Progress Report: Active, Real-time Prototype Recognition in a Conceptual Network

Bowen Xu

CIS, Temple University

bowen.xu@temple.edu

May 2, 2024

Motivation

Modeling the sensorimotor procedure in a logical way

Fig. 1. Task-oriented Top-Down Attention. The eye trajectories vary in different task-hints [6].[3]

Theory

Assimilation:

- An intelligent system tries to explain and predict sensations via the existing concepts in its memory;
- the system tries to change its perceptive field via actions, for the sake of confirming its explanations and predictions, as well as achieving its desires.

Accommodation:

- The system changes its memory to explain past experiences.

Working cycle:

- The system accepts sensory input, processing it within a relatively constant time, and decides where to see next, subsequently executing an action to shift its perceptive field.

Theory

- A **prototype** is a combination of compositions, each of whose locations can change slightly.
  - A special prototype (i.e., atomic prototype contains no composition inside it (in contrast to a compound prototype that contains at least two compositions), so that it can be the base of the recursive structure of prototypes and compositions

- A **composition** is a special concept composed of a prototype as part and a prototype as whole. A composition is attached with an attribute, *relative location*, which indicates the location of a part *relative* to its whole.
  - It is *relative* in the sense that the displacement between two locations can be computed without defining an absolute, original point.

Theory

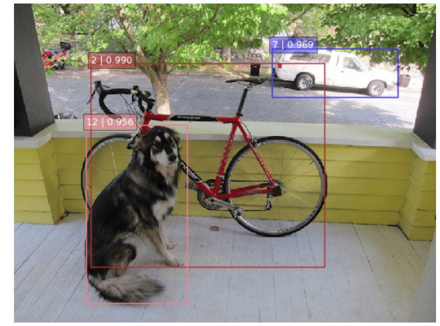
- **Implications**

  - The system works with an endless loop; there is no “final report” on “an image’s categorization(s)” .
  - The system continuously perceives its sensations, so that different concepts may catch its attention at different moments.
  - The system may gradually get a better and better understanding of scenery as time goes by, but it probably loses many details when it is in a hurry.

  Different interpretations from traditional computer vision:

  - ***Feature*** is the alias of *prototype*.
  - An ***object*** is an instance of a prototype. In this sense, object here does not mean “a thing as it is”, but rather the summary of relations among prototypes.
  - ***Recognition*** is the process in a system to retrieve its memory (i.e., the conceptual network) and pay attention to active concepts.

Representation & Inference Rule



Composition has a different meaning from *intersection/union*.

- *E.g.*, A bicycle is composed of two wheels and a frame.
- Concept $bicycle$ is not the intersection or union of concepts $wheel$ and $frame$.
- Concept $bicycle$ is the *extensional intersection* of concepts $vehicle$ and $machine$.
- Concept $bicycle$ is the *intentional intersection* of concepts $human\text{-}powered\text{-}vehicle$ and $two\text{-}wheeled\text{-}vehicle$.

"$P \mapsto W$" means P is a part of W

- *Is transitivity* valid? – Can $P \mapsto W, W \mapsto S$ derive $P \mapsto S$?
- It may break the hierarchy.
- For example, when recognizing a bicycle, a bearing in a wheel does not directly contribute evidence to bicycle.
- *Deduction* here is not valid, but another type of 'transitivity' is valid (see next pages).

Definition 1 *If P and W are events, composition statement "$P \mapsto W$" is true if and only if P's occurrence provides a piece of positive evidence for W's occurrence. The first term P is called part, and the second term W is called whole.*

Representation & Inference Rule

(IL) $(P_i \mapsto W)[l_i] \wedge (P_j \mapsto W)[l_j]$ is true if and only if "$(P_i, \Uparrow move(l_j - l_i), P_j) \leftrightarrow W$" and "$(P_j, \Uparrow move(l_i - l_j), P_i) \leftrightarrow W$", where l_i, l_j are local and *relative* to W.

