From Learning To Discovery

Prof. Shou-de Lin

The Progress of Human Intelligence

Memorizing

- kids, adults, animals can all memorize things
- such capability diminishes through age
- Machine does better than humans

Learning

- A process of generalizing. Can be performed by most adults.
- Machine can do as good as humans

Discovering

- A highly-intelligent mind process
- Not all human can perform discovery, but generally they are better than machines
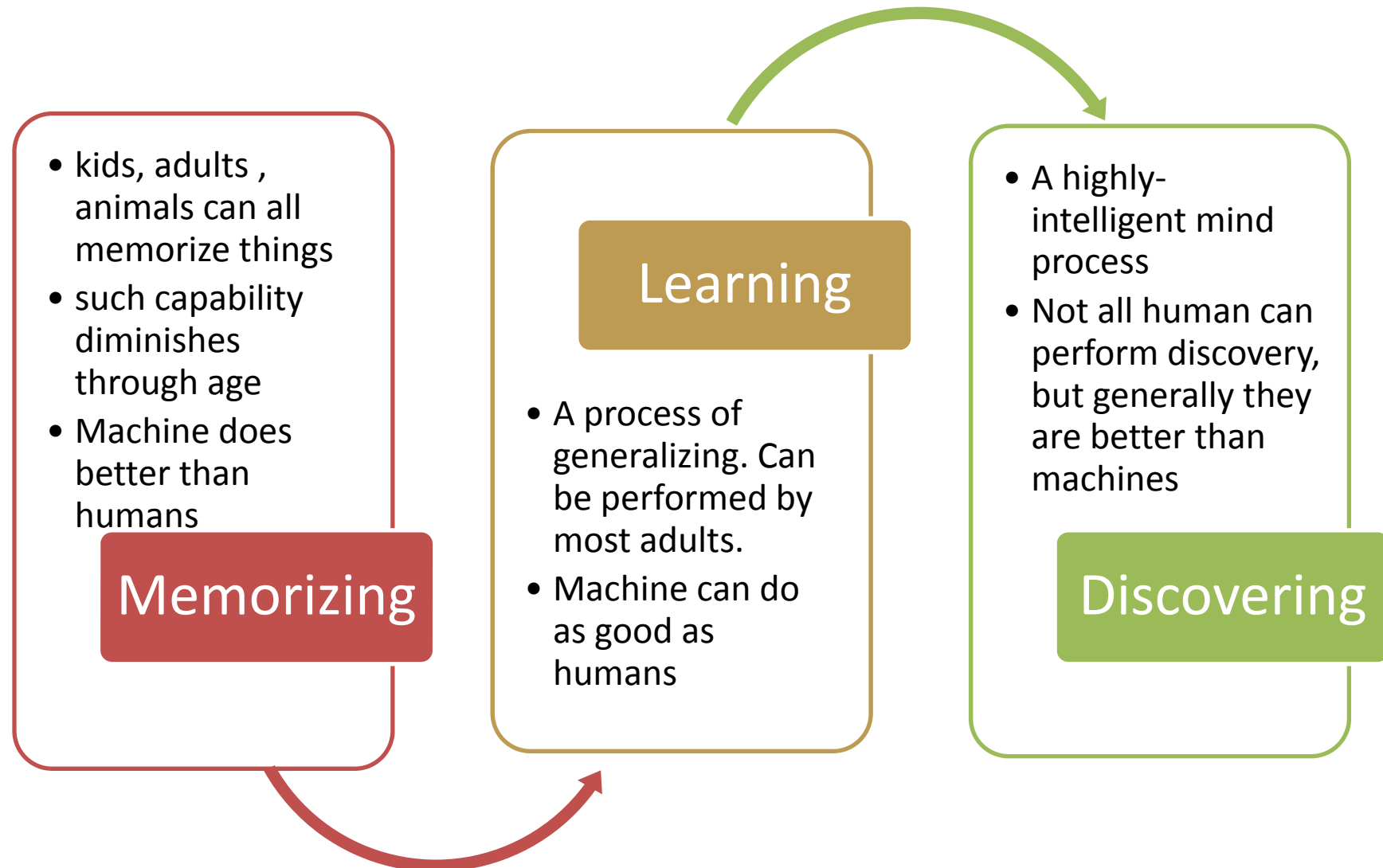
From HL to HD

