Human like
learning & thinking machine

2021. 12. 06

김보관[1], 진승완[1], 이종현[2]

1. 한양대학교 인텔리전스컴퓨팅학과
2. 한양대학교 디지털의료융합학과

Introduction

NEUROSCIENCE

Using neuroscience to develop artificial intelligence

Combining deep learning with brain-like innate structures may guide network models toward human-like learning

By **Shimon Ullman**

Introduction

“Can machines think?” – Alan Turing (1950)

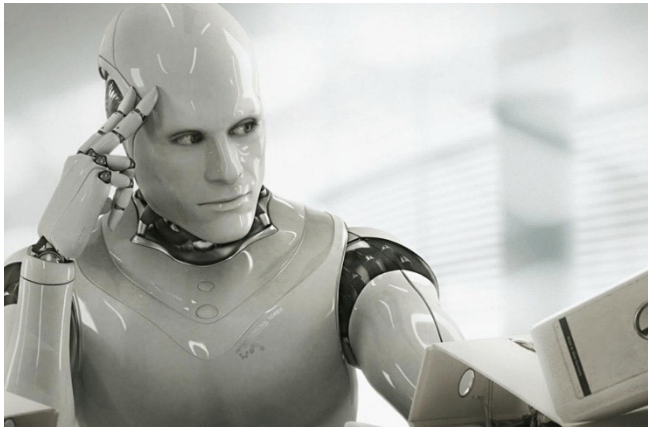
Biological nervous systems
(Brain circuits)

- The only known systems carrying out complex computations
- Modeling cortical circuitry (recently)

Deep net

Introduction

Cognition and general AI



The authors discussed **how additional aspects of brain circuitry could supply cues for guiding network models toward broader aspects of cognition and general AI**.

Introduction

What is the problem in "deep nets"?

- **The key problem ➔ learning** which is the adjustment of the synapses to produce the desired outputs to their input patterns.
  (e.g., supervised learning)

- Successful learning ➔ beyond memorizing the training examples, and be able to generalize, and provide correct outputs to new input patterns.

Introduction

Reinforcement learning (RL)

- In addition to deep nets, **AI models incorporated** another major aspect of **brain-like computations** (i.e., **reinforcement learning**)

- Reinforcement Learning
  - Reward signals in the brain are used to modify behavior
  - **The goal:** is to learn an optimal "policy", which is a mapping from states to actions - maximize an overall measure of the reward obtained over time.

- RL + Deep nets
  - applied in video games, chess and produced stunning results in game playing.

Problem Statement

From the standpoint of using neuroscience to guide AI,,,

- The success of **deep net-based AI is surprising**, given the **highly reduced form** of the network models **compared with cortical circuitry**.
- Almost everything that we know about neurons – structure, types, interconnectivity, and so on – was left out of current deep-net models.
- It is currently **unclear which aspects of the biological circuitry are computationally essential** and could be useful for AI systems.

➔ However, the differences in structure are prominent.

Problem Statement

From the standpoint of using neuroscience to guide AI,,,

- For example, biological neurons are highly complex and diverse in terms of their morphology, physiology, and neurochemistry.
    - The inputs to a typical excitatory pyramidal neuron
      - over complex, highly branching basal and apical dendritic trees
    - Inhibitory cortical neurons
      - come in a variety of different morphologies → perform different functions

➔ None of this heterogeneity and other complexities are included in typical deep-net models.

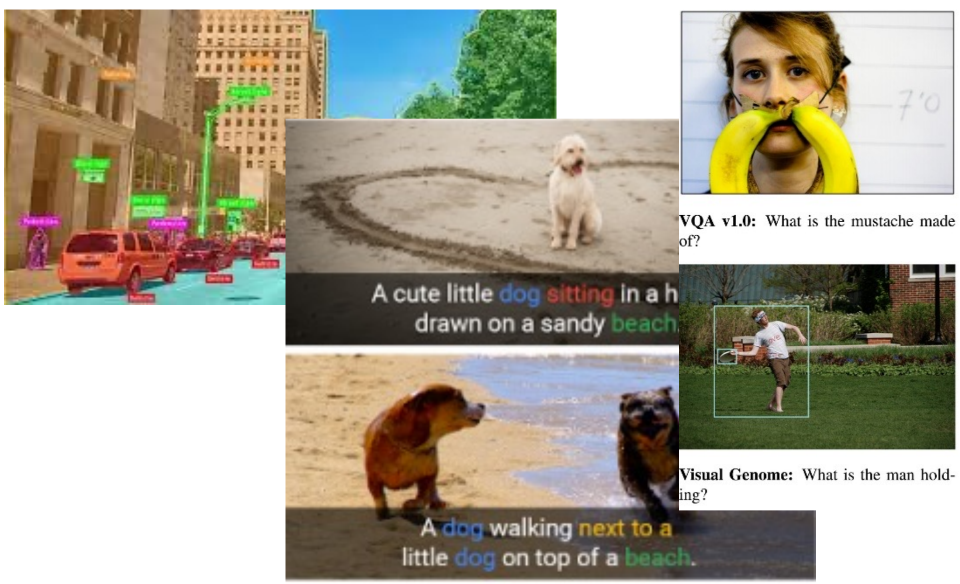
➔ Deep-net models use a limited set of highly simplified homogeneous artificial neurons.