$P_1 \mapsto W.[l_1]$
...
$P_n \mapsto W.[l_n]$

can be abbreviated as "$W: \{| P_1[l_1], \ldots, P_n[l_n] |\}$", meaning that W is composed of $P_1, \ldots, P_n$ attached with their corresponding *relative* locations $l_1, \ldots, l_n$.

Implication:

- Spatial representation in an intelligent system is *relative* rather than absolute.
  - *The origin of a coordinate is meaningless (in the meta-level).*
  - *The distance between any two points/locations is determined.*

Representation & Inference Rule

$$\{P \mapsto W, P\} \vdash W.F_{prt} \Rightarrow \{P \mapsto W, W \mapsto S, P\} \vdash S.F_{prt}F_{prt} \Rightarrow \{P \mapsto W, W \mapsto S\} \vdash P \mapsto S.\widehat{F_{prt}}$$

(C, P, Q are *events*)

$$\{(P \mapsto W)[l_1], P[l_2]\} \vdash W[l_2].F_{spj}F_{prt}$$

$F_{spj}F_{prt}$:
$$f, c' = F_{prt}(f_1, f_2, c_1, c_2)$$
$$c = F_{spj}(c', l_1, l_2)$$

$$\{W[l_1], W[l_2]\} \vdash W[l].F_{srv}F_{rev}$$

$F_{srv}F_{rev}$:
$$f, c = F_{rev}(f_1, f_2, c_1, c_2)$$
$$l = F_{srv}(f_1, f_2, c_1, c_2, l_1, l_2)$$

type	inference	name	function (a tentative proposal)
weak syllogism	*part-whole*	F_{prt}	$f = and(f_1, f_2)$ $w = and(f_1, f_2, c_1, c_2)$ similar to deduction, but provides no more than 1 piece of evidence
immediate inference	*spatial projection*	F_{spj}	$c = bell(l_1 - l_2) \times c'$
local inference	*spatial revision*	F_{srv}	$l = (l_1 f_1 c_1 + l_2 f_2 c_2)/(f_1 c_1 + f_2 c_2 + \epsilon)$

Representation & Inference Rule

Is NAL1-9 enough to represent the *composition* relation?

$(P \times W) \rightarrow partof.$
$P.$

$((P \times W) \rightarrow partof) \wedge W. F_{int}$

$((\$P \times \$W) \rightarrow partof) \wedge \$P \Rightarrow \$W.$
$((P \times W) \rightarrow partof) \wedge P.$

$W. F_{ded}$

strong syllogism	deduction	F_{ded}	$f = and(f_1, f_2)$ $c = and(f_1, f_2, c_1, c_2)$
term composition	intersection	F_{int}	$f = and(f_1, f_2)$ $c = and(c_1, c_2)$

Representation & Inference Rule

Is NAL1-9 enough for representing a prototype? Yes.

$$W \leftrightarrow (P_1, \Uparrow move(l_2 - l_1), \ldots, \Uparrow move(l_n - l_{n-1}), P_n)$$
$$W \leftrightarrow W_1, \ldots, W \leftrightarrow W_m$$

It is trajectory-dependent – a good attribute.
It is trajectory-dependent only – a bad attribute since humans can generalize to objects following an unusual and even novel observational trajectory. In this case, a trajectory-independent representation is more appropriate. Both coexist in the system.

I argue that trajectory matters and it is more relevant to the control mechanism.

