# THOUGHTS ON THE APPLICABILITY OF MACHINE LEARNING TO SCIENTIFIC DISCOVERY AND POSSIBLE FUTURE RESEARCH DIRECTION [PERSPECTIVE]

Shiro Takagi (Independent Researcher)

Profile Paper

**Automation of general scientific processes** ⇄ **Automation for specific scientific problems**

**Hypothesis candidate generation** → **Hypothesis space exploration** → • • • →

## 1. Analogy

**Why Analogy?**

- Generating hypotheses needs to map a known structure to an unknown one.
- The main goal of science is to understand the laws of nature, which requires abstract and systematic relations.

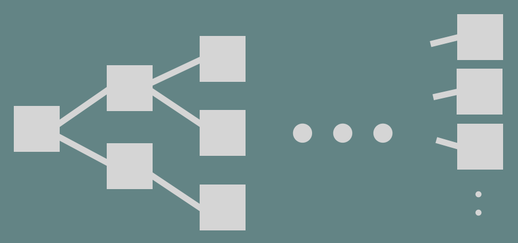
**What to do?**

- Benchmarks like ImageNet and GLUE are driving forces for research in ML.
- We should create datasets and tasks of analogies specific to scientific discovery.
- E.g. regarding research papers as pairs of “problem” and “solution,” learn the relations from the paper dataset.
- Record all intermediate outputs of a research process on a publicly available server to create research process datasets.

*[Chan+ 2018] *A Mixed Initiative System for Finding Analogies Between Research Papers*

## 2. Goal-Conditioned Planning

- To elucidate an unknown, you must study a sequence of sub-unknowns.
- Thus, we should develop a way to break down a goal into a chain of sub-goals.
- Applying RL to ATP might give an insight into learning reasoning in science.
- Estimating the importance of a study is necessary.



## 3. Language Understanding

- Scientific findings in human history are in texts.
- Grounded language learning is necessary to map understanding in language to that of nature.
- Systematic generalization is vital for systematic thinking in science.