データベースからの知識発見システム DB-Amp
設計と実装とエキスパートシステム開発への応用
KDD System, DB-Amp
Design, Implementation and Its Application to an Expert System

鷹津 恵子
Graduate School of Media and Governance, Keio University. 
kshimazu@mag.keio.ac.jp

古川 康一
furukawa@sf2.keio.ac.jp

Keywords: KDD, ER modeling, inductive logic programming, e-mail, expert system.

Summary

We developed a Progol/Database integrated system called DB-Amp to realize data mining lightly against structured data. A new data model RER (Refined Entity Relationship model) is introduced to define a target concept to be learned as well as background knowledge. A data conversion system DAISY (Data Arrangement Interface System) is developed to automate the preparation of input data to Progol as far as possible. We concentrate on the following issues: (1) given a database and an inductive inference problem, how to design target concept representation to be learned by the ILP system Progol, (2) how to identify necessary background knowledge (BK) and how to define predicates constituting the BK, and (3) how to restrict the Progol search space to a finite set. We applied DB-Amp to build a rule base for an automatic e-mail classification system in real business environment.

1. はじめに

本論文は、データベースと帰納論理プログラミングの結合による「データベースからの知識発見」(KDD: Knowledge Discovery in Database) [Fayyad 96, 西尾 93] の一方法を提案するものである。われわれはその両者を結合したシステム DB-Amp を設計・試作し、それを実際のビジネスデータベースからの知識獲得に応用した。その後結果として得られた知識ベースは、実運用しているエキスパートシステムの参照先として利用した。


分類問題に限定しても、ID3 を用いて決定木を生成する例が実用分野において数多く報告されており、より表現力の高い帰納論理プログラミング (ILP: Inductive Logic Programming) [古川 92, 古川 97a, 山本 97a] の利用例は、ビジネスデータからの応用領域では依然として多くない [溝口 97, 沼尾 97b]。ただし、実験段階の応用領域は幅が広く、文書データベース [Slattery 98] や Web のコンテンツ [Craven 98] など非定型のデータベースを対象とするものも存在する。

つまり、帰納論理プログラミングは、その有用性は認識されていたが、データベースからの知識発見への応用としては、その多くが実験データからの知識獲得による報告であり、実稼動システムの一モジュールとして組み込まれたものはほとんどみられなかった。