Control

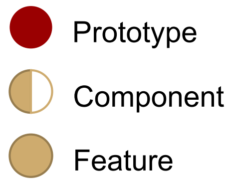
Organization Mechanism — Performance

Hypothesize Recycle → Revise → Retrieve → Recognize

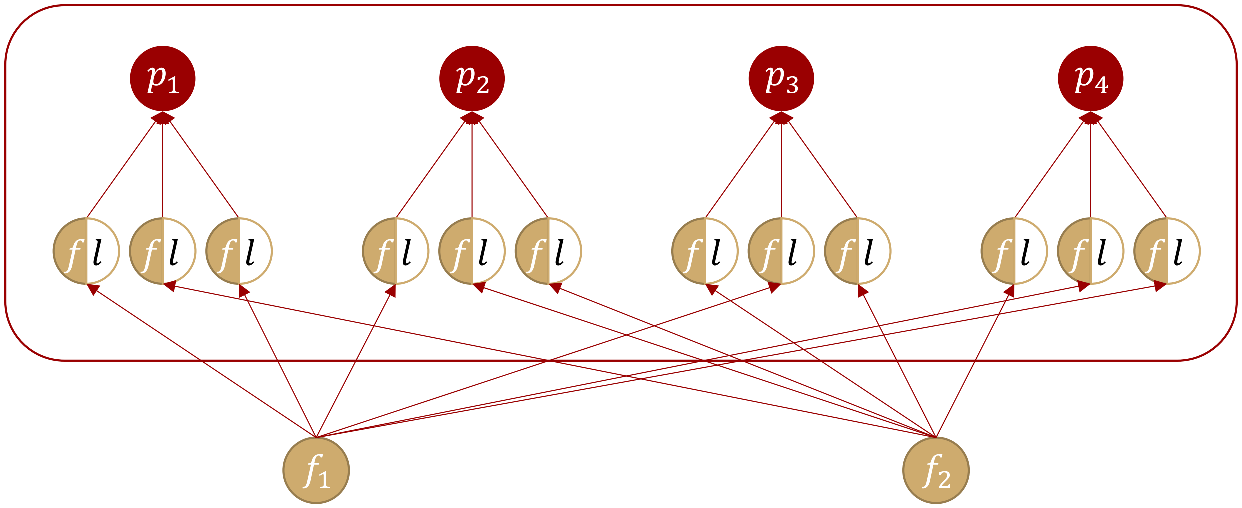
Retrieve! ← Recognize!

Control

- **Memory**

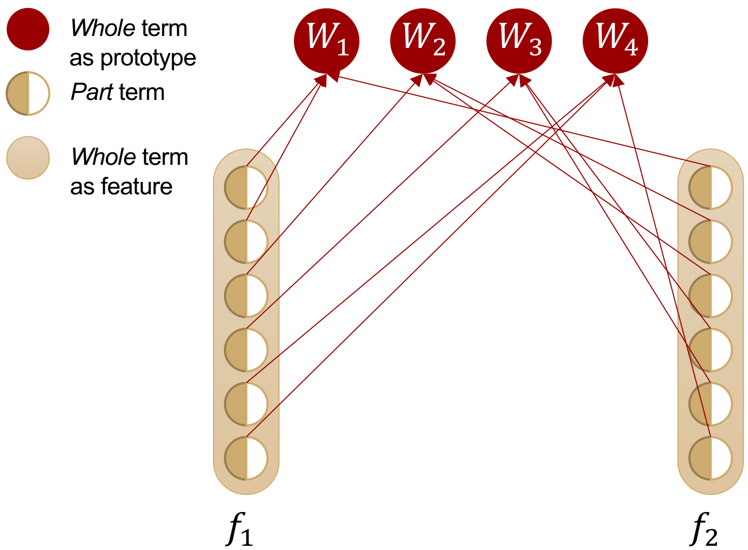


Conceptual Network (Sensorimotor Layer)



