

属性指向データマイニングサーバKnodiasの試作

安田 智, 和田 信義, 石川 洋, 石井 篤, 小幡 康, 和田 雄次[†], 三石 彰純, 杉崎 元[‡]

2 R - 3

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

1 はじめに

近年、データベースから情報を発掘する技術であるデータマイニングが注目されている。中でも相関ルールと呼ばれるデータ間の相関関係を発見しようとする分野では、その処理方式等について、さまざまな提案がなされている。著者らも、これまでに、効果的な相関ルール抽出処理のためのひとつ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）を適用したマイニング手法を検討し、その構想について紹介してきた[1][2]。（業務ノウハウについては後述する。）現在検討中のデータマイニングサーバ（開発コード名:Knodias）は、この業務ノウハウを背景知識として相関ルール抽出に活用する。また、オントロジに記述された属性情報（用語の概念的階層）を用いることによる属性指向の柔軟なデータマイニングを目指している。今回我々は、これらKnodiasの構想の一部を機能として実装する試作を行った。本稿では、このKnodiasの特長と実現方式について報告する。

2 Knodiasのコンセプト

Knodiasの基本的なコンセプトである「業務ノウハウの活用」「オントロジの利用」の2つについて説明する。

2.1 業務ノウハウ（既得知識）の活用

業務ノウハウは、既に有効性が確認されているルールや、各分野の専門家の有する知識、以前に行ったデータマイニングの結果等を、ルールの形式で蓄積したものである。Knodiasは、これをデータマイニングに利用する。

一般にデータマイニングで機械的に規則性を掘り起こした場合、その結果得られるルールの数は膨大であり、かつその多くはユーザが求めているものではないと考えられる。即ち、偶然に成り立っている意味のない相関であったり、公然またはユーザにとって既知のルールが、

その多くを占める結果になる。従ってユーザは、この大量に抽出された相関ルールの中から、そのユーザにとって価値あるルールを見つけ出す「宝探し」を行うことになるが、この過程において、ルール取捨の判断の基準となるのは、ユーザの持つ背景知識であると考えられる。

Knodiasは、主にこのユーザの背景知識を補うものとして業務ノウハウを活用する。さらにこの業務ノウハウは、前述のルールの取捨といった後処理だけでなく、相関をとる項目の組合せの絞り込み等にも活用可能であると考えられる。

2.2 オントロジ（属性情報）の利用

項目間の相関を求める場合、個々の項目そのもの同士の相関を求めるのではなく、ある括りによって上位概念を形成し、この上位概念間の相関に着目した方がよい場合があると考えられる。例えば、POSデータとして記録されるデータは具体的な商品そのもののIDであるが、これをそのまま「〇〇製サラダ油500ml」という最も詳細なレベルで扱うよりも、「油」「サラダ油」といった、上位概念で括って扱った方がよい場合もあろう。そこで、Knodiasは、マイニング対象データ中のデータ項目に関する属性情報を定義したオントロジを想定し、必要に応じてデータ項目を上位概念語へ置き換えて扱うことを可能にする。また、オントロジには、各データ項目の、データ中の出現頻度情報を持たせておくことにより、相関をとる項目の組合わせの絞り込みといった処理に利用する。

3. 実現方式

3.1 システム構成

図1は今回試作したKnodiasのシステム構成を表した図である。Knodiasは、イントラネット接続された

Trial Implementation of the Attribute-Oriented Data Mining Server, "Knodias".

