Fuzzy-Teacher-based Learning for an Ensemble of Neural Networks

Neamat El Gayar

University of Cairo
Faculty of Computers and Information

Int. School on Neural Nets
7th Course: Ensemble Methods for Learning Machines
Vietri sul Mare, September 2002

Outline of the Talk

- Introduction
- Fuzzy Labels and Fuzzy Teaching Signals
- Fuzzy Teacher Based Learning Model (FTBL)
- Fuzzy Teacher Based Model for and Ensemble of Neural Networks (EFTBL)
- Experiments with a two Network Model
- Summary and Conclusions
- Future Work

Background

Joint ideas/problems from several disciplines:

- Neuroscience
- Pattern Recognition
- Sensor Fusion
- Multiple Classifier Systems

Learning Paradigms

Supervised

- Require labels (no learning if no correct response is offered)
- Explicit labeling: tedious, costly and sometimes impossible.
- Biologically not plausible.
- Do not take advantage of natural grouping present in the data.

Unsupervised

- No actual labels are required.
- Process large data efficiently.
- A much harder problem
- Can be inefficient in some applications that require external guidance.
- Constraints to succeed in classification

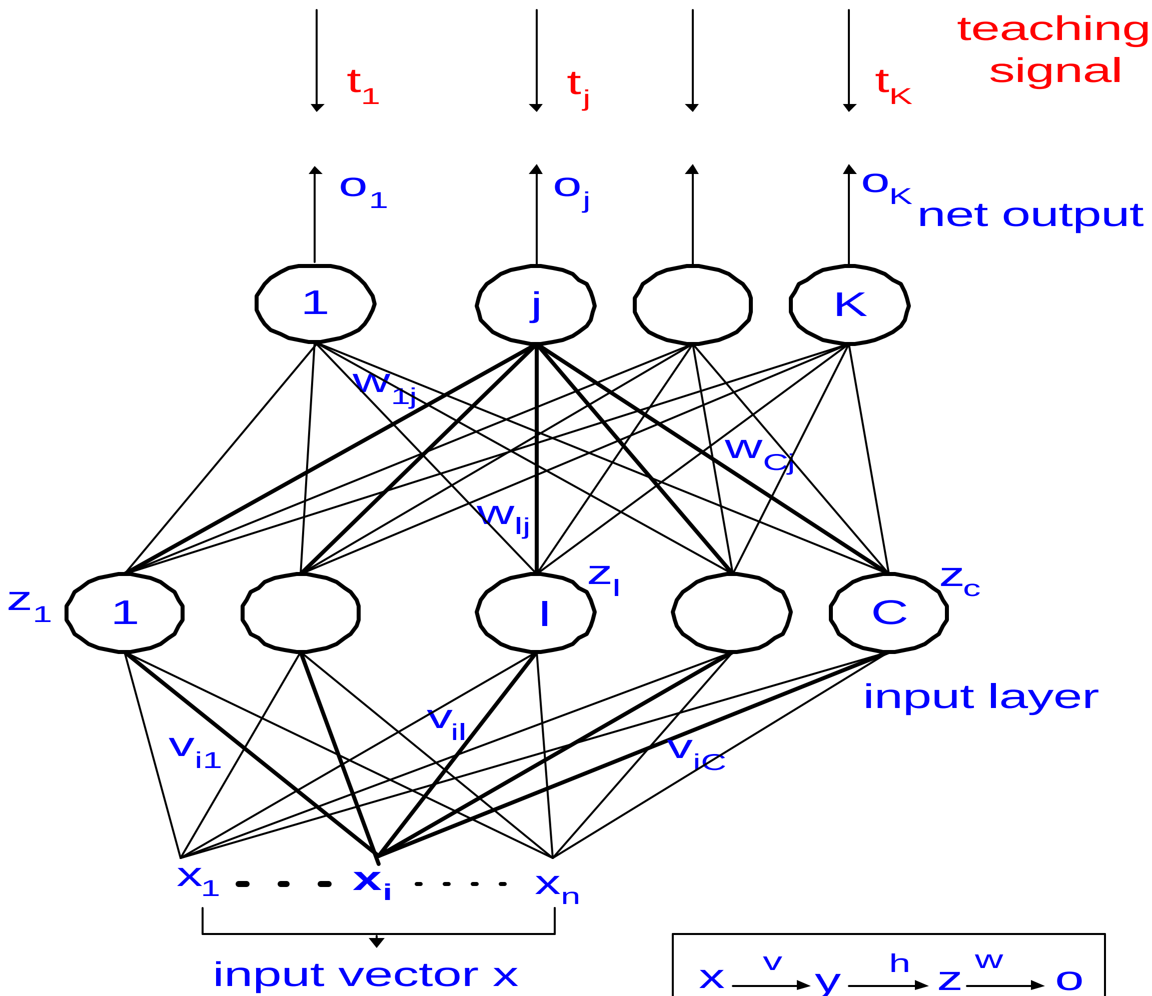
Alternative Learning Models:

•Reinforcement Learning.
•Partial Supervised Models.
•Self-Supervised Models.
•Models using Cross-Modal Information.
•Fuzzy Teacher-Based Learning Model
•Fuzzy Teacher Based Model for and Ensemble of Neural Networks.

Fuzzy Labels and Fuzzy Teaching Signals

- More realistic:
  - To model overlapping or ill-defined clusters
  - To accommodate uncertainty of the teacher
  - To deal with linguistic features
- Hard vrs. Fuzzy Labels:
  - **Hard Label:** $l_{\mathrm{x}} = (0, 0, \ldots, \underbrace{1}_{j}, 0, \ldots)$
  - **Fuzzy Label:** $l_{\mathrm{x}} = (\mu_1, \mu_2, \ldots, \underbrace{\mu_j}_{j}, \mu_{j+1}, \ldots)$
- How Fuzzy Labels can be obtained?
  - No class information - Partial class information
  - Class information available
- Objective of Learning with Fuzzy Labels.
- Validity: misclassifications - average certainty

Architecture of the FTLB Net



$$o_j^{'} = o_j / (\sum_q o_q)$$

$$o_j = \sum_{l=1}^{C} w_{lj} z_l$$

$$z_l = \exp \frac{-\|\mathbf{x}^{\mathrm{m}} - \mathbf{v}_{\mathrm{l}}\|^2}{\sigma_l^2}$$

$x \xrightarrow{v} y \xrightarrow{h} z \xrightarrow{w} o$

Net Terminology

Fuzzy Teacher-Based Learning Model (FTLB)

- **Objective Function:**

$$F = \alpha_u \underbrace{\sum_{l=1}^{c} z_l \|\mathbf{x^m} - \mathbf{v_l}\|^2}_{Unsupervised} + \alpha_s \underbrace{\sum_{j=1}^{K} (o_j^{'} - t_j)^2}_{Supervised}$$