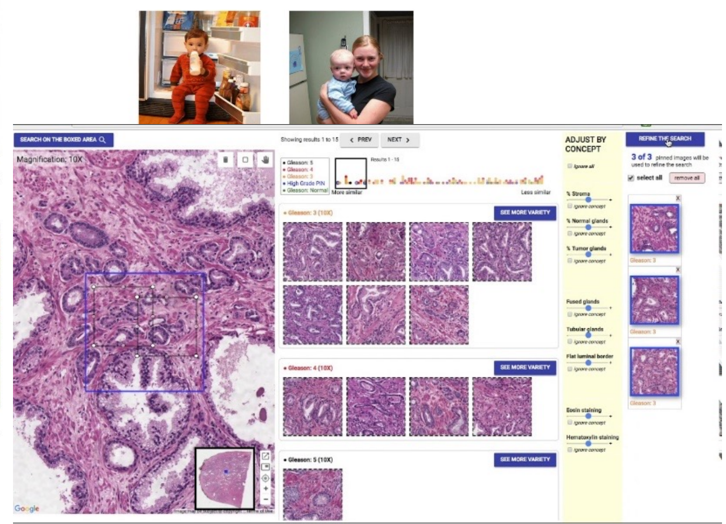
➔ **Cortical circuits in the brain are more complex! (connectivity)**

Problem Statement

The success of deep network-based learning methods

- The notable successes in problems related to real-world perceptual data. (e.g., computer vision & natural language understanding, speech analysis, …)
- Increasing efforts to confront problems that are more cognitive in nature.





Discussion

"Current approaches will be able to produce "real" and human-like understanding results?"

- Radically different directions will be needed to deal with broad aspects of cogn ition, and artificial general intelligence (AGI).
- If the success of current deep network-based models in producing human-like cognitive abilities proves to be limited?

➔ A natural place to look for guidance is **again neuroscience**!

"Can aspects of brain circuitry, over-looked in AI models so far, provide a key to AGI?"

"Which aspects of the brain are likely to be particularly important?"

Discussion

"Current approaches will be able to produce "real" and human-like understanding results?"

- Radically different directions will be needed to deal with broad aspects of cogn ition, and artificial general intelligence (AGI).
- If the success of current deep network-based models in producing human-like cognitive abilities proves to be limited?

  ➔ A natural place to look for guidance is **again neuroscience**!

"Can aspects of brain circuitry, over-looked in AI models so far, provide a key to AGI?"

"Which aspects of the brain are likely to be particularly important?"

**There are at present no obvious answers, yet.
Because our understanding of cortical circuitry is still limited.**

Discussion

Fundamentally difference between brain & deep network-based AI

Brain

- often accomplish complex behavioral tasks with limited training.
  - pre-encoded in the circuitry prior to learning
  - animals' complex navigation task (innate domain-specific mechanisms)
  - development of infants' complex perceptual and cognitive skills in the first months of life, with little or no explicit training.
- complex neural network
  - connectivity in cortical networks (rich sets of connections)
  - local and long-range lateral connectivity
  - top-down connections from high to low levels of the hierarchy

AI modeling

- leans heavily toward the empiricist side.
  - relying primarily on extended learning
  - using large sets of training data
- uses simple & uniform network
- tries to mimic the human brain
  - discovering similar mechanisms

Discussion

Fundamentally difference between brain & deep network-based AI

Brain

- often accomplish complex behavioral tasks with limited training.
  - pre-encoded in the circuitry prior to learning
  - animals' complex navigation task (innate domain-specific mechanisms)
  - development of infants' complex perceptual and cognitive skills in the first months of life, with little or no explicit training.
- complex neural network
  - connectivity in cortical networks (rich sets of connections)
  - local and long-range lateral connectivity

AI modeling

- leans heavily toward the empiricist side.
  - relying primarily on extended learning
  - using large sets of training data
- uses simple & uniform network
- tries to mimic the human brain
  - discovering similar mechanisms

Fast and unsupervised learning is possible because the human cognitive system is equipped, through evolution, with basic innate structures.

Discussion

The superiority of human cognitive learning and understanding

- The innate components → "Proto concepts"
  – leads to the progressive acquisition and organization of complex concepts (with little or no explicit training)

- For example, a particular pattern of image motion can provide a reliable intern al teaching signal for hand recognition.

The detection of hands, and their engagement in object manipulation

Learning system toward detecting direction of gaze

Learning to infer peoples' goals.

guide

play a role in

Conclusion

Combination of preexisting structures & AI models

- Useful preexisting structures could also be adopted in artificial network models to make their learning and understanding more human-like.
  (e.g., mimicking related brain mechanisms)

- In general, the computational problem of "*learning innate structures*" is different from current learning procedures.

➡ ***"Combining the empirical and computational approaches to the probl em is likely to benefit in the long run both neuroscience and AGI."***

Thank you

Contents

Introduction

Introduction

RESEARCH ARTICLES

COGNITIVE SCIENCE

Human-level concept learning through probabilistic program induction

Brenden M. Lake,[1*] Ruslan Salakhutdinov,[2] Joshua B. Tenenbaum[3]

In press at *Behavioral and Brain Sciences.*

Building Machines That Learn and Think Like People

Brenden M. Lake,[1] Tomer D. Ullman,[2,4] Joshua B. Tenenbaum,[2,4] and Samuel J. Gershman[3,4]
[1]Center for Data Science, New York University
[2]Department of Brain and Cognitive Sciences, MIT
[3]Department of Psychology and Center for Brain Science, Harvard University
[4]Center for Brains Minds and Machines

