Lifelong Context Recognition Via Online Deep Feature Clustering

Combining Deep Learning with Adaptive Resonance

Sasha Petrenko, Donald C. Wunsch II,

Missouri University of Science and Technology
Electrical and Computer Engineering Department
Kummer Institute Center for Artificial Intelligence and Autonomous Systems

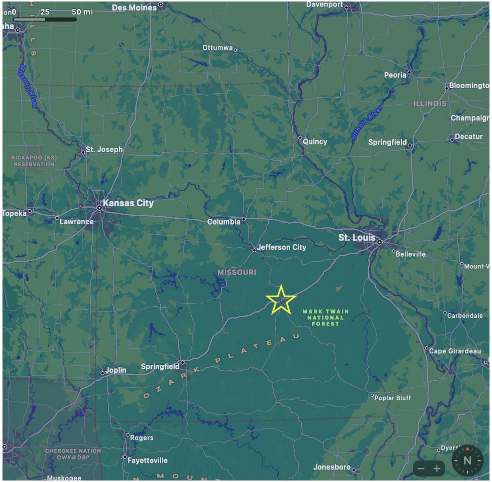
14th May, 2025

Overview

About

About

- Missouri University of Science and Technology
  - Missouri S&T - Rolla, MO
- Kummer Institute Center for Artificial Intelligence and Autonomous Systems (CAIAS)
  - **Director** - Dr. Donald C. Wunsch II
- **2016/19** - B.S./M.S. Aerospace Engineering
- **2024** - Ph.D. Computer Engineering
- S&T ECE/KICAIAS Assistant Research Professor



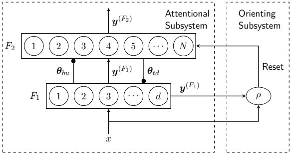
Executive Summary

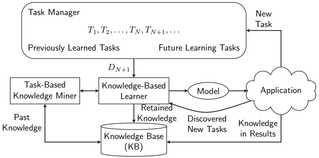
Papers

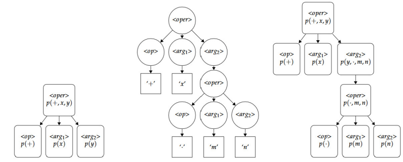
- **DeepART** - ART + Deep Local Learning
- **START** - ART + Symbols/Graphs
- **DFCCR** - ART + Deep Feature Clustering

Themes

- Combining Deep Learning and Adaptive Resonance Theory
- Adaptive Resonance Algorithms for Lifelong Machine Learning
- Novel Adaptive Resonance Algorithms and Architectures



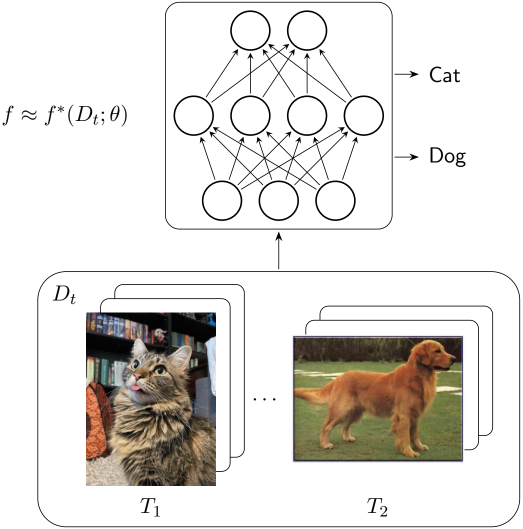




(a) START prototype before update. (b) Relation parse tree of $x + m \cdot n$. (c) START prototype after update.

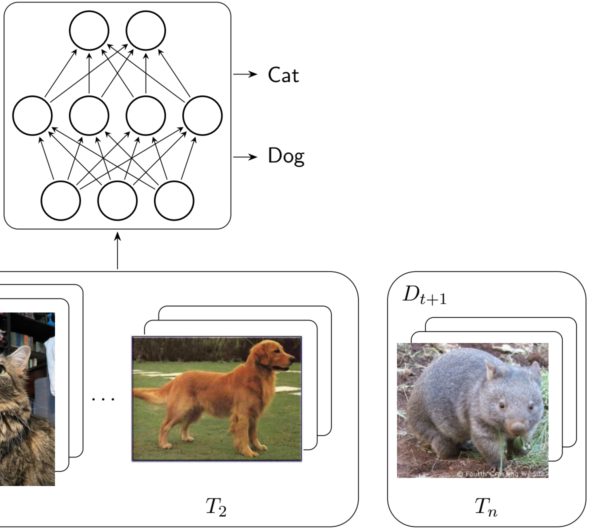
Table of Contents

Deep Learning and L2

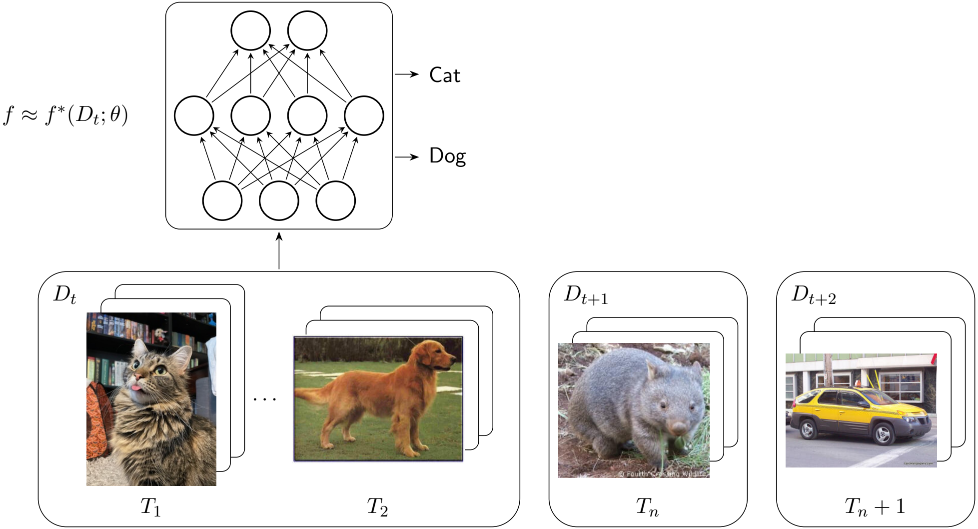


Deep Learning and L2

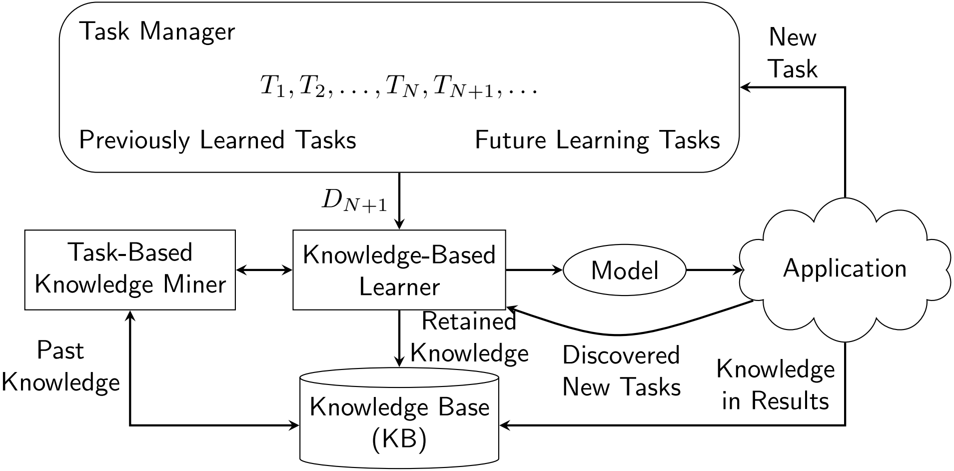
$f \approx f^*(D_t; \theta)$



Deep Learning and L2



L2 Diagram


Task Manager
$T_1, T_2, \ldots, T_N, T_{N+1}, \ldots$
Previously Learned Tasks
Future Learning Tasks
New Task
D_{N+1}
Task-Based Knowledge Miner
Knowledge-Based Learner
Model
Application
Retained Knowledge
Discovered New Tasks
Knowledge in Results
Past Knowledge
Knowledge Base (KB)

Catastrophic Forgetting and L2

L2

- Brute force insufficient
  - Real world is nonstationary, real engineered systems
  - Energy cost
  - Transferrability and democratization
- New model considerations
  - Add new classes?
  - Regularize?
  - Learning scenarios?