- **Confidence-Based Learning:**

$$\alpha_s = \alpha_3(\beta_1 + \gamma\beta_2)$$

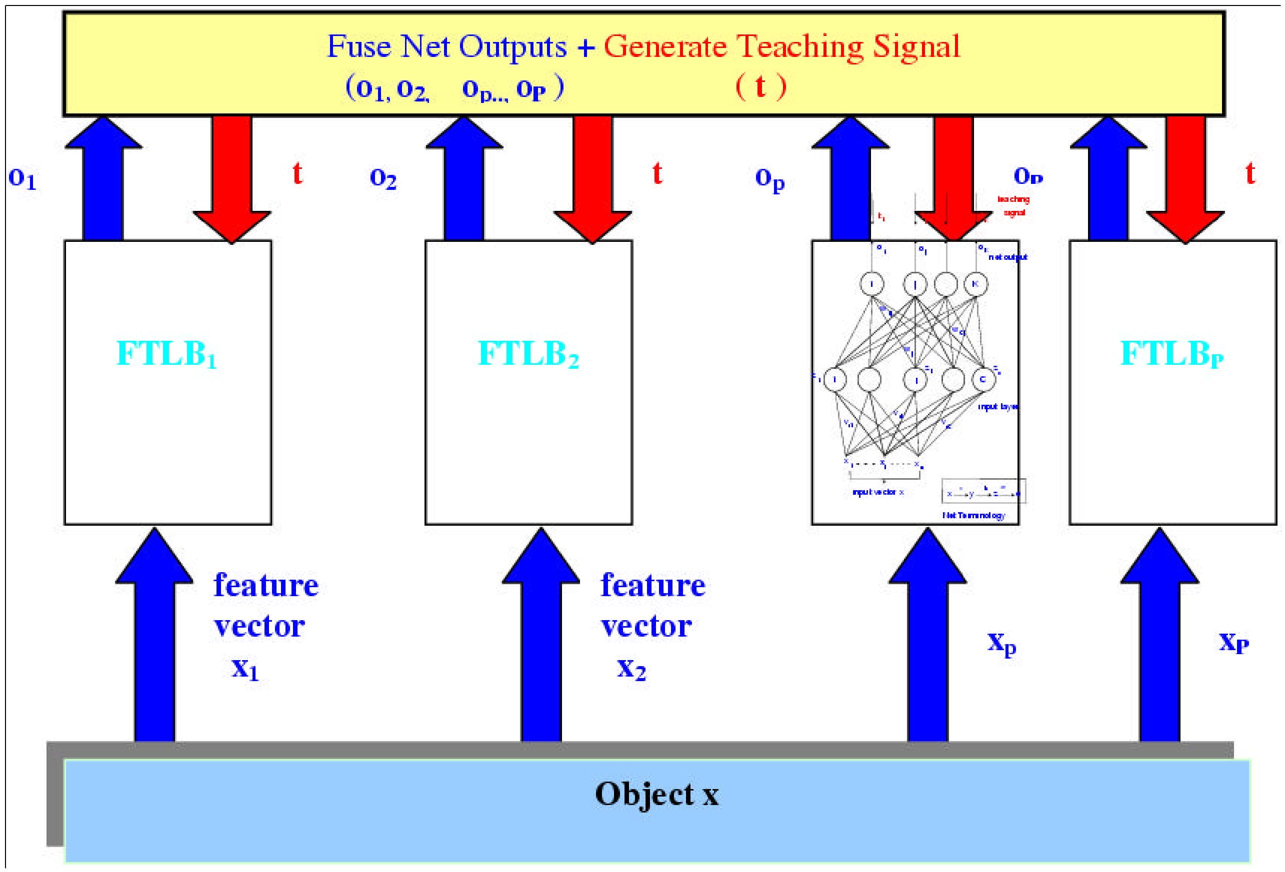
$$\begin{aligned} \beta_1 &= \log_2 K + \sum_j t_j \log_2 t_j \\ \beta_2 &= 1 + (\sum_j o_j^{'} \log_2 o_j^{'})/(\log_2 K) \end{aligned}$$

- **The $FTBL$ Families**

$$\alpha_s = \begin{cases} \alpha_3(\beta_1 + |\gamma|\beta_2) & \gamma > 0 \quad (FTBL+) \\ \alpha_3\beta_1 & \gamma = 0 \quad (FTBL0) \\ \alpha_3(\beta_1 - |\gamma|\beta_2) & \gamma < 0 \quad (FTBL-) \end{cases}$$

- **Learning Rules: Calculate** $\Delta w_{lj} = -\eta_1 \frac{\delta F}{\delta w_{lj}}$, $\Delta v_{il} = -\eta_2 \frac{\delta F}{\delta v_{il}}$ **and** $\Delta \sigma_l = -\eta_3 \frac{\delta F}{\delta \sigma_l}$

