Learning and Inference in Cartoon Videos using Deep Hypernetworks Concept Structure

2016-11-01

Introduction to Machine Learning

Byoung-Tak Zhang, Kyung-Min Kim

Definition – What is Concept?

Concept ‘Pororo’

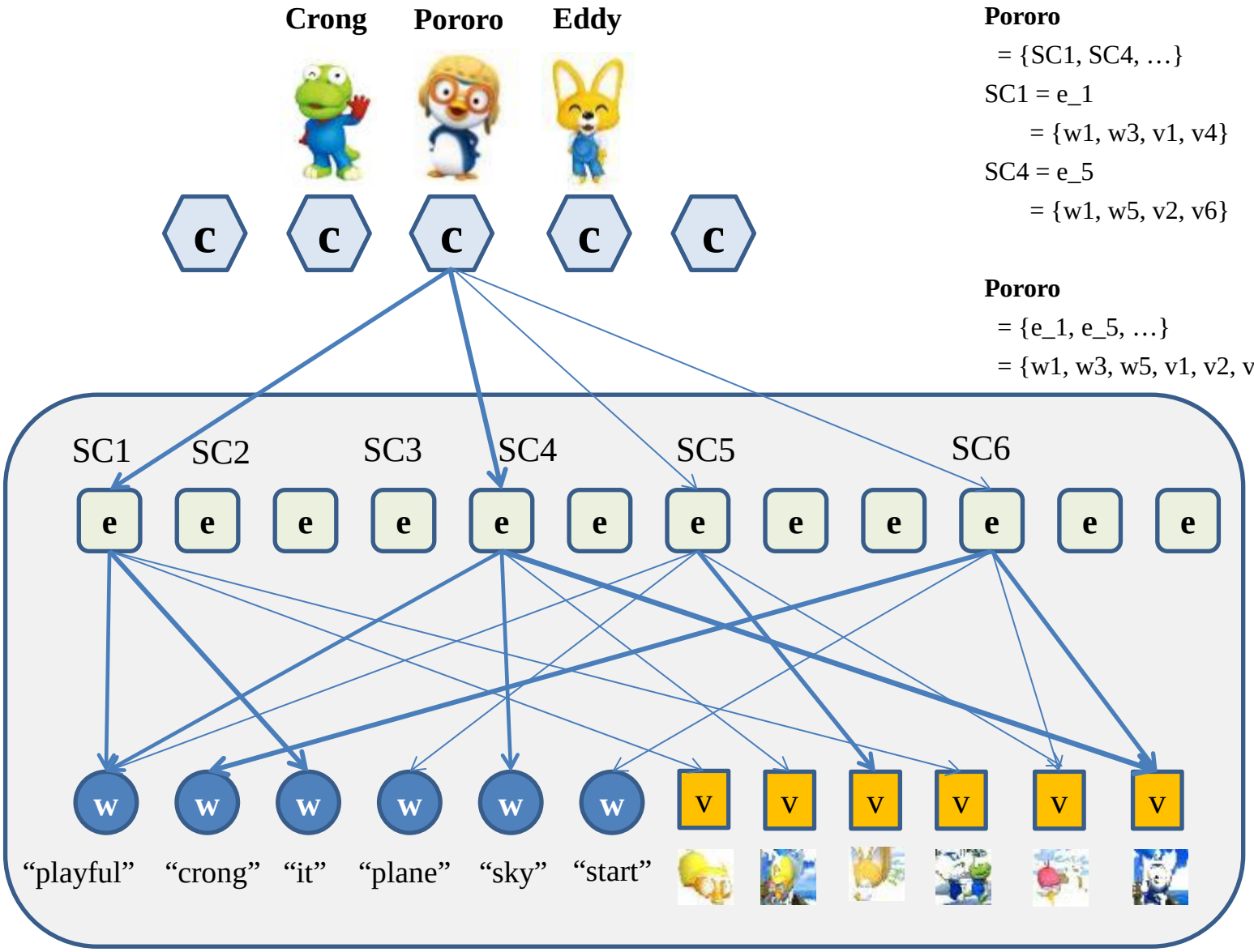
High-level

blue, penguin,
pororo, cute, playful

Low-level

Visual and linguistic representation

Sparse Population Coding model[1] (SPC)



Pororo
= {SC1, SC4, …}
SC1 = e_1
= {w1, w3, v1, v4}
SC4 = e_5
= {w1, w5, v2, v6}

Pororo
= {e_1, e_5, …}
= {w1, w3, w5, v1, v2, v4, v6, …}

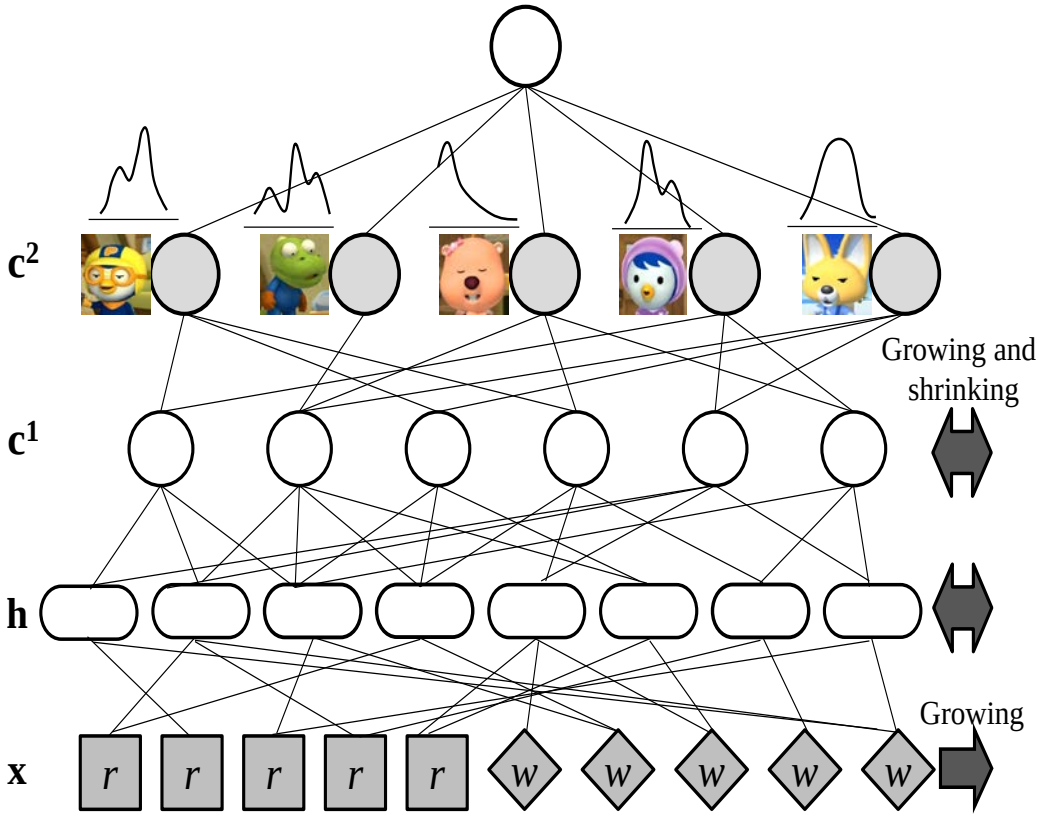
Microcodes (sparse)

