

## 業務ノウハウを活用するデータマイニングサーバ Nodias

石井 篤, 田中 秀俊, 安田 智, 白石 将, 小幡 康, 石川 洋, 和田 信義, 和田 雄次<sup>†</sup>, 杉崎 元<sup>‡</sup>

3R-2

<sup>†</sup>三菱電機（株） 情報技術総合研究所<sup>‡</sup>三菱電機東部コンピュータシステム（株）

## 1 はじめに

近年の計算機性能の向上、記憶装置の大容量、低価格化により、企業では膨大なデータを蓄え、処理することが可能となってきた。また、エンドユーザコンピューティング環境の浸透により、エンドユーザによる情報系業務に対するニーズが高まっている。このような、大量のデータをエンドユーザが意思決定支援などの情報系業務に活用しようとする手法としてOLAP (On-Line Analytical Processing) やデータマイニング (Data Mining) がある。

データマイニングとは、データベースに蓄えられた大量のデータから企業活動において価値のあるルールを発掘 (Mining) しようという技術である。データマイニング技術の実用化のキーワードとして、ルール発掘に要する時間の問題と、結果として獲得されるルールの価値に関する問題がある。前者は、言うまでもなく大量のデータ処理に費やされる時間の問題であり、後者は、重要度の低いルールを排斥し、いかにして有効なルールのみを獲得するか、という問題である。

我々は有効なルールを獲得するためのアプローチとして、業務ノウハウやオントロジ (概念辞書) を利用したデータマイニング手法を検討している。ここで言う業務ノウハウとは、既に有効性が確認されているルールや専門家の知識を蓄積したものであり、対象とするルールは相関ルールである。相関ルールとは事象Aと事象Bがあったとき、事象Aが起きたとき同時に事象Bが起こる確率がY%であるとき、「事象AのときY%の確率で事象Bが同時に起こる。」というルールを相

関ルールという。

本稿では、この業務ノウハウを活用したデータマイニングの手法についての検討内容を述べる。

## 2 相関ルール発掘の課題

相関ルールを効率よく求めるアルゴリズムとしてAprioriアルゴリズムがある [1]。このAprioriアルゴリズムでデータの相関ルールを発掘すると、得られるルールは大量で、冗長なルールや意味のないルールを多数含んでいることが多い。さらに、この意味のないルールをふるい分けて得られるルールのうち有効なルールの大半は既知のルールであったりする。つまり、未知の有効なルールが、得られるルールの中にさらに埋もれている状況が発生するという問題がある。これは、ルールを発掘しようとするデータに対して、何も付加情報を与えないで、一律に処理を施すために起こると考えられる。

## 3 データマイニングサーバ Nodiasの概略

先に述べたようにデータマイニングの結果得られるルールは既知のものであることが多い。この既知であるルールというのは、人間がその業務において経験的に得た知識、専門家などにより確立されている知識、OLAP やデータマイニングの結果得られた知識などが考えられる。我々はこれらの知識を業務ノウハウとして蓄積し、これを参照してデータマイニングを行う方式を検討している。図1にその構成の概略を示す。図1において、業務ノウハウとは {シャツ、ネクタイ} や {雨、傘} のように既に獲得された相関のある項目を集

---

Using business know-how for data mining

Atsushi ISHII, Hidetoshi TANAKA, Satoshi YASUDA, Masashi SHIRAIISHI, Yasushi OBATA, Hiroshi ISHIKAWA, Nobuyoshi WADA, Yuji WADA<sup>†</sup>

Gen SUGIZAKI<sup>‡</sup>

<sup>†</sup>Mitsubishi Electric Corporation

<sup>‡</sup>Mitsubishi Electric Computer System (TOKYO) Corporation

合の形に表現したものである。また、オントロジとは、「靴 = シューズ」のような同義語や、「衣類 > シャツ」のような概念語を格納する辞書である。

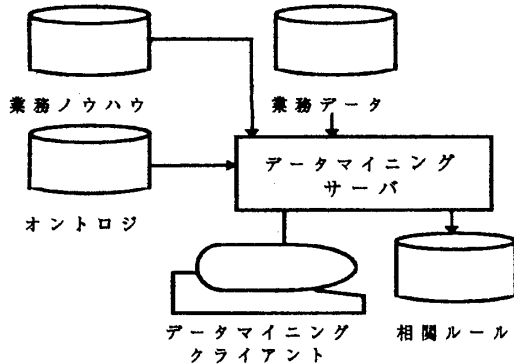


図1 Nodiasの構成の概略

## 4 業務ノウハウを用いたルール発見

次に、業務ノウハウとオントロジを用いたルール発見の手法について説明する。

### 4.1 マスクとしての業務ノウハウの利用

データマイニングの結果得られるルールは既知のルールを多数含み、その結果、得られるルールが大量になり有効なルールが分かりづらくなってしまうことが多い。そこで、業務ノウハウが既知のルールであることを利用し、獲得したルールから既に業務ノウハウに蓄積されているルール以外のものをルールとして採用することにより、有効なルールを発見しやすくすることができる。

また、逆に業務ノウハウにあるルールが発見されない場合も考えられる。この場合、発見されないルールが専門家などにより確立された普通のルールであれば、その既知のルールが成り立っていないという問題提起にも利用できると思われる。

### 4.2 トレンド分析のための業務ノウハウの利用

一般に、流通業や小売業などにおいては社会的環境の変化などによるデータの傾向の変化がある。この傾向の変化（トレンド）を捕らえることは大変意義のあることである。業務ノウハウに蓄積されたルールを過去のルールとしてとらえれば、

それを時系列に扱いルールの変化からトレンドを分析することが可能となる。

### 4.3 候補属性抽出のための業務ノウハウの利用

関連ルールを生成する際、データベースのすべての属性について一律に関連を求めるのではなく、関連のありそうな属性に“あたり”をつけて関連を求めることにより有効なルールを発見することを考える。そこで、業務ノウハウに存在するルールの項目に着目し、その項目とルール発掘するデータのスキーマの属性とを比較する。ここで、業務ノウハウに一致する属性に着目してその属性のデータ間について関連分析することにより有効なルールの発見を目指す。

ここで、業務ノウハウとスキーマの属性を比較する際、必ずしも業務ノウハウの項目名とスキーマの属性が一致するとは限らない。そのため、オントロジを参照して、同義語として定義されているものについては同一であるとみなすことにより、業務ノウハウを活用することが可能となる。

### 4.4 関連ルール抽出へのオントロジの利用

関連ルールを求める場合、個々のデータでは関連がないように見えるが、それらを包含する上位概念で関連を調べると有効な関連ルールが得られる場合が考えられる。そこで、関連ルールを発掘する際、データをオントロジで定義してある上位の概念に置き換えて、その置き換えたデータから関連ルールを発掘することにより有効なルールを発見することが期待される。

## 5 まとめ

以上、有効な関連ルールを発見するために業務ノウハウとオントロジを利用したデータマイニングの手法について述べた。今後は実装を進めてその有効性を検証することと改良点を抽出することが課題である。

## 参考文献

- [1] R.Agrawal, R.Srikant: “Fast Algorithm for Mining Association Rules”, Proceedings of the 20<sup>th</sup> VLDB Conference 1994.