On the Complexity of Bayesian Generalization

Yu-Zhe Shi *,[1], Manjie Xu *,[1,2], John E. Hopcroft [3], Kun He [4], Joshua B. Tenenbaum [5]
Song-Chun Zhu [1,2], Ying Nian Wu [6], Wenjuan Han [7,8], Yixin Zhu [1]

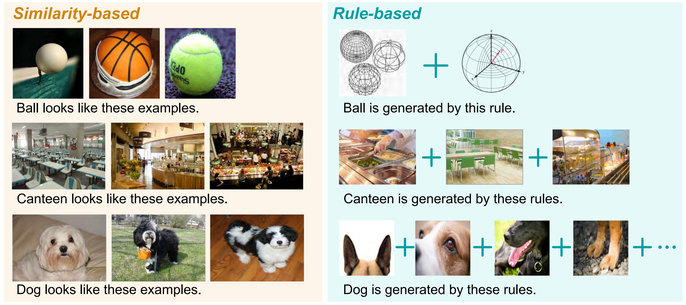
[1]Peking University, [2]National Key Laboratory of General Artificial Intelligence, BIGAI, [3]Cornell University,
[4]Huazhong University of Science and Technology, [5]MIT, [6]UCLA, [7]Beijing Jiaotong University, [8]CUPK

ICML 40 Years — International Conference On Machine Learning

Concept Representation and Generalization

> Concepts can be described either directly by similar examples or indirectly by a set of related rules. Here, we demonstrate this intuition through the concepts of ball, canteen, and dog.

> An illustration of similarity- and rule-based generalization. The former is similar to word learning (Xu & Tenenbaum, 2007; Jiang et al., 2023): Given very few examples of known concept dax, tell which is most likely to be dax in unseen examples. The latter is akin to concept learning (Salakhutdinov et al., 2012; Zhang et al., 2019a): Given a rule tufa over two known concepts, tell how tufa generates the examples of unknown concepts.



Taking-away Messages

Problem: When the problem space scales up (e.g., using data collected from the natural world), is there a unified concept representation that combines the two established modes (i.e., rule- and similarity-based)?If it does, how does the generalization shift between the two modes w.r.t. the complexity of concepts?

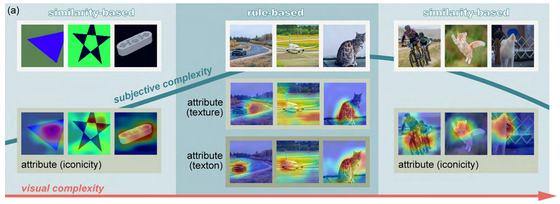
Solution: We disassemble the problem into two on the basis of Marr's representational level and computational level, respectively:
(i) How does the complexity change when a visual concept is mapped to the representation space?
(ii) How does the complexity of representation affect the shift between rule- and similarity-based generalization?

Method: We analyze the complexity of concept generalization in Marr's representational and computational level respectively, with the rational analysis of Representativeness of Attribute (RoA).

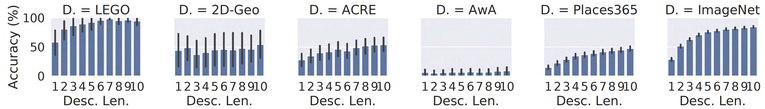
Result: (i) Representation: the subjective complexity significantly falls in an inverted-U relation with the increment of visual complexity.
(ii) Computation: rule-based generalization is significantly positively correlated with the subjective complexity of the representation, while the trend is the opposite for similarity-based generalization.

Significance: We provide first pieces of evidence that people may name natural visual concepts in a rational fashion according to the representativeness of attributes.

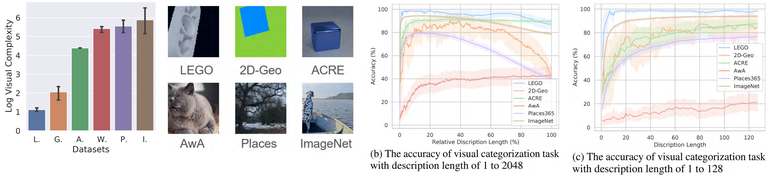
Main Results



> As concept visual complexity increases, concept subjective complexity first increases, then decreases—and the computation mode shifts from similarity to rules as subjective complexity increases.



