

アクティブマイニングの構想と展開

The Active Mining Project

元田 浩

Hiroshi Motoda

大阪大学産業科学研究所

Institute of Scientific and Industrial Research, Osaka University
motoda@sanken.osaka-u.ac.jp, <http://www.ar.sanken.osaka-u.ac.jp/>

沼尾 正行

Masayuki Numao

東京工業大学大学院情報理工学研究科計算工学専攻

Department of Computer Science, Tokyo Institute of Technology
numao@cs.titech.ac.jp, <http://www.nm.cs.titech.ac.jp/>

山口 高平

Takahira Yamaguchi

静岡大学情報学部情報科学科

Affiliation2 in English
yamaguti@cs.inf.shizuoka.ac.jp, <http://www.ai-gakkai.or.jp/~user2/>

津本 周作

Shusaku Tsumoto

島根医科大学医学部医学科医療情報学講座

Affiliation1 in English
tsumoto@shimane-med.ac.jp, <http://www.ai-gakkai.or.jp/~user5/>

keywords: keywords in lowercase English

Summary

The need for 1) gathering the relevant information, 2) mining useful knowledge and 3) promptly reacting to situation change is ever increasing in this digital society of information overload. Active mining project explores core technologies of 1) active information gathering to effectively search and preprocess necessary information, 2) user-centered active mining to effectively mine various forms of information sources, and 3) active user reaction to make the mined knowledge easily accessible for prompt feedback. The project further aims at exhibiting the interleaved and spiral effect of these three by challenging to analyze medical and chemical datasets as a common test bed.

1. はじめに

周知の通り、通信技術を含む計算機ハードウェアの急速な進歩により、大量情報が各種ネットワークを飛び交い個人がアクセスできる情報量が飛躍的に増加している。この状況は、情報洪水にたとえることができ、実際、1) 膨大な情報空間のどこを見ればよいのかが分からない、2) 見る場所が同定できても、その中から目的にかなった価値ある知識を簡単には取り出せない、3) 状況変化に即応できず、頻繁な知識の更新に対応できない、などの大きな問題がクローズアップされている。情報収集・データ解析・目的設定変更のサイクルが高速回転し、個人も組織も情報洪水の中で疲弊しているのが現状である。このような状況を打破するために、新しいマイニングの枠組み「アクティブマイニング技術」を実現することを目的として、特定領域研究(B)「情報洪水時代におけるアクティブマイニングの実現」(略称:アクティブマイニング)が、平成13年度から平成16年度までの4年間の研究プロジェクトとしてスタートした。「アクティブ」の名が示す通り、システム側からの情報源への積極的な働きかけ、目的に合致した質の高い知識の効率的な発掘と効果的な提示、ユーザ側からのシステム側への迅速な

フィードバックの実現を標榜している。本稿はアクティブマイニングプロジェクトの目指すものと現時点までの成果を紹介したものである [元田 01, Motoda 02, AMh]。

2. アクティブマイニングの目指すもの

従来からデータマイニングのプロセスは複数のステップからなり、少なくとも次の5つ^{*1}: 1) 対象領域の理解、2) データの準備、3) パタン(知識)の発見、4) パタンの事後処理(視覚化、解釈など)、5) 結果の活用、が含まれ [Mannila 96]、さらに、これらが繰り返されると説明されている。しかし、実態は3)のパタン(知識)の発見に関する研究が中心であり、全体を有機的に結合する仕組みに関する本格的な研究は少ない。その理由は、あまりにも多くの要素が絡みあい、個人レベルの研究で全体をカバーできないことにある。

アクティブマイニングプロジェクトでは、上記のプロセスを1章で述べた3つの課題から眺め直し、これらを克服するために、対応して以下の3つの研究項目を掲げる。図1に示すように各研究項目がさらに3, 4, 3個の計画研究から構成されている。