Catastrophic Forgetting

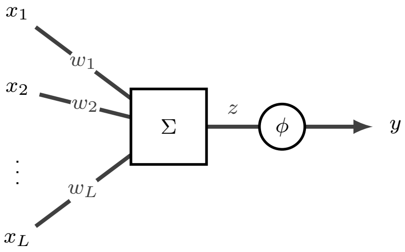
- Shared basis $f^*(\phi(x))$
- Nonstationary data
  - New classes
  - Shift in previous classes
- Stability-plasticity dilemma
  - Stable enough to retain previous knowledge
  - Plastic enough to learn new knowledge

Table of Contents

McCulloch-Pitts Neuron

Problems

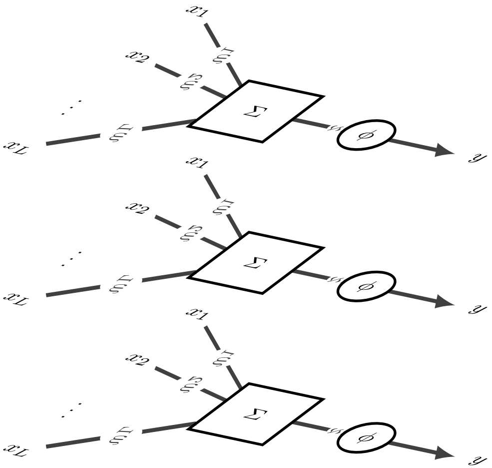
- Want to become pattern/feature detector/filter
- ...but which pattern?
  - Random starting weights?
  - Uniqueness with respect to neighbors?
  - **What if it is a moving target?**
- ...and how?
  1. **Someone tells me how to change**
  2. **I figure it out myself**



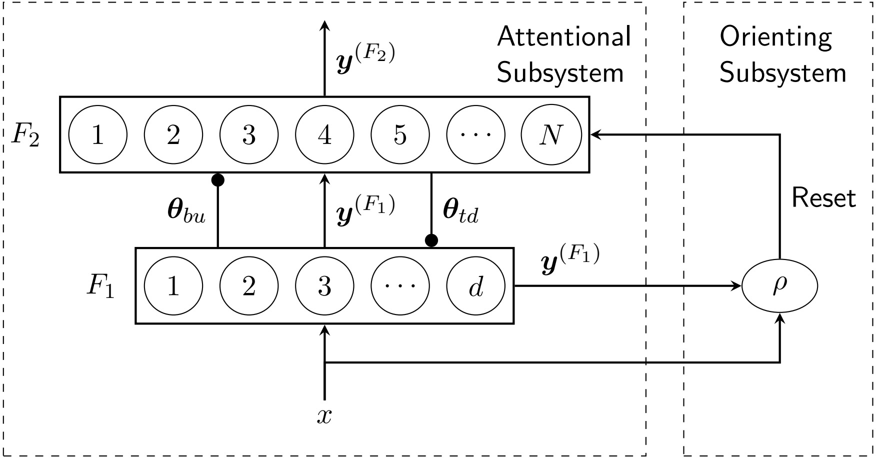
Multiple Neurons

Problems

- **Deep Learning**: Patterns of filters are useful as inputs for *other filters*
- We don't want to converge on the same filter pattern
  1. **Prescribed Order**: Weight assigning "oracle" knows each role
  2. **Natural Order**: Competition, "might-makes-right", "who gets to learn, and how much"?
- **Interference and stability**



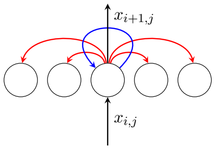
Adaptive Resonance

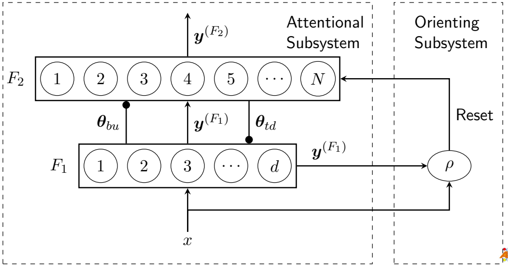


Adaptive Resonance

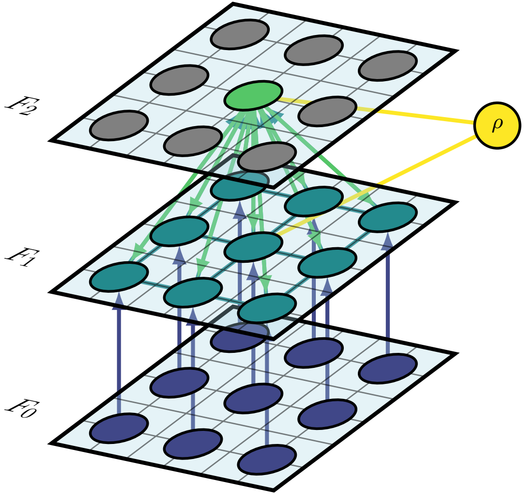
Optimal Assignment

- Stability-plasticity dilemma addressed through winner-take-all (WTA) competitive learning dynamics [1]
- Steady-state analysis of recurrent networks realizes activation (T) and match (M) functions, vigilance criterion, and WTA learning rule.
- **Fuzzy** combination $\mathbf{x} \wedge \mathbf{w}$ rather than inner product $\mathbf{x} \cdot \mathbf{w}$





FuzzyART: Visualized



$$F_2(F_1(F_0(\mathbf{x}))) = M_J = \frac{\|\mathbf{x} \wedge \mathbf{w}_J\|_1}{\|\mathbf{x}\|_1}$$

$$F_1(F_0(\mathbf{x})) = T_j = \frac{\|\mathbf{x} \wedge \mathbf{w}_j\|_1}{\alpha + \|\mathbf{w}_j\|_1}$$

$$F_0(\mathbf{x}) = [\bar{1} - \mathbf{x}, \mathbf{x}]$$

Table of Contents

stART: Overview

Overview

- *Analyzing Biomedical Datasets with Symbolic Tree Adaptive Resonance Theory*
- MDPI Information - Grossberg Special Issue
- Paper and Supplementary Materials:
  - MDPI Information (Open): `https://www.mdpi.com/2078-2489/15/3/125`
  - IEEE TechRxiv Preprint DOI: 10.36227/techrxiv.24542782

information MDPI

Article

Analyzing Biomedical Datasets with Symbolic Tree Adaptive Resonance Theory

Sasha Petrenko [1,*], Daniel B. Hier [1,2], Mary A. Bone [3], Tayo Obafemi-Ajayi [4], Erik J. Timpson [5], William E. Marsh [5], Michael Speight [5] and Donald C. Wunsch II [1]

1 Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA; hierd@mst.edu (D.B.H.); dwunsch@mst.edu (D.C.E.II)
2 Department of Neurology and Rehabilitation, University of Illinois at Chicago, Chicago, IL 60607, USA
3 Department of Science and Industry Systems, University of Southeastern Norway, 3616 Kongsberg, Norway; mary.bone@drmarybone.com
4 Engineering Program, Missouri State University, Springfield, MO 65897, USA; tayoobafemiajayi@missouristate.edu
5 Honeywell Federal Manufacturing & Technologies, Kansas City, MO 64147, USA; etimpson@kcnsc.doe.gov (E.J.T.); wmarsh@kcnsc.doe.gov (W.E.M.)
* Correspondence: petrenkos@mst.edu

Abstract: Biomedical datasets distill many mechanisms of human diseases, linking diseases to genes and phenotypes (signs and symptoms of disease), genetic mutations to altered protein structures, and altered proteins to changes in molecular functions and biological processes. It is desirable to gain new insights from these data, especially with regard to the uncovering of hierarchical structures relating disease variants. However, analysis to this end has proven difficult due to the complexity of the

