Adaptive Resonance Theory

Farhad

Shenzhen University

2018.12.24

Outlines

Outlines

Introduction

One of the main challenges of batch-learning models,i.e., MLP and RBF, is to overcome the **stability-plasticity** dilemma. Stability-plasticity means:

- The model should be **stable** enough to remember previous learned samples, and
- **Plastic** enough to absorb (learn) new input(s).

Introduction

One of the main challenges of batch-learning models,i.e., MLP and RBF, is to overcome the **stability-plasticity** dilemma. Stability-plasticity means:

- The model should be **stable** enough to remember previous learned samples, and
- **Plastic** enough to absorb (learn) new input(s).

Introduction

To solve the problem of **stability-plasticity** dilemma, **online ANNs** that are able to learn incrementally have been proposed, e.g. Adaptive Resonance Theory (ART)and Fuzzy Min-Max(FMM) networks.

The characteristics of an online learning model:

- The training samples are presented **one-by-one** for learning,
- Able to learn **only new input samples**, instead of **re-learning** all previously learned samples,
- Able to learn new knowledge without **disturbing or forgetting** existing knowledge, and
- Able to **predict** the label (target) of a new input sample **during learning**.

Introduction

To solve the problem of **stability-plasticity** dilemma, **online ANNs** that are able to learn incrementally have been proposed, e.g. Adaptive Resonance Theory (ART)and Fuzzy Min-Max(FMM) networks.

The characteristics of an online learning model:

- The training samples are presented **one-by-one** for learning,
- Able to learn **only new input samples**, instead of **re-learning** all previously learned samples,
- Able to learn new knowledge without **disturbing or forgetting** existing knowledge, and
- Able to **predict** the label (target) of a new input sample **during learning**.

Introduction

To solve the problem of **stability-plasticity** dilemma, **online ANNs** that are able to learn incrementally have been proposed, e.g. Adaptive Resonance Theory (ART)and Fuzzy Min-Max(FMM) networks.

The characteristics of an online learning model:

- The training samples are presented **one-by-one** for learning,
- Able to learn **only new input samples**, instead of **re-learning** all previously learned samples,
- Able to learn new knowledge without **disturbing or forgetting** existing knowledge, and
- Able to **predict** the label (target) of a new input sample **during learning.**

Introduction

To solve the problem of **stability-plasticity** dilemma, **online ANNs** that are able to learn incrementally have been proposed, e.g. Adaptive Resonance Theory (ART)and Fuzzy Min-Max(FMM) networks.

The characteristics of an online learning model:

- The training samples are presented **one-by-one** for learning,
- Able to learn **only new input samples**, instead of **re-learning** all previously learned samples,
- Able to learn new knowledge without **disturbing or forgetting** existing knowledge, and
- Able to **predict** the label (target) of a new input sample **during learning**.

Outlines

Adaptive Resonance Theory

- **Adaptive Resonance Theory (ART)** is known as a human cognitive information processing theory which has led to evolve many online neural network models.
- Proposed by Gail Carpenter and Stephen Grossberg (Boston University) in 1980s.
- ART models incorporate new data by measuring the **similarity level** between the existing prototype nodes and a new input sample against **a threshold**, i.e.,the vigilance test. If the vigilance test is not satisfied, a new prototype node can be added to learn the new input sample.

Adaptive Resonance Theory

- **Adaptive Resonance Theory (ART)** is known as a human cognitive information processing theory which has led to evolve many online neural network models.
- Proposed by Gail Carpenter and Stephen Grossberg (Boston University) in 1980s.
- ART models incorporate new data by measuring the **similarity level** between the existing prototype nodes and a new input sample against **a threshold**, i.e.,the vigilance test. If the vigilance test is not satisfied, a new prototype node can be added to learn the new input sample.

Adaptive Resonance Theory

- **Adaptive Resonance Theory (ART)** is known as a human cognitive information processing theory which has led to evolve many online neural network models.
- Proposed by Gail Carpenter and Stephen Grossberg (Boston University) in 1980s.
- ART models incorporate new data by measuring the **similarity level** between the existing prototype nodes and a new input sample against **a threshold**, i.e.,the vigilance test. If the vigilance test is not satisfied, a new prototype node can be added to learn the new input sample.

