

P040 共起成分の含意関係に基づく

データマイニングの実験と考察

二木 克也, 湊 真一

目標 アイテムの組み合わせの集合であるデータベースより、従来手法では、発見できなかった意味のある関係を抽出する。

共起成分の含意関係 → 直接的共起関係に基づく手法では、発見できない関係が抽出できる

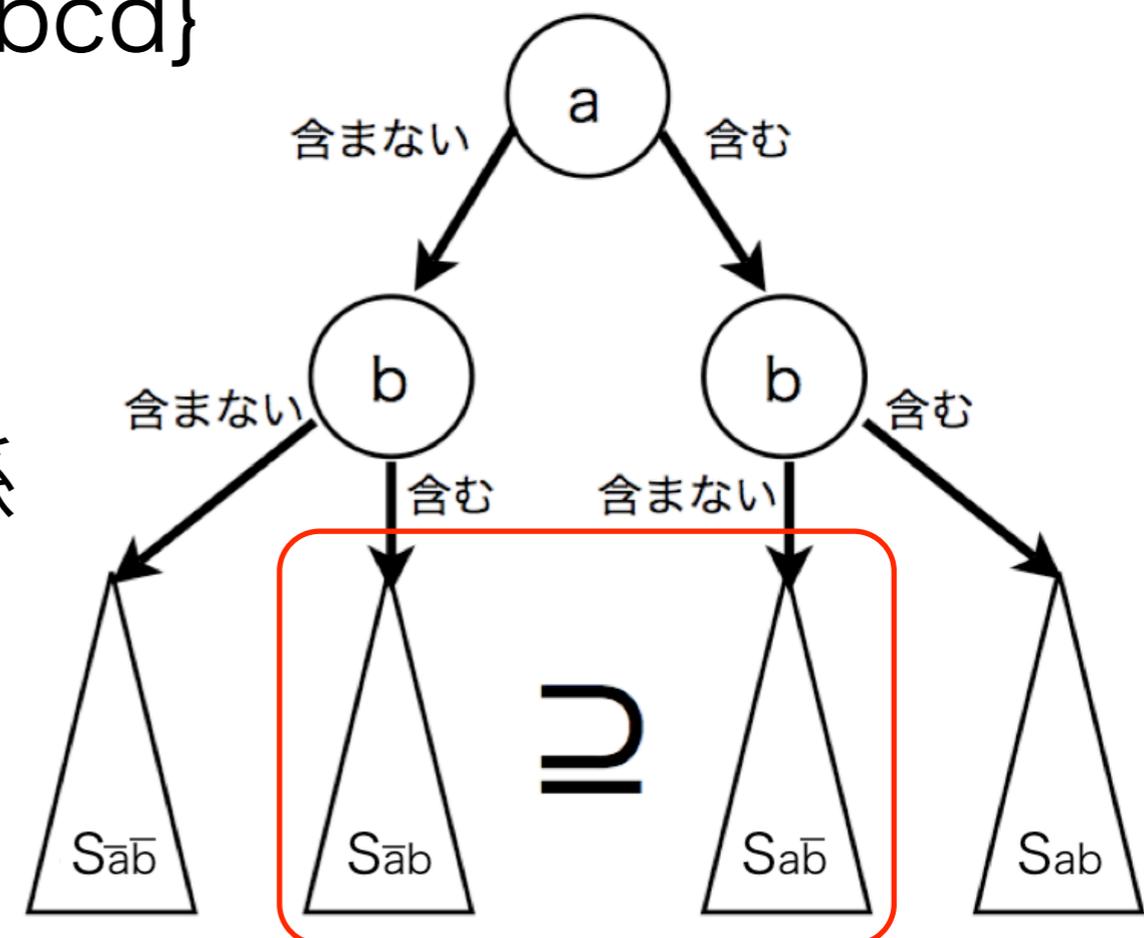
例: $S = \{bc, acd, ce, ac, abe, bd, bcd\}$

$$\left. \begin{array}{l} S_{\bar{a}b} = \{c, d, cd\} \\ S_{a\bar{b}} = \{cd, c\} \end{array} \right\} S_{\bar{a}b} \supseteq S_{a\bar{b}}$$