stART

num	gene_locat	disease	disease_MI	gene	gene_MIM	inheritance	protein	uniprot	chromoson	chromoson	protein_cla	biologic_process
1	1p36.22	Charcot_M	609260	MFN2	608507	AD	Mitofusin_2	O95140	1	11980181_	Disease_rel	Apoptosis|Autophagy|Unfolded_protein_response
2	1p36.22	Charcot_M	617087	MFN2	608507	AD	Mitofusin_2	O95140	1	11980181_	Disease_rel	Apoptosis|Autophagy|Unfolded_protein_response
3	22q12.2	Charcot_M	616924	NEFH	162230	AD	Neurofilam	P12036	22	29480218_	Disease_rel	none
4	1p13.1	Charcot_M	618036	ATP1A1	182310	AD	ATPase_Na	P05023	1	11637266	Cancer_rel	Ion_transport|Potassium_transport|Sodium_transport|Sodium_pot
5	2p23.3	Charcot_M	618400	MPV17	137960	AR	Mitochond	P39210	2	27309492_	Disease_rel	none
6	7q11.23	Charcot_M	606595	HSPB1	602195	AD	Heat_shock	P04792	7	76302673_	Cancer_rel	Host_virus_interaction|Stress_response
7	1q23.2	Charcot_M	619519	CADM3	609743	AD	Cell_adhes	Q8N126	1	15917160	Predicted_i	Cell_adhesion
8	10q24.32	Charcot_M	606483	GBF1	603698	AD	Golgi_brefe	Q92538	10	102245371	Plasma_pro	Host_virus_interaction|Protein_transport|Transport
9	8q13_q23	Charcot_M	607731	CMT2H	607731	AR	none	none	none	none	none	none
10	20p12.2	Charcot_M	619574	JAG1	601920	AD	Jagged_can	P78504	20	10637684_	CD_marker	Notch_signaling_pathway
11	15q14	Charcot_M	620068	SLC12A6	604878	AD	Solute_carr	Q9UHW9	15	34229784_	Disease_rel	Ion_transport|Potassium_transport|Symport|Transport
12	8q21.11	Charcot_M	607831	JPH1	605266	AD|AR	Junctophili	Q9HDC5	8	74234700_	Human_dis	none
13	8q21.11	Charcot_M	607831	GDAP1	606598	AD|AR	Gangliosid	Q8TB36	8	74320613_	Disease_rel	none
14	12q24.23	Charcot_M	608673	HSPB8	608014	AD	Heat_shock	Q9UJY1	12	11917155	Disease_rel	Stress_response
15	19p13.2	Charcot_M	606482	DNM2	602378	AD	Dynamin_2	P50570	19	10718079_	Cancer_rel	Endocytosis|Phagocytosis
16	16q22.1	Charcot_M	613287	AARS1	601065	AD	Alanyl_tRN	P49588	16	70251983_	Disease_rel	Protein_biosynthesis
17	14q32.31	Charcot_M	614228	DYNC1H1	600112	AD	Dynein_cyt	Q14204	14	10196457	Disease_rel	Cell_cycle|Cell_division|Mitosis|Transport
18	9q33.3_q3	Charcot_M	614436	LRSAM1	610933	AD|AR	Leucine_ric	Q6UWE0	9	12745148	Disease_rel	Autophagy|Protein_transport|Transport|Ubl_conjugation_pathway
19	10p14	Charcot_M	615025	DHTKD1	614984	AD	Dehydroge	Q96HY7	10	12068954_	Disease_rel	Glycolysis
20	11q13.3	Charcot_M	616155	IGHMBP2	600502	AR	Immunoglo	P38935	11	68903863_	Disease_rel	Transcription|Transcription_regulation
21	3q25.2	Charcot_M	617017	MME	120520	AD|AR	Membrane	P08473	3	155024124	Cancer_rel	none
22	12q13.3	Charcot_M	616280	MARS1	156560	AD	Methionyl_	P56192	12	57475445_	Disease_rel	Protein_biosynthesis
23	17q21.2	Charcot_M	616491	NAGLU	609701	AD	N_acetyl_a	P54802	17	42536241_	Disease_rel	none
24	5q31.3	Charcot_M	616625	HARS1	142810	AD	Histidyl_tR	P12081	5	none	Disease_rel	Protein_biosynthesis
25	15q21.1	Charcot_M	616668	SPG11	610844	AR	SPG11_ves	Q96JI7	15	44562696_	Disease_rel	none
26	22q12.2	Charcot_M	616688	MORC2	616661	AD	MORC_fam	Q9Y6X9	22	30925130_	Disease_rel	Fatty_acid_metabolism|Lipid_metabolism
27	8q21.11	Charcot_M	607706	GDAP1	606598	AR	Gangliosid	Q8TB36	8	74320613_	Disease_rel	none
28	8q21.13	Charcot_M	618279	PMP2	170715	AD	Peripheral_	P02689	8	81440326_	Disease_rel	Transport
29	14q32.12	Charcot_M	619764	FBLN5	604580	AD	Fibulin_5	Q9UBX5	14	91869411_	Disease_rel	Cell_adhesion
30	12q23.3	Charcot_M	619742	POLR3B	614366	AD	RNA_polym	Q9NW08	12	10635774	Disease_rel	Antiviral_defense|Immunity|Innate_immunity|Transcription
31	6p21.31	Charcot_M	620111	ITPR3	147267	AD	Inositol_1_	Q14573	6	33620365_	Human_dis	Calcium_transport|Ion_transport|Transport
32	19p13.2	Charcot_M	606482	DNM2	602378	AD	Dynamin_2	P50570	19	10718079_	Cancer_rel	Endocytosis|Phagocytosis
33	1p35.1	Charcot_M	608323	YARS1	603623	AD	Tyrosyl_tR	P54577	1	32775237_	Disease_rel	Protein_biosynthesis
34	1q23.3	Charcot_M	607791	MPZ	159440	AD	Myelin_pro	P25189	1	16130473	Disease_rel	none
35	14q32.33	Charcot_M	614455	INF2	610982	AD	Inverted_fo	Q27J81	14	10468114	Disease_rel	none
36	3q26.33	Charcot_M	615185	GNB4	610863	AD	G_protein_	Q9HAV0	3	17939608	Disease_rel	none
37	8p21.2	Charcot_M	617882	NEFL	162280	AD|AR	Neurofilam	P07196	8	24950955_	Disease_rel	none
38	8q21.11	Charcot_M	608340	GDAP1	606598	AR	Gangliosid	Q8TB36	8	74320613_	Disease_rel	none

Figure 1: stART Biomedical Dataset: 81 Charcot-Marie-Tooth phylogenetic variants

⟨*S*⟩ ::= ⟨*attribute*⟩+ ;

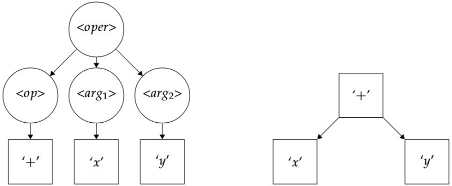
⟨*attribute*⟩ ::= ⟨*num*⟩ | ⟨*gene_location*⟩ | ⟨*disease*⟩ | ⟨*disease_MIM*⟩ | ⟨*gene*⟩ | ⟨*gene_MIM*⟩ | ⟨*inheritance*⟩+ | ⟨*protein*⟩ | ⟨*uniprot*⟩ | ⟨*chromosome*⟩ | ⟨*chromosome_location*⟩ | ⟨*protein_class*⟩+ | ⟨*biologic_process*⟩+ | ⟨*molecular_function*⟩+ | ⟨*disease_involvement*⟩+ | ⟨*MW*⟩ | ⟨*domain*⟩+ | ⟨*motif*⟩+ | ⟨*protein_location*⟩+ | ⟨*length*⟩ | ⟨*disease_MIM2*⟩ | ⟨*phenotype*⟩+ | ⟨*weight_tag*⟩ | ⟨*length_tag*⟩ ;

⟨*phenotype*⟩ ::= 'ataxia' | 'atrophy' | 'auditory' | 'autonomic' | 'behavior' | 'cognitive' | 'cranial_nerve' | 'deformity' | 'dystonia' | 'gait' | 'hyperkinesia' | 'hyperreflexia' | 'hypertonia' | 'hypertrophy' | 'hyporeflexia' | 'hypotonia' | 'muscle' | 'pain' | 'seizure' | 'sensory' | 'sleep' | 'speech' | 'tremor' | 'visual' | 'weakness' ;

⟨*biologic_process*⟩ ::= 'Apoptosis' | 'Mitosis' | 'Lipid_metabolism' | 'Symport' | 'Ubl_conjugation_pathway' | 'Glycolysis' | 'Glucose_metabolism' | 'Ion_transport' | 'Unfolded_protein_response' | 'Cell_division' | 'DNA_repair' | 'Cell_adhesion' | 'Notch_signaling_pathway' | 'Protein_biosynthesis' | 'Stress_response' | 'Endocytosis' | 'Transcription' | 'Sodium_potassium_transport' | 'Transcription_regulation' | 'Fatty_acid_metabolism' | 'Host_virus_interaction' | 'Antiviral_defense' | 'Lipid_degradation' | 'Autophagy' | 'Sodium_transport' | 'Immunity' | 'none' | 'Protein_transport' | 'Nucleotide_biosynthesis' | 'Calcium_transport' | 'Transport' | 'Phagocytosis' | 'Inflammatory_response' | 'DNA_damage' | 'Potassium_transport' | 'Carbohydrate_metabolism' | 'Cell_cycle' | 'Innate_immunity' ;

Figure 2: stART EBNF grammar production rules example

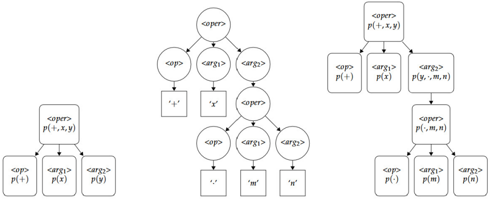
stART



(a) START relation parse tree TreeNode.

(b) Gram-ART syntax tree TreeNode.

Figure 3: stART Relation Parse Syntax Trees for $x + y$.



(a) START prototype before update.

(b) Relation parse tree of $x + m \cdot n$.

(c) START prototype after update.