本論文では、現在最もよく知られている帰納論理プログラミングの一部である Progol[Muggleton 95] とデータベースを結合して、より高度な機能をもったデータベースからの知識発見システムの提案をし、その設計
原理と試作結果およびその応用について報告する。
われわれはデータベースの設計にER（Entity-Relationship）モード[Chen 76]を拡張したRER（Refined Entity-Relationship）モードと呼ばれるデータモデルを
定義し、それによってPrologが必要とする入力情報の
同定が可能であることを示した。ER モデルは古典的で
はあるがオブジェクト指向モデルの基礎として用いられ
た[鈴木 98]だけでなく、今日企業のシステムを構築す
る際のIE（Information Engineering）モード[Martin
89]と並んで最も頻繁に用いられている手法である。特に電子商取引（EC: Electronic Commerce）の実現を目
指した商用CALS（Commerce At Light Speed あるいは
Continuous Acquisition and Lifecycle Speed）[片岡
96, 岡本 96]において提案されたものにER モデルに基づ
く企業の情報交換のための標準化手法がある[福田 98]。
そこでわれわれはER モデルによるデータベースの解釈
を取り入れたシステムを設計することで、適応領域の広
いKDDシステムを実現することができる考えた。
また入力ファイルを同定する際、処理の効率化を達成
するために、定数の探索空間を、データベースから定義
される有限領域に限定する方法を提案する。
さらに本論文では設計したDB-Ampを実装し、その出
力を知識ベースに格納することで、知識が増幅的に増加
するエキスパートシステムの実現に成功したことを報告
する。われわれはこのエキスパートシステム導入前と導
入後の実業における作業時間の差を比較した実験結果
から、データベースからの知識獲得がビジネス領域にお
いて効率化を図れるという意味で有効であることを示す。
1990年代の半ば以降インターネットを代表とするコ
ンピュータネットワークの普及にともない、日常的に流
通する情報量が激増している。今日ネットワーク
上では、多数の人が多様なコミュニケーションを支
援することが求められている[森田 96, 上田 98]と[塚
本 96]は、固定的なメールリストに頼らず情報配信
先として最適な利用者にメールを分散配信するシステム
(MILD)を提案し、[Mizoguchi 99]は配信されたメール
をそれぞれの属性値をもとに分類し、優先度によるソー
トや利用者ごとに異なるメールの分類などを支援する仕
組みを示した。これに対して、われわれはさらにメール
本文に注目し、その内容に基づいてメールを分類するた
めのルール群をDB-Ampを用いて獲得した。
以下に、本論文の構成を示す。2 章から5 章で、われ
われが提案するKDDシステムであるDB-Ampについ
て述べる。具体的には、2 章ではDB-Ampの全体像を
構成するモジュールとKDDプロセス中における位置づ
けを用いて示す。3 章に、われわれが提案するRER モデ
ルを示す。この手法によるデータの解釈を従って、デー
タベースからの知識がILPシステムへの入力データに変換さ
れる。4 章では今回データマイニングエンジンとして採
用しているILPシステムPrologの特徴と基本機能を紹
介する。5 章で、6 章のフレームワークを基礎に置いた
入力データ生成モジュールであるDAISYを説明する。6
章と7 章では、DB-Ampをパッチ処理モジュールとして
内蔵するエキスパートシステムであるAUTOMAILの
仕様と運用結果を述べる。最後に8 章と9 章にまとめと
今後の課題を示す。
2. DB-Amp
2.1 DB-Ampの設計思想
DB-Amp[Shimazu 97]は、データベースと帰納論理
プログラミングシステムPrologを統合することによって
実現したデータベースからの知識発見を行うシステムで
ある。われわれは、Prologをデータマイニングエンジン
とする汎用のKDDシステムを目指した。帰納論理プログラ
ミングシステムは一階述語論理に基づいているため、
一般には高効率性を望めない。一方、データベース構築
技術は、膨大な量のレコードに対する効率的な格納と参
照能力の強化に焦点を当てている。したがって、実運用
に耐えるKDDシステムを望むのであれば、二つの
技術を統合する際に効率性を損なわないように帰納論
理プログラミングを機能させる必要がある。そこでわれ
われはDB-Ampを設計するためにあたり以下の点を考慮し、
この問題を解決することを試みた。
① データベースの全データから最適な部分集合を抽
出し、探索空間の絞り込みを行う。
② Prologの入力ファイルの仕様を最大限に活用し、
絞り込んだ探索空間に領域を限定する。
③ 直接Prologに与えることができる入力ファイル
を生成する。
もう一つの設計上の選択は、システムの汎用性である。
DB-Ampを多くの領域へ適用させるためにさらに以下の
2点を考慮した。
④ すべてのERモデル解釈が可能なデータベースに
適用可能とする。
⑤ 可能な限り自動的に入力ファイルを生成し、エン
ドユーザによるマニュアル操作の編集作業等を最小
限にする。
また、このとき実運用に耐え得る汎用性の実現を目指
し、われわれはPrologの入力データのうち完全に自動
生成できるものとそうでないものを明確に分けることに
した。
2.2 DB-Ampの構成と役割
データベースからの知識獲得作業の全工程は、データ
マイニングエンジンとその前処理部で構成されていると
見えることができる。図1にはこの二つのモジュールと
KDDプロセス[Rayyad 96]の関係付けを上段に示し、さ
らにDB-Ampの構成を下段に表している。
データマイニングの上流工程である前処理部分として、
われわれは“DAISY”（Data Arrangement Interface SYstem）を設計した。DAISYの目的は直接データベースからデータを抽出し、Prologの入力形式でデータ表現を変換することである。このとき利用者が獲得したい知識を指定することで、DAISYはこれを目標概念として設計し、データベースから抽出したデータを加工してこの概念の正例を生成するとともに、この概念の説明に必要であると思われる背景知識を特定する。一方、データマイニングエンジンとしては、最も強力な帰納論理プログラミングシステムの一つであるPrologを採用した。