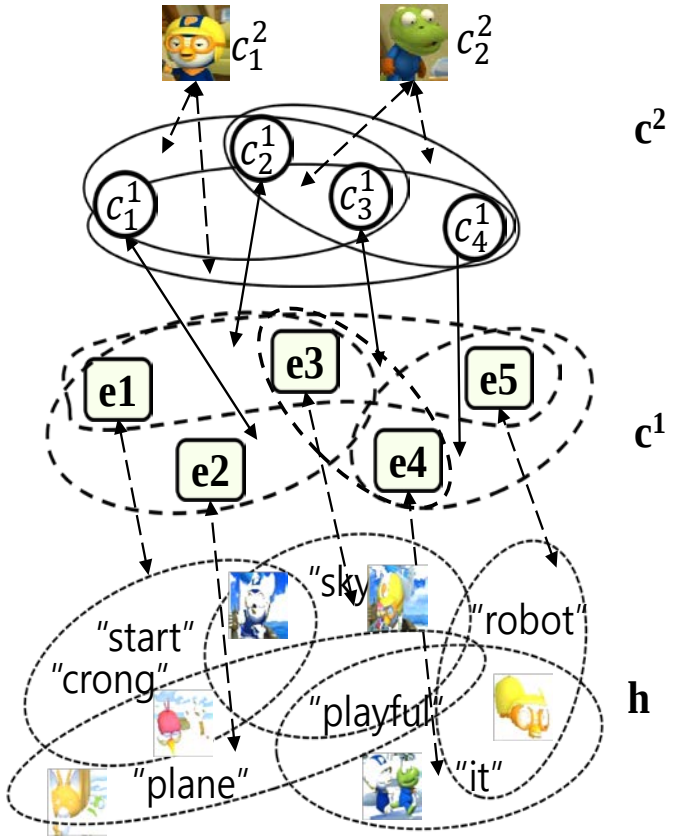
Population code = a collection of sparse microcodes

1. Zhang B.-T., Ha J.-W., Kang M., Sparse Population Code Models of Word Learning in Concept Drift, *Cogsci* 2012

Deep Concept Hierarchies

$$P(\mathbf{r},\mathbf{w}\,|\,\mathbf{c}^1,\mathbf{c}^2)=\sum_{\mathbf{h}} P(\mathbf{r},\mathbf{w}\,|\,\mathbf{h},\mathbf{c}^1,\mathbf{c}^2)P(\mathbf{h}\,|\,\mathbf{c}^1,\mathbf{c}^2)$$





(a) Example of deep concept hierarchy learned from Pororo videos (b) Hypergraph representation of (a)

Algorithm: Learning of Concept Layers

- **Three issues**
  - Determining the number of the nodes of the concrete concept layer $\mathbf{c}^1$ ($\mathbf{c}^1$-nodes)
  - Associating $\mathbf{c}^1$-nodes and the modality layer **h**
  - Associating $\mathbf{c}^1$-nodes and the abstract concept nodes ($\mathbf{c}^2$-nodes)
- **Number of $\mathbf{c}^1$-nodes and association of the $\mathbf{c}^1$ layer and h**
  - A c^1-node is associated with a subgraph of the hypernetwork
  - A subgraph = a cluster of hyperedges
  - The number of clusters changes depending on the closeness centrality using word2vec:
    (Mikolov et al. 2013)

$$Sim(\mathbf{h}^m) = \frac{Dist(\mathbf{h}^m)}{|\mathbf{h}^m|} \quad (\mathbf{h}^m : \text{the hyperedge cluster associated with } c_m^1)$$

  - If $Sim(\mathbf{h}^m) > \theta_{\max}$ or $Sim(\mathbf{h}^m) < \theta_{\min}$ then $\mathbf{h}^m$ is split or merged.
- **Association of $\mathbf{c}^1$ and $\mathbf{c}^2$ layers**
  - A hyperedge contains the c2 node information

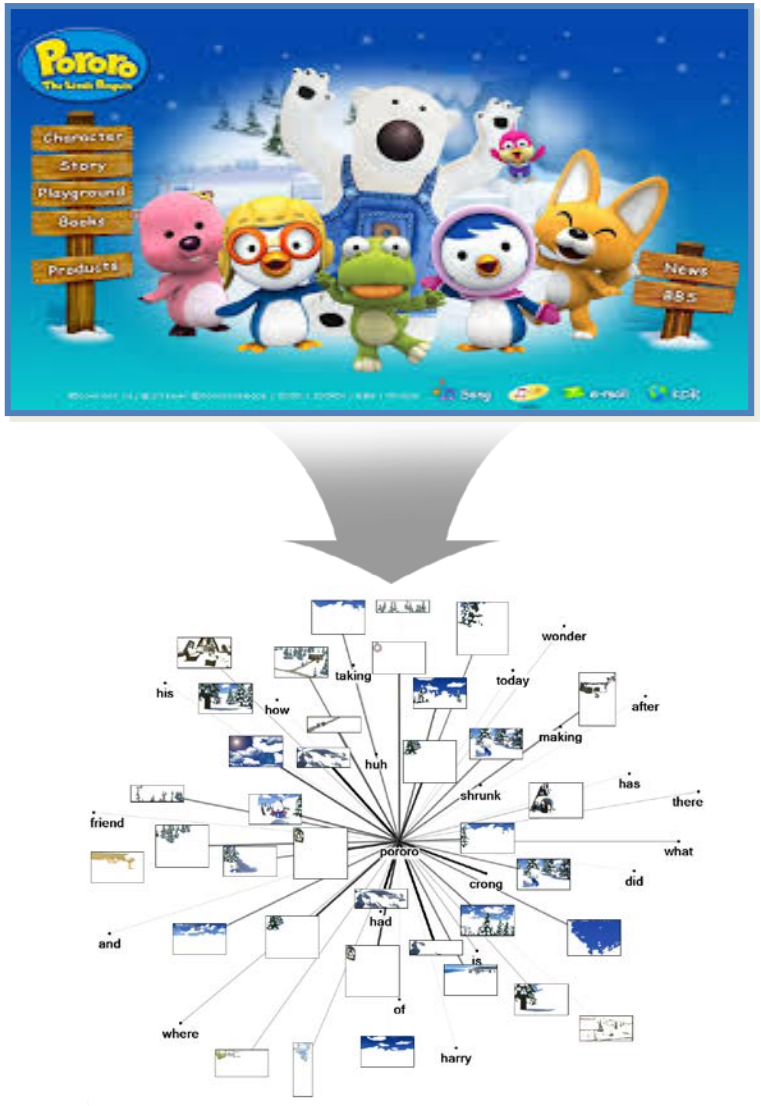
$$\omega(c_i^1, c_j^2) = \frac{\sum_{h_m \in \mathbf{h}^i} \alpha_m C(c_j^2, h_m)}{\sum_{h_m \in \mathbf{h}^i} \alpha_m}$$

Why Cartoon Videos?