Figure 4: stART Statements and Prototypes for pretrained $x + y$ prototype updated with $x + m \cdot n$

stART

Algorithms

- **stART**: Symbolic Tree Adaptive Resonance Theory
- Six variants:
  - stART, Dual-Vigilance stART, Distrubuted Dual-Vigilance stART
  - Supervised modification (Lifelong Context Recognition)

Algorithm 1: START algorithm. A set of symbolic statements under a formal context-free grammar are parsed into their syntax trees. Prototypes are defined as learning dynamics otherwise follow the activation, competition, match, update, and initialization rules of unsupervised ART algorithms [19]. ART dynamics notation here largely follow the elementary ART algorithm outlined in [19]. Inference during classification follows the same match rule dynamics without the instantiation of new categories; in the case of complete mismatch, either an "unknown" label or the best matching unit (the category that maximizes the match criterion) may be returned. Please see Table 1 for full notation.

Data: Symbolic statements $\mathbf{S}$; CFG grammar $\mathbf{G}$ with terminal symbols $\mathbf{T}$, non-terminal symbols $\mathbf{N}$, production rules $\mathbf{P}$, and statement entry symbol $\mathcal{S}$.
Result: Cluster labels $\mathbf{Y} \in \mathbb{N}^n$

`/* Parse statements into constituency parse trees */`
1 $\mathbf{X} \leftarrow \text{Parser}(\mathbf{S}, \mathbf{G})$
`/* Iteration over parsed statement trees */`
2 **foreach** $\mathbf{x} \in \mathbf{X}$ **do**
  `/* Compute activations for all nodes */`
3   $T_j \leftarrow f_T(\mathbf{x}, \mathcal{R}_j), \forall j \in \mathcal{C}$
  `/* Perform WTA competition for active nodes */`
4   $J \leftarrow \arg\max_{j \in \Lambda} (T_j)$
  `/* Compute match for the winning category */`
5   $M \leftarrow f_M(\mathbf{x}, \mathcal{R}_J)$
  `/* Vigilance test */`
6   **if** $M > \rho$ **then**
    `/* Update category */`
7     $\mathcal{R}_J \leftarrow f_L(\mathbf{x}, \mathcal{R}_J)$
8   **else**
    `/* Deactivate category */`
9     $\Lambda \leftarrow \Lambda - \{J\}$
10   **if** $\Lambda \neq \emptyset$ **then**
      `/* Continue match search */`
11   Goto Line 4
12   **else**
      `/* Create and initialize new category */`
13       $K \leftarrow \|\mathcal{C}\|_1 + 1$
14       $\mathcal{R}_K \leftarrow f_N(\mathbf{x}, \mathbf{G})$

stART

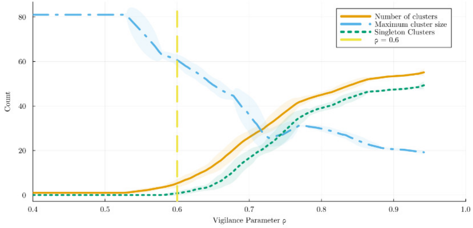


Figure 3. Effect of vigilance parameter ρ on number of clusters. A Monte Carlo of shuffled sample presentation order was run to generate 1σ intervals of the results at each vigilance parameter value. As ρ was increased from 0.0 to 1.0, the maximum cluster size decreased, the number of clusters increased, and the number of singleton clusters increased. A value of $\rho = 0.6$ (yellow dashed line) was selected to yield 9 clusters with only two singleton clusters. Larger ρ values gave too many singleton clusters, and smaller ones put too many cases into one cluster.

Figure 6: stART Parameter Sweep

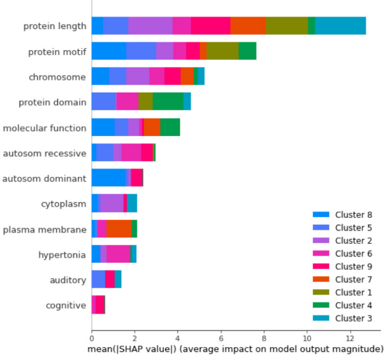


Figure 12. SHAP cluster summary plot for the 9 clusters derived from CMT dataset with $\rho = 0.6$. The SHAP plot shows which features contributed the most to the cluster configuration by cluster. Important features are protein length, chromosome, mode of inheritance (autosomal dominant and recessive), protein location (cytoplasm and plasma membrane), and certain phenotypes (auditory, cognitive, and hypertonia). The domain expert rated these features as highly biologically plausible. SHAP plots were created using the method of Lundberg et al. [68].

Figure 7: stART SHAP Analysis

Table of Contents

Lifelong Context Recognition: Overview

Overview

1. *Lifelong Context Recognition via Deep Feature Clustering*
2. Components
   (a) ARTSCENE [2]
   (b) YOLOv10 [3]
   (c) DDVFA [4]
   (d) Simplified FuzzyARTMAP [5]

Lifelong Context Recognition via Online Deep Feature Clustering

Sasha Petrenko[a,*] (Researcher), Andrew Brna[b,**] (Researcher), Mario Aguilar-Simon[b] (Researcher) and Donald C. Wunsch II[a] (Researcher)

[a]Missouri University of Science and Technology, 1201 N State St, Rolla, 65401, MO, USA
[b]Teledyne Scientific, 5001 S Miami Blvd, Durham, 27703, NC, USA

ARTICLE INFO

Keywords:
Adaptive resonance theory
deep learning
lifelong machine learning
online clustering
context recognition

ABSTRACT

Context recognition for lifelong learning (L2) agents is an open-ended problem whereby aggregate features in an environment are utilized to signal the active context in which the agent is operating. The ability to recognize context is necessary in L2 agents to engage modulatory signals to account for significant changes in the input state space associated with a given context or task, such as altering learning dynamics or shifting attention to more relevant features. Context recognition is itself an L2 problem due to the ever-increasing number of distinct contexts that an agent might encounter, requiring incrementally learning novel contexts while prescribing them to supervised task labels when available. This paper demonstrates an algorithm based on clustering of deep-extracted features with adaptive resonance theory methods that satisfies these requirements on the behalf of an embodied L2 agent in a computer vision environment. The strength of this algorithm lies in its flexibility, being capable of online incremental learning in supervised, realistic semi-supervised, and unsupervised scenarios while demonstrating continual learning in its own right.

1. Introduction

Context recognition is an open-ended problem that concerns attentional and adaptation mechanisms for lifelong learning (L2) agents. Within an embodied lifelong learning agent, the objective of context recognition is to identify patterns in an environment or underlying problem structure that should influence task execution. This would prompt appropriate situational changes in the agent's operation, such as shifting attention to context-relevant features or switching to more context-appropriate goals. Such functionality is strictly necessary in lifelong learning agents to allow them to adapt to an ever-expanding set of environments or tasks. This work demonstrates an algorithm designed to perform online, multimodal context recognition for an L2 agent with context-specific tasking on a visual-wavelength computer vision problem.

*This work was supported by the Air Force Research Laboratory (AFRL)

Lifelong learning algorithms and the agents that utilize them are tasked with learning on an indeterminate number of tasks in a sequential manner Kudithipudi, Aguilar-Simon, Babb, Bazhenov, Blackiston, Bongard, Brna, Chakravarthi Raja, Cheney, Clune, Daram, Fusi, Helfer, Kay, Ketz, Kira, Kolouri, Krichmar, Kriegman, Levin, Madireddy, Manicka, Marjaninejad, McNaughton, Miikkulainen, Navratilova, Pandit, Parker, Pilly, Risi, Sejnowski, Soltoggio, Soures, Tolias, Urbina-Meléndez, Valero-Cuevas, van de Ven, Vogelstein, Wang, Weiss, Yanguas-Gil, Zou and Siegelmann (2022). Accomplishing this necessitates more intelligent methods of both storing and utilizing accumulated knowledge. Towards this end, Chen and Liu Chen and Liu (2018) describe several key characteristics for L2 algorithms, which summarize an L2 algorithm as one that continually learns on the job, augmenting future learning with existing knowledge and mitigating catastrophic forgetting, all while being self-guided in its discovery of new tasks and robustly managing its own

ARTSCENE: Gist

ARTSCENE: Architecture

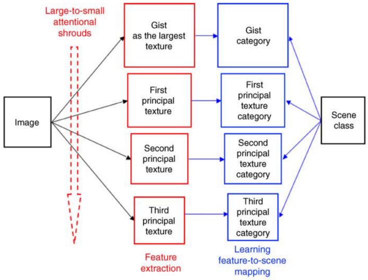


Figure 8: Supervised Scene Learning

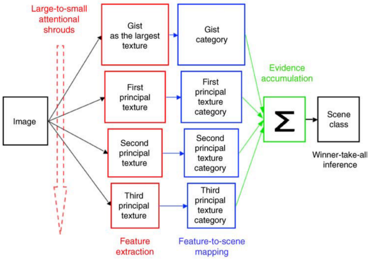


Figure 9: Inference

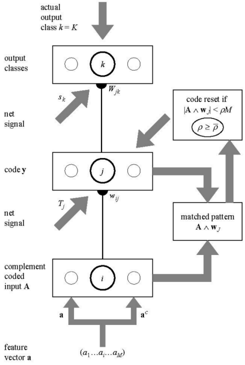


Figure 10: Default ARTMAP 2 [6]