Introduction

Human like learning & thinking machine

- 인간의 학습 및 추론 과정은 단 몇 개의 샘플만으로도 가능
- 기존의 AI 모델은 많은 양의 데이터와 시간을 요구

i)

예시) 빨간 사각형 안에 있는 글자와 동일한 글자를 찾는 과제

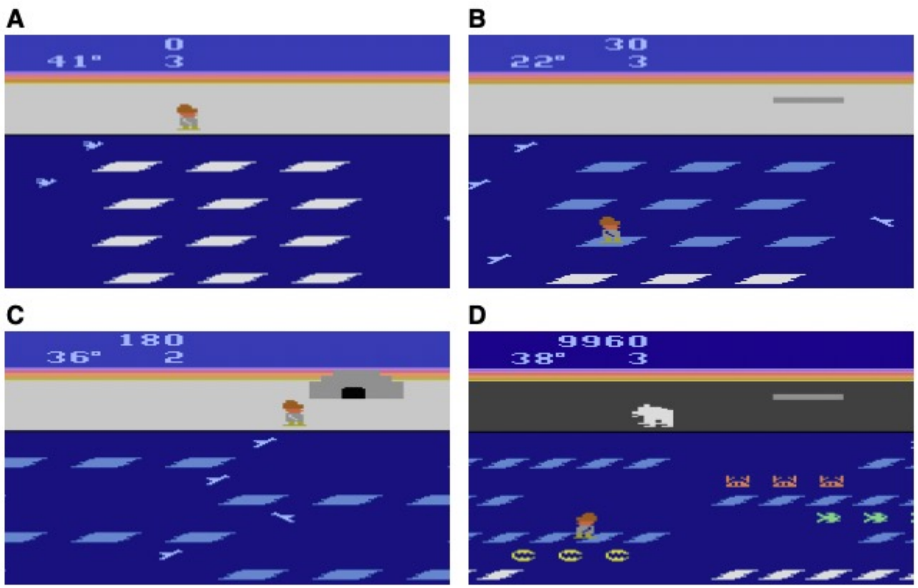
인간은 one-shot 으로 바로 찾아낼 수 있으나 similarity 계산을 통해 과제를 수행하는 Siamese network 같은 경우, 반복을 통한 학습 과정이 요구됨

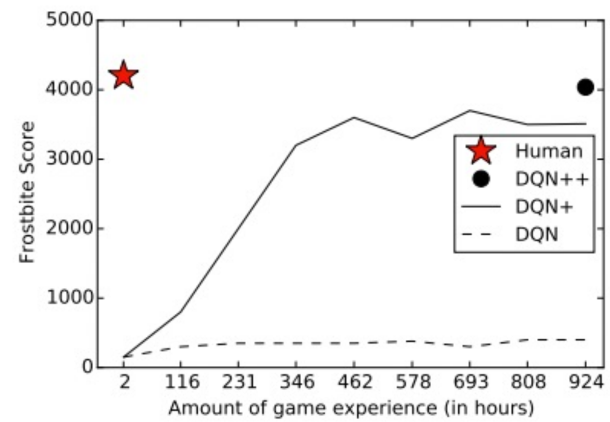
Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). *Human-level concept learning through probabilistic program induction. Science, 350(6266), 1332-1338.* doi:10.1126/science.aab3050

Introduction

Human like learning & thinking machine

- 인간의 학습 및 추론 과정은 단 몇 개의 샘플만으로도 가능
- 기존의 AI 모델은 많은 양의 데이터와 시간을 요구





ing learning speed for people versus Deep Q-Networks (DQNs). Test performance game "Frostbite" is plotted as a function of game experience (in hours at a frame hich does not include additional experience replay. Learning curves (if available) own from different networks: DQN (V. Mnih et al., 2015), DQN+ (Schaul et al., 2016), and DQN++ (Wang et al., 2016). Random play achieves a score of 66.4. The "human starts" performance measure is used (van Hasselt et al., 2016).

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). *Building machines that learn and think like people. Behavioral and Brain Sciences, 40.* doi:10.1017/s0140525x16001837

Introduction

Core ingredients of human intelligence

1. Intuitive physics

2. Intuitive psychology

3. Compositionality

4. Causality

5. Learning to learn

Introduction

Intuitive physics

- 자신의 주변 환경, 즉 물리 시스템에 대한 이해
- 물리 시스템을 기반으로 model base simulation
  - 물체는 위에서 아래로 떨어진다
  - 얼음 위에서는 물체가 미끄러진다