- **Children's cartoon videos**
  - Multimodal
  - Vision + language
  - Simple grammars
  - Explicit story lines
  - Image processing
  - Pseudo-real
  - Educational
  - Cognitive

Comparison to Previous Approaches

- **Previous approaches to knowledge acquisition and representation**
  - Semantic networks (Steyvers 2006)
  - WordNet (Fellbaum 2010)
  - Single-modality (usually linguistic)
- **Here: Automated construction of conceptual knowledge from videos**
  - Visually-grounded knowledge
  - Dynamic
  - Concept drift
  - Multimodal



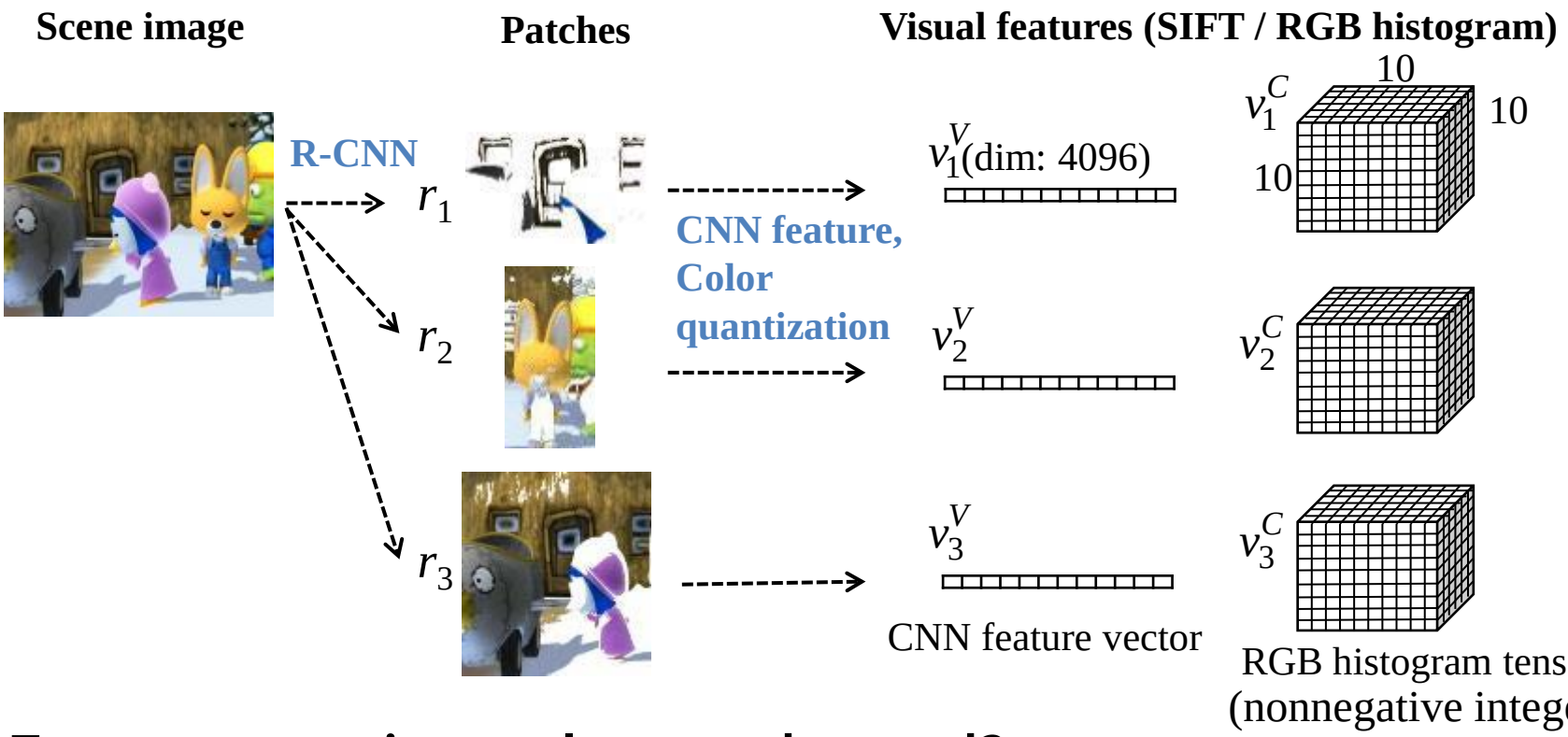
Data Description (1/2)

- **Cartoon video data**
  - 17 'Pororo' DVDs, 183 episodes, 1232 minutes, 16000 scene-subtitle pairs
- **Image**
  - R-CNN for extracting image patches
  - CNN feature vector (4096-D vector)
  - Color histogram (1000-D vector)
- **Text**
  - Word set including functional words for sentence generation
  - Total 3452 words
  - Represent each word as a real vector by word2vec[1]

1 Mikolov T., Sutskever I., Chen K., Corrado G., Dean J., Distributed Representation of Words and Phrases and Their Compositionality, NIPS 2013

Data Description (2/2)

- **Image preprocessing: segmentation + descriptor**
  - Each scene image → a set of image patches (MSER)
  - Image patch → CNN feature / RGB histogram

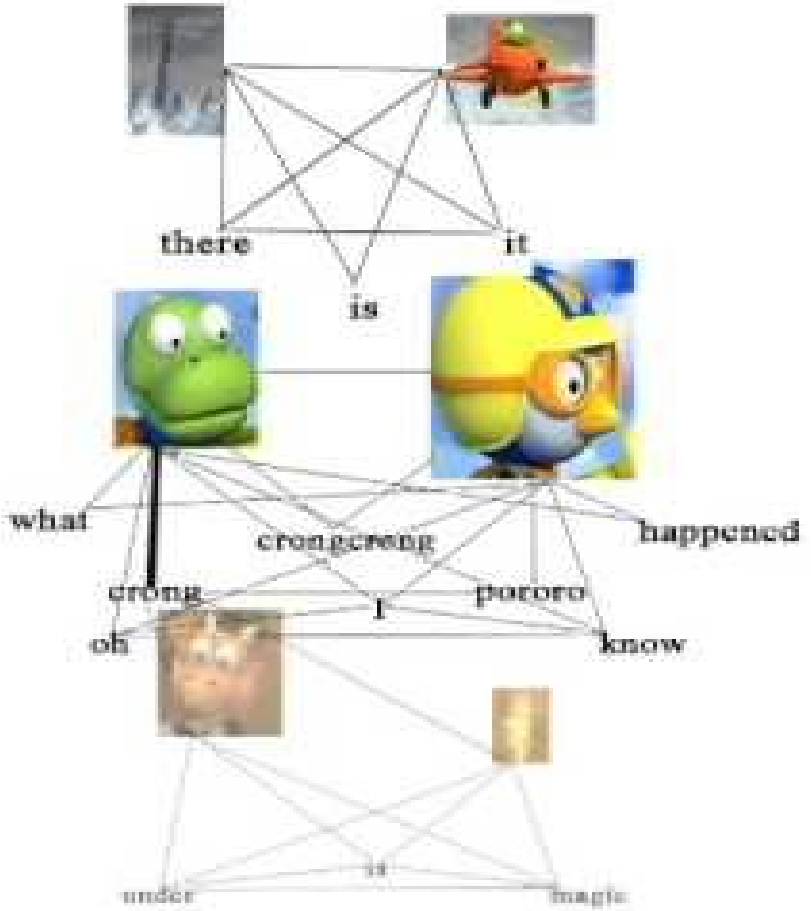


- **Text preprocessing: real vectors by word2vec**

Concept Learning

Video Source

Concept Map



Data Pipeline

Unit of input : an episode

i-th episode

Scene #1

Scene #2

....