ARTSCENE: Filters

Stages

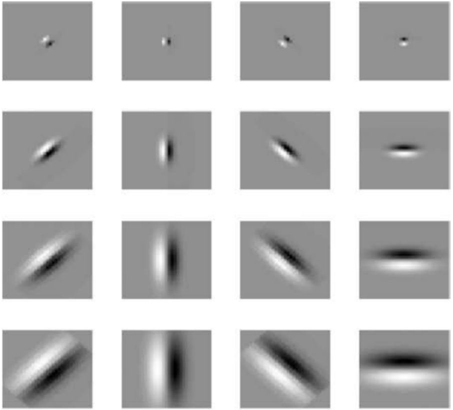
1. Color-to-gray image transformation
2. Contrast normalization
3. Contrast-sensitive oriented filtering
4. Contrast-insensitive oriented filtering
5. Oriented competition at the same position
6. Gist feature vector
7. Attentional shrouds

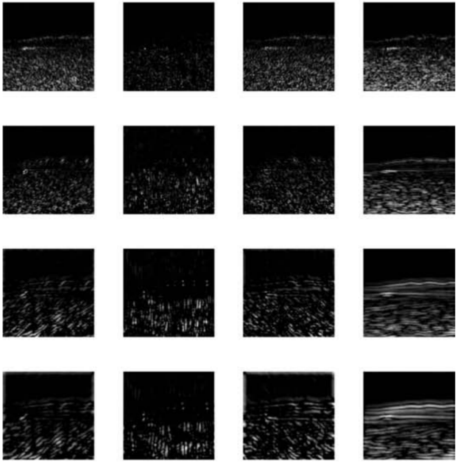
$$O_k^{\pi g} = \frac{1}{|\pi|} \sum_{(i,j)\in\pi} Z_{ijk}^g$$

$$C^{\pi\omega} = \frac{1}{|\pi|} \sum_{(p,q)\in\pi} I_{p,q}^\omega$$

$$G = \left(\frac{O_k^{\pi g}}{\sum_{\ell=1}^4 O_\ell^{\pi g}}, \frac{C^{\pi\omega}}{\sum_{v\in\{R,G,B\}} C^{\pi v}} : k \in \{1,2,3,4\}; \omega \in \{R,G,B\}; \pi \in \{1,\ldots,16\} \right)$$

ARTSCENE: Filters

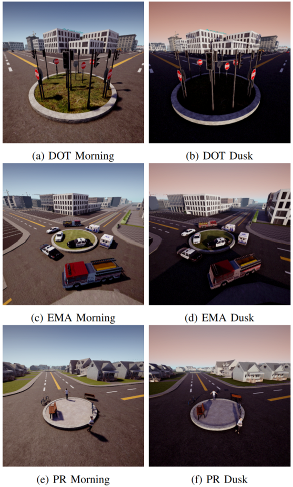




Method

Experimental Method

- **Multi-modal deep feature clustering approach**
- AirSim: Custom environment, assets, simulate multiple flight paths
- Extract full mid-layer activations with pretrained YOLOv10
  - Multi-object classifier
  - Deep CNN
- Compute average activations along pixel dimensions
- Z-score feature statistics on training dataset
- Normalize, sigmoid, DDVFA clustering
- Modulated, label-mapped Simplified FuzzyARTMAP module for supervised



(a) DOT Morning (b) DOT Dusk
(c) EMA Morning (d) EMA Dusk
(e) PR Morning (f) PR Dusk

Method

Experiments

1. **Randomized Sampling**: full shuffle, train/test
2. **Lifelong Learning Condensed Scenario Permutations**:
   (a) Train T1, test T1, train T2, test T1 and T2, train T3, etc.
   (b) Permute $(6! = 720)$
3. **Semi-Supervised Training**: Train 80% supervised, train 20% unsupervised

Results: Monte Carlo

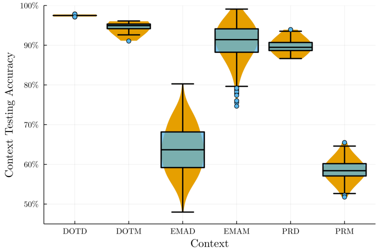


Figure 11: Shuffled Samples Performance

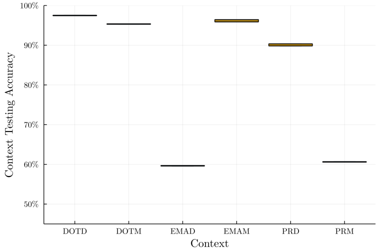


Figure 12: Permuted Contexts Performance

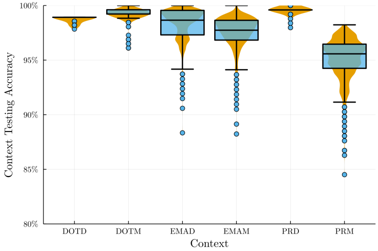


Figure 13: Semi-Supervised Performance

Results: Single L2 Condensed Scenario

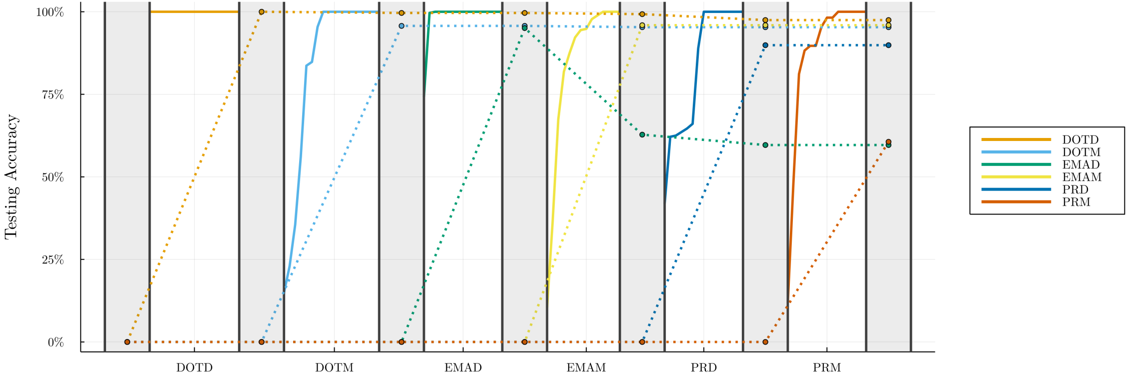


Figure 14: A single lifelong learning scenario of alternating training and testing blocks. Gray areas are Experience Blocks (EBs), where performance is tested for each class. White areas denote Learning Blocks (LBs). Solid curves during learning blocks indicate learning validation results during training. Dotted lines interpolated between EBs indicate the global trends of performance per class as new classes are sequentially introduced

Table of Contents

DeepART

Overview

- *Deep Adaptive Resonance*
- Combining deep learning and ART
- Challenges:
  - Deep learning is model of architectures, not necessarily learning rules (i.e. gradient descent with backpropagation, loss penalties such as EWC)
  - Deep Hebbian literature

DeepART: Deep Gradient-Free Local Learning with Adaptive Resonance

Sasha Petrenko, *Member, IEEE*, Leonardo Enzo Brito da Silva, *Member, IEEE*, and Donald C. Wunsch II, *Fellow, IEEE*

***Abstract*—This article presents DeepART, a new, gradient-free technique for the training of deep Hebbian neural networks using the dynamics of Adaptive Resonance Theory (ART) algorithms. In the presented model, layers of a deep neural network are interpreted as modified Fuzzy ART modules with complement coded inputs and soft winner-take-all learning that are simultaneously updated with a Fuzzy ARTMAP module providing feature-category-label mapping to enable supervised learning. Moreover, local weight update rules are derived for fully-connected and convolutional layers. DeepART provides a significant performance boost, reduction of category proliferation, and subsequent improved scalability to comparable ART-based methods processing high-dimensional datasets in lifelong learning scenarios, because it combines the nonlinear feature representation learning of deep neural networks with the lifelong learning properties of ART algorithms. This paper is thus a primer on how to exchange deeper networks for other ART models while still retaining the properties of ART throughout the architecture.**

***Index Terms*—Adaptive resonance theory, deep Hebbian learning, lifelong machine learning, task-incremental learning**

I. Introduction

DEEP neural networks trained with backpropagation have achieved wonders in the realm of parametric model design for difficult computer vision and natural language processing tasks [1], [2], [3], [4], [5], [6], demonstrating the power and usefulness of hierarchical, function-composed models. Nonetheless, backpropagation is a fragile design decision for the training of deep neural network architectures, suffering from three broad categories of flaws: sample efficiency (deep models require very many iterations on very large datasets the size of the networks, increasing the power and compute costs of training deep models on such large datasets, and ultimately probing pretrained models *a posteriori* to decipher their black-box reasoning.