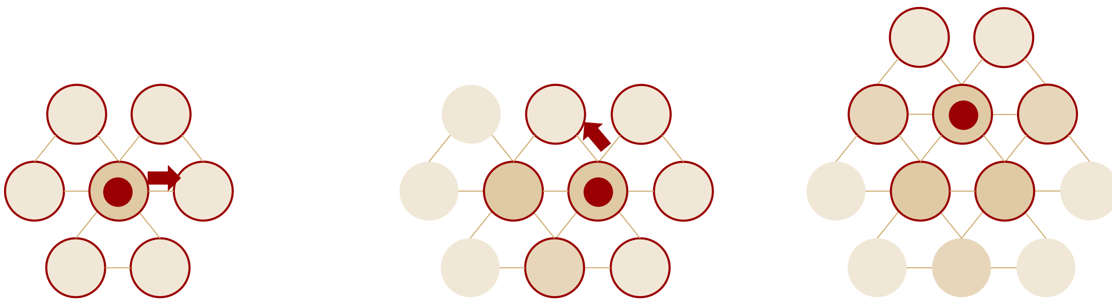
Control

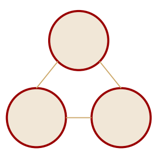
- **Memory**



Conceptual Network (Sensorimotor Layer)

Control



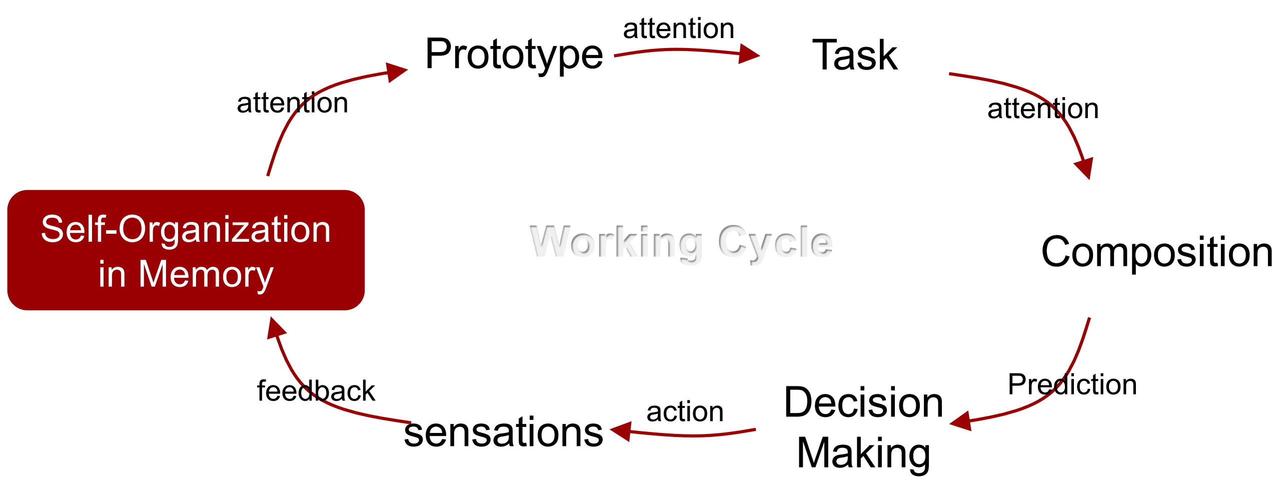
Task: 

Current Location:

Action:

Control

- **Working Cycle**



Working Example

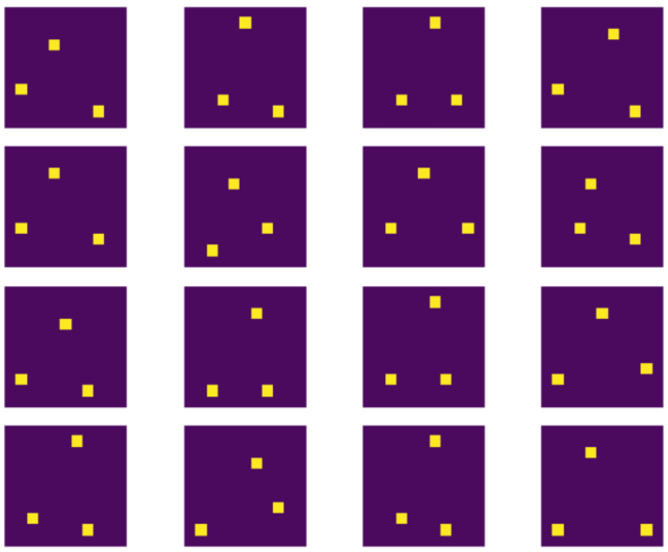
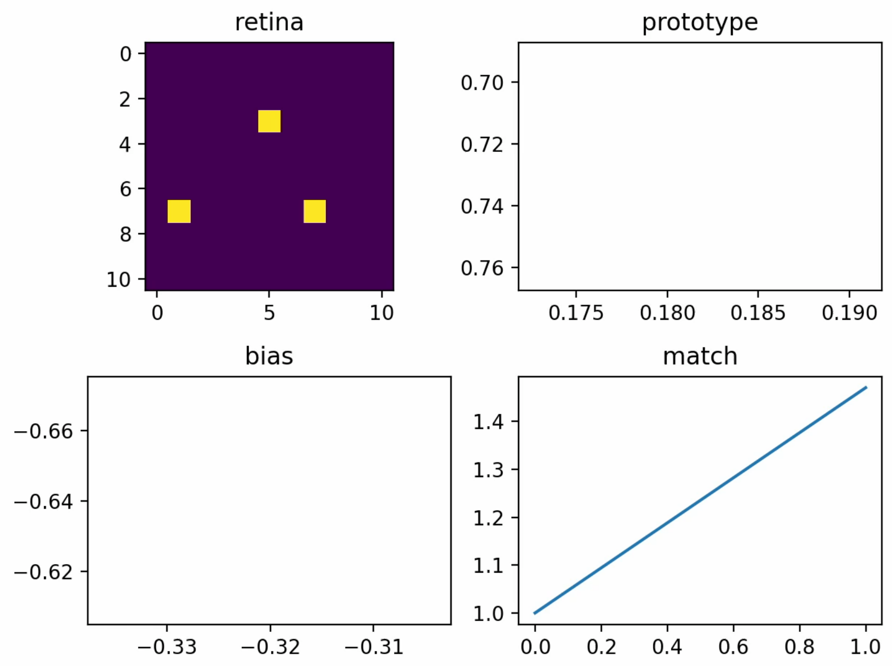
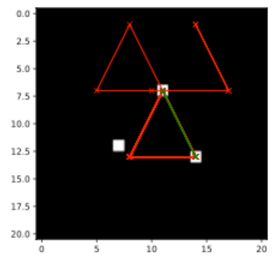


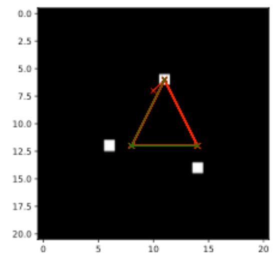
Fig. 4. Input Examples



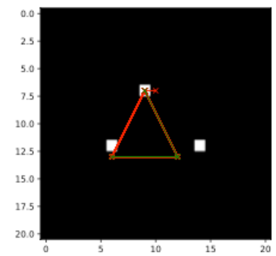
Working Example



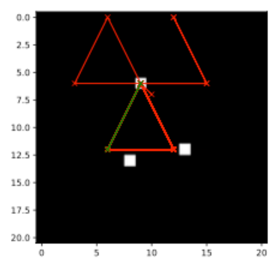
(a)



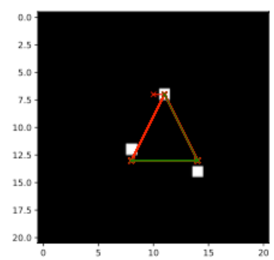
(b)



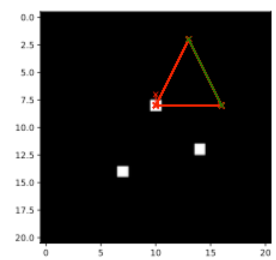
(c)



(d)



(e)



(f)

Fig. 6. (a)~(d) success examples; (f) a failure example

Future Work

$\overbrace{\text{Hypothesize Recycle} \rightarrow \text{Revise} \rightarrow \text{Retrieve}}^{\text{Organization Mechanism}} \rightarrow \overbrace{\text{Recognize}}^{\text{Performance}}$

Hypothesize Recycle! ← Revise! ← Retrieve! ← Recognize!

