

## XMLまたは関係表現の変形による前処理

沼尾正行

東京工業大学 大学院情報理工学研究科 計算工学専攻

Email: numao@cs.titech.ac.jp

### 概要

データマイニングの前処理には、文字単位のクレンジングも含まれるが、ある程度まとまった単位での構造変換や属性の選択などにむしろ困難がある。特に前者についてはデータベースの構築手法が利用されているだけで、データマイニング向けの研究はあまり行われていない。本発表ではそのような手法について、議論を深めたい。さらに、前処理を半自動的に進めるために、各種ヒューリスティックを導入する試みについても述べる。前処理の後段はデータマイニングであり、その結果からフィードバックをかけることも有効であり、それが複数ステップ学習の枠組みで取り扱えることも示す。

## Preprocessing for datamining in XML or relational representations

Masayuki NUMAO

Department of Computer Science, Tokyo Institute of Technology

Email: numao@cs.titech.ac.jp

### Abstract

We present a comparison between preprocessing for datamining and the frameworks of multi-step learning.