This motivates research into alternative learning rules for the training of deep networks with biologically-plausible methods using the self-organizing principles that biology has discovered to solve relevant problems such as the stability-plasticity dilemma; such techniques include the use of local Hebbian rules [9], [10], [11], [12], [13], competitive local learning rules [14], [15], and even hybrid learning rules [16], [17], [18]. Adaptive Resonance Theory (ART) algorithms are a class of prototype-based biologically-inspired [19] neural networks designed to optimally address the stability-plasticity dilemma in unsupervised scenarios [20], [21], but they suffer from their own memory complexity due to the potential for category proliferation and are not designed for deep hierarchical models.

This article demonstrates a novel local weight update methodology, named DeepART, that combines the hierarchical feature representation learning of deep networks with the lifelong learning capabilities of ART algorithms for the incremental multimodal learning of both fully-connected multilayer perceptrons (MLP) and convolutional deep neural networks (CNN). This method addresses the memory scalability of ART-based systems by mitigating their category proliferation when handling complex high dimensional data in lifelong learning scenarios while also circumventing the scalability issues of gradient-based neural network optimization by using incremental local weight update rules. This method is derived in the space of deep Hebbian learning, and the method is

Problem Statement

Problems

- Catastrophic forgetting
- Stability-plasticity dilemma
- Deep loss landscape fragility
- Non-local learning with gradients
- ART optimal category learning, assuming a stable, complete feature space

Objectives

- Manifold learning from deep architectures
- Lifelong learning
  - Performance maintenance
  - Forward transfer
  - Backward transfer
- **Combine ART with feature learning**

Deep Hebbian Learning

- No gradients
- Local learning
- $\Delta \mathbf{w} = f(\mathbf{x}, \mathbf{w}, y)$
- **Incremental learning**

Adaptive Resonance Theory

- Prototype-based
- WTA learning
- Neurogenesis
- **Incremental learning**

Problem Statement

Distillation

1. Multi-layer, function-composed ART algorithm
2. MLP and CNN derivations
3. Supervised modification
4. Stability/convergence criterion

Table of Contents

Learning Rules

Backpropagation

$$\Delta \mathbf{w}_{ij} = -\eta \frac{\delta e}{\delta \mathbf{w}_{ij}}$$

$$e = \mathcal{L}(t, y)$$

- Pros:
  1. Statistical/optimization guarantees
- Cons:
  1. Unexplainable
  2. Data/energy inefficient

Hebbian

$$\Delta \mathbf{w}_{ij} = f\left(y_j, \mathbf{x_{ij}}, \mathbf{w}_{ij}\right)$$

1. Pros:
   - Explainable
2. Cons:
   - Weight instability
   - Limited research with deep networks

Local Learning Rules

Notation

$$y_{ij} = \mathbf{w}_{ij}\mathbf{x}_j$$
$$z_{ij} = \phi\left(\mathbf{w}_{ij}\mathbf{x}_j\right)$$

- i: layer index, n layers
- j: neuron index, m neurons
- y_{ij}: neuron output
- t_{ij}: neuron target
- $\mathbf{w}_{ij}$: weight vector
- $\mathbf{x}_j$: input vector
- ϕ: (nonlinear) transfer function

Rule	$\Delta\mathbf{w}$
Unsupervised	
Hebb's	$\eta y \mathbf{x}$
Oja's	$\eta \cdot y\left(\mathbf{x} - y\mathbf{w}\right)$
Instar	$\eta \cdot y\left(\mathbf{x} - \mathbf{w}\right)$
Supervised	
Widrow-Hoff	$\eta\left(t - y\right)\mathbf{x}$

Table 1: Local neuron update rules. Shorthand: $y = y_{ij}$, $\mathbf{x} = \mathbf{x}_j$, $\mathbf{w} = \mathbf{w}_{ij}$

Local Learning Rules

What about η?

- Varying η learning rules
- $\eta = \beta$ in ART literature
- DeepART revelation:
  1. Each category gets a different feature vector $\mathbf{x}$
  2. Each category shares the same feature vector $\mathbf{x}$ but compete at the level of β

Rule	β_j	
	$j = J$	$j \neq J$
WTA	1	0
SOM	1	h_{Jj}
Softmax	$\sigma(\mathbf{y})$	$\sigma(\mathbf{y})$
Contrast	1	$-\sigma(\mathbf{y})$

Table 2: Local neuron learning rate rules.

Table of Contents

Learning Rule Derivation

Derivation

Hebb	$\Delta w = \eta y x$	(1)
Layer	$\Delta \mathbf{W} = \eta \mathbf{y} \mathbf{x}^T$	(2)
Decay	$\Delta w = \eta y x - \gamma$	(3)
Decay Ratio	$\gamma = \eta y w$	(4)
Instar	$\Delta w = \eta y \left(x - w\right)$	(5)

FuzzyART

Activation	$T_j = \dfrac{\lVert \mathbf{x} \wedge \mathbf{w}_j \rVert_1}{\alpha + \lVert \mathbf{w}_j \rVert_1}$	(6)
Match	$M_J = \dfrac{\lVert \mathbf{y}^{(F_1)} \rVert_1}{\lVert \mathbf{x} \rVert_1} = \dfrac{\lVert \mathbf{x} \wedge \mathbf{w}_J \rVert_1}{\lVert \mathbf{x} \rVert_1}$	(7)
Update	$\mathbf{w}_J \leftarrow \left(1 - \beta\right) \mathbf{w}_J + \beta (\mathbf{x} \wedge \mathbf{w})$	(8)
Delta Rule	$\Delta \mathbf{w}_J = \beta \left(\left(\mathbf{x} \wedge \mathbf{w}_J \right) - \mathbf{w}_J \right)$	(9)

Modifications

Forward
$$\mathbf{x}_i = CC\left(\mathbf{y}_{i-1}\right) \tag{10}$$
$$= CC\left(\phi\left(Z_i\left(\|\mathbf{x_{i-1}} \wedge \mathbf{w}_{i,j}\|_1\right)\right)\right) \tag{11}$$

Sigmoid
$$\phi(z) = \frac{1}{1+e^{-z}} \tag{12}$$

Update
$$\Delta\mathbf{w}_{i,J} = \beta_{i,J}\left(\left(\mathbf{x}_i \wedge \mathbf{w}_{i,J}\right) - \mathbf{w}_{i,J}\right) \tag{13}$$

Learn Rate
$$\beta_{i,J} = \beta_d\sqrt{m_i} \tag{14}$$

Convolution
$$\mathbf{x}_i = CC\left(\phi\left(Z_i\left(\mathbf{W}_{i-1}\hat{*}\mathbf{x}_{i-1}\right)\right)\right) \tag{15}$$

Supervised Component

Option 1: **FuzzyARTMAP**

- Feature - category - label map

Option 2: **Leader Neuron**

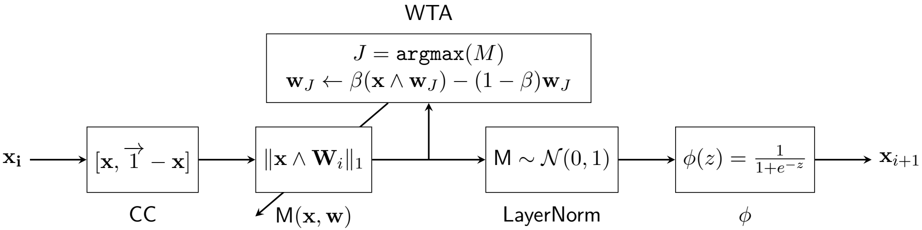
Target: $1_\omega(j) = \begin{cases} 1 & \text{if } y_{L,j} = \omega \\ 0 & \text{if } y_{L,j} \neq \omega \end{cases}$ (16)

Widrow-Hoff: $\Delta\mathbf{w} = \eta\left(t-y\right)\mathbf{x}$ (17)

Algorithm

DeepART

- Deep, nonlinear function-composed hierarchical model
- Alternating normalization (F0), complement-coding (F1), and competition layers (F2)
- Local weight update rule, **no gradients**
- Default unsupervised, FuzzyARTMAP head supervised
- Feedforward MLP and CNN variations



Stability

Convergence Criterion

- Cohen-Grossberg theorem for shunting network component
- ART1, FuzzyART: winning activation monotonically increases
- Hypothesis:
  - Monotonic decrease in $M_{i,J}$ effect nullified by normalization
  - Filter specialization and stability driven by convergence to **fuzzy subset** of initial weights

$$\mathbf{x}_i \leftarrow [\mathbf{1} - \mathbf{x}_i, \mathbf{x}_i] \tag{18}$$

$$M_{i,J} = \frac{\|\mathbf{x}_i \wedge \mathbf{w}_{i,J}\|_1}{\|\mathbf{x}_i\|_1} \tag{19}$$

$$= \frac{\|\mathbf{x}_i \wedge \mathbf{w}_{i,J}\|_1}{2d_i} \tag{20}$$

$$\Delta \mathbf{w}_{i,J} = \beta_{i,J} \left((\mathbf{x}_i \wedge \mathbf{w}_{i,J}) - \mathbf{w}_{i,J} \right) \tag{21}$$

$$\frac{dM_{i,J}}{d\Delta \mathbf{w}_{i,J}} \leq 0 \tag{22}$$

Table of Contents

Experimental Setup

Dataset	Feature Dimension	# Train	# Test	# Total	# Classes	# Tasks × # Classes per Task
MNIST	$28 \times 28 \times 1$	$60,000$	$10,000$	$70,000$	10	5×2
Fashion MNIST	$28 \times 28 \times 1$	$60,000$	$10,000$	$70,000$	10	5×2
CIFAR-10	$32 \times 32 \times 3$	$50,000$	$10,000$	$70,000$	10	5×2
CIFAR-100 (Fine)	$32 \times 32 \times 3$	$50,000$	$10,000$	$60,000$	100	5×20
CIFAR-100 (Coarse)	$32 \times 32 \times 3$	$50,000$	$10,000$	$60,000$	20	5×4
USPS	$16 \times 16 \times 1$	$7,281$	$2,007$	$9,298$	10	5×2