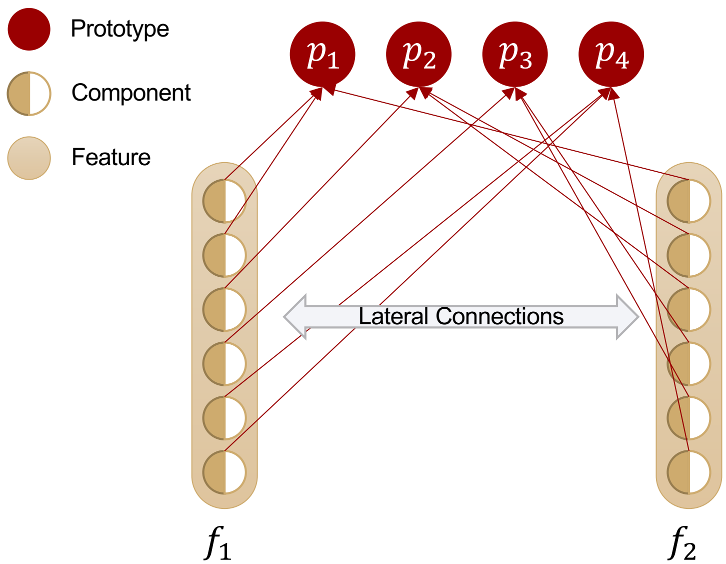
Single layer → Multiple layers

Future work

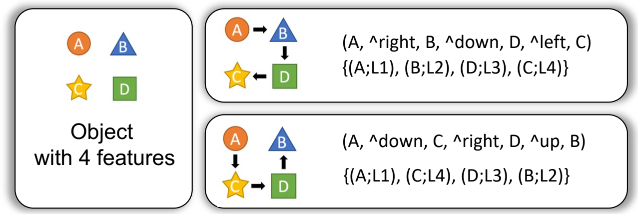
Given a large amount of spatiotemporal information as input,

- how to actively organize them into memory and recall knowledge efficiently
- how to achieve goals

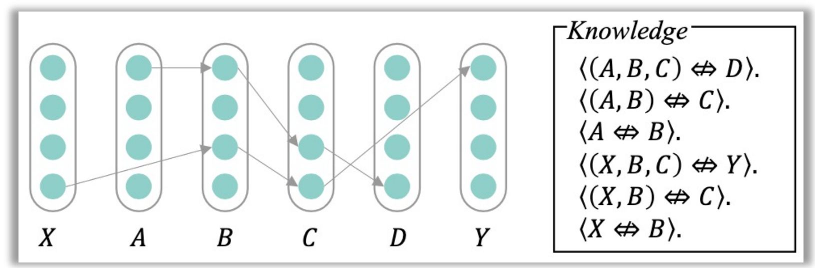
Future Work



Conceptual Network (Sensorimotor Layer)



The unity of the two types of representation



Sequence Learning Model (Xu, 2023)

Xu, B. (2023). *A Brain-Inspired Sequence Learning Model based on a Logic* (arXiv:2308.12486). arXiv. https://doi.org/10.48550/arXiv.2308.12486

Future Work

sight 👀

vision

hearing 👂

audition

synesthesia 🔯

touch ☝️

somatosensation

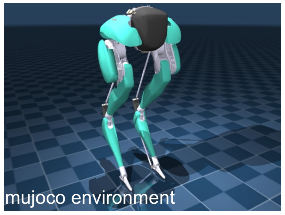
smell 👃

olfaction

taste 👅

gustation

motor control 🦵



control mechanism

NAL + Narsese

Thank you!