Scene #n

Preprocessing

Kids Videome

HE Sampling

v1 v2 v5

v2 v3 v7

v3 v4 v6

Network construction

Large-scale video ⟶ **Preprocessing** ⟶ **Hyperedge** ⟶ **Concept Networks**

Evolution of Concept Map

Image 개수 : 20000
Word 개수 : 1000
Episode 개수 : 800

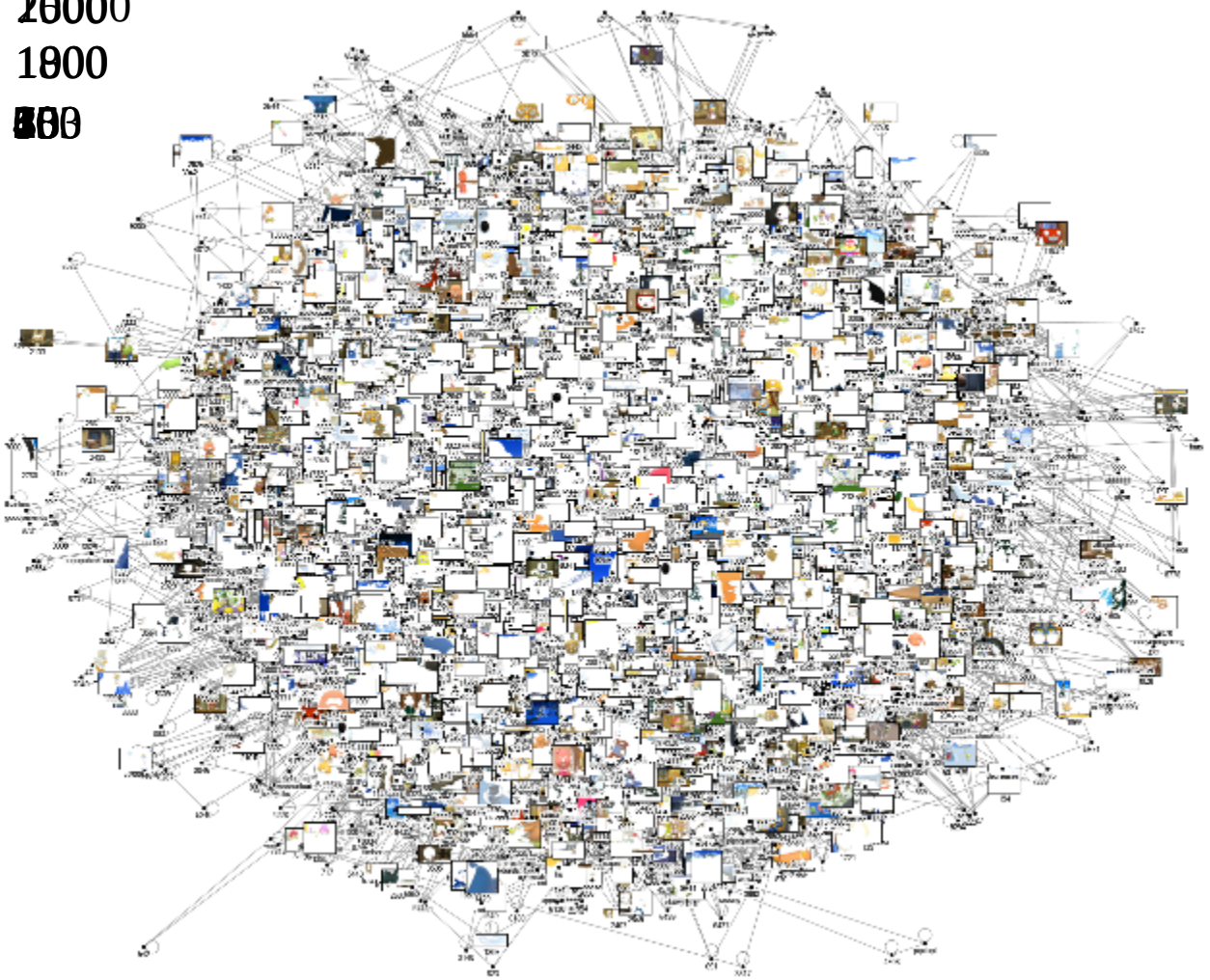


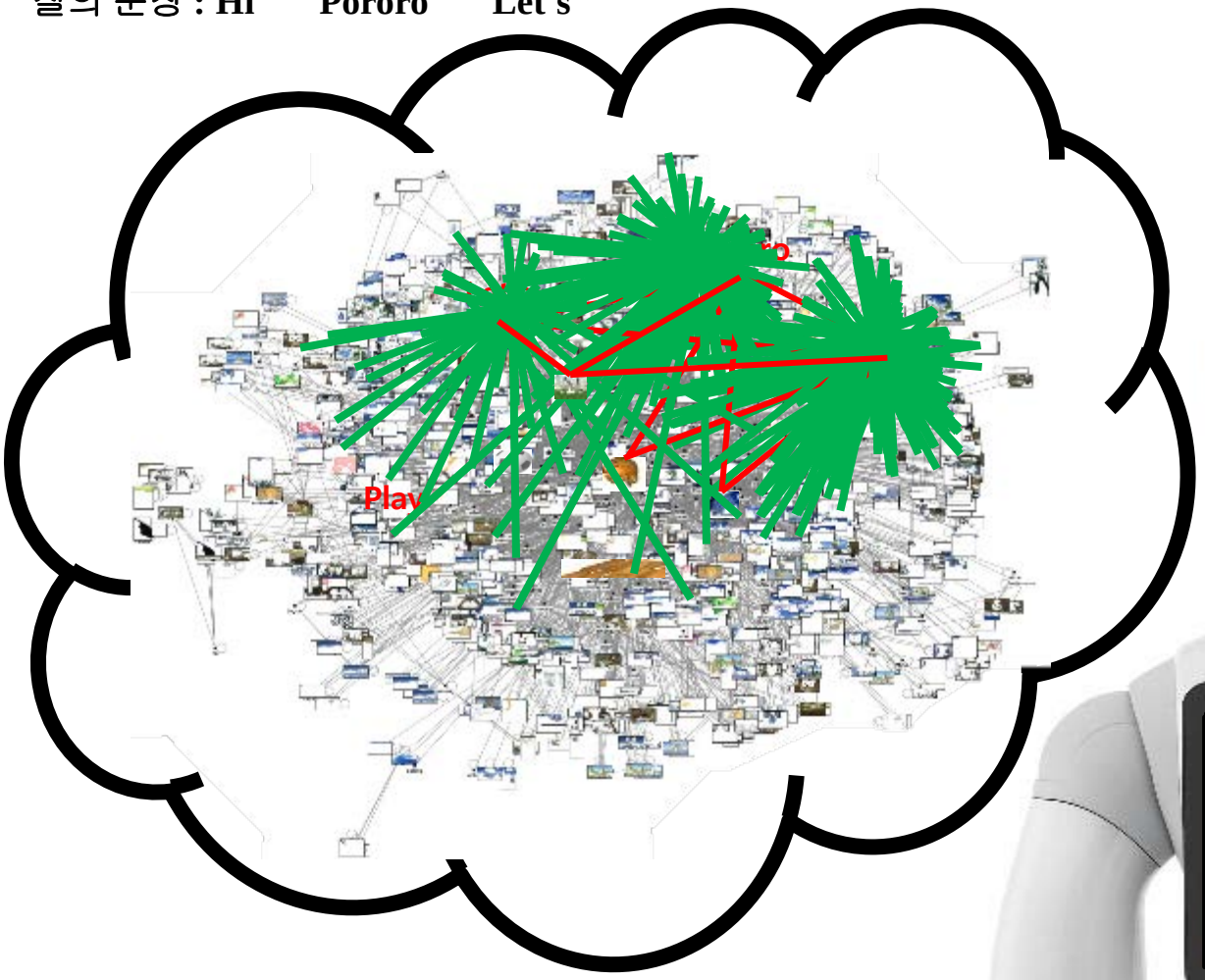
Image Generation from Sentence

- **Intermediate images generated from query sentences**

Query sentences	Episodes 1~52 (1 season)	Episodes 1~104 (2 seasons)	Episodes 1~183 (all seasons)
• Tongtong, please change this book using magic. • Kurikuri, Kurikuri-tongtong!			
• I like cookies. • It looks delicious • Thank you, loopy			