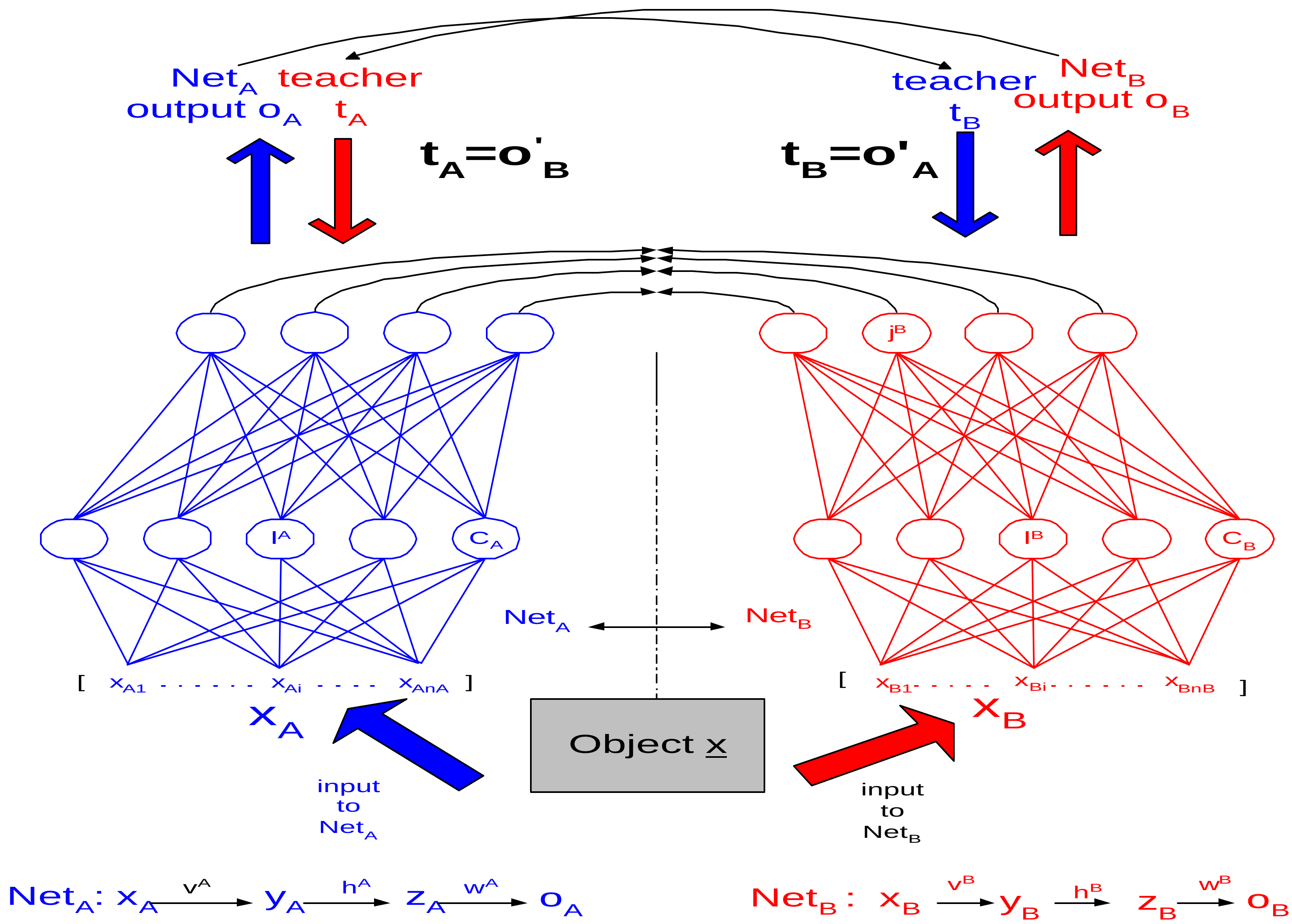
Fuzzy Teacher-Based for an Ensemble of NN (EFTLB)



Fuzzy Teacher-Based for an Ensemble of NN (EFTLB)