3. 拡張ERモデル

3.1 データモデル

従来、KDDの上流工程は統計解析手法によるデータ解析やDBMSの問い合わせ機能を用いた対話的処理で実現されていた。われわれは、情報に関する定量的なルールの発見よりもむしろ、定性かつ構造的なルールの獲得を目指している。このため、意味情報を用いて定性的な概念的構造的側面を捉えることを試みた。そこで、われわれは代表的なデータのモデリング手法の一つであるERモデルを拡張したRERモデルを定義し、それによるデータベースの意味情報の記述を行った。

データベース設計の観点に立つと、データのモデルには構文モデルと意味モデルの2種類が存在する[近松92]。典型的な構文モニールには、関係モデルやネットワークモデルが挙げられ、一方意味モデルとしてはERモデルがある。最近では実用的なデータベースの設計をめざし、ERモデルで表現したものを利用的に構文モデルが表現に変換することを目的としたCASEツールが数多く開発されている。

図2の上側に示すようにERモデルは実体、属性、関連の三つの要素から成り立っている。ERモデルにおける最も基本的な要素は実体であり、それはモデリングを行っている対象の実世界中の本質的な構成要素を指し、したがってこれ以上分割して表現することができないオブジェクトもしくはエージェントを表現するものである。各実体と関連は属性を持つことができる。また実体を関係づける関連はそれぞれの実体とリンクが張られている。

一方われわれが提案するRERモデルは、属性および関連を基本と派生の2種類にわけて表現するものである（図2の下側の図を参照）。基本属性と基本関連は、他のそれを参照することなく値を独立に決定できるものであり、派生属性と派生関連はそれぞれ他のそれを参照しないと値を決定できないものである。

3.2 RERモデル表現例

図3にRERモデリング表現の例を示す。上段に図形情報データを対象としたRERモデリング表現の例を、下段にキーワードのリストから成るいわゆる意味論的データを対象とした例を示している。

図形情報データの場合、node属性は各lineの端の座標値を持っている。そのためlineのlength属性の値はそれぞれnode属性の値から計算できる。キーワードのリストの例も同様である。

4. Prolog

4.1 帰納論理プログラミングシステム

Prolog[Muggleton93]は、われわれの設計したDB-Amp中でデータマイニングエンジンとして採用されているシステムであるが、PROLOG言語のインタプリタを
持ち、この言語表現を使って入力ファイルを与えるとシステムが動作する。すなわちこの入力ファイルは他の帰納論理プログラミングシステムと同様に正例、負例、背景知識から構成され、それらはすべてPROLOGプログラムとして与えられる。

Prologは逆対応法と呼ばれるアルゴリズムを採用しており、それによって[Yamamoto 97b]あるいは[Furukawa 97b]が示した条件の下で、正例を説明するすべての仮説を生成できることが知られている。

4.2 决定木による分類器システムとの違い

帰納論理プログラムは、ID3やC4.5などの命題論理に基づく機械学習システムとは異なり、述語論理を表現言語として用いることにより、学習目標概念に関連する知識を述語間の関係によって定義して、背景知識として利用できる。この機能により目標概念を無理なく帰納学習せずに、ある程度の知識を前提として学習することが可能となった。

命題論理に基づく機械学習システムの多くが決定木の形式で形式化を生成する。その際、データベースのデータの単純な交換操作だけで、システムへ与える入力データを生成することができる点が実用に際して最も大きい特徴であるが、2次元の単一テーブルで表現されるデータベースしか処理できない。

これに対し、帰納論理プログラミングシステムは、データベース中のデータ（もしくはデータから算出される属性）を、正例と背景知識に変換しようと必要に応じて生成する工程を発生させるが、これらが利用できることにより関係データベースで表現された複数の2次元テーブルから成るデータからルールを抽出することが可能となっている。

