance changes, in the sens that now the positive class becomes the majority class. The updated frequencies and fuzzy sets for each class are displayed in (B_2).

TABLE II

EXPERIMENT 3: THE DATA DISTRIBUTION OF THE ONLINE BATCHES.

Batch	DataInPos	DataNotInPos	DataInNeg	DataNotInNeg
1	5	0	10	0
2	0	3	0	5
3	17	0	65	0

$$\begin{pmatrix} x^- \\ f^- \\ FS^- \end{pmatrix} = \begin{pmatrix} 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 1 & 4 & 4 & 3 & 1 & 1 \\ 0.4 & 0.4 & 1 & 1 & 0.8 & 0.4 & 0.4 \end{pmatrix}$$
$$\begin{pmatrix} x^+ \\ f^+ \\ FS^+ \end{pmatrix} = \begin{pmatrix} 8 & 9 & 10 & 11 & 12 \\ 1 & 6 & 6 & 5 & 2 \\ 0.2 & 1 & 1 & 0.9 & 0.4 \end{pmatrix} \qquad (B_2)$$

After receiving the last batch (consisting exclusively of negative examples) the negative class becomes again the majority class (see (B_3)). The obtained fuzzy sets are displayed in Figure 7.

$$\begin{pmatrix} x^- \\ f^- \\ FS^- \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 4 & 5 & 5 & 8 & 17 & 14 & 7 & 3 & 2 \\ 0.1 & 0.4 & 0.5 & 0.5 & 0.8 & 1 & 0.9 & 0.6 & 0.4 & 0.2 \end{pmatrix}$$
$$\begin{pmatrix} x^+ \\ f^+ \\ FS^+ \end{pmatrix} = \begin{pmatrix} 8 & 9 & 10 & 11 & 12 \\ 1 & 6 & 6 & 5 & 2 \\ 0.2 & 1 & 1 & 0.9 & 0.4 \end{pmatrix} \qquad (B_3)$$

Experiment 3: For the third experiment we consider the problem in which examples as well as counterexamples for a class can arrive online. The task is to **learn and unlearn** the class with the purpose of classification. The Table II shows the distributions of the examples for each batch. First batch consist of five positive and ten negative examples. The obtained frequencies and the fuzzy sets are exactly as the ones listed in (B_1). Further, in the second batch arrive counter examples for both the classes, as follows:

- For the positive class the last three data (from the first batch) are received as *not belonging to the positive class*. However, this does not imply that they belong to the negative class. Thus, this data must be unlearned buy the positive class.
- Similarly, the negative class must unlearn the last five examples from the first batch.

The obtained frequencies and fuzzy sets are listed in (B_4). At this point, after receiving first two batches of data, the first two examples of the positive class and the first five examples of the negative class are learned. On the last step the remaining data are received and the fuzzy sets are equivalent with the ones obtained otherwise offline using all data (see $(B3)$ and Figure 7). Of course, the online batches may consist of both example and counter examples for the same class.

$$\begin{pmatrix} x^- \\ f^- \\ FS^- \end{pmatrix} = \begin{pmatrix} 5 & 7 & 9 \\ 3 & 1 & 1 \\ 1 & 0.6 & 0.6 \end{pmatrix}$$
$$\begin{pmatrix} x^+ \\ f^+ \\ FS^+ \end{pmatrix} = \begin{pmatrix} 12 \\ 2 \\ 1 \end{pmatrix} \qquad (B_4)$$

V. Multi-class and multi-dimensional data

The online fuzzy classifier presented here can be extended in a straightforward manner for multi-class classification problems as well, since each class is modeled separately in a corresponding fuzzy sets.

For multidimensional data, classifier components along each dimension are derived first. This is justified first, by the fact that the construction of the classifier relies on frequency distributions. When the number of dimensions increases, data tend to be sparse and hence it is difficult to construct meaningful frequency distributions. At the same time, assuming that such a construction were made, the testing step would entail multidimensional interpolation. As an alternative to this, each component is modeled and tested separately and then the results are aggregated. Roughly speaking this is analogous to the use of a Naive Bayesian approach in a probability based approach. Various aggregation methods [9] can be used to infer the final classification result.

VI. Conclusions

The following summarizes the features of fuzzy classifiers that make them suitable for online learning and classification:

- The OFC does not store the data coming online, rather requires a limited storage of the data frequencies;
- It is not order dependent and it is faster than other approaches (e.g. neural networks);
- Since the OFC is scalable, efficient and can deal with imbalanced data [6], it can be further applied for distributed data. In this case, the required frequencies can be computed at peers and centralized or exchanged afterwards;
- From the set of frequencies, different fuzzy sets may be derived [2];
- The OFC models well imbalanced data sets [8], [9] ;
- The OFC can learn from both, examples and counterexamples;
- The online version of the classifier is equivalent to the offline one obtained using all data at once. This is a great advantage over other learning methods: for example the decision trees must recalculate the tree each time a new data arrives and the neural network must retrain.

Acknowledgments

This work was partially supported by the Ohio Board of Regents, Grant N000140310706 from the Department of the Navy, and the Rindsberg Graduate Fellowship of the College of Engineering, University of Cincinnati.

References

[1] Visa S. and Ralescu A., Learning Imbalanced and Overlapping Classes using Fuzzy Sets, Proceedings of the ICML-2003 Workshop: Learning with Imbalanced Data Sets II, Washington, 97-104, 2003.

[2] Ralescu A. L., Quantitative Summarization of Numerical Data Using Mass Assignment Theory, Invited lecture at the SOFT Kansai Meeting, May, 1997.

[3] Baldwin J., Martin T. and Pilsworth B., FRIL-Fuzzy and Evidential Reasoning in Artificial Intelligence, Research Studies Press, 1995.

[4] Fern A. and Givan R., Online Ensemble Learning: An Empirical Study, Machine Learning, 53, 71-109, 2003.

[5] Lane T. and Brodley C., Approaches to Online Learning and Concept Drift for User Identification in Computer Security, Knowledge Discovery and Data Mining, 259-263, 1998.

[6] Chan P., Fan W., Prodromridis A., and Stolfo S., Distributed Data Mining in Credit Card Fraud Detection, IEEE Intelligent Systems, 14(6), 67-74, 1999.

[7] Visa S., Ralescu A., Yeung S. and Genaidy, A., A word is worth many numbers - Modelingthe meaning of linguistic description of numerical data, Proceedings of the 13th Midwest Artificial Intelligence and Cognitive Science Conference, 48-55, 2002.

[8] Visa S. and Ralescu A., Fuzzy Classifiers for Imbalanced, Complex Classes of Varying Size, Proceedings of the IPMU Conference, Perugia, 393-400, 2004.

[9] Visa S. and Ralescu A., The Effect of Imbalanced Data Class Distribution on Fuzzy Classifiers - Experimental Study, Proceedings of the FuzzIEEE Conference, Reno, 2005.