Image Generation from Sentence

질의 문장 : Hi Pororo Let's



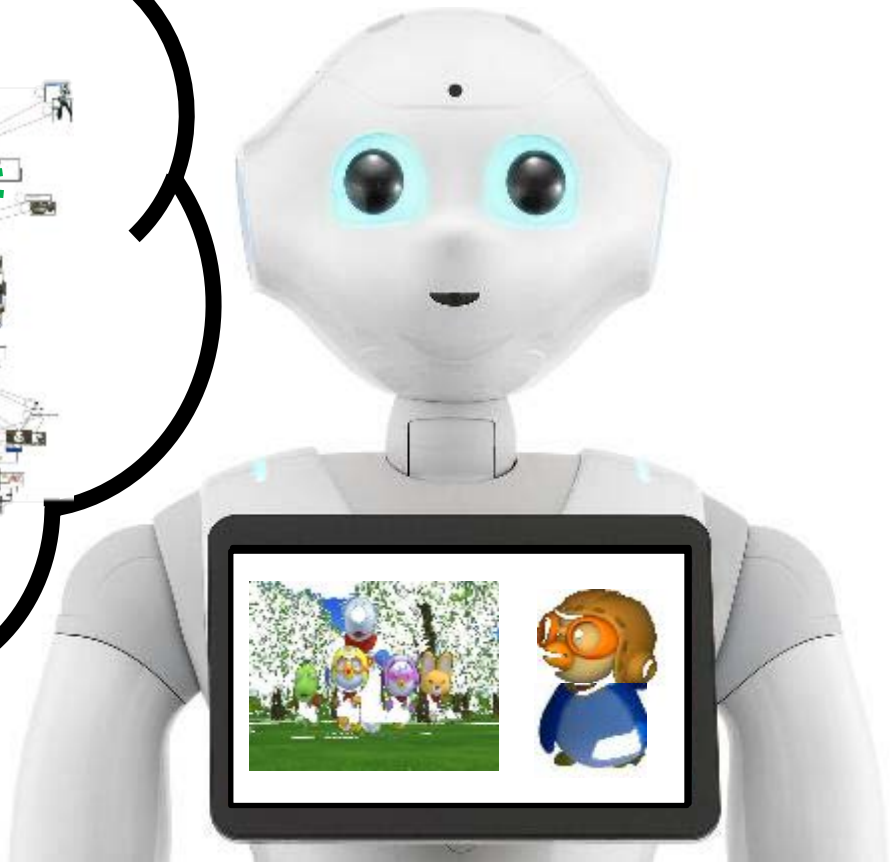
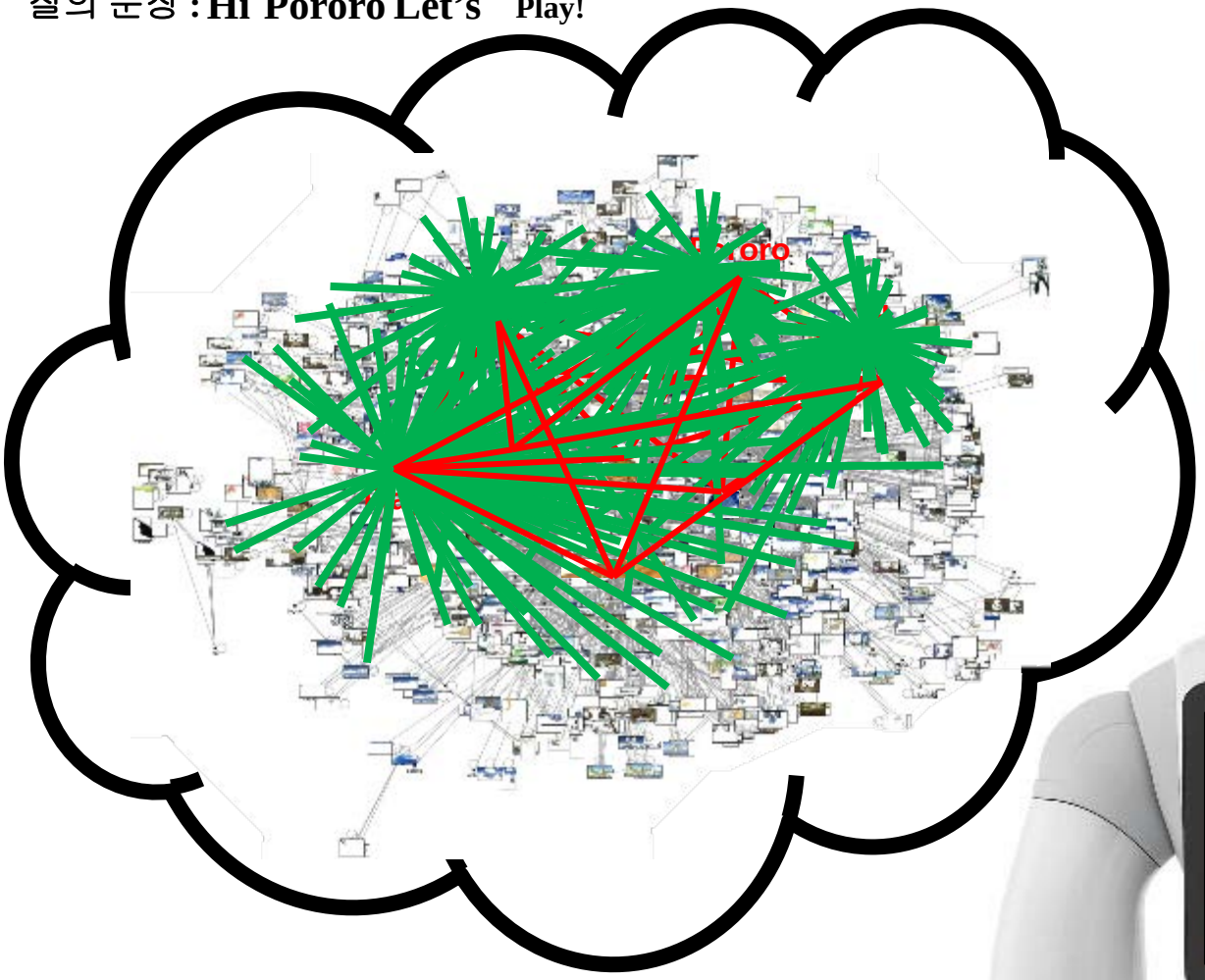
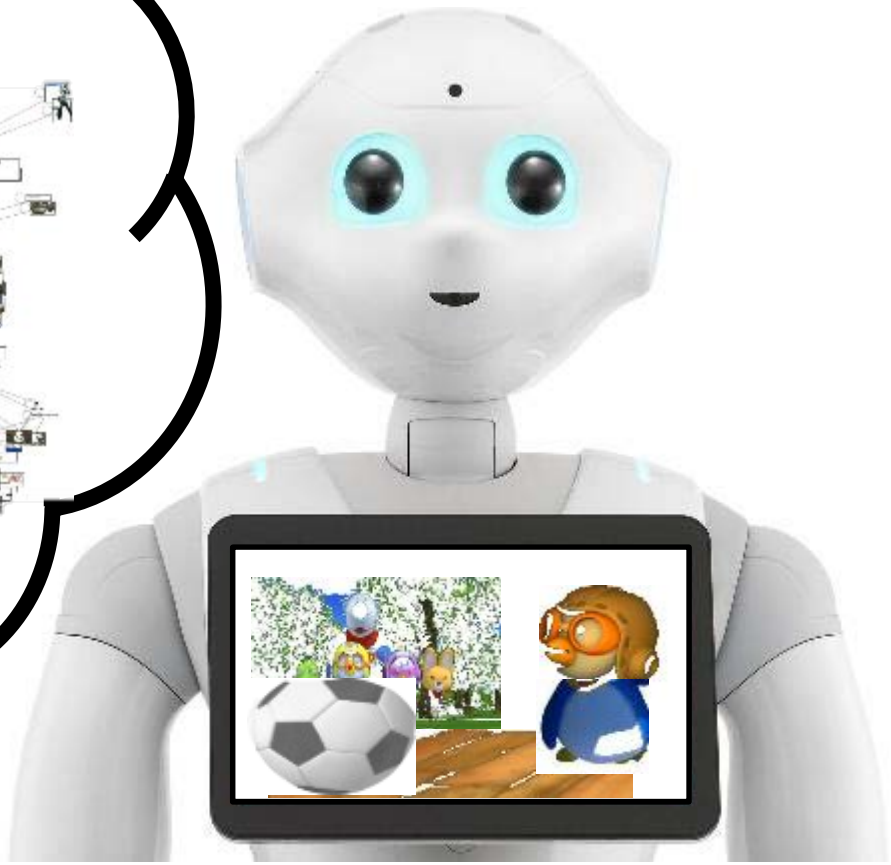


