On Semantic Cognition, Inductive Generalization, and Language Models

Kanishka Misra (kmisra@purdue.edu)
Purdue University, USA

AAAI 2022 Doctoral Consortium

Overview

Question: To what extent do models that only rely on language experience learn about everyday concepts and categories?

Approach: Study the **synthetic semantic knowledge of language models** by investigating how they perform *property induction*: generalization of novel information about concepts and properties.

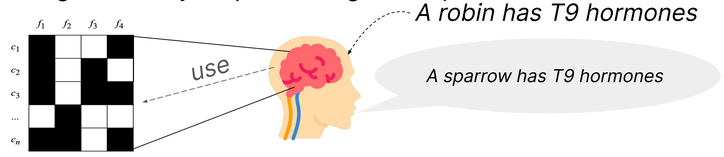
Motivation: Property-inductions made by humans have provided context within which cognitive scientists have explored the nature and organization of human conceptual knowledge (see below)

Inductive Generalizations
about Concepts and Properties

- Humans readily go beyond available data to project novel information about concepts and properties, by relying on their knowledge about the world.

- E.g., when told ***robins*** *have T9 hormones*, humans prefer projecting it to **all birds** more strongly than to **all animals**.

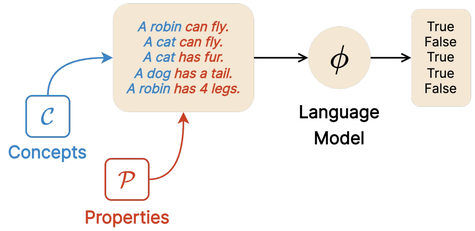
- Cognitive Scientists have assimilated numerous phenomena, revealing insights about how humans use the semantic knowledge that they acquire during development.



Method: A Property Induction Framework for Language Models

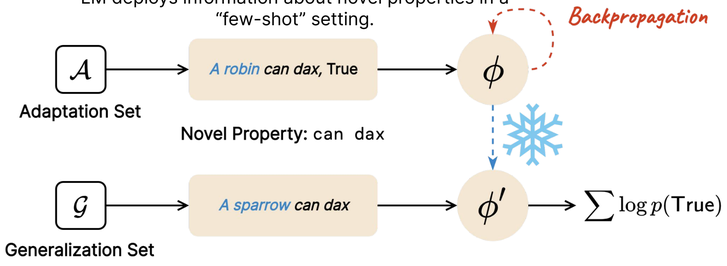
Stage 1: Property Judgment

LM is trained to assess the compatibility of properties with concepts, by assessing truths of "property knowledge" sentences



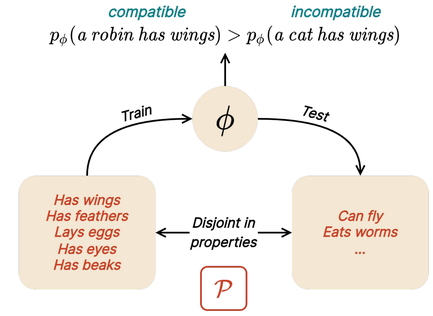
Stage 2: Inductive Generalization

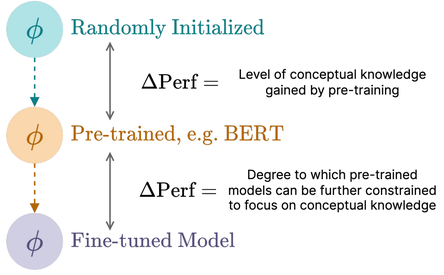
LM deploys information about novel properties in a "few-shot" setting.



Base Property Knowledge in LMs
Research Question 1

To what extent do LMs learn the compatibility of concepts and properties?

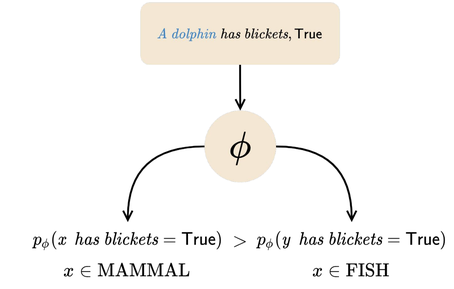




Characterizing Inductions made by LMs
Research Question 2

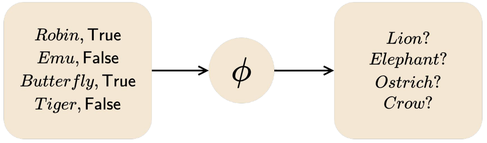
How do LMs deploy novel information about concepts and their properties?

Taxonomic Category-membership vs. Feature overlap



Reasoning about Implicit Property Knowledge

Implicit Salient Property: *can fly/can be airborne*

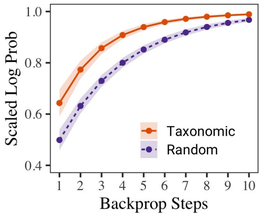


Can the LM detect presence of can fly implicitly and generalize based on it?

Dynamics of LM Inductions
Research Question 3

How do the dynamics of induction relate to the representational space of LMs?

Dynamics of how *has blickets* is generalized:



From **Cow** to other **mammals**

From **Cow** to other **non-mammals**

Representational Geometry of *cow*, a random *mammal*, and a random *non-mammal*