

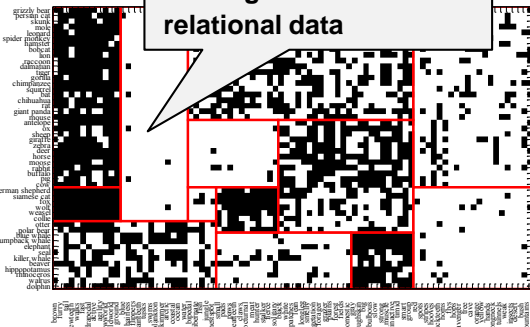
Abstract

We address the problem of infinite data analysis. Big data analysis has been a very active area of research in machine learning, but so far, as implied by the term 'big data' itself, the amount of input data has been assumed to be finite. However, many types of data often grow infinitely in size, and therefore, the observed data must be a part of a potentially infinite amount of data. This is the reason why we think the machine learning systems must be able to handle unlimited size of data. As an example of our results, this presentation deals with relational data represented by matrices, where the rows indicate instances and the columns represent values attributed to the instances.

Relational data analysis

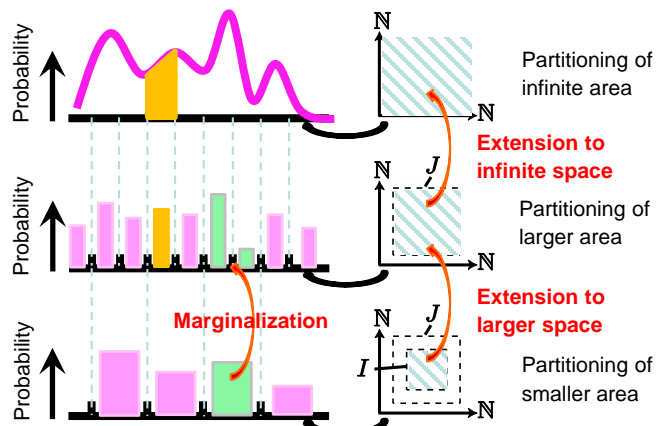
Example: Analysis of 55 animals based on 85 features.

Discovery of hidden rectangle clusters in the relational data



Infinite extension of models

Infinite model as the limit of finite models.



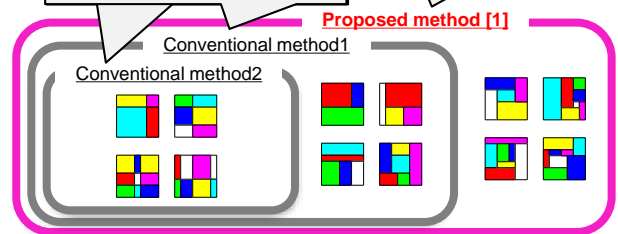
Stochastic block models

Strategy: Fit probabilistic models of rectangular partitioning to the input data.

Contribution: Construct stochastic arbitrary rectangular partitioning. (Conventionally, there are limited classes of stochastic rectangular partitioning.)

Limited classes of rectangular partitioning.

Arbitrary rectangular partitioning, which leads to analysis free of model biases.



Related works

[1] M. Nakano, K. Ishiguro, A. Kimura, T. Yamada, and N. Ueda, "Rectangular Tiling Process," in Proc. International Conference on Machine Learning, 2014.

Contact

Masahiro Nakano Media Information Laboratory
E-mail : nakano.masahiro(at)lab.ntt.co.jp