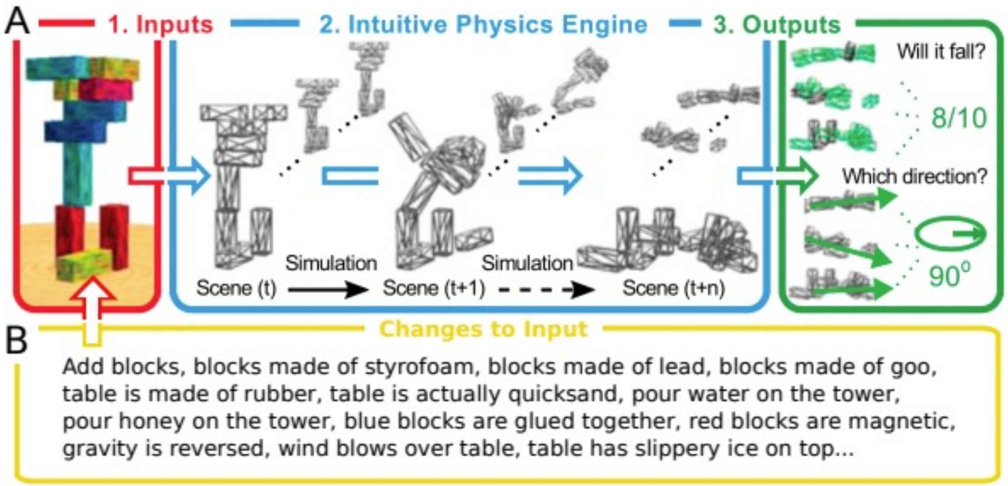
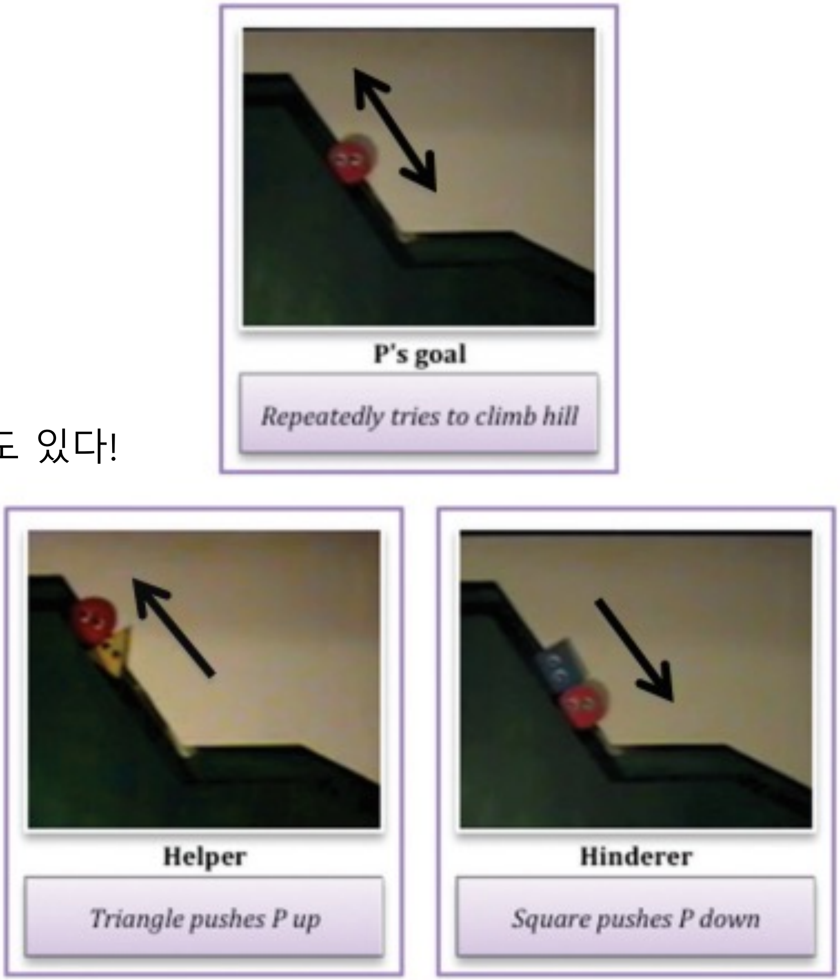


Figure 4: The intuitive physics-engine approach to scene understanding, illustrated through tower stability. (A) The engine takes in inputs through perception, language, memory and other faculties. It then constructs a physical scene with objects, physical properties and forces, simulates the scene's development over time and hands the output to other reasoning systems. (B) Many possible 'tweaks' to the input can result in much different scenes, requiring the potential discovery, training and evaluation of new features for each tweak. Adapted from Battaglia et al. (2013).

Introduction

Intuitive psychology

- Theory of mind (마음 이론)
- 타인에 대한 공감과 이해 능력
- 타인에게도 나와 같은 마음이 있음을 인지
  - 아이들이 싸울 때: 너 마음만 있냐? 내 마음도 있다!



Hamlin, J. K. (2013). *Moral Judgment and Action in Preverbal Infants and Toddlers. Current Directions in Psychological Science, 22(3), 186-193.* doi:10.1177/0963721412470687