Adaptive Resonance Theory

Some models of ART:

ART1	binary input, unsupervised
ART2	continuous input, unsupervised
Fuzzy ART	continuous input, unsupervised
ARTMAP	binary input, supervised
Fuzzy ARTMAP	continuous input, supervised

Adaptive Resonance Theory

Some models of ART:

ART1	binary input, unsupervised
ART2	continuous input, unsupervised
Fuzzy ART	continuous input, unsupervised
ARTMAP	binary input, supervised
Fuzzy ARTMAP	continuous input, supervised

Adaptive Resonance Theory

Some models of ART:

ART1	binary input, unsupervised
ART2	continuous input, unsupervised
Fuzzy ART	continuous input, unsupervised
ARTMAP	binary input, supervised
Fuzzy ARTMAP	continuous input, supervised

Adaptive Resonance Theory

Some models of ART:

ART1	binary input, unsupervised
ART2	continuous input, unsupervised
Fuzzy ART	continuous input, unsupervised
ARTMAP	binary input, supervised
Fuzzy ARTMAP	continuous input, supervised

Adaptive Resonance Theory

Some models of ART:

ART1	binary input, unsupervised
ART2	continuous input, unsupervised
Fuzzy ART	continuous input, unsupervised
ARTMAP	binary input, supervised
Fuzzy ARTMAP	continuous input, supervised

Adaptive Resonance Theory

- Among different models of ART, **Fuzzy ART** and **Fuzzy ARTMAP(FAM)**are two popular unsupervised and supervised models,
- Both models merge the capability of ART in solving the stability-plasticity dilemma with the capability of fuzzy set theory in handling vague and imprecise human linguistic information.

Adaptive Resonance Theory

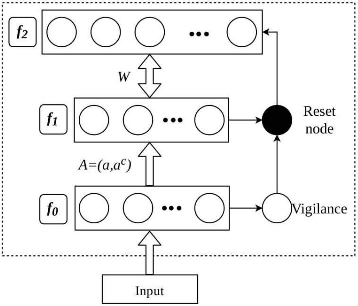
- Among different models of ART, **Fuzzy ART** and **Fuzzy ARTMAP(FAM)**are two popular unsupervised and supervised models,
- Both models merge the capability of ART in solving the stability-plasticity dilemma with the capability of fuzzy set theory in handling vague and imprecise human linguistic information.

Outlines

Fuzzy ART

Fuzy ART consists of three layers:

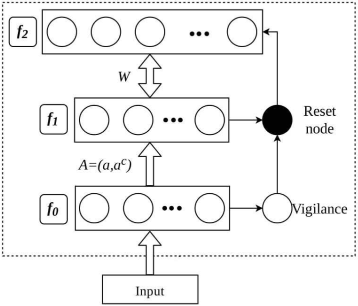
- f_0 is pre-processing layer,
- f_1 is the input layer, and
- f_2 is is the recognition layer.



Fuzzy ART

Fuzy ART consists of three layers:

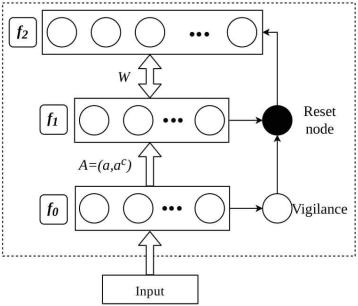
- f_0 is pre-processing layer,
- f_1 is the input layer, and
- f_2 is is the recognition layer.



Fuzzy ART

Fuzy ART consists of three layers:

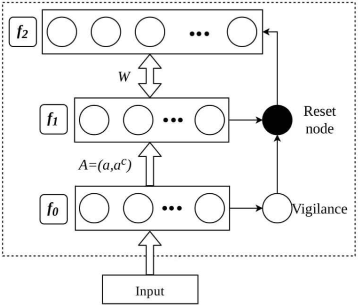
- f_0 is pre-processing layer,
- f_1 is the input layer, and
- f_2 is is the recognition layer.



