関係データベースシステムを結合した論理データマイニングの実装	知能系	和田・犬塚研究室
	No. 16115137	牧野 敏行

1　はじめに

データマイニングとは，大量のデータから隠された知識や，新しい規則を発見するプロセスである．論理プログラムの形式でデータマイニングをおこなう手法 **MAPIX**[1] は，事例からそれが持つ性質を抽出してデータマイニングに用いる手法である．

現在までの MAPIX の手法で実装されていたものは，Prolog プログラムとして主記憶上にある述語論理で記述されたデータをマイニングするものだった．そこで，本研究では，データベースに接続し，データベースシステム内の表形式で表現されたデータをマイニングするように拡張したアルゴリズムを示す．

2　MAPIX

例えば，ある家族関係のデータベースにおいて事例 grandfather(koji) に関して，次の式が成り立つとする．

$$\mathbf{grandfather}(\mathbf{koji}) \leftarrow \mathbf{parent}(\mathbf{koji}, \mathbf{yozo}) \wedge \mathbf{parent}(\mathbf{yozo}, \mathbf{kyoichi}) \wedge \mathbf{male}(\mathbf{kyoichi}).$$

これは「koji が kyoichi という孫息子をもつ」という事実である．MAPIX の目的は，このようなパターンの頻度を調べ，頻出なものを枚挙していくことである．

MAPIX の概要は以下のようになる．

1. 与えられた事例集合からいくつかの事例を選ぶ
2. 選んだ事例から得られる性質を全て抽出する
3. すべての性質を用いて，閾値以上の頻度を持つアイテムセットを興味深いものとして枚挙する

3　SQL_MAPIX

このアルゴリズムでは，図 1 のように Prolog 側とデータベース側の操作に分け，データベース側に SQL 文によるクエリーを出すことによりマイニングを行う．

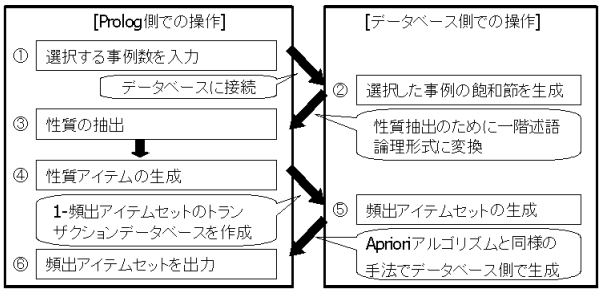

図 1: 提案アルゴリズムの手順

まず，図 1 の　ではデータベースシステム内にある事例のテーブルから，いくつかの事例を選択したテーブルを新たに作り，その事例を説明し得るリテラルを全て結合したものである飽和節のテーブルを生成している．また，性質を取り出すためにテーブルの表形式のデータを Prolog 側に持ってきて述語論理の形式に変換する必要がある．

また，図 1 の　では Prolog 側で得られた各性質を変数化してパターン化したものである性質アイテムから，データベース側で事例がその性質アイテムを持つかどうかのテーブルを作成し，そのテーブルから性質アイテムの頻度を計算し，1-頻出アイテムセットのトランザクションデータベースを作成する．

さらに，作成したトランザクションデータベースから Apriori アルゴリズムと同様の手法でデータベース側で閾値以上の頻度を持つ頻出アイテムセットを生成していく．この部分は SQL 文による Apriori アルゴリズムの手法を提案した S.Sarawagi らによる文献 [2] に従う．

この 2 つ以外の図 1 の左側の部分は既存の MAPIX と同様である．

4　実験結果

英文の構造を表形式で表現した 3369 個の事例に対して，1-頻出アイテムセットのトランザクションデータベースの作成の段階までのプログラムの平均実行時間 (試行回数 10 回) を記録した．ただし，MAPIX については事例数が 1000，3369 個の場合は時間がかかるので，試行回数を 5 回とした．

	100	1000	3369
MAPIX	00’11’55	01’57’16	06’36’06
SQL_MAPIX	00’02’36	00’11’18	00’46’32

図 2: 抽出に使用した事例数に対する平均計算時間

図 2. より，既存の MAPIX と比較しても，計算速度は比較的良くなっていることがわかる．

5　まとめと今後の課題

このアルゴリズムがデータベース内の表形式で表現されたデータに対しても有効であることが確認された．今後の課題としては，現段階のアルゴリズムでは既存の MAPIX で実行可能だったものが全て実行できるものではなく，制限されたものとなっているため，この制限を取り除く必要がある．

参考文献

[1] J.Motoyama, S.Urazawa, T.Nakano and N.Inuzuka: A Mining Algorithm Using Property Items Extracted from Sampled Examples. ILP 2006, Revised Selected Papers, pp.335-350, 2007.

[2] S.Sarawagi, S.Thomas, R.Agrawal: “ Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications ”, Data Mining and Knowledge Discovery,Vol.4,Nos.2-3, pp. 89-125, Springer, 2004.