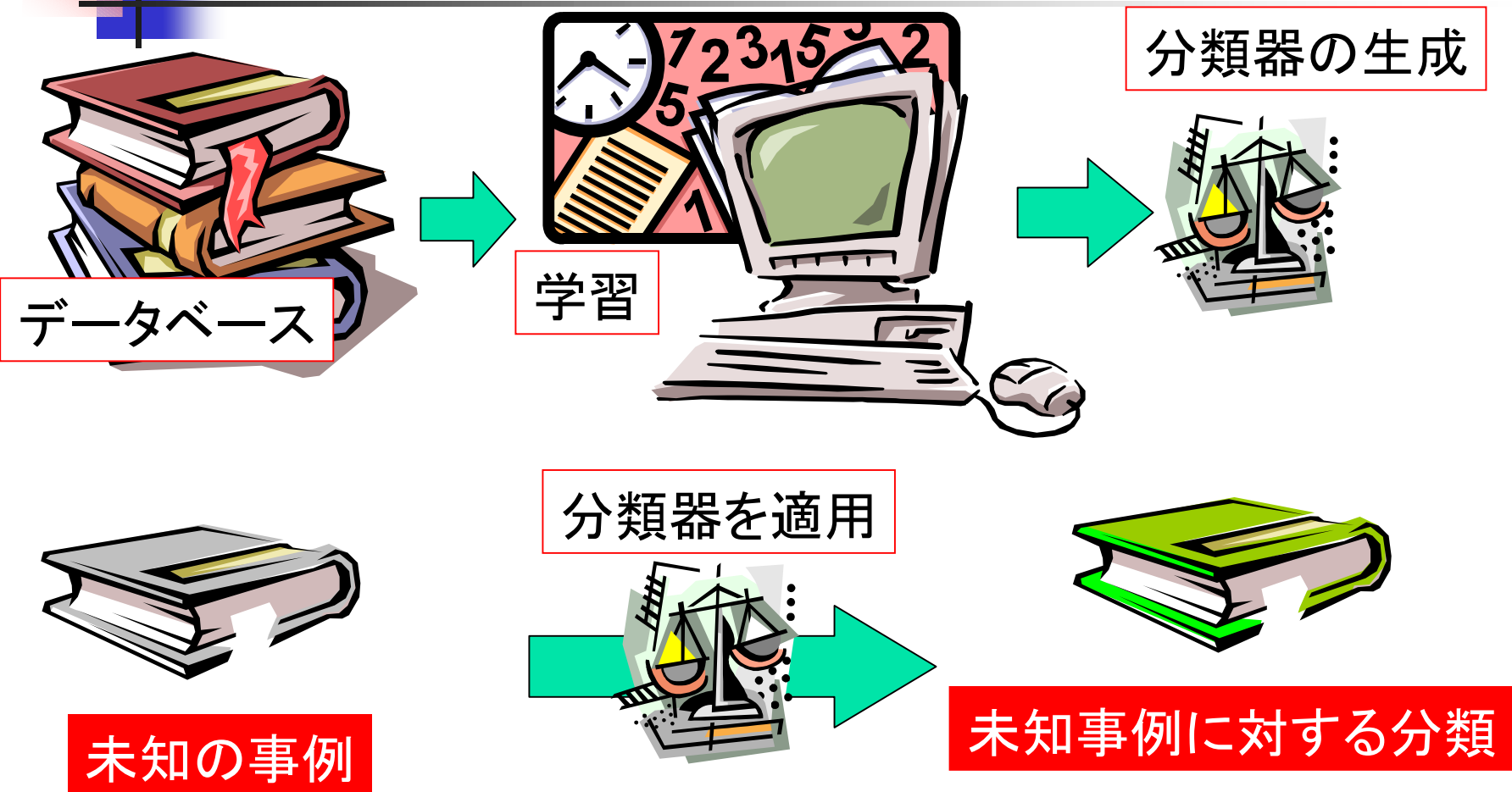


学習のための特徴選択 についての研究

世木研究室

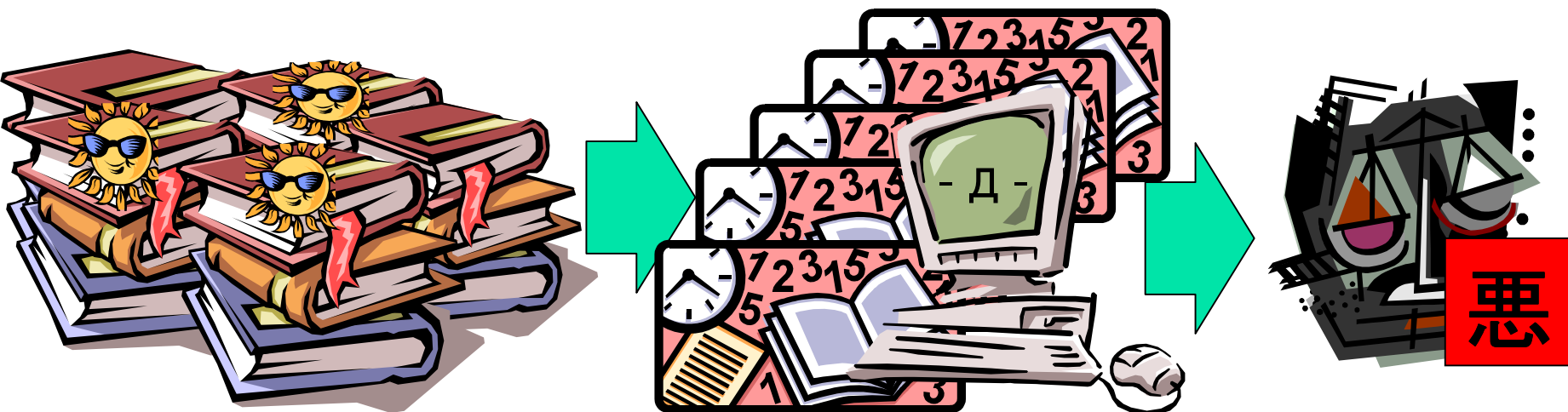
M1 小川智也

学習（機械学習）



過多・冗長な特徴による弊害

- データベースは、事例とそれを表す特徴からなる
- 特徴の過多、冗長な特徴は学習に問題
 - 計算時間、Overfittingによる精度の悪化



機械学習のための特徴選択

- 不要な特徴を取り除き、分類器の精度の低下を抑える
- 計算時間の抑制にもなる





去年、卒論でやったもの

exchangeability(a, c) =

$$\sum primary(a, b) * primary(b, c)$$

$b \in$ All Attributes except a and c

- ある類似度尺度に基づいた、特徴間の可換性について定義
- 精度をあまり下げずに、計算量を低下
- 機械学習と関係ない



今後

- 特徴間の可換性に用いた尺度が特徴間の距離尺度にも使えそう
→類似度尺度に応用など