Fuzzy ART

Fuzy ART consists of three layers:

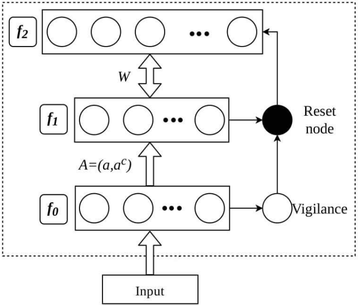
- f_0 is pre-processing layer,
- f_1 is the input layer, and
- f_2 is is the recognition layer.



Fuzzy ART

Fuzy ART consists of three layers:

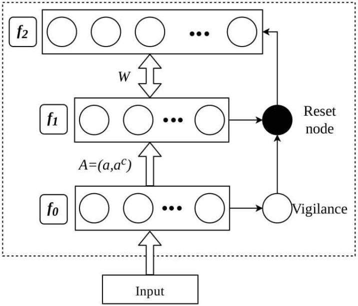
- f_0 is pre-processing layer,
- f_1 is the input layer, and
- f_2 is is the recognition layer.



Fuzzy ART

Fuzy ART consists of three layers:

- f_0 is pre-processing layer,
- f_1 is the input layer, and
- f_2 is is the recognition layer.



Fuzzy ART

The pre-processing layer performs complement coding in order to avoid the problem of category proliferation, as follows:

$$A = (a_1, ..., a_m, 1 - a_1, ..., 1 - a_m) \tag{1}$$

Then, it receives the complement-coded of input sample (A) to determine the similarity level between current input A and the j-th prototype node in f_2, as follows:

$$T_j = \frac{|A \wedge W_j|}{\alpha + |W_j|} \tag{2}$$

where $\alpha > 0$ is the learning parameter, $W_j \equiv (w_{j,1}, ..., w_{j,2M})$ is the weight vector of the j-th prototype node in f_2, and $\wedge$ indicates the fuzzy *and* operator:

$$(u \wedge v)_i \equiv min(u_i, v_i) \tag{3}$$

Fuzzy ART

The prototype node with the highest choice score is chosen as the winning node, denoted as node J, as follows:

$$T_J = max(T_j : j = 1, 2, ..., N) \tag{4}$$

If node J satisfies the vigilance criterion, resonance is said to occur.

$$\frac{|A \wedge W_J|}{|A|} > \rho \tag{5}$$

where ρ is the vigilance parameter of Fuzzy ART. However, if the condition in Eq. (5) is not satisfied, a mismatch occurs, whereby the selected node J is deactivated, and a new search cycle is triggered in Fuzzy ART to select a new winning node.

Fuzzy ART

This search cycle repeats until one of the existing prototype nodes satisfies the condition in Eq. (5), or a new prototype node is created in f_2 to encode the current input sample. Then, learning take place in order to update J-th weight vector (W_J) in f_2 as follows:

$$W_J^{(new)} = \beta(A \wedge W_J^{(old)}) + (1 - \beta) W_J^{(old)} \tag{6}$$

where β is learning rate of Fuzzy ART.

Fuzzy ART

Steps of Fuzzy ART:

For each training sample **do**

1. Do complement coding (Eq. 1).
2. Compute the choice value of all prototype nodes (Eq. 2).
3. Find the winner prototype node (Eq. 4).
4. Perform vigilance test (Eq. 5).
5. If the vigilance test is not satisfied, deactivate the winner prototype node,go to step 2 (select other prototype node or add new one).
6. Update the winner node (Eq. 10).

Fuzzy ART

Steps of Fuzzy ART:

For each training sample **do**

1. Do complement coding (Eq. 1).
2. Compute the choice value of all prototype nodes (Eq. 2).
3. Find the winner prototype node (Eq. 4).
4. Perform vigilance test (Eq. 5).
5. If the vigilance test is not satisfied, deactivate the winner prototype node,go to step 2 (select other prototype node or add new one).
6. Update the winner node (Eq. 10).