Image Generation from Sentence

질의 문장 : **Hi Pororo Let's Play!**





Sentence Generation from Image

Query = {i, try}

Original subscript: clock, I have made another potion come and try it

Generated subscripts

- **as i don't have the right magic potion come and try it was nice**
- **ah, finished i finally made another potion come and try it we'll all alone?**

Query = {thanks, drink}

Original subscript: Oh, thanks Make him drink this

Generated subscripts

- **oh, thanks make him drink this bread**
- **oh, thanks make him drink this forest**

Query = {he, take}

Original: Tong. Tong Let. Me. Take. It. To. Clock

Generated

- **take your magic to know what is he doing?**
- **take your magic to avoid the house to know what is he keeps going in circles like this will turn you back to normal**

Query = {ship, pulled}

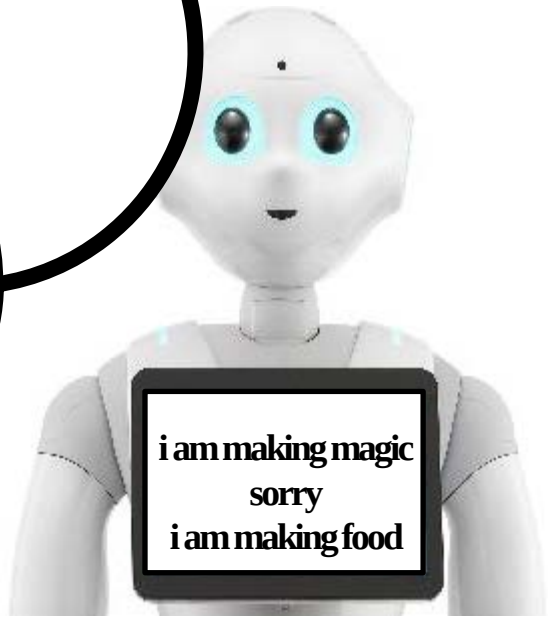
Original: The ship is being pulled

Generated

- **wow looks as if that's the ship is being pulled**
- **the ship is being pulled**

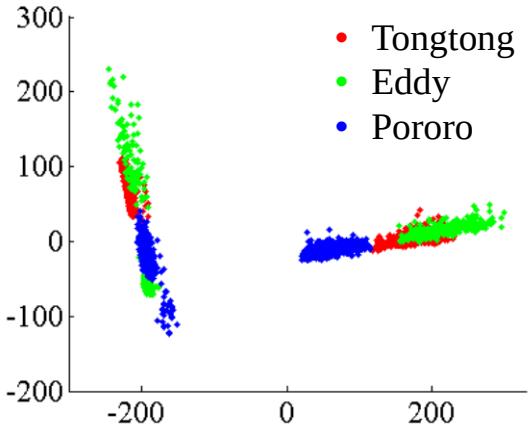
Sentence Generation from Image

질의 이미지

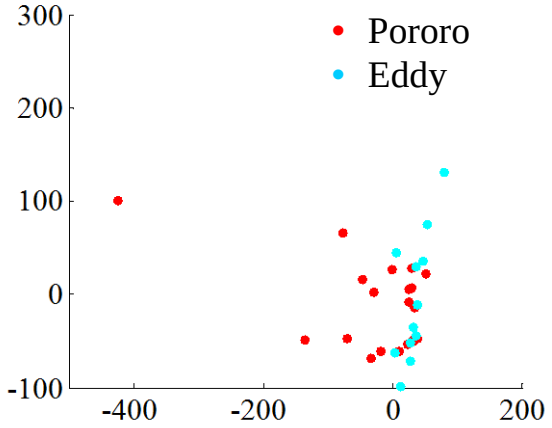


Analysis of Constructed Concept Structures

- **Distinguishable concept nodes**

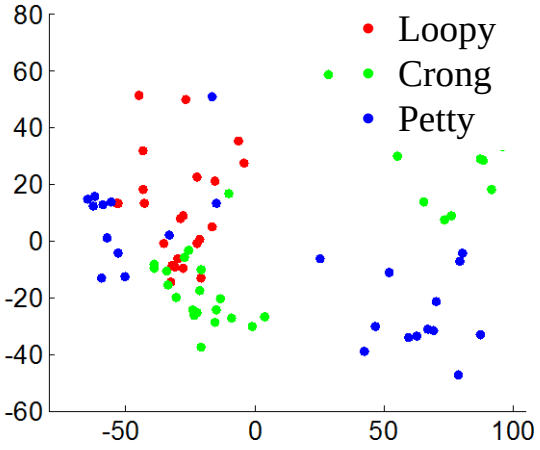


(a) Microcodes

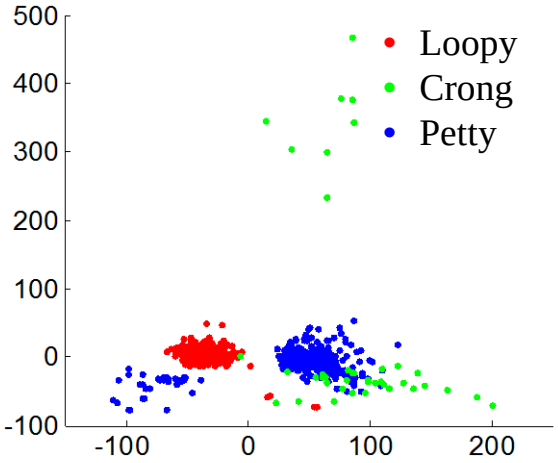


(b) Centroids of $\mathbf{c}^1$-nodes

- **Formation of the character properties (Petty and Loopy)**



(a) Episode 1



(b) Episodes 1~52

Loopy:
- A female beaver
- Girlish and shy
- Likes cooking

Petty:
- A female penguin
- Boyish and active
- Likes sports

Additional Experiment 1

- Proposing 3 different graph search methods

Graph Monte-Carlo Method

- **Learning strategy: The probability of selecting a vertex, $P(v(x))$, for generating hyperedges**
- **Uniform graph MC (UGMC)**
  - Same probability for all vertices

$$P(v(x)) = \frac{1}{\left|\{x \mid x \in \mathbf{x}_+^{(n)}\}\right|}$$ ($\mathbf{x}_+^{(n)}$: the set of variables with the positive value of the n-th instance)

  - Random selection
- **Poorer-richer graph MC (PRGMC)**
  - Prefer more frequently appearing vertices in graphs

$$P(v(x)) = \frac{R^+\{d(v(x))\}}{|\mathbf{x}|},\ d(v(x)) = \sum_{e_i \in G_t} \alpha_i h(v(x), e_i)$$ ($R^+(.)$: a ranking function of d(v(x)) in the ascending order)