4.3 入力データの形式

図4はPrologに対する入力データとその出力結果の例を示している。この例では、家族関係のデータベースからgrandfather_ofを目標概念としてルール抽出を行っている。モード宣言では、目標述語とこれを記述するのに必要な述語が、それぞれmode1、mode2として宣言される。学習された結果は入力データ表記と同様にPROLOG言語表記で記述されるが、前者の目標述語は頭部に、後者は本体部に登る。それぞれの述語の引数は、入力力をそれぞれ"+"、"-"によって指定する。タイプ情報は、数値の値が与えられるドメインを定義するものであり、探索空間を表している。したがって、実際のデータベース内のデータと、このタイプ情報が示す領域を効率的に関連づけることが、データベースからの知識獲得の際の探索空間の絞り込みの実現に貢献することになる。

5. DAISY

DAISYはわれわれが開発したKDDシステムDB-Ampの中で、データマインニングエンジンであるPrologとデータベースを直接つなぐインタフェースモジュールを実現している部分である。具体的にはデータベースの構造をRERモデルに従って解釈し、自動的にデータをProlog入力ファイルの一部へと変換する。

5.1 目標概念・正例の設計と生成

RERモデルにおけるすべての実体を
\[ e_1, e_2, \ldots, e_m \]
とし、それに対応しているすべての基本属性を
\[ P_{n_1}, P_{n_2}, \ldots, P_{n_m} \]
としたとき、目標概念を表す述語\( t \)の構造を
\[ t(P_{n_1}, P_{n_2}, \ldots, P_{n_1}, P_{n_2}, \ldots, P_{n_{m-1}}, P_{n_m}); \]
と定める。すなわち述語\( t \)は、その引数としてすべての実体のすべての基本属性を持つものとする。

目標概念の正例となるデータをデータベース中からSQLにより抽出し、これらを述語\( t \)の構造を満たすと仮定する。これが正例の生成となる。

5.2 背景知識の設計と生成

各実体の各名指称及び属性関係に対して、その各名指称及び属性名を名指称とする述語を導入する。述語の引数として、その名指称及び属性関係が依存する基本属性、あるいは他の名指称及び属性を用いる。これらは背景知識の一端となり、Prologが帰納学習を行う際の目標概念を記述するための部品として使用される。

これらの各名指称の定義は、入力により与えられる。
5.3 有限探索空間の設計と生成

有限関数内の効果的な学習を行うには、探索数値の増大を防ぐ工夫が必要である。このために、正例に現れるデータのみを述語の引数のドメインとする方式を採用した。

具体的には有限領域を与えるタイプ情報を正例中的データのみによって定義した。

例えば任意の要素からなるリストは

\[
\text{list}([]).
\]

\[
\text{list}(\text{H}\text{T})\text{.}=\text{list}(\text{T}).
\]

と定義される。探索空間を絞り込むために事例に現れるデータのみに限定しようとする

\[
\text{list}(\text{H}\text{T})\text{.}=\text{db\_instance}(\text{H}), \text{list}(\text{T}).
\]

とする記述が採用される。

ここで \(\text{db\_instance} \) は、各正例

\[
\begin{align*}
& \text{list}(p_{11}, p_{12}, \ldots, p_{m1}, p_{21}, p_{22}, \ldots, p_{m2}, \ldots, p_{mn}, p_{m1}, p_{m2}, \ldots, p_{mn}). \\
& \text{db\_instance}(p_{11}). \ldots \text{db\_instance}(p_{m1}). \\
& \text{db\_instance}(p_{12}). \ldots \text{db\_instance}(p_{m2}). \\
& \vdots \\
& \text{db\_instance}(p_{mn}). \ldots \text{db\_instance}(p_{mn}).
\end{align*}
\]

のようにタイプ情報として生成されるものとする。

これにより探索空間が有限領域に限定される。

DAISY によって生成された入力ファイルの例は、6-4節で与える（図 4）。

6. エキスパートシステム -AUTOMAIL

われわれは今回の実験において、DB-Amp をパッチ処理モジュールとして内蔵するエキスパートシステム AUTOMAIL を開発し、知識ベース内のルールを既存的に増加させることを試みた。

