

Adapting to Drift in Continuous Domains (Extended Abstract)

Miroslav Kubat¹ and Gerhard Widmer²

¹ Institute for Systems Sciences, Johannes Kepler University Linz,
A-4040 Linz, Austria

² Department of Medical Cybernetics, University of Vienna, and
Austrian Research Institute for Artificial Intelligence, A-1010 Vienna, Austria

1 Introduction: The Algorithm FRANN

Perhaps the first systems capable of tracking concept drift in supervised incremental learning were STAGGER (Schlimmer and Granger, 1986), FLORA (Kubat, 1989), and IB3 (Aha et al., 1991). Learning in time-varying environments has also been studied in the framework of genetic algorithms (Smith, 1987) and in adaptive-control applications of neural networks (Narendra & Parthasarathy, 1990). The latest descendant of the FLORA family is FLORA4 (Widmer, 1994).

Whereas most drift-tracking algorithms were oriented towards learning with symbolic attributes, the system FRANN, to be briefly described here, was developed for *numeric* domains with complex decision boundaries. Its algorithm is summarized in Table 1: a time window slides over a series of examples. From the window examples, a *radial-basis function (RBF) network* is created. At each step, N new examples are added to the window. Concept drift is recognized by decreased classification accuracy, as measured on the last M incoming examples.

Like the FLORA systems, FRANN possesses a *window-size* heuristic which decides about the potential need for shrinking the window (i.e., deleting from it some of the oldest examples) when concept drift is suspected. It combines two variables: the current window size and the number of erroneous classifications on the last M examples. Basically, old examples are deleted when the window is too large and/or when the number of errors grows (signalling concept drift).

2 Experimental Design

To assess the performance of FRANN, we compared its classification accuracy to that of FLORA4 (Widmer, 1994) on several artificial data sets, all of which simulate more or less extreme forms of concept drift. For FRANN, symbolic attributes are turned into numeric ones by replacing each attribute-value pair with a boolean variable and substituting 0 for *false* and 1 for *true*. To be able to deal with numeric attributes, FLORA4 was extended with a simple generalization operator that implements Michalski's *closing interval rule*. FLORA4 thus approximates numeric concepts by axis-parallel hyperrectangles.

Below is the characterization of the test domains. For each of them, 10 training sequences of 200 random examples were generated; the underlying concept

Table 1. Algorithm of the system FRANN

-
1. Put N new examples in the window;
 2. Apply hill-climbing search to find the best subset of the examples in the window, and create an RBF network from them; determine the output-layer weights;
 3. Use the network for the classification of both training and testing data and collect the respective statistics;
 4. Use the window-size heuristic to decide about deletions of older examples from the window, carry out the deletions and return to step 1.
-

was made to change after every 50 examples. Predictive accuracy was tested after every learning step on a set of 100 independent test examples. The reported results are averages over the 10 sequences. To ensure a stable learning environment, the positive and negative examples in the training sets alternated.

1. SINE. *Abrupt concept drift, noise-free examples.* The examples were uniformly distributed in the 2-dimensional unit square $[0, 1] \times [0, 1]$, with coordinates $[x, y]$. In the first context, all points lying below the curve $y = \sin(x)$ are classified as positive and the rest as negative. In the second, the classification is reversed;
2. CIRCLES. *Gradual concept drift, noise-free examples.* Four contexts are determined by four circles partially overlapping each other. Points inside the circles are positive and the points outside the circles are negative;
3. GAUSS. *Abrupt concept drift, noisy examples.* Positive examples from the domain $R \times R$ are normally distributed around the center $[0, 0]$ with standard deviation 1. Negative examples are normally distributed around $[2, 0]$ with standard deviation 4. The overlap of the two classes models noise. At each point of context change, the classification is reversed;
4. SYMBOLIC. *Abrupt concept drift, symbolic noise-free examples.* Here we used the series of symbolic concepts from Schlimmer and Granger (1986);
5. MIXED. *Abrupt concept drift, mixed boolean/numeric noise-free examples.* The examples are described by two boolean attributes and two numeric attributes from the domain $[0, 1]$. After each context change, the classification is reversed.

3 Results

Figures 1 through 5 compare the accuracies of FRANN vs. FLORA4 in the above domains. Points where the target concept changes are marked by vertical lines. Figures 4 and 5 give the results of different versions of FRANN: the curves marked *FRANN 1/3* represent the accuracy of FRANN when the number of hidden units in the RBF is 1/3 of the current window size; *FRANN 2/3* and *FRANN 3/3* are to be interpreted analogously. All other figures plot *FRANN 1/3*. FRANN is superior to FLORA4 in most of the numeric problems because FLORA4 approximates numeric concepts by axis-parallel hyperrectangles, while FRANN's RBF networks represent concepts by a set of weighted prototypes.