Introduction

Intuitive psychology

- Theory of mind (마음 이론)
- 타인에 대한 공감과 이해 능력

a woman riding a horse on a dirt road

an airplane is parked on the tarmac at an airport

a group of people standing on top of a beach

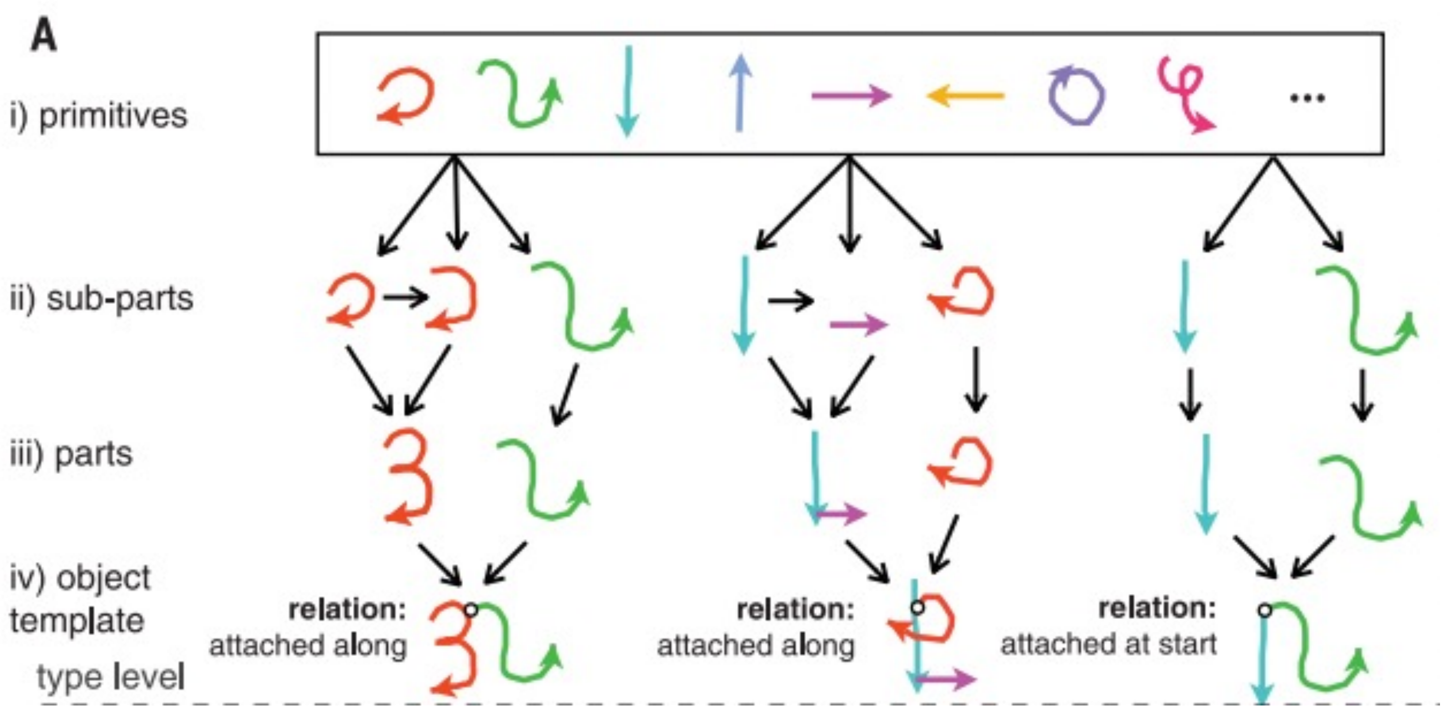
Figure 6. Perceiving scenes without intuitive physics, intuitive psychology, compositionality, and causality. Image captions are generated by a deep neural network (Karpathy & Fei-Fei 2017) using code from github.com/karpathy/neuraltalk2. Image credits: Gabriel Villena Fernández (left), TVBS Taiwan/Agence France-Presse (middle), and AP Photo/Dave Martin (right). Similar examples using images from Reuters news can be found at twitter.com/interesting_jpg.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). *Building machines that learn and think like people. Behavioral and Brain Sciences, 40.* doi:10.1017/s0140525x16001837

Introduction

Compositionality

- 유한 개의 primitive 속성들의 조합으로 무한 개의 조합을 만들어 낼 수 있음
- CNN layer 의 filter 들을 조합해서 abstract information을 처리하는 것과 유사한 맥락



Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). *Human-level concept learning through probabilistic program induction. Science, 350(6266), 1332-1338.* doi:10.1126/science.aab3050

Introduction

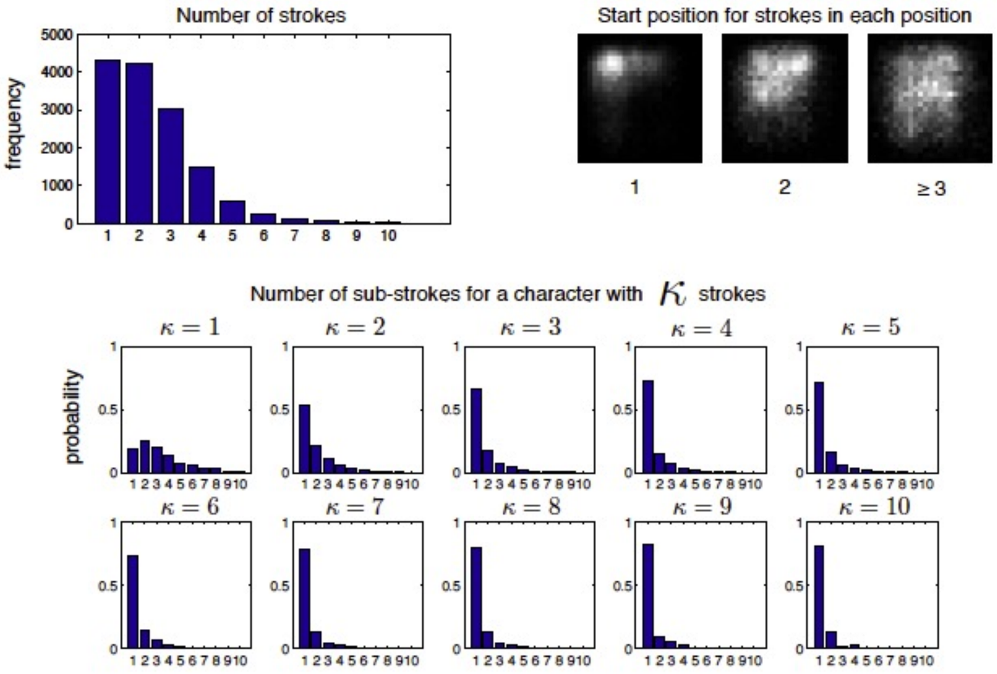
Causality

- 각각의 속성들이 서로 인과적인 관계를 맺고 있음
- 한자의 획순

Introduction

Causality

- 각각의 속성들이 서로 인과적인 관계를 맺고 있음
- 선호되는 모양새, 좌우/상하 선호 방향, 결합되는 조합들 관계



Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). *Human-level concept learning through probabilistic program induction. Science, 350(6266), 1332-1338.* doi:10.1126/science.aab3050