6.1 電子メール分類問題

実験に使用したデータベースは、インターネットを介して入手した電子メールとそれらに対する返信メールの格納結果である。現在多くの日本企業に、電子メールが激しい進度で浸透しているが、電子メールの利用普及はコミュニケーションのあり方を変革した一方、情報の氾濫という新たな問題を引き起こした要因となっていることも見逃さない。このような問題に対処するために、われわれは過去に入手した問い合わせの事実データからで典型質問パターンを抽出し、それを基に新たな質問カタログを構成した。このシステムの主な機能である。

6.2 AUTOMAIL の機能

図 5 は、AUTOMAIL を構成する処理の流れとその中での DB-Amp の位置づけを示したものである。

担当者が新たに問い合わせメールを入手すると、このエキスパートシステムの持つ推論エンジンが自動的に知識ベースを参照し、メール内容がどの典型質問パターンのルールに合致するかを判定し、典型質問文に相当する内容を格納されているそれに対する最適回答文を表示することがシステムの主な機能である。

6.3 データベース設計

図 6 中の“電子メール応答記録” データベースの設計は図 6 に示すとおりであり、商用データベースに実装されている。過去に入手した 2085 件すべての問い合わせメールが、専門家により 86 のグループに分類され、それぞれに属しているかが対応する典型質問文番号の欄 (Corresponding Prototypical Sentence NO#) に入力されている。またすべての問い合わせメールの内容は、このシステム用に開発した辞書 (図 5 中の“キーワードシ

※1 日本経済新聞社が行った日本の大手 2500 社に対する情報化に関する調査では、電子メールの会社への普及率は 1996 年 11
月、1997 年 11 月、1999 年 5 月で 31.5%、48.8%、96.6%と
推移している。
図6 “電子メール応答記録” データベースの設計

ソーラース）を用いてキーワードを要素とするリストに変換され、図7中のQKeywordの欄に格納される。Answer to the Inquiry欄には、Contextに入力されている入力したメールの問い合わせ内容に対し、実際に返信した回答文が格納される。このデータベースがDB-Ampによって帰納学習の対象となる。

図7 “フォーマット変換（DAISY）” が生成した入力ファイルのことによって生成した。

2 背景知識

RERモデル表現中の派生属性名と関連名を述語名とする背景知識を生成した。

3 モード宣言とタイプ情報

モード宣言に関しては、modebはRERモデル表現から、modebは関数依存性を用いて生成した。またデータベース中のキーワードのリストのすべての要素をdb_instance述語を用いてタイプ情報に加えた。

これらの内で、最適化された正例と背景知識とmodebは、入力によって実装した。

6.5 Prologによるデータベースからの知識獲得結果

図5の“ルールの抽出（Prolog）”処理は、データマイニングエンジンの機能を活用している。6.4節で生成した入力ファイルを与えた結果、RERのすべてのクラスに対しこの当該特性ルールを獲得した。たとえば典質問文85に該当する問い合わせメールが持つ特徴は以下のように出力された。

sentence85(A):- have(A, '方法'),
in.order(A, 'MPFm', 'モード'),
not.have(A, 'NewMPFm'),
sentence85(A):- in.order(A, 'MPFm', '自動操作'),
sentence85(A):- in.order(A, '電源切'),

このテスト単語のルールは、Aで示される問い合わせメールの文章中に“方法”が出現（have(A,'方法') か
つ "MPFm"="モード"が該当の順で出現（in.order(A, 'MPFm', 'モード')）し、さらに "NewMPFm"が文章中に出現しない（not.have(A,'NewMPFm')）場合は、Aが典質問文85に該当することを表現している。

なおこの典質問文は、MPFmという機能の操作モードを自動的に切り替え方法の問い合わせである。この機能は後継機であるNewMPFmには反映されていないため、同様の問い合わせの内容であったとしても対象が
NewMFPtn である場合はまったく異なる返信を送付する必要がある。