aはbに対して共起成分の含意関係

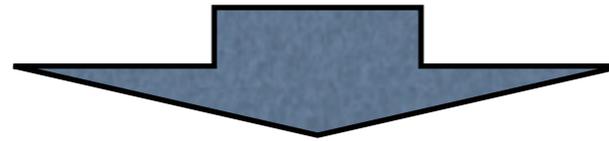
全てのアイテムの組について実行

今まで知られていない
有意な関係を抽出



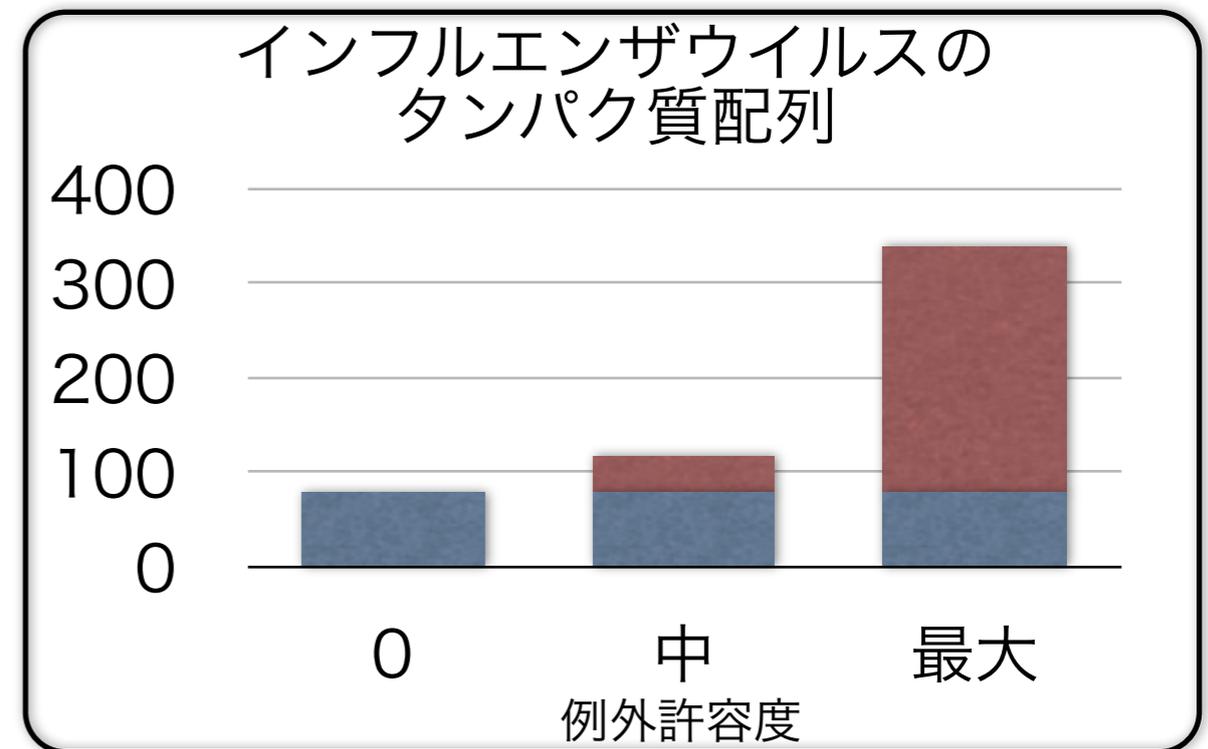
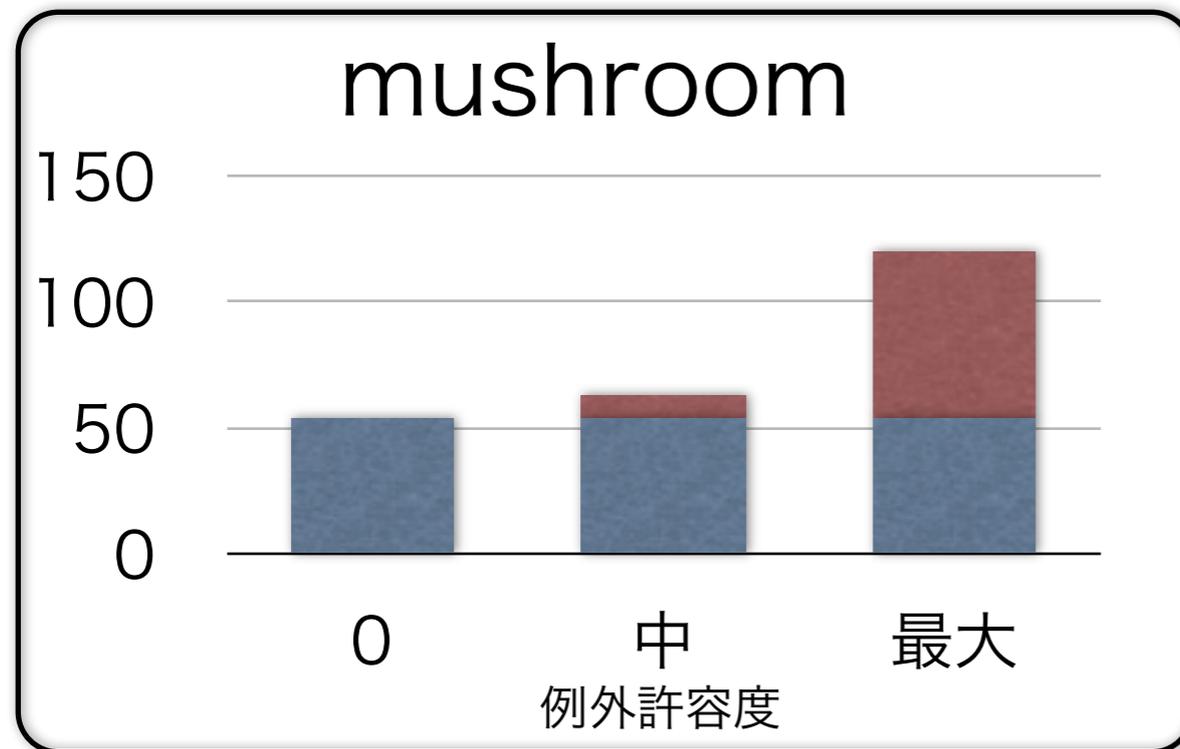
アイテム数の二乗回 → ZDDで高速化

ところで、従来手法では、共通成分がどれほど多くても、例外が1件でもあれば、抽出できなかった



例外を許容してみる

実験結果



例外を一定の水準まで許容した場合

意味のある新たな関係が抽出された

例外を最大限許容した場合

n者択一型のデータベースに関しては

設問ごとのアイテムの集合が抽出された