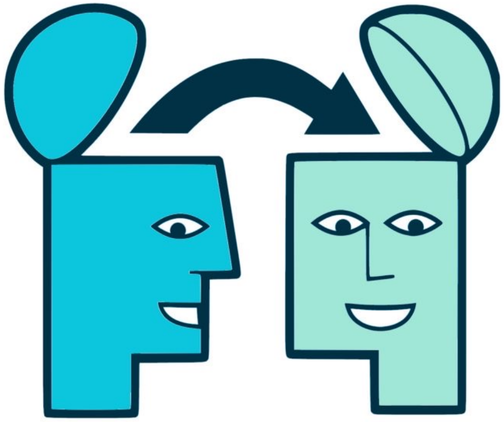
Introduction

Learning to learn

- Transfer learning, pretraining, multi-task learning, representational learning

- 사전에 학습하는 방법을 미리 배운 상태에서 다른 과제를 새로이 배우는 것을 총칭

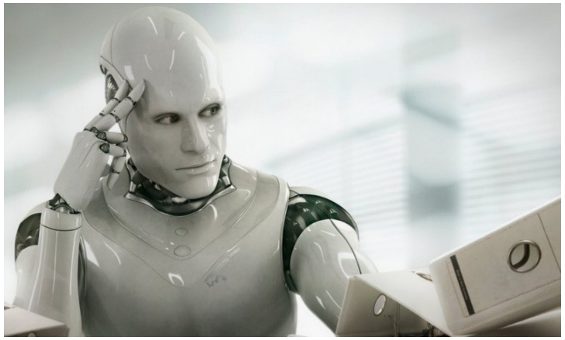
- 유사한 경험과 축적된 지식을 바탕으로 새로운 과제에 빠르게 적응



Introduction

Human like learning machine

1. Intuitive physics
2. Intuitive psychology
3. Compositionality
4. Causality
5. Learning to learn



Experiments

RESEARCH ARTICLES

COGNITIVE SCIENCE

Human-level concept learning through probabilistic program induction

Brenden M. Lake,[1*] Ruslan Salakhutdinov,[2] Joshua B. Tenenbaum[3]

Experiments

Bayesian program learning (BPL)

- Omniglot dataset 을 학습하는 모델

핵심 파트

1. Generating character types
2. Generating character tokens
3. Learning-to-learn motor programs

Experiments

Bayesian program learning (BPL), **Generating character types**

A character type $\psi = \{\kappa, S, R\}$ is defined by a set of κ strokes $S = \{S_1, ..., S_\kappa\}$ and spatial relations $R = \{R_1, ..., R_\kappa\}$ between strokes. The joint distribution can be written as

$$P(\psi) = P(\kappa) \prod_{i=1}^{\kappa} P(S_i) P(R_i | S_1, ..., S_{i-1}), \tag{S1}$$

Algorithm 1 Generate a new character type

procedure GENERATETYPE
- $\kappa \leftarrow P(\kappa)$ ▷ Sample the number of strokes
- **for** $i = 1 \ldots \kappa$ **do**
  - $n_i \leftarrow P(n_i | \kappa)$ ▷ Sample the number of sub-strokes
  - $S_i \leftarrow$ GENERATESTROKE(i, n_i) ▷ Sample stroke
  - $\xi_i \leftarrow P(\xi_i)$ ▷ Sample relation to previous strokes
  - $R_i \leftarrow P(R_i | \xi_i, S_1, ..., S_{i-1})$ ▷ Sample relation details
- **end for**
- $\psi \leftarrow \{\kappa, R, S\}$
- **return** @GENERATETOKEN(ψ) ▷ Return program handle

end procedure

procedure GENERATESTROKE(i, n_i)
- $z_{i1} \leftarrow P(z_{i1})$ ▷ Sample the identity of the first sub-stroke
- **for** $j = 2 \ldots n_i$ **do**
  - $z_{ij} \leftarrow P(z_{ij} | z_{i(j-1)})$ ▷ Sample the identities of the other sub-strokes
- **end for**
- **for** $j = 1 \ldots n_i$ **do**
  - $x_{ij} \leftarrow P(x_{ij} | z_{ij})$ ▷ Sample a sub-stroke's control points
  - $y_{ij} \leftarrow P(y_{ij} | z_{ij})$ ▷ Sample a sub-stroke's scale
  - $s_{ij} \leftarrow \{x_{ij}, y_{ij}, z_{ij}\}$
- **end for**
- $S_i \leftarrow \{s_{i1}, \ldots, s_{in_i}\}$ ▷ A complete stroke definition
- **return** S_i

end procedure

Experiments

Bayesian program learning (BPL), **Generating character tokens**

The token-level variables, $\theta^{(m)} = \{L^{(m)}, x^{(m)}, y^{(m)}, R^{(m)}, A^{(m)}, \sigma_b^{(m)}, \epsilon^{(m)}\}$, are distributed as

$$P(\theta^{(m)}|\psi) = P(L^{(m)}|\theta^{(m)}_{\setminus L^{(m)}}, \psi) \prod_i P(R_i^{(m)}|R_i) P(y_i^{(m)}|y_i) P(x_i^{(m)}|x_i) P(A^{(m)}, \sigma_b^{(m)}, \epsilon^{(m)}).$$