獲得されたルールは、図 5 の“ルール”に出力されたあと、人手による確認が行われ、経済的なルールに置かれる。そして、システムが毎日 24 時間に更新されたルールをプロテクションルールに変換し、“知識ベース”に格納することで反映される。

6.6 推論エンジン
推論エンジンの役割は、システムが新たに電子メールを入手しキーワードのリストを生成した後、それを“知識ベース”内のすべてのルールに照らし合わせ、それぞれのルールにどのくらい適合しているかを算出することである。この数値が各典型質問文に該当する度合い（適合度）を表す。この処理は、図 5 の“知識ベースの参照と各ルールに対する適合度の計算”モジュールが実現している。

本推論エンジンは、適合度を各ルールの本体部の各選言肢の中で真となるものの割合から求めている。

6.7 推論結果の表示と返信メールの送付
図 5 の“典型質問文とその最適回答文と適合度の高い順に表示”では 6.6 節で算出した適合度の値の高い順に、該当する典型質問文とそれらに対する最適回答文を表形式で表示する。初期設定では、求めた適合度の値が最高のものが選択されているが、およびラ、多用途性で欠検査を避けるために、各質問文の選択肢をマニュアルで行うようになっており、最終判断をオペレータが実施したものを返信の送付として“回答文の編集と送信”で行うように実装されている。

実際にオペレータが参照する画面イメージを図 8 に示す。図観の“メール本文の欄に入力した問い合わせメールの本文を表示し、回答文の編集と、該当する典型質問文と最適回答文（操作画面の名称は、それぞれ“該当する典型質問文”“推奨回答文”）を送信する。

7. 実験と評価
われわれはこれまでに、DB-Amp の汎用性を確認するために、IGES データから菱形などの特徴的なパターンを表示するルールを獲得する問題に応用し、その有用性を確かめた [Shimazu 96]。
また今回の実験では、DB-Amp と AUTO MAIL の評価を、End-User であるオペレータの作業時間の変化で計ることにした。

図 9 は、無作為抽出した 10 人のオペレータによる 1 日の電子メールによる問い合わせに対する作業時間（電子メールの受信の確認から返信メール送付まで）の平均を取ったごとの変化を示したものである。AUTO MAIL 導入時に瞬時に作業時間が増加しているが、これは新操作の習得に対するものであると考えられる。最終的にすべてのオペレータがシステム導入前と比較し、作業時間をおよそ 1/3 に減少させることに成功した。

8. まとめ
われわれは今回の実験で、データベースとデータマイニングのエンジンを結合することで、KDD システム DB-Amp を実現することを試みた。データマイニングエンジンには、推論過程プログラムシステムである Progol を採用し、特にこれとデータベースを直接つなぐモジュールとして DAISY を設計した。ここでは、データベースの内容を活用的に構成し、データマイニングエンジンに渡すためのデータモデルと、Progol とデータベースを効率的につなぐ方法を提案した。
データモデルは従来の ER モデルの拡張であり、つなぐ方法として Progol が持つモード宣言とタイプ情報の仕組みを利用した。DAISY は、Progol に対する入力ファイルを自動的に生成することに成功した。
また DB-Amp の実用性を検証するために、この学習結果を知識ベースの内容に反映させるエキスパートシステム AUTOMAIL を開発し、知識ベース内のプロックワークノートを既存のところに増加させてきることで、システム全体の品質向上を実現できることを示した。今回の実験は、帰納するプログラミングによるデータベースからの知識発見の応用領域の拡大につながるものである。

9. 今後の課題

9.1 DB-Amp の強化

DB-Amp の課題は、データマイニングエンジンの性能向上と DAISY の機能充実である。

今回の電子メール対応記録のデータベースを対象とした実験では、データマイニングエンジンとして Progol を採用したが、ルールを抽出しやすい一昼夜以上を費やした例が数多く見られた。

変更はお願いすると、全体のクオリティを改善することを目指すので、 pct パック処理であろう。今回のサイズ（サービスレコード）のデータの処理でこの時間を要することは、実用システムの応用を考えると大きな課題となる。これに対する解決のためには、新たなエッジオの開発や記号データと定義データを扱うことを考えるエンジン（C.4.5 等）との統合、また定性データの定量化のアプローチの開発を行い、相関ルールに利用しあ事前にあらかじめ考慮する処理 [喜連川 97] などの追加が必要である。

