

01

Automatic tailor-made data analysis

Generating probabilistic models using structure information

Abstract

Probabilistic latent variable models have successfully captured the intrinsic characteristics of various data. Understanding them is helpful for discovering latent rules and facts behind data. However, it is nontrivial to design appropriate models for given data because both machine learning and domain-specific knowledge are required.

We propose an **automatic model generation method** for data with hierarchical structure. Our method constructs an appropriate model for given data by **extracting important hierarchies** and preserves hierarchical and sequential information if needed or desired. We automatically extract latent structures of given data and **discover hidden rules and facts** behind the data.

Data become complex

- Becoming diverse and massive
- Consisting of hierarchical structures
- Indwelling **latent structures** User Clusters

Examples of hierarchical structures

Blogs : **Users** - Articles - Sentences - **Words**
 Music : Artists - Tracks - Phrases - Notes
 Purchases : Dates - Consumers - Shops - Items
 Sensors : Positions - Days - Times - Amounts

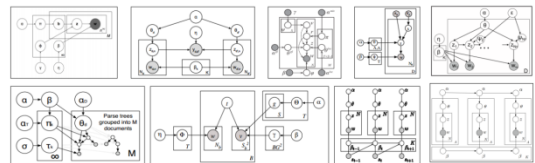
Finding **latent structures** helps find **rules behind data**.

Analysis considering hierarchy is difficult ☹

Models also become complex

- Extracting **latent structures** by **latent variables**
- Represented as complex relations of variables
- Requiring expertise in both data and machine learning

Examples of latent variable models

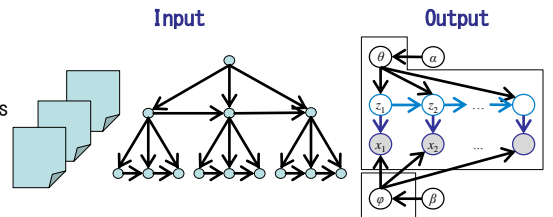


Only experts can design appropriate models ☹

Automatic model generation method using hierarchy

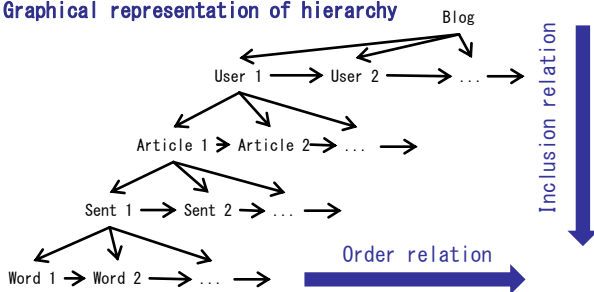
- Automatically extracting **important hierarchy**
- Providing **a universal learning method** for all possible models
- Employing **dynamic programming** for efficient learning

- ➡ **Rederivation and reimplementations are unnecessary.**
- ➡ **Latent structures can be automatically extracted.**



Dataset + graph for hierarchy Latent variable model

Graphical representation of hierarchy



Both inclusion and order relations are representable ☺

Efficient learning algorithm based on dynamic programming



Recursive computation makes learning efficient ☺

Related works

[1] M. Ishihata and T. Iwata, "Generating structure of latent variable models for nested data," in *Proc. The 30th Conference on Uncertainty in Artificial Intelligence (UAI-2014)*, 2014.

Contact

Masakazu Ishihata Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
 E-mail: ishihata.masakazu(at)lab.ntt.co.jp

