

1 序論

データマイニングの目的とは、データベースから、妥当で新規性があり、役に立ち、かつ分かりやすい知識を見つけることである [3]。MAPIX[2] は、事例に関する事実を取り出し、それらをアイテムとしてマイニングを行う手法である。MAPIX からは数千ものパターンが出力されるが、その中でどのようなパターンが有用であるのかという考察は行われてこなかった。大量のパターンが出力されてしまう問題に対して、Dong らは形式概念分析の観点から極小生成系の中で簡潔なもの (SSMG)[1] を出力するアルゴリズムを提案した。本研究では、MRDM の分野における手法である MAPIX に対して SSMG に基づいた極小生成系を出力するアルゴリズムについて提案する。

2 MAPIX

例 1 $gf(koji) \leftarrow p(koji, yozo) \wedge m(yozo)$.

事例 $gf(koji)$ に現れている事実として「 $koji$ が $yozo$ という息子を持つ」ことがある。このような事実を一般化して「息子を持つ」という性質として扱うことにより、実際にマイニングを行うことを可能としている。性質をアイテムとして、与えられたデータベースに対し、事例がその性質を持っているかどうかでその事例をパターンとして枚挙する手法である。

3 極小生成系のマイニング

あるアイテム集合 A に対する出現集合を $\mathcal{T}(A)$ と表記する。極小生成系は次の様に定義される。

定義 1 (極小生成系) あるアイテム集合 P が極小生成系であるとは $\mathcal{T}(P') = \mathcal{T}(P)$ を満たす様な $P' \subset P$ が存在しないことである。

例 2 SSMG_MINER[1] は表 1 の様な TDB に対して、図 1 に示す SE-tree という探索木を生成し深さ優先探索を行う。各ノードは注目しているアイテム集合を表す。それぞれのノードに対しての飽和集合を求め、その飽和集合に対する極小生成系として今訪れているノードを登録する。

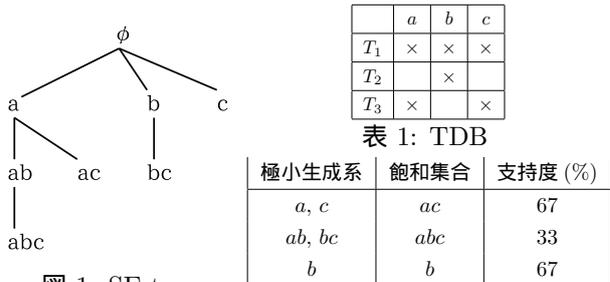


図 1: SE-tree

一般に極小生成系は一つの飽和集合に対し複数存在するが SSMG では極小生成系の出現集合を考慮して分類を行い、その中で冗長でないものを出力している。

4 提案手法

提案手法では、MRDM における極小生成系のマイニングを SSMG に基づいて提案する。MRDM ではアイテムは述語論理の形式で表現されているため、アイテム間に論理的包含関係 (包摂関係) が存在することが特徴である。この様な特徴を踏まえ選定を行った極小生成系を、論理的包含関係に基づく極小生成系 (Logical Minimal Generator: LMG) と呼ぶ。

SSMG で分類された極小生成系の中で、アイテム集合間に包摂関係が成り立つ場合がある。提案手法ではこの時、より一般的なアイテム集合を出力とする。例えば、表 2 においてアイテム a, c 間に $c \geq a$ という包摂関係が成り立つ場合には、飽和集合 ac の LMG として c のみを出力する。

5 実験とまとめ

MAPIX と提案手法の間で出力したパターン数を比較するために突然変異性を持つ化学物質のデータ Mutagenesis に対して実験を行った。図 2 の横軸は支持度 (%), 縦軸は出力したパターン数 (log 個) である。

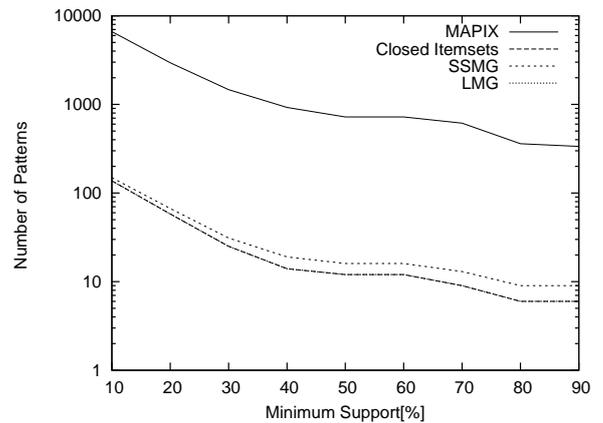


図 2: 出力パターン数の比較 (LMG は提案手法)

MAPIX の出力した頻出パターンより、提案手法で出力したパターンは飽和パターンと同等に出力するパターン数を抑えられていることが分かる。しかし、処理時間についてはまだ検証が不十分であり、また検討が必要である。

参考文献

[1] Dong, G., Jiang, C., Pei, J., Li, J. and Wong, L. Mining succinct systems of minimal generators of formal concepts. In *DASFAA*, pp.175-187, 2005

[2] Yusuke, N. and Nobuhiro, I. Multi-relational pattern mining based-on combination of properties with preserving their structure in examples. In *ILP*, 2010.

[3] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. From data mining to knowledge discovery: an overview. In *Advance in Knowledge Discovery and Data Mining*, pp.1-34, 1996