Satoshi YASUDA, Nobuyoshi WADA, Hiroshi ISHIKAWA, Atsushi ISHII, Yasushi OBATA, Yuji WADA[†]

Akitoshi Mitsuishi, Gen SUGIZAKI[‡]

[†]Mitsubishi Electric Corporation

[‡]Mitsubishi Electric Computer System (TOKYO) Corporation

データベースサーバ上に基本部（本体）を置き、ユーザはクライアントから WWW ブラウザを介してアクセスする方式をとった。これにより、ユーザが利用環境の違いによって制約を受けないようにしている。

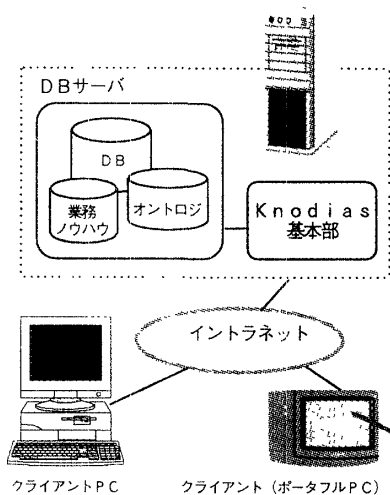


図1. Knodiasのシステム構成

3. 2 Knodias 基本部の構成

図2は、サーバ上のデータマイニング基本部の構成の概略図である。

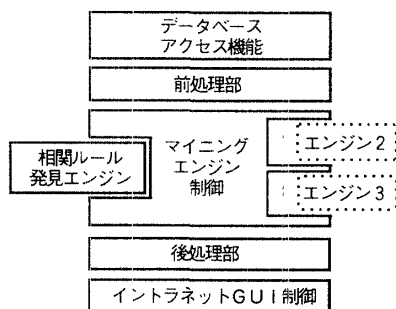


図2. Knodias基本部の構成

基本部は、マイニングエンジン制御部、前処理部、後処理部、DB アクセス機能群、GUI 制御部からなる。実際にマイニングを行うエンジン（アルゴリズム）は独立に作成し、マイニングエンジン制御部にプラグインのイメージで組み込むような形で用いられる。前処理部では、オントロジを用い、ユーザの指定により相関を求める階層水準の決定等を行う。後処理部は、業務ノウハウを用い、結果としてユーザに提示するルールの取捨を行う。

4 Knodiasによるデータマイニングの特長

今回試作した Knodias で相関ルール抽出を行う場合の特長について述べる。

4. 1 オントロジの利用による効果

ユーザは、まずマイニングしようとするデータに含まれるデータ項目について、それぞれ、オントロジに定義されたどの用語階層で扱うかを指定することができる。この時、オントロジに格納されている各データ項目の頻度情報が、マイニング時に用いる足切り基準として用いられる support 値なる最小出現頻度を下回るものは、表示によってそれがわかるので、ユーザはこれを参考にして適切な用語階層を決定することが可能である。この前処理により、上位概念への置き換えを指定されたデータ項目は、マイニングエンジン制御部にてその上位概念に置き換えられて、エンジン部で扱われる。

4. 2 業務ノウハウの利用による効果

業務ノウハウに含まれる個々のルールには、そのルールの扱いに関する情報が付与されている。その情報とは、例えば、これは優先的に扱うべきルールであるとか、不要なルールである、といった類のものである。Knodias では、この情報を用い、データマイニング後の後処理として、結果として表示するルールの整理を行うことができる。「優先」のルールは前の方に表示する、「不要」のルールは表示しないようにすることで、ユーザの「宝探し」を支援する。また、データマイニングの結果得られたルールに対し、結果表示の画面上で行った操作（削除など）の履歴は、ユーザノウハウとして、業務ノウハウと同じ形式で自動的に登録され、以降、業務ノウハウと併せて使用することが可能である。

5 おわりに

現在、この試作版評価のための準備を進行中である。今後はこの評価作業と問題点の抽出とその改良、ならびに未実装機能の検討と実装を進めることが課題である。

参考文献

- [1]石井, 他: 業務ノウハウを活用するデータマイニングサーバ Knodias., 第53回情処全国大会 1996.
- [2]和田, 他: データマイニングサーバKnodiasにおける属性情報利用方式の検討, 情処DBS(1996.10.24)
- [3]Srikant, R., Agrawal, R.: "Mining Generalized Association Rules", Proc. VLDB '95.