Fuzzy ART

Steps of Fuzzy ART:

For each training sample **do**

1. Do complement coding (Eq. 1).
2. Compute the choice value of all prototype nodes (Eq. 2).
3. Find the winner prototype node (Eq. 4).
4. Perform vigilance test (Eq. 5).
5. If the vigilance test is not satisfied, deactivate the winner prototype node,go to step 2 (select other prototype node or add new one).
6. Update the winner node (Eq. 10).

Fuzzy ART

Steps of Fuzzy ART:

For each training sample **do**

1. Do complement coding (Eq. 1).
2. Compute the choice value of all prototype nodes (Eq. 2).
3. Find the winner prototype node (Eq. 4).
4. Perform vigilance test (Eq. 5).
5. If the vigilance test is not satisfied, deactivate the winner prototype node,go to step 2 (select other prototype node or add new one).
6. Update the winner node (Eq. 10).

Fuzzy ART

Steps of Fuzzy ART:

For each training sample **do**

1. Do complement coding (Eq. 1).
2. Compute the choice value of all prototype nodes (Eq. 2).
3. Find the winner prototype node (Eq. 4).
4. Perform vigilance test (Eq. 5).
5. If the vigilance test is not satisfied, deactivate the winner prototype node,go to step 2 (select other prototype node or add new one).
6. Update the winner node (Eq. 10).

Fuzzy ART

Steps of Fuzzy ART:

For each training sample **do**

1. Do complement coding (Eq. 1).
2. Compute the choice value of all prototype nodes (Eq. 2).
3. Find the winner prototype node (Eq. 4).
4. Perform vigilance test (Eq. 5).
5. If the vigilance test is not satisfied, deactivate the winner prototype node,go to step 2 (select other prototype node or add new one).
6. Update the winner node (Eq. 10).

Outlines

Fuzzy ARTMAP

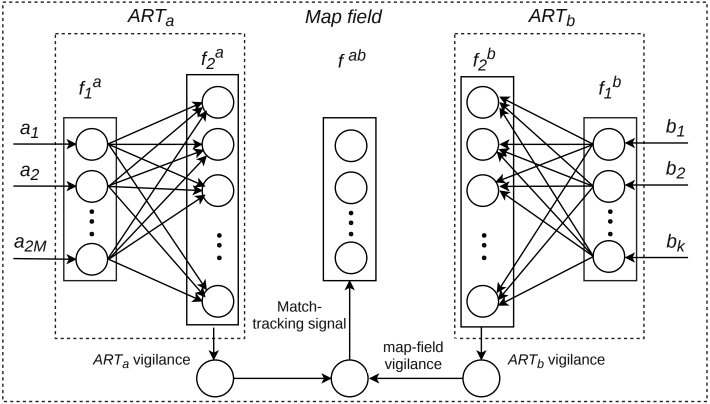


Figure 1: The structure of FAM.

Fuzzy ARTMAP

- Consists of two Fuzzy ART models,i.e., ART_a and ART_b, and a map field.
- ART_a and ART_b receive the complement-coded input sample (A) and its corresponding target class (B), respectively.
- Map field is used to map input samples into their corresponding outputs.

Fuzzy ARTMAP

- Consists of two Fuzzy ART models,i.e., ART_a and ART_b, and a map field.
- ART_a and ART_b receive the complement-coded input sample (A) and its corresponding target class (B), respectively.
- Map field is used to map input samples into their corresponding outputs.

Fuzzy ARTMAP

- Consists of two Fuzzy ART models,i.e., ART_a and ART_b, and a map field.
- ART_a and ART_b receive the complement-coded input sample (A) and its corresponding target class (B), respectively.
- Map field is used to map input samples into their corresponding outputs.

Fuzzy ARTMAP

When the winning nodes in both ART_a and ART_b are selected, the map-field vigilance test is applied as follows:

$$\frac{|y^b \wedge W_J^{ab}|}{|y^b|} > \rho_{ab} \tag{7}$$

where W_J^{ab} is the weight vector from f_2^a to f^{ab}, ρ_{ab} is the map-field vigilance parameter, and y^b indicates the output vector of f_2^b, which is defined as follows:

$$y^b = \begin{cases} 1, & k = K \\ 0, & otherwise \end{cases} \tag{8}$$

where K is the winning ART_b prototype node.