DAISY の実現に中心的役割を果たしている RER データモデル表現については、入力ファイル生成に自動化が実現されていない部分に関し、これが可能になるよう精練化を行う予定である。

また帰納強化プログラムのデータベースへの応用にとって負荷の自動生成は大きな課題である。今回の実験では相互接続数の大规模クラスに分類する問題を扱ったので、他のクラスの例を負荷とした、より一般的な領域に対応できるよう [Shimizu 96] の負荷の自動生成を統合する。

さらに、帰納的強化プログラムシステムと関係データベース（RDB）を直接接続する試みは [Brockhausen 97] にも報告されているが、企業の中で運用されているシステムのあらゆる形態を検討するとき、精密な関係モデルに従って実装されていることがほとんど無く、すべてのシステムをこの提案を応用できるとは限らない。そこで [Anderson 95] に示されたリバースエンジニアリングの手法を用いて RDB の記述を ER モデルの派生表現形（ERC-+）に変換し、さらに [McBrien 97] の方法で標準の ER モデル表現への補正を行った後に、RER モデルに変換する方法を現在検討中である。これにより現在の業務システムのほとんどに対し、今回提案する KDD システムを利用することが可能になると考えられる。

9.2 AUTOMAIL の改善

今回試作した AUTOMAIL は、新たにメールを入手すると本文を形態素解析し、専用に開発した辞書を参照することで生成されるキーワードリストをもとにその後の処理を行う。この方法に限れば、新たな状況（新製品の発売等）が発生するたびにこの辞書の内容を人手で更新しなければならない、日常的に適用されるシステムを目指すのであれば、推論結果の品質の低下は最も避けなければならない課題である。現在のまでは更新作業の手間とのトレードオフになったに、これを解決するために、[中山 98] の知識ベースとソノハウスベースに分割したデータ構造による手作業の効率化を考慮に際した。

9.3 広域領域

現在インターネットの利用が一般家庭にまで浸透しつつある。また多くの企業ではこの技術を利用して、インタラネットを社内の情報システム基盤とし、すべての情報システムをオールで扱うことができるように改革を進めていている。こういった背景からますます個人が処理しなくてはならない情報量は増加する。電子メールの効率的な処理だけでなく、インターネットやインパートラネットの効果的な活用を図った広域領域への貢献を目指して、具体的には Web サーバ内のコンテンツや利用者のアクセス記録データベースを対象とした知識獲得を試みる予定である。

謝辞

本研究の機会を与えてくださった慶應義塾大学 大学院政策研究科研究科 隼藤充男 教授および現職 サイバーセンター 中山啓子 主任研究員との機会を通してデータベースのあり方に関する議論は、本論文を構成するにあたり大変有意義なものとなりました。また最終稿完成に際し、慶應義塾大学 環境情報学部 隼藤功主にお手伝いいただき、感謝の意を表したい。

◇ 参考文献◇


[SunUltra30, メインメモリ 528Mbyte である。]
データベースからの知識発見システム DB-Amp


[担当委員: 有村博紀]
1999年11月1日 受理

著者紹介

鈴木 憲子 (正会員)
1982年富士タンポックス (株)入社, 1985年からAI事
業部にてエクスペリメントシステムの開発に従事, 1996年慶
應義塾大学大学院政策・メディア研究科修士課程修了, 情
報処理学会, AAAI各会員.

古寺 廣一 (正会員)
1965年東京大学工学部計数工学科卒業, 1977年同大学
院修士課程 (計数工学) 修了, 同年, 電気通信研究所 (現電気
通信研究所) 入所, 1982年8月博士 (工学) 進位, 同年慶應義塾大
学大学院政策・メディア研究科博士課程修了, 情報処理会
計数工学科, 日本情報処理学会, 日本ソフェア学会, 日本
認知科学学会, 日本バイオメカニクス学会, 各会員.