  - Smaller and denser graph → fast but premature convergence
- **Fair graph MC (FGMC)**
  - Prefer less frequently appearing vertices

$$P(v(x)) = \frac{R^-\{d(v(x))\}}{|\mathbf{x}|}$$

  - Larger and sparse graph → a diverse representation but slow convergence

($R^-(.)$: a ranking function of d(v(x)) in the descending order)

Scene-to-Subtitle Translation

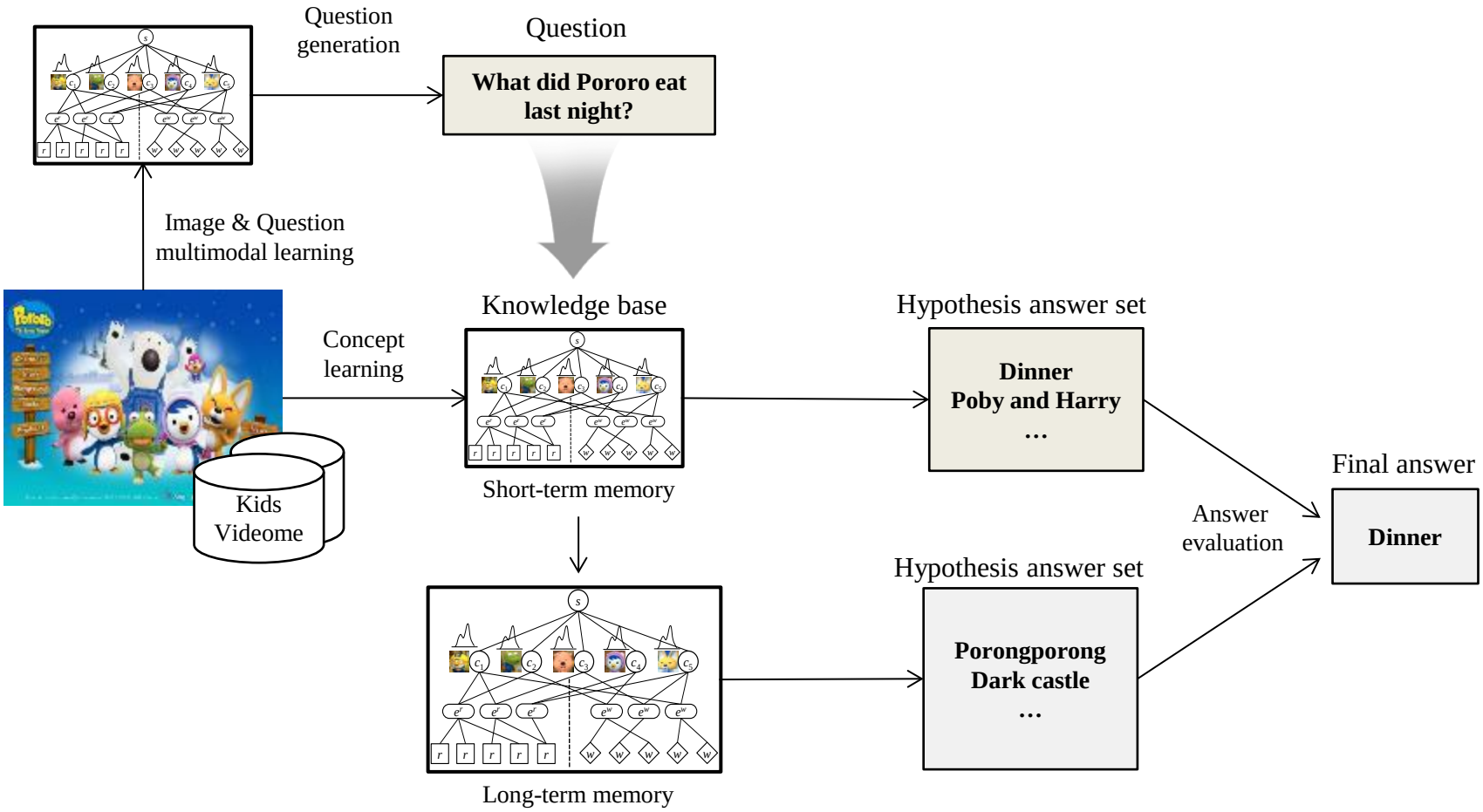
Scenes	Episodes 1~13		Episodes 1~36	
	Original	**And petty taught loopy how to ski.**		
	UGMC	- did you ask me how to swim. - the end how grateful I think she is coming.	UGMC	- Wow petty that how that is not always so loopy taught if i can do fly it well. - How did you have to be that I could ski just.
	PRGMC	- end how was so happy - the end how did you I would	PRGMC	- How did you pororo you. - How about now you can I do not worry.
	FGMC	- To show how big you found - The end how grateful I am petty nice to lose careful	FGMC	- Harry realized that how that is it is dangerous - I thought that how that I could ski just
	SPC	- But how do someone stop. - The end how was it.	SPC	- How about now you can you give me that how that is great. - I will see let see how big.
	Original	**Wow poby, you caught so many already.**		
	UGMC	- Has been caught	UGMC	- Come out if you go in to hear you guys you have got a lot of fish I caught. - You have caught a lot today did you see you later.
	PRGMC	- Has been caught	PRGMC	- Everyone has caught a fish for dinner. - You have caught a lot today did you ask me how.
	FGMC	- What are you guys you have caught a lot. - What happened to ten everyone has caught a lot.	FGMC	- Poby caught a boat a secret that all the wind is so big. - You have caught a fish for the art diving.
	SPC	- Pororo no pororo has caught - She caught the first place	SPC	- You come with his new friend has caught a very interesting book recently - What about pororo has caught a lot of fish