In the noisy GAUSS experiment, FLORA4 scores well in the first context, but worse in the second; this is probably because the first concept—positive examples

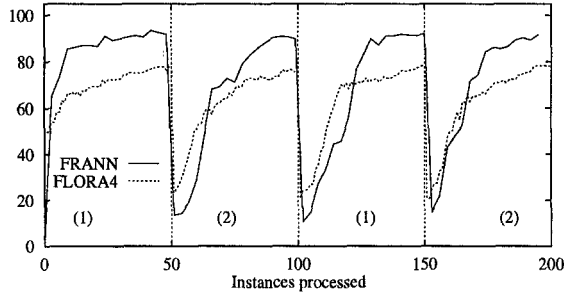


Fig. 1. Experiment 1 — Abrupt drift, noise-free examples.

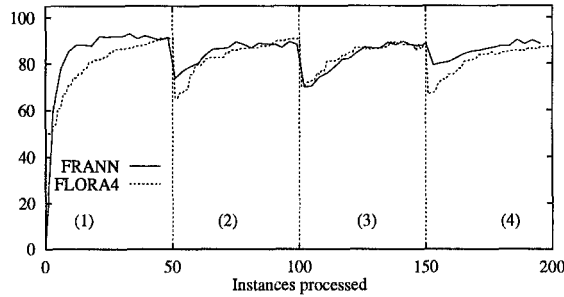


Fig. 2. Experiment 2 — Gradual drift, noise-free examples.

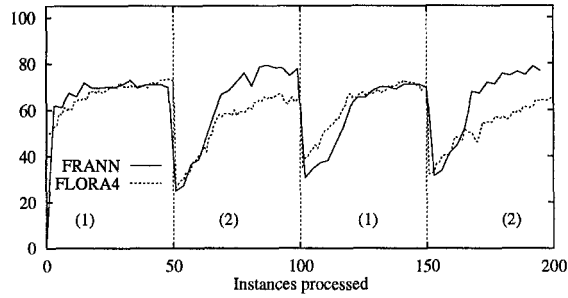


Fig. 3. Experiment 3 — Abrupt drift, noisy examples.

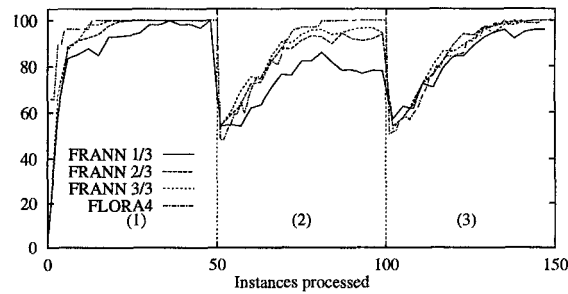


Fig. 4. Experiment 4 — Abrupt drift, symbolic noise-free examples.

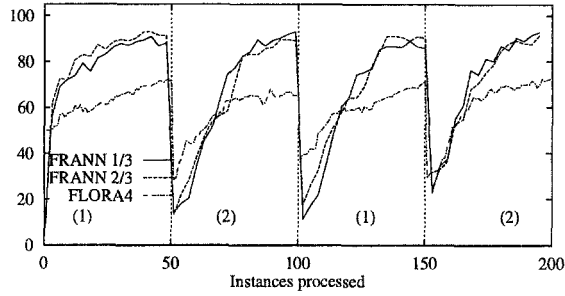


Fig. 5. Experiment 5 — Abrupt drift, mixed symbolic-numeric examples.

around a center with standard deviation 1—is ‘denser’ and easier to represent in closed form. FRANN, on the other hand, represents both classes explicitly via prototypes and thus learns both the concept and its inversion equally easily.

In the symbolic domain, FLORA4 outperforms FRANN because of its crisp concept representation and the more sophisticated windowing mechanism.

In the experiments with the ‘mixed data’, FRANN is significantly superior. Here, the underlying target concept is rather complex in relation to the number of attributes and training examples. This causes problems for FLORA4, whose window adjustment heuristic considers the relative complexity of hypotheses.

4 Conclusion

The experiments demonstrate that FRANN compares favourably with FLORA4 in the presence of concept drift. Learning is possible from examples described by symbolic as well as by numeric attributes, and because of its representation formalism (RBF networks, which realize a kind of prototype weighting scheme) FRANN is particularly effective in capturing concepts with nonlinear boundaries.

References

- Aha, D., Kibler D., and Albert, M.K (1991). Instance-Based Learning Algorithms. *Machine Learning* 6:37–66.
- Kubat, M. (1989). Floating Approximation in Time-Varying Knowledge Bases. *Pattern Recognition Letters*, 10:223–227.
- Narendra, K.S. and Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks* 1:4–27.
- Schlimmer, J.C. and Granger, R.H. (1986). Incremental Learning from Noisy Data. *Machine Learning* 1, 317–354.
- Smith, R.E. (1987). Diploid genetic algorithms for search in time varying environments. *Proceedings of an International Conference on Genetic Algorithms and their Applications*, 202–206.
- Widmer, G. (1994). Combining Robustness and Flexibility in Learning Drifting Concepts. In *Proceedings of the 11th European Conference on Artificial Intelligence, ECAI94*, Amsterdam. Chichester: Wiley & Sons.