Table 3: Datasets

Hyperparameter	Value
ρ	0.6
β_s	1.0
β_d	0.1
η	1×10^{-3}
α	1×10^{-7}
FuzzyARTMAP F1 Dimension	m

Table 4: Hyperparameters

Layer	**Size**	**Transfer Function**
CC	$\lvert \mathbf{x}_0 \rvert$	Complement Code
Dense	512	sigmoid
Normalize	-	LayerNormalize
CC	-	Complement Code
Dense	256	sigmoid
Normalize	-	LayerNormalize
CC	-	Complement Code
Dense	784	sigmoid

Table 5: Dense network

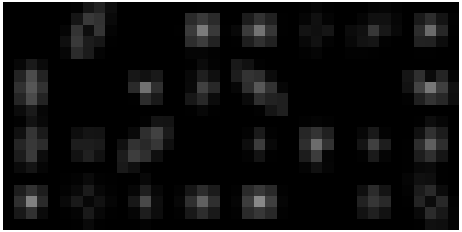
Layer	**Size**	**Transfer Function**
Conv2D	$3 \times 3 \times 2 \times 8$	Sigmoid
MaxPool	2×2	-
Normalize	-	LayerNormalize
CC	-	Complement Code
Conv2D	$5 \times 5 \times 2 \times 16$	Sigmoid
MaxPool	4×4	-
Flatten	-	-
Normalize	-	LayerNormalize
CC	-	Complement Code
Dense	784	Sigmoid

Table 6: Convolutional network

Filter Visualization

ST vs. MT

- **ST**: Single task
  - Train/test
- **MT**: Multi task
  - L2 condensed scenario
  - Train T_1, test T_1
  - Train T_2, test T_1, T_2,
  - Train T_3, test T_1, T_2, T_3
  - ...



Performances

Method	Dataset			
	MNIST	Fashion MNIST	USPS	CIFAR-10
FuzzyARTMAP (ST)	0.7239 ± 0.0401	0.6109 ± 0.0214	0.7010 ± 0.0421	0.1891 ± 0.0131
FuzzyARTMAP (MT)	0.6886 ± 0.0443	0.5751 ± 0.0379	0.6648 ± 0.0288	0.1767 ± 0.0155
MLP DeepART (ST)	0.7027 ± 0.1231	0.6434 ± 0.0604	0.8292 ± 0.0213	$\mathbf{0.2196 \pm 0.0120}$
MLP DeepART (MT)	0.5650 ± 0.2667	0.5399 ± 0.1643	0.8244 ± 0.0340	0.2138 ± 0.0136
CNN DeepART (ST)	$\mathbf{0.7529 \pm 0.0592}$	$\mathbf{0.6616 \pm 0.0282}$	$\mathbf{0.8545 \pm 0.0132}$	0.2192 ± 0.0363
CNN DeepART (MT)	0.7309 ± 0.0930	0.6082 ± 0.1219	0.8405 ± 0.0536	0.2066 ± 0.0247

Category Analysis

Method	Dataset			
	MNIST	Fashion MNIST	USPS	CIFAR-10
FuzzyARTMAP (ST)	841.4 ± 746.9	2474.7 ± 2308.5	755.4 ± 410.3	1575.5 ± 512.7
FuzzyARTMAP (MT)	972.8 ± 904.9	2724.7 ± 2611.7	829.4 ± 472.5	884.3 ± 278.7
MLP DeepART (ST)	118.8 ± 75.1	187.8 ± 125.3	45.8 ± 10.1	1613.7 ± 1309.7
MLP DeepART (MT)	$\mathbf{27.0 \pm 4.7}$	$\mathbf{35.6 \pm 12.7}$	$\mathbf{23.4 \pm 4.6}$	254.0 ± 195.1
CNN DeepART (ST)	106.4 ± 56.6	224.3 ± 163.9	59.2 ± 19.6	2182.4 ± 1958.7
CNN DeepART (MT)	38.8 ± 14.0	48.5 ± 22.3	30.8 ± 7.5	$\mathbf{248.4 \pm 168.8}$

Condensed Scenario

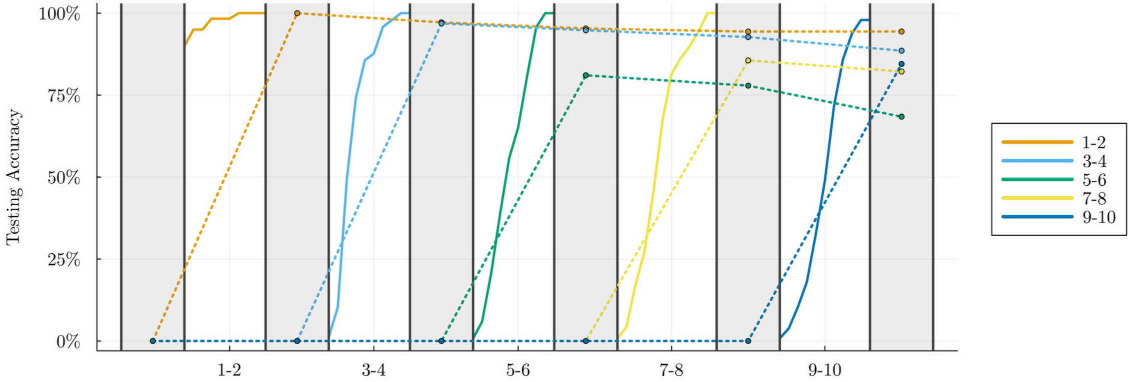
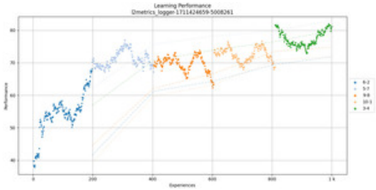


Figure 15: An example condensed lifelong learning scenario for illustration using the deep MLP DeepART algorithm trained on the 2-class 5-task Split MNIST handwritten digits dataset variant.

L2Metrics

Lifelong Learning Metrics

- l2logger
- l2metrics
- **Metrics**: Performance, Activation, Match
- **Metametrics**: Performance Maintenance, Forward Transfer Ratio, Backward Transfer Ratio