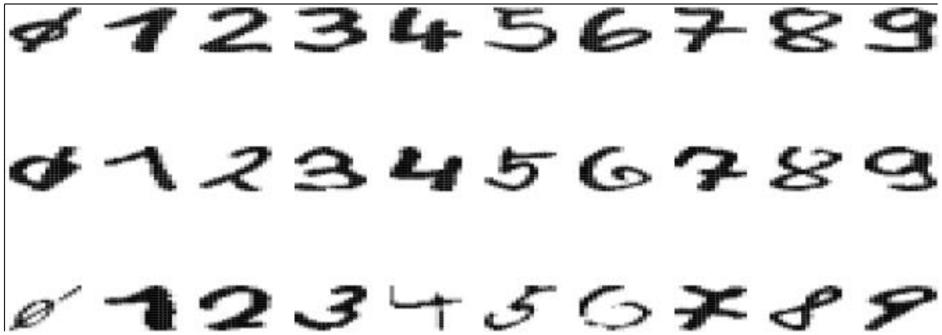
- Idea:
  - Object is observed by different nets processing different features.
  - A pattern is better classified in one module than the other.
  - Net being able to classify a pattern better is more certain
  - Each net learns using the output of the other net -> is only affected by the confident decisions (FTBL learning)
- Net Architecture
- Main Issues:
  - Feature selection to modules
  - Prototype initialization
  - Initialization of output layer
  - Learning (FTBL)
- Validity: Judge the combined decision
- Remark: Training is performed with few labels

Example with a two Network Model



Data Set

- Data Set and Features:
  - Subset of 10,000 handwritten digits (1000 samples per class)
  - Digits are normalized in height and width
  - Digits represented through a 16x16 gray scale matrix
  - Input to modules: upper and lower half of the digit



Experiments

- Fuzzy teachers:
  - *Teachers provided to the upper module:*

    •t1 (more miss.+less cer.) • t2 (more miss.+more cer.) • t3(actual labels)

  - *Teacher provided to the lower module:*

    •t1 (less miss.+less cer.) • t2 (less miss.+more cer.) • t3(actual labels)

- Learning: Using different versions of *FTBL* models.
- Testing: Evaluate classification rate of Upper, Lower and combined net on training and test data sets.
- Comparison to other methods:
  - DeSa's self-supervised model –Kohonen's supervised OLVQ

Comments on the FTBL

- Learning with a Hard Teacher:
  - Better results when supervised part is emphasized
  - Combined model avoids misclassifications done by either modules
  - FTBL models yield close results when supervised part is emphasized.
- Learning with a Fuzzy Teacher
  - Type of the teacher influences how much is learned (improvement)
  - Unsupervised part of the objective can be given larger values
  - Combined decision of lower and upper module improves results
  - Choose $\alpha^{FTLB+} < \alpha^{FTLB0} < \alpha^{FTLB-}$

Results on Training Data

	Upper	**Lower**	**Combined**
FTLB(t1)	87.0% (3.5%)	86.6% (9.5%)	89.3% (1.3%)
FTLB(t2)	86.5% (3.0%)	90.0% (13.0%)	91.2% (6.2%)
FTLB(t3)	99.2% (15.7%)	99.2% (22.2%)	100% (15.0%)
DeSa	80.0% **(-5%)**	84.5% (6%)	
OLVQ	94.0% (9.0%)	85.5% (7.0%)	

Results on Test Data

	Upper	**Lower**	**Combined**
FTLB(t1)	83.7% (1.7%)	84.7% (4.7%)	87.8% (4.8%)
FTLB(t2)	83.7% (1.7%)	86.5% (6.5%)	89.5% (6.5%)
FTLB(t3)	94.3% (12.3%)	93.7% (13.7%)	98.5% (15.5%)
DeSa	78.0% **(-7%)**	82.5% (4.0%)	
OLVQ	88.5% (3.5%)	87.5% (9.0%)	

Summary of Results

- FTBL models produce better classification rate than known self-supervised (DeSa's) model on both training and test data.
- Learning is influenced by the type of the teacher. (confidence+miss)
- Combined decision of module always improves results:
  - Mistakes not correlated
  - correct classification -> more certain
- Combined model on the trained without the actual labels:
  - close results to the supervised OLVQ on training data.
  - better results than the supervised OLVQ on test data.
- FTBL models provide much better results when trained with the actual labels than the OLVQ model on both training and test data.

The FTBL models are more efficient for training when either fuzzy or hard labels are available.

Current Work

- Experimenting with the multiple network paradigm.
- Studying different strategies of feature selection as input to modules.
- Testing effect of different output combination strategies.
- Investigating different classifier types for combination.
- Testing the ensemble learning model using real cross modal data, generated from sensors, such as audio-visual signals and others.