(a) The accuracy of visual categorization task with description length from 1 to 12



(b) The accuracy of visual categorization task with description length of 1 to 2048

(c) The accuracy of visual categorization task with description length of 1 to 128

Methods

We rewrite the category prediction considering the attribute as a latent variable:

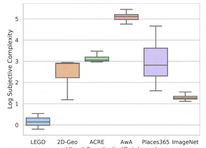
$$P(c|X;\theta) = \sum_{z\in\mathcal{Z}} P(c,z|X;\theta) = \sum_{z\in\mathcal{Z}} P(c|z)P(z|X;\theta).$$

We formally define the RoA as:

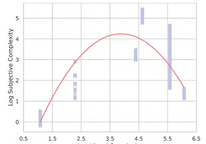
$$\mathrm{RoA}(z,c) = \log \frac{P(z|c;\theta)}{\sum_{\hat{c}\neq c} P(\hat{c}|z;\theta)} \approx \log \frac{P(z|c)}{\sum_{\hat{c}\neq c} P(\hat{c})P(z|\hat{c})}$$

The generalization function is given by:

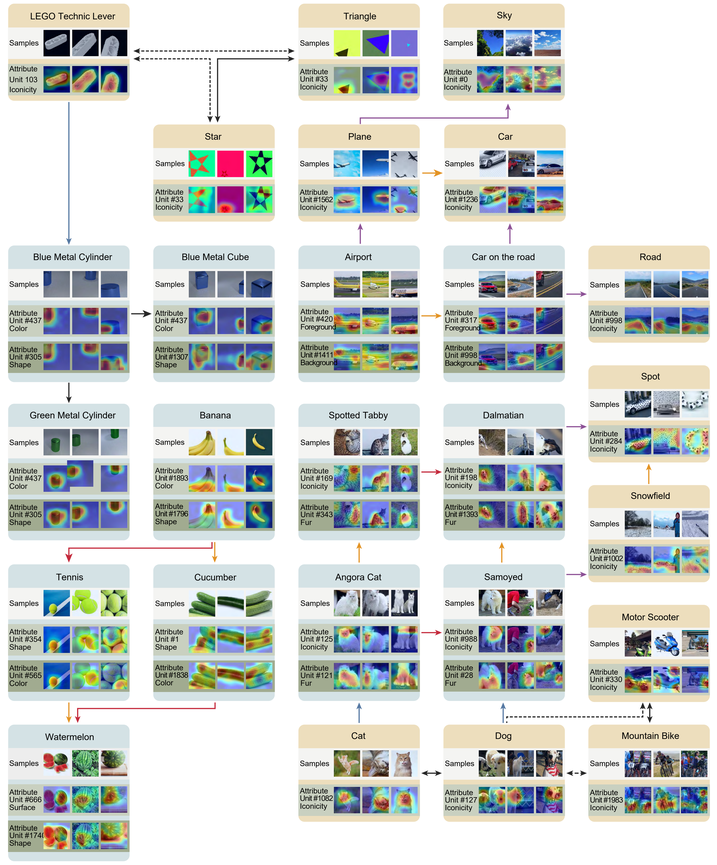
$$P(c'|\hat{X}) = \sum_{z\in\mathcal{Z}} P(c'|z)P(z|\hat{X};\theta) = \sum_{z\in\mathcal{Z}} \frac{P(c')P(z|c')}{\sum_{c\in\mathcal{C}} P(c)P(z|c)} P(z|\hat{X};\theta)$$
$$\propto \underbrace{P(c')}_{\text{uninformative prior}} \sum_{z\in\mathcal{Z}} \underbrace{\exp\left(\mathrm{RoA}(z,c')\right)}_{\text{informative prior}} P(z|\hat{X};\theta)$$



(d) The average subjective complexity on different datasets



(e) The estimated inverted-U relation between visual and subjective complexity



A landscape of similarity- and rule-based generalization over concepts with relatively high and low subjective complexity, considering both concept complexities and concept hierarchy. Bidirectional arrows denote the similarity judgment between concepts, wherein concepts linked by solid lines are more similar than those linked by dashed lines. Arrows denote rules over concepts. Rule-based generalization in basic-level generalizes given rules to unknown rules. Similarity shifts to rules when the sample hierarchy goes from superordinate-level to subordinate-level (e.g., from "block" to "blue cylinder", from "cat" to "angora cat"). Rules shift to similarity as the sample hierarchy goes from subordinate-level to superordinate-level (e.g., from "car on the road" to "car", from "dalmatian" to "spot"). We also note a confusing similarity judgment between blue cylinder, blue cube, and green cylinder.