Fuzzy ARTMAP

If condition in Eq. (7) fails, it means the current J-th winning node in f_2^a makes an incorrect predicted class in ART_b. To correct this erroneous prediction, a match-tracking procedure is triggered to update the vigilance parameter of ART_a, as follows:

$$\rho_a = \frac{|A \wedge W_J^a|}{|A|} + \delta \tag{9}$$

where $\delta > 0$. Then, a new search cycle with the updated ρ_a setting ensues. This process ends once the map-field vigilance test is satisfied.

Fuzzy ARTMAP

As such, learning takes place in which the J-th node in f_2^a is updated to:

$$W_J^{a(new)} = \beta_a (A \wedge W_J^{a(old)}) + (1 - \beta_a) W_J^{a(old)} \tag{10}$$

where β_a is learning rate of ART_a.

Fuzzy ARTMAP

Algorithm 1 The learning phase of FAM

Input: Parameters of FAM and training samples
Output: Parameters of trained FAM

1: **for** each training sample $(a_1, ..., a_m)$ **do**
2:     Complement-coding (Eq.1).
3:     Calculate the choice function for all prototype nodes (W^a) (Eq. 2).
4:     Select the winning node using Eq. 3 (J-th node in W^a).
5:     Perform the vigilance test (Eq. 4) for the winning node.
6:     **while** the vigilance test is not satisfied **do**
7:         Deactivate the winning node ($T_j = 0$).
8:         **if** all prototype nodes in W^a are deactivated **then**
9:             Add new node.
10:         **end if**
11:         Go to Step 3.
12:     (The same cycle occurs for the target vectors to identify the winning prototype node (K-th), simultaneously).
13:     Perform the map-filed vigilance test (Eq. 7).
14:     **while** the condition in Eq. 7 is not satisfied **do**
15:         Perform match-tracking (Eq. 9).
16:         Deactivate the winning node (J-th) in W^a.
17:         Go to Step 3.
18:     Update the winning node using Eq. 10.
19: **end for**

Outlines

Numerical Example

Suppose in a tw0-class problem, FAM receives a sequence of input-output patterns, as follows:
a_1=[0.1, 0.2] belongs to b_1=c_1=[1]
a_2=[0.8, 0.4] belongs to b_2=c_2=[0]
a_3=[0.6, 0.7] belongs to b_3=c_2=[0]

The network parameters are set to their basic values:
α_a=0, ρ_a=0 and β_a=1.
Suppose there are two uncommitted nodes in F_2^a:
$W_1^a = W_2^a$=[1, 1, 1, 1] and $W_1^{ab} = W_2^{ab}$=[1, 1].

Numerical Example

Step 1: the complement-codded input vectors are as follows:
A_1=[0.1, 0.2, 0.9, 0.8] belongs to b_1=c_1=[1, 0]
A_2=[0.8, 0.4, 0.2, 0.6] belongs to b_2=c_2=[0, 1]
A_3=[0.6, 0.7, 0.4, 0.3] belongs to b_3=c_2=[0, 1]

Numerical Example

Input A_1:
Propagate A_1 to f_2^a. Since there is no previous learning, node J=1 is selected as winner.

$$\frac{|A_1 \wedge W_1|}{\alpha + |A_1|} = \frac{|[0.1, 0.2, 0.9, 0.8]| \wedge [1, 1, 1, 1]|}{0 + |2|} = 1 \geq \rho_a(Passed)$$

Map field activity: Since C_1 is the target class, y^b=[1, 0].
Map field vigilance test:

$$\frac{|y^b \wedge w_1^{ab}|}{|y^b|} = \frac{|[1, 0]| \wedge [1, 1]|}{|[0, 1]|} = 1 \geq \rho_{ab}(Passed)$$

Numerical Example

Learning:

$$W_1^{a(new)} = \beta_a(A \wedge W_1^{a(old)}) + (1 - \beta_a) W_1^{a(old)}$$

$$= 1 \times [0.1, 0.2, 0.9, 0.8] \wedge [1, 1, 1, 1] + 0 = [0.1, 0.2, 0.9, 0.8]$$