Algorithm 2 Run the stochastic program of type ψ to make an image

procedure GENERATETOKEN(ψ)
- **for** $i = 1 \ldots \kappa$ **do**
  - $R_i^{(m)} \leftarrow R_i$ ▷ Directly copy the type-level relation
  - **if** $\xi_i^{(m)} =$ 'along' **then**
    - $\tau_i^{(m)} \leftarrow P(\tau_i^{(m)}|\tau_i)$ ▷ Add variability to the attachment along the spline
  - **end if**
  - $L_i^{(m)} \leftarrow P(L_i^{(m)}|R_i^{(m)}, T_1^{(m)}, ..., T_{i-1}^{(m)})$ ▷ Sample stroke's starting location
  - **for** $j = 1 \ldots n_i$ **do**
    - $x_{ij}^{(m)} \leftarrow P(x_{ij}^{(m)}|x_{ij})$ ▷ Add variability to the control points
    - $y_{ij}^{(m)} \leftarrow P(y_{ij}^{(m)}|y_{ij})$ ▷ Add variability to the sub-stroke scale
  - **end for**
  - $T_i^{(m)} \leftarrow f(L_i^{(m)}, x_i^{(m)}, y_i^{(m)})$ ▷ Compose a stroke's pen trajectory
- **end for**
- $A^{(m)} \leftarrow P(A^{(m)})$ ▷ Sample global image transformation
- $\epsilon^{(m)} \leftarrow P(\epsilon^{(m)})$ ▷ Sample the amount of pixel noise
- $\sigma_b^{(m)} \leftarrow P(\sigma_b^{(m)})$ ▷ Sample the amount blur
- $I^{(m)} \leftarrow P(I^{(m)}|T^{(m)}, A^{(m)}, \sigma_b^{(m)}, \epsilon^{(m)})$ ▷ Render and sample the binary image
- **return** $I^{(m)}$

end procedure

Experiments

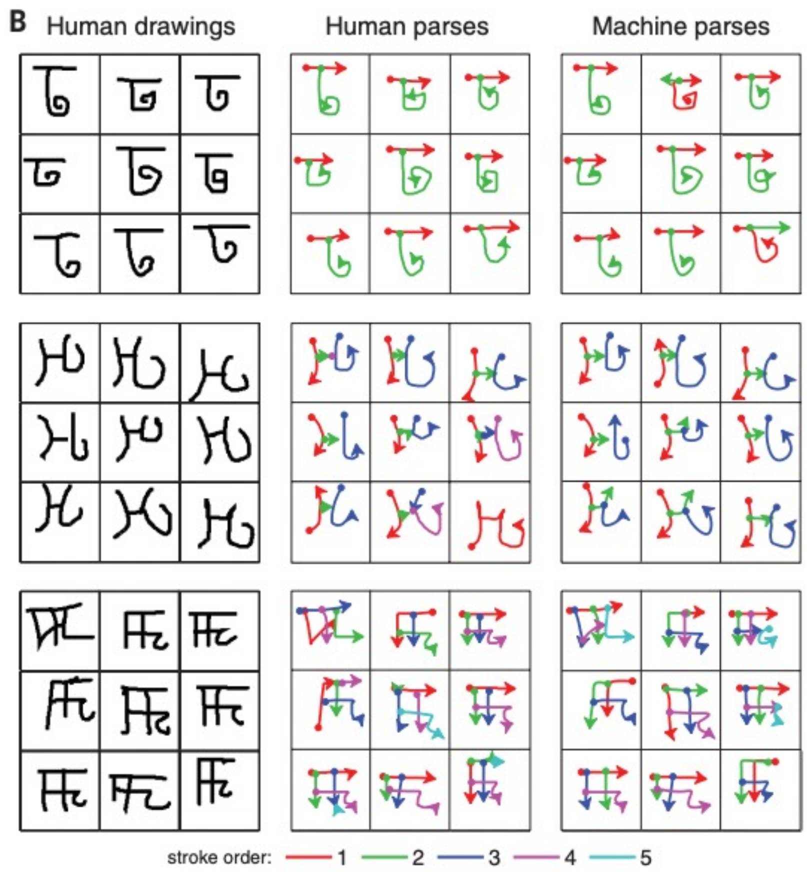
Bayesian program learning (BPL), **Learning-to-learn motor programs**

- Learning primitives
- Learning start positions
- Learning relations and token variability
- Learning image parameters

Results

Results

Bayesian program learning (BPL)



Results

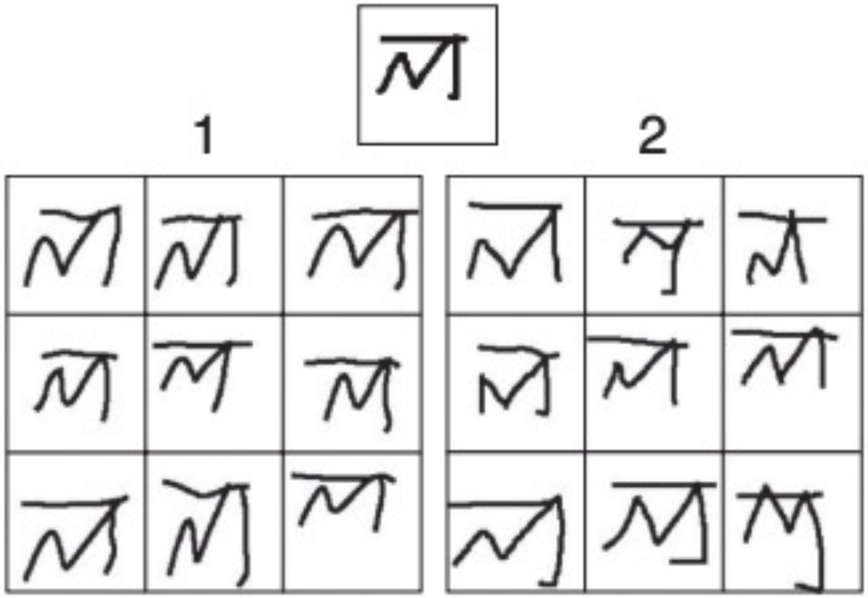
Bayesian program learning (BPL)