Human Learning (HL)：

- "The act, process, or experience of gaining knowledge".
- E.g. You learned something about Bayesian Learning through attending a lecture

Human Discovery (HD):

- To obtain knowledge or awareness of something not known before."
- E.g. Shao-Chuan found the connection between Bayesian Prior and add-one smoothing (nobody told him that)

From ML to MD

Machine Learning (supervised)

- Rule-based Learning
- Supervised Learning

Machine Discovery (unsupervised)

- Discover useful and interesting things out of data (KDD)
- Typical technique used: clustering, outlier detection, association rule finding, EM (MLE or MAP)

電腦發明，Prof. Shou-de Lin

Why Machine Discovery?

- Def: Exploit machines to perform or assist human beings to perform discovery
- Benefits:
  - High reward
  - Tolerates low precision and recall
- Challenge:
  - Much more difficult than learning
  - Hard to evaluate the results
    - It is sometimes hard to tell whether the discovered results are correct (because something unknown is found, and there is no gold standard)

Machine Discovery System in History

Math: Graph Theory（GRAFFITI，1990 - , Fajtlowicz）

- it makes conjectures in Graph Theory such as
  chromatic_number(G) + radius(G) =< max_degree(G) + frequency_of_max_degree(G)
- It has discovered > 800 conjectures, which have lead to 60 papers and 1 thesis
  publications to proof or disproof them

Physica: ASTRA (1998 - , Kocabas and Langley)

- suggested novel and interesting fusion reactions, and generated reaction pathways for helium, carbon, and oxygen that do not appear in the scientific literature

Astronomy: SKICAT (1995 - , Kennefick)

- aiming at detecting quasars in digital sky surveys, and has found 5 new quasars.
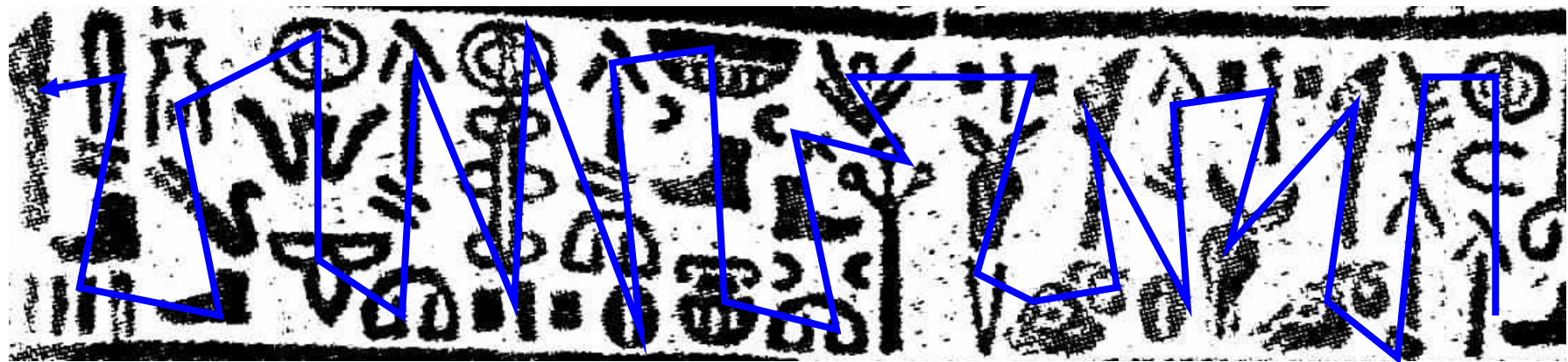
Discovering the Linear Writing Order of a Two-Dimensional Ancient Script

Prof Shou-de Lin

sdlin@csie.ntu.edu.tw

NTU CSIE/GINM

The Luwian script

It's a Hieroglyphic script used in city-states of Southern Anatolia and Northern Syria, from 1300 BCE to 700 BCE.

It's two dimensional!!

Why?

- We want to test how well the machine can do while very little knowledge is available

- It is fun

- It's a real problem that hasn't been solved by anybody.

What data do we have?

24 pages of non-digitized scripts

On this page, we have:

9 lines (avg 80 symbols/line)

73 distinct symbols