Input A_2:
Propagate A_2 to f_2^a.
Compute choice value:

$$T_1 = \frac{|A_2 \wedge W_1^a|}{\alpha + |W_1^a|} = \frac{1.1}{2} (Winner)$$

$$T_2 = \frac{|A_2 \wedge W_2^a|}{\alpha + |W_2^a|} = \frac{2}{4}$$

Numerical Example

$$\frac{|A_2 \wedge W_1^a|}{\alpha + |A_2|} = \frac{1.1}{2} \geq \rho_a \;\; (\textit{Passed})$$

Map field vigilance test: Since C_2 is target class, y^b= [0, 1].

$$\frac{|y^b \wedge w_1^{ab}|}{|y^b|} = \frac{|[0,1]| \wedge [1,0]|}{|[0,1]|} = 0 \leq \rho_{ab} \;\; (\textit{Faild})$$

Match tracking: Assume δ=0.0001, thus ρ_a is raised to:

$$\rho_a = \frac{|A_2 \wedge W_1^a|}{|A_2|} + \delta = 0.5501$$

Note that after increasing ρ_a, w_1^a fails:

$$\frac{|A_2 \wedge W_1^a|}{\alpha + |A_2|} = \frac{1.1}{2} \leq \rho_a \;\; (\textit{failed})$$

Numerical Example

In F_2^a, node J=1 is inhibited, it is de-activated and input A_2 is re-propagated to F_2^a, and node J=2 is selected as winner node.

$$\frac{|A_2 \wedge W_2^a|}{\alpha + |A_2|} = \frac{|[0.8, 0.4, 0.2, 0.6]| \wedge [1, 1, 1, 1]|}{0 + |2|} = 1 \geq \rho_a = 0.5501(Pass$$

Map field vigilance test: F_2^b output vector is still y^b= [0, 1].

$$\frac{|y^b \wedge w_2^{ab}|}{|y^b|} = \frac{|[0, 1]| \wedge [1, 1]|}{|[0, 1]|} = 1 \geq \rho_{ab} \ \ (passed)$$

Learning:

$$W_2^{a(new)} = \beta_a(A \wedge W_2^{a(old)}) + (1 - \beta_a) W_2^{a(old)} = [0.8, 0.4, 0.2, 0.6]$$

$$W_2^{ab} = [0, 1]$$

Numerical Example

Input A_3:
Propagate A_3 to f_2^a.
Compute choice value:

$$T_1 = \frac{|A_3 \wedge W_1^a|}{\alpha + |W_1^a|} = \frac{1}{2}$$

$$T_2 = \frac{|A_3 \wedge W_2^a|}{\alpha + |W_2^a|} = \frac{1.5}{2} \quad (\mathit{Winner})$$

Vigilance test:

$$\frac{|A_3 \wedge W_2^a|}{\alpha + |A_2|} = \frac{1.5}{2} \geq \rho_a = 0 \quad (\mathit{Passed})$$

Numerical Example

Map field vigilance test: Since C_2 is target class, y^b= [0, 1].

$$\frac{|y^b \wedge w_2^{ab}|}{|y^b|} = \frac{|[0,1]| \wedge [0,1]|}{|[0,1]|} = 1 \geq \rho_{ab} \quad (Passed)$$

Learning:

$$W_2^{a(new)} = \beta_a (A_3 \wedge W_2^{a(old)}) + (1 - \beta_a) W_2^{a(old)}$$

$$= [0.6, 0.7, 0.4, 0.3] \wedge [0.8, 0.4, 02, 0.6] = [0.6, 0.4, 0.2, 0.3]$$

W_2^{ab} unchanged.