One shot classification task

Results

Bayesian program learning (BPL)

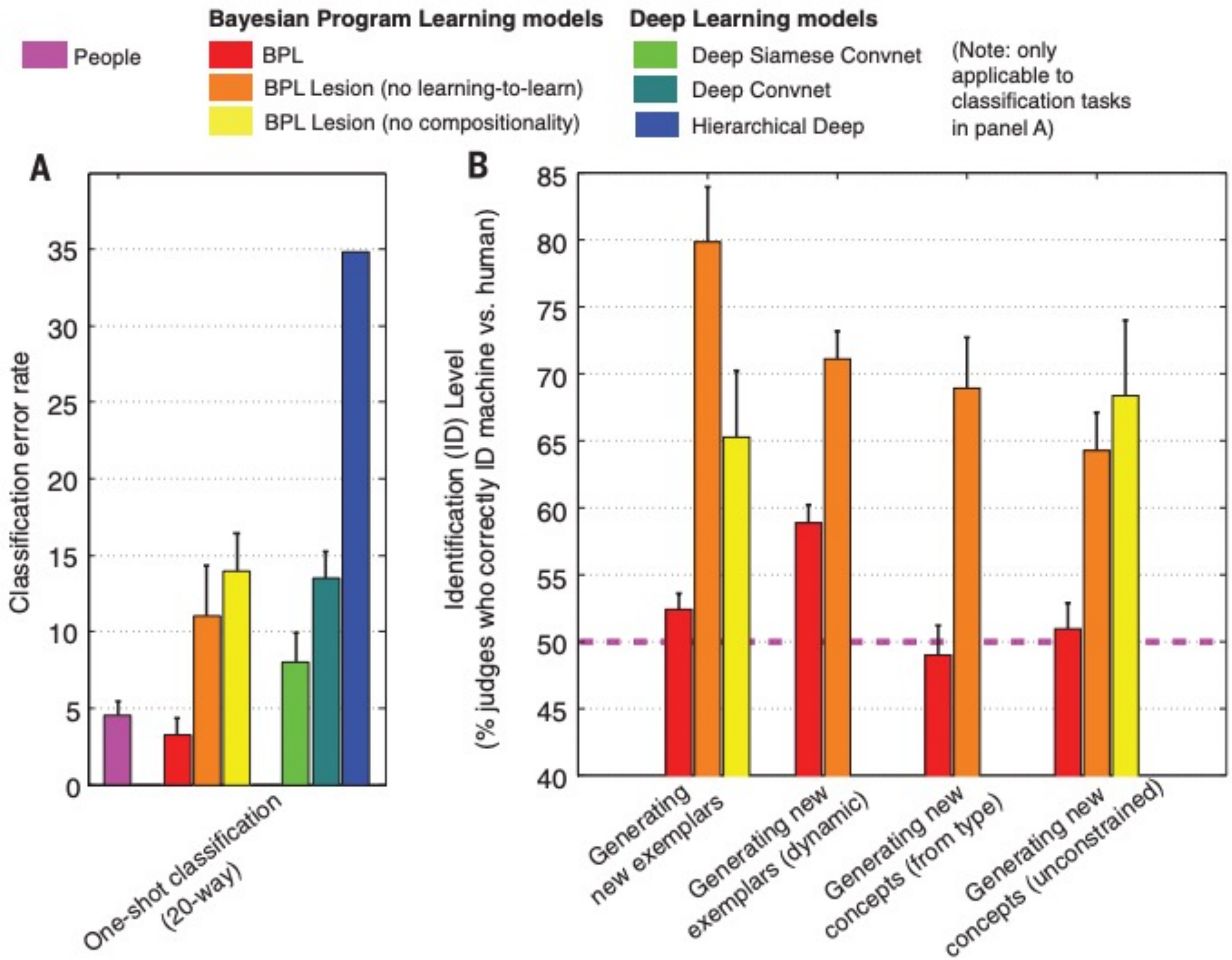
Visual turing test, generation task



Human or Machine?

Results

Bayesian program learning (BPL)



Discussion

Discussion

Human like learning & thinking machine 의 가능성

- Physics
    - 행동의 제약 조건, 효율적 움직임 등에 대한 이해
- Psychology
    - 다른 agent의 의도, 행동 이해 / game 등의 task 에서 유리

Discussion

Human like learning & thinking machine 의 가능성

- Composition
  - 일부의 primitives 를 생성하고 이를 조합
- Causality
  - Primitive 간 관계, 제약, 순서 학습
- Learning to learn
  - 과거의 경험, 노하우를 이용해 다른 과제에 빠르게 적용

Q & A

Thank you for listening