Additional Experiment 2

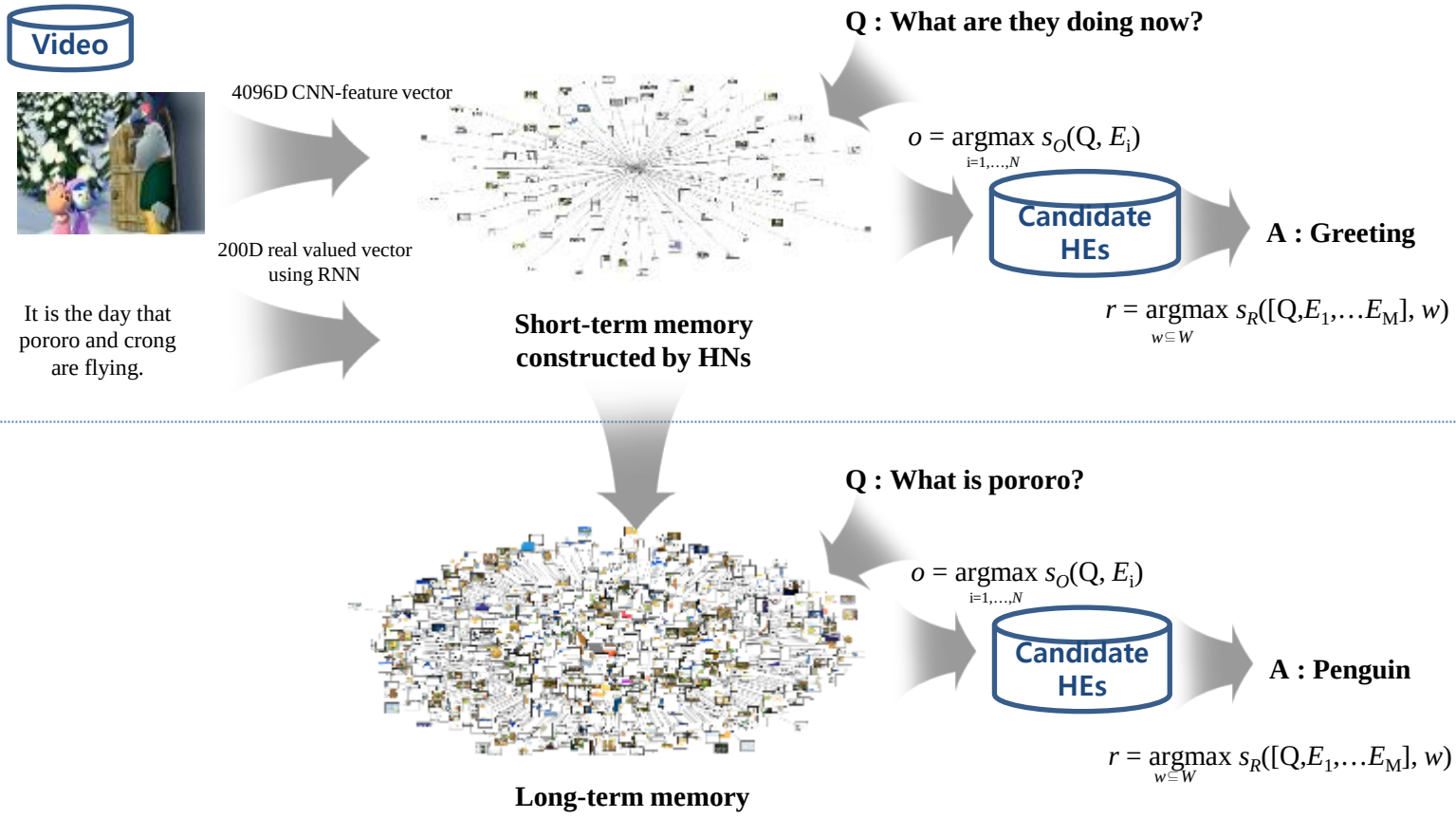
- Proposing Dual Memory for Video Q&A

Video Q&A System (1/2)

- **Overview of Video Q&A System**

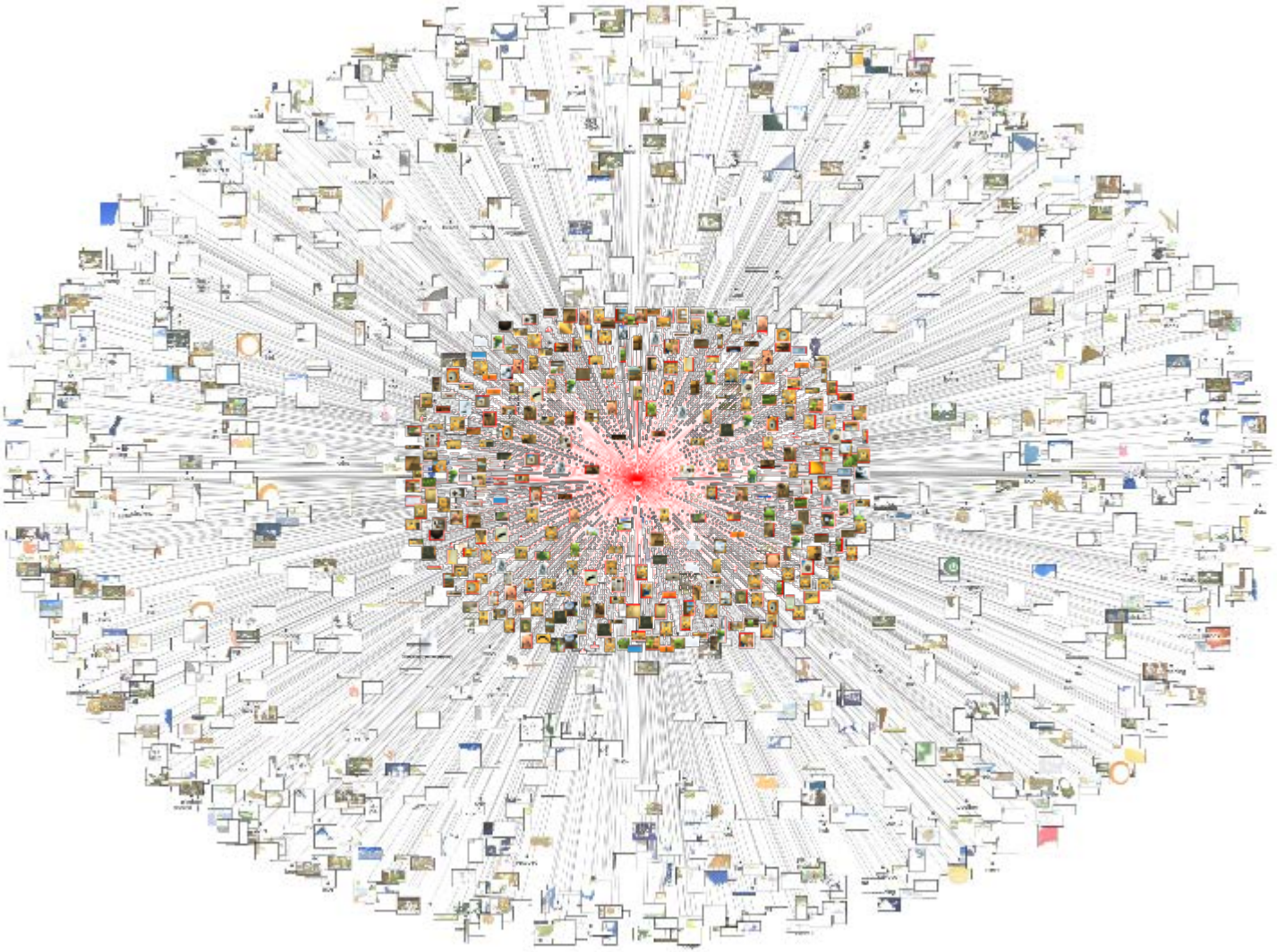


Video Q&A System (2/2)



- s_O is a scoring function to get a candidate hyperedge-answer set
- In this work, cosine distance measuring between words in the question and words in hyperedges will be used as s_O
- s_R is a retrieval function to give answers.

Dual Memory Networks



Evaluation

- **Examples of generated questions & retrieved answers**

* S and L indicate short-term memory and long-term memory

Sequence of Images

Questions	Answers (S/L)
Can pororo swim out too far?	Yes / Yes
How can pororo swim well?	Because they were so loud / His tall height and great strength

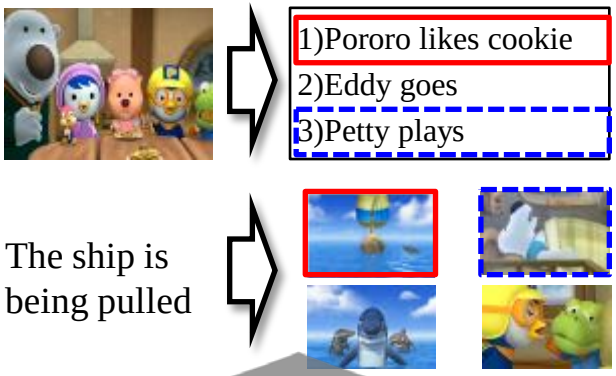
Sequence of Images

Questions	Answers (S/L)
What did eddy trying to go to the playground all day?	Baking / Making a new toy
What does eddy find in her sleep?	Stars / Ball

- **Video turing test**

Dataset	Video Turing Test		
	Pass	Fail	Pass Rate(%)
Pororo	258	542	32.25

Future Works : Human-machine Interactive system



교육용 비디오

비디오 학습

아이

학습 패턴 관찰에 따른 사용자 모델링

피드백

비디오 개념 학습

교육용 로봇