L2 Metric	Method	Dataset: MNIST	Fashion MNIST	USPS	CIFAR-10	CIFAR-100 (Coarse)	CIFAR-100 (Fine)
Performance							
PM	FuzzyART	0.9940 ± 0.0091	1.0019 ± 0.0267	1.0018 ± 0.0051	1.0000 ± 0.0043	0.9995 ± 0.0107	0.9912 ± 0.0209
	MLP DeepART	0.9612 ± 0.0402	0.9587 ± 0.0401	0.9822 ± 0.0143	0.9513 ± 0.0349	0.9237 ± 0.0633	0.9592 ± 0.0283
	CNN DeepART	0.9703 ± 0.0291	0.9646 ± 0.0459	0.9931 ± 0.0045	0.9611 ± 0.0293	0.9258 ± 0.0607	0.9475 ± 0.0359
FTR	FuzzyART	0.9940 ± 0.0091	1.0019 ± 0.0267	1.0018 ± 0.0051	1.0000 ± 0.0043	0.9995 ± 0.0107	0.9912 ± 0.0209
	MLP DeepART	0.9612 ± 0.0402	0.9587 ± 0.0401	0.9822 ± 0.0143	0.9513 ± 0.0349	0.9237 ± 0.0633	0.9592 ± 0.0283
	CNN DeepART	0.9703 ± 0.0291	0.9646 ± 0.0459	0.9931 ± 0.0045	0.9611 ± 0.0293	0.9258 ± 0.0607	0.9475 ± 0.0359
BTR	FuzzyART	0.9957 ± 0.0043	0.9991 ± 0.0029	0.9996 ± 0.0021	1.0002 ± 0.0007	0.9994 ± 0.0002	0.9997 ± 0.0007
	MLP DeepART	0.9884 ± 0.0044	0.9769 ± 0.0131	0.9962 ± 0.0032	0.9657 ± 0.0088	0.9359 ± 0.0257	0.9535 ± 0.0330
	CNN DeepART	0.9931 ± 0.0048	0.9778 ± 0.0140	0.9986 ± 0.0009	0.9544 ± 0.0172	0.8931 ± 0.0326	0.9465 ± 0.0476
Activation							
PM	FuzzyART	1.0553 ± 0.1236	1.0536 ± 0.1091	0.9922 ± 0.0137	0.9929 ± 0.0171	1.0271 ± 0.0550	1.0092 ± 0.0136
	MLP DeepART	1.0033 ± 0.0037	1.0150 ± 0.0315	1.0064 ± 0.0141	1.0021 ± 0.0012	1.0058 ± 0.0029	1.0095 ± 0.0049
	CNN DeepART	1.0042 ± 0.0046	1.0113 ± 0.0223	1.0100 ± 0.0181	1.0013 ± 0.0006	1.0074 ± 0.0056	1.0094 ± 0.0078
FTR	FuzzyART	1.0553 ± 0.1236	1.0536 ± 0.1091	0.9922 ± 0.0137	0.9929 ± 0.0171	1.0271 ± 0.0550	1.0092 ± 0.0136
	MLP DeepART	1.0033 ± 0.0037	1.0150 ± 0.0315	1.0064 ± 0.0141	1.0021 ± 0.0012	1.0058 ± 0.0029	1.0095 ± 0.0049
	CNN DeepART	1.0042 ± 0.0046	1.0113 ± 0.0223	1.0100 ± 0.0181	1.0013 ± 0.0006	1.0074 ± 0.0056	1.0094 ± 0.0078
BTR	FuzzyART	1.0001 ± 0.0004	1.0001 ± 0.0005	1.0001 ± 0.0004	1.0001 ± 0.0003	1.0001 ± 0.0002	1.0002 ± 0.0004
	MLP DeepART	0.9971 ± 0.0018	0.9977 ± 0.0012	0.9973 ± 0.0016	0.9991 ± 0.0006	0.9953 ± 0.0016	0.9891 ± 0.0035
	CNN DeepART	0.9990 ± 0.0008	0.9982 ± 0.0014	0.9996 ± 0.0004	0.9997 ± 0.0002	0.9928 ± 0.0029	0.9852 ± 0.0056
Match							
PM	FuzzyART	0.9254 ± 0.2470	0.9710 ± 0.0683	1.0602 ± 0.0996	1.0627 ± 0.0788	0.9369 ± 0.1063	0.9793 ± 0.0224
	MLP DeepART	0.9620 ± 0.0280	0.9560 ± 0.0427	1.0434 ± 0.1425	1.0156 ± 0.1334	0.9721 ± 0.0660	0.9844 ± 0.0756
	CNN DeepART	0.9754 ± 0.0176	0.9163 ± 0.1051	1.0152 ± 0.0594	1.0386 ± 0.1227	0.9868 ± 0.0812	1.0085 ± 0.0863
FTR	FuzzyART	0.9254 ± 0.2470	0.9710 ± 0.0683	1.0602 ± 0.0996	1.0627 ± 0.0788	0.9369 ± 0.1063	0.9793 ± 0.0224
	MLP DeepART	0.9620 ± 0.0280	0.9560 ± 0.0427	1.0434 ± 0.1425	1.0156 ± 0.1334	0.9721 ± 0.0660	0.9844 ± 0.0756
	CNN DeepART	0.9754 ± 0.0176	0.9163 ± 0.1051	1.0152 ± 0.0594	1.0386 ± 0.1227	0.9868 ± 0.0812	1.0085 ± 0.0863
BTR	FuzzyART	0.9995 ± 0.0024	0.9972 ± 0.0093	0.9987 ± 0.0040	0.9996 ± 0.0007	1.0000 ± 0.0010	0.9999 ± 0.0001
	MLP DeepART	0.9481 ± 0.0271	0.9714 ± 0.0165	0.9749 ± 0.0066	0.9338 ± 0.0303	0.9018 ± 0.0384	0.9478 ± 0.0364
	CNN DeepART	0.9647 ± 0.0245	0.9554 ± 0.0294	0.9935 ± 0.0049	0.9400 ± 0.0290	0.8624 ± 0.0471	0.9472 ± 0.0484

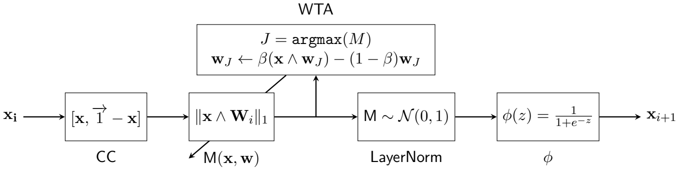
TABLE VII: The Performance Maintenance (PM), Forward Transfer Ratio (FTR), and Backward Transfer Ratio (BTR) L2 meta-metrics computed on the trendlines of the single-task performances, FuzzyARTMAP best-matching-unit (BMU) activation, and FuzzyARTMAP BMU match for each method and dataset tested. The BMU during inference is the winning category that is reported during the FuzzyART WTA classification whether by its normal match rule or as the highest-activated category in the case of a mismatch whereby no one category satisfies the current vigilance criterion. The subsequent activation and match

Table of Contents

Conclusion

Summary

1. First **deep, function-composed** adaptive resonance algorithm
2. MLP and CNN layer derivations, arbitrary network architecture
3. Performance improvement over FuzzyARTMAP with same head dimensionality
4. Significant category proliferation reduction at head layer over FuzzyARTMAP



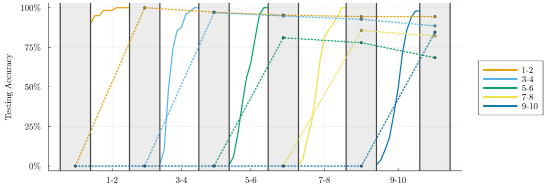


Table of Contents

Conclusion

What Was Learned

- L2 and learning stable feature detectors on nonstationary data are the **same problem**
- ART viable for alternative architectures altogether i.e.:
  - DeepART and deep neural networks
  - stART and graphs
  - NLP via structured syntactic parsing
  - Deep transfer learning via deep feature clustering

Future Work

1. Expand capabilities and applications of ART
2. Lateral thinking to succeed in modern ML landscape
   - Cannot compete with big industry and tech on existing ML methods; outpaced, outscaled, outbudgeted
   - Succeed by tackling new problems, rejecting trends

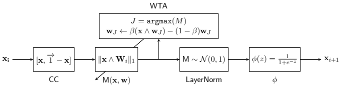
Thank You!

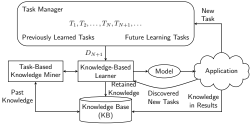
Contributions

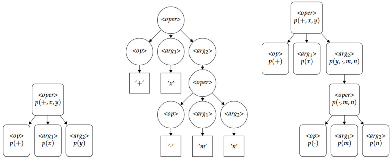
1. Combining Deep Learning and Adaptive Resonance Theory
2. Adaptive Resonance Algorithms for Lifelong Machine Learning
3. Novel Adaptive Resonance Algorithms and Architectures

Papers

- **stART**: *Analyzing Biomedical Datasets with Symbolic Tree Adaptive Resonance Theory*, Accepted *MDPI Information*[7]
- **Deep Context Recognition**: *Context Recognition with Deep Feature Clustering*, Under Review *IEEE SMC Transactions*[8]
- **DeepART**: *Deep Gradient-Free Local Learning with Adaptive Resonance*, Accepted *Neural Networks*







(a) START prototype before update.

(b) Relation parse tree of $x + m \cdot n$.

(c) START prototype after update.

References I

- L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch, "A Survey of Adaptive Resonance Theory Neural Network Models for Engineering Applications," *Neural Networks*, vol. 120, no. xxxx, pp. 167–203, 2019.
- S. Grossberg and T. R. Huang, "ARTSCENE: A neural system for natural scene classification," *Journal of Vision*, vol. 9, no. 4, pp. 1–19, 2009.
- J. Redmon and A. Farhadi, "YOLO v.3," *Tech report*, pp. 1–6, 2018.
- L. E. Brito da Silva, I. Elnabarawy, and D. C. Wunsch, "Distributed dual vigilance fuzzy adaptive resonance theory learns online, retrieves arbitrarily-shaped clusters, and mitigates order dependence," *Neural Networks*, vol. 121, pp. 208–228, 2020.
- T. Kasuba, "Simplified fuzzy artmap, ai expert," 1993.

References II

- G. P. Amis and G. A. Carpenter, "Default ARTMAP 2," *IEEE International Conference on Neural Networks - Conference Proceedings*, vol. 2, no. September 2007, pp. 777–782, 3 2007.
- S. Petrenko, D. B. Hier, M. A. Bone, T. Obafemi-Ajayi, E. J. Timpson, W. E. Marsh, M. Speight, and D. C. Wunsch, "Analyzing biomedical datasets with symbolic tree adaptive resonance theory," *Information*, vol. 15, no. 3, 2024.
- S. Petrenko, A. Brna, M. Aguilar-Simon, and D. Wunsch, "Lifelong context recognition via online deep feature clustering," Jul. 2023.