Numerical Example

Therefore, two prototype nodes are created to learn A_1, A_2 and A_3, as follows:
W_1^a=[0.1, 0.2, 0.9, 0.8] , W_1^{ab}=[1,0] belongs to C_1
W_2^a= [0.6, 0.4, 0.2, 0.3], W_1^{ab}=[0,1] belongs to C_2

Test trained Fuzzy ARTMAP:
Assume a_4=[0.2, 0.3] belongs to C_1.
Therefore, A_4=[0.2, 0.3, 0.8, 0.7]

Numerical Example

Compute choice value:

$$T_1 = \frac{|A_4 \wedge W_1^a|}{\alpha + |W_1^a|} = \frac{1.8}{2} \quad (Winner)$$

$$T_2 = \frac{|A_4 \wedge W_2^a|}{\alpha + |W_2^a|} = \frac{1}{1.5}$$

Vigilance test:

$$\frac{|A_4 \wedge W_1^a|}{\alpha + |A_4|} = \frac{1.8}{2} \geq \rho_a = 0 \quad (Passed)$$

W_1^a is winner, which is belong to C_1. Therefore, the predicted class for A_4 is C_1 which is correct with its actual class.

Fuzzy ARTMAP

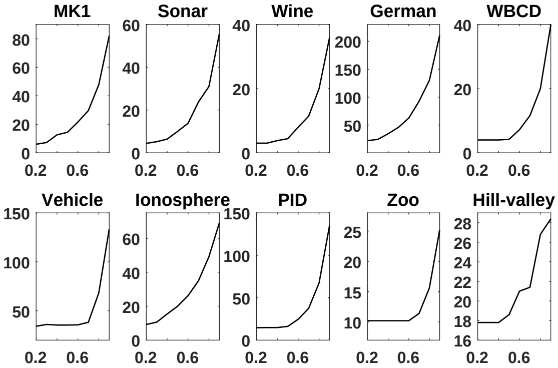
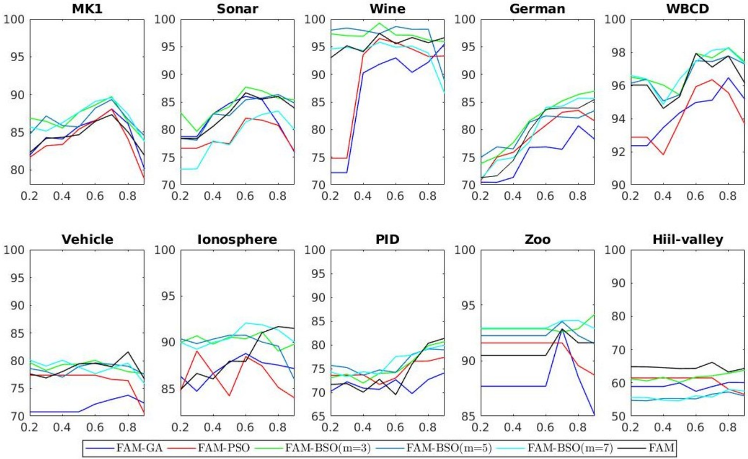


Figure 2: The number of created prototype nodes with different ρ_a setting.

Fuzzy ARTMAP



Fuzzy ARTMAP

- The **number of created prototype nodes** increases with increasing ρ_a from 0.1 to 1.
- The accuracy of FAM increases with increasing ρ_a from 0.1 to 1.

Fuzzy ARTMAP

- The **number of created prototype nodes** increases with increasing ρ_a from 0.1 to 1.
- The accuracy of FAM increases with increasing ρ_a from 0.1 to 1.

Outlines

Summary

- ART models are online learning ANNs with the capability of incremental learning.
- Fuzzy ART and Fuzzy ARTMAP (FAM) are two popular unsupervised and supervised learning models, respectively.

Summary

- ART models are online learning ANNs with the capability of incremental learning.
- Fuzzy ART and Fuzzy ARTMAP (FAM) are two popular unsupervised and supervised learning models, respectively.

Outlines

References

[1] G. A. Carpenter, S. Grossberg, **A massively parallel architecture for a selforganizing neural pattern recognition machine**, Computer vision, graphics, and image processing 37 (1) (1987) 54-115.

[2] G. A. Carpenter, S. Grossberg, D. B. Rosen, **Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system**, Neural networks 4 (6) (1991) 759-771.

[3] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds and D. B. Rosen, **Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps**, in IEEE Transactions on Neural Networks, vol. 3 (5) (1992) 698-713.

Thanks!