*1 [Fayyad 96] では9個のプロセスに細分されている。

アクティブマイニング =
 アクティブ情報収集 + ユーザ指向マイニング
 + ユーザリアクション

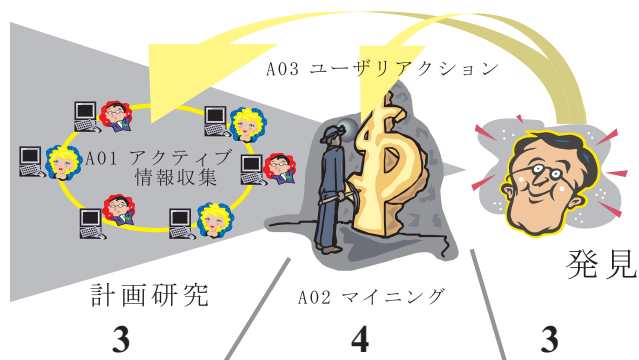


図 1 アクティブマイニングプロジェクトの全体構成

A01: アクティブ情報収集

不特定・非常常・大規模・分散知識源の中から、ユーザの目的や興味に合致するデータやそれらの関連を効率良く探索し前処理するための情報収集技術を、メタ情報源の活用、ヒューリスティック探索知識の活用、機械学習法の活用など、最新の情報処理技術を駆使して開発する(図2)。

目的: ユーザの目的や興味に合致する情報の効率的な収集

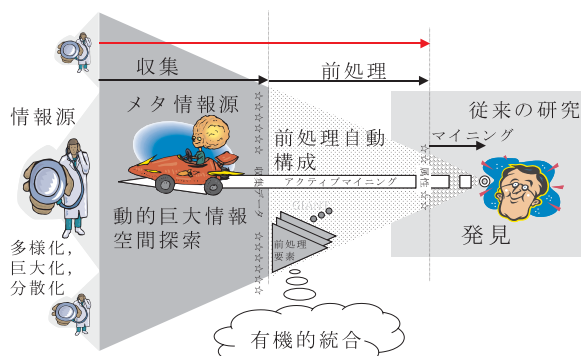


図 2 アクティブ情報収集の主要研究項目

A02: ユーザ指向アクティブマイニング

多様な形式や多種の情報源に対応でき、汎用性と状況の変化にも対応できる柔軟性を持つマイニング手法を開発する。とくに、テキスト情報に代表される半構造化データ、分子化学情報・ネットワーク情報に代表される構造化データからのマイニング、これら個別のデータに最適なマイニング手法の自動構築、状況変化検知に強力な例外性の発見技術に注力する(図3)。

A03: アクティブユーザリアクション

具体的な問題領域を対象にマイニングシステムを構築し、発掘した知識を、ユーザにとって有用なものとするための仕組み(知識の表示法, 評価手法, ユーザからの効果的なフィードバックの手法)を具体化する(図4)。

目的: ユーザの目的を考慮した有用知識候補の発掘

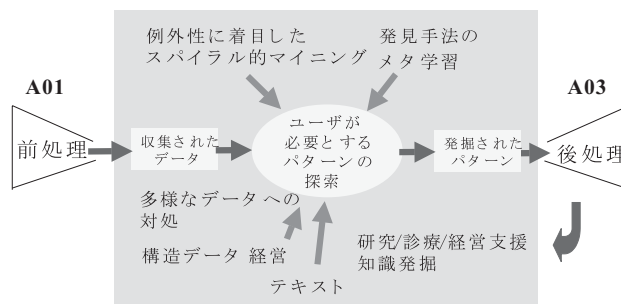


図 3 ユーザ指向アクティブマイニングの主要研究項目

目的: アクティブユーザリアクションによる専門家の発見プロセスの促進

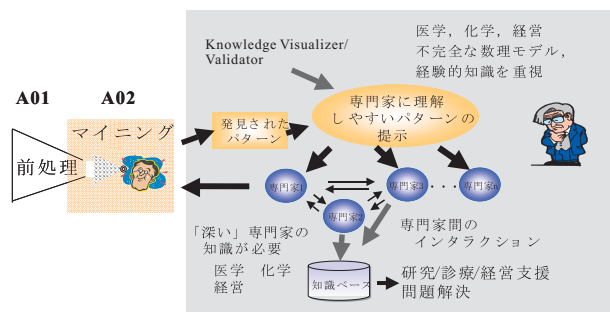


図 4 アクティブユーザリアクションの主要研究項目

具体的な問題としては医療データと化学薬品を取り上げる。とくに前者に関しては千葉大学医学部より提供される肝炎のデータを共通医療データとして各計画研究が取り上げ、上記の3つの機能を統合した”科学発見のらせんモデル”による相乗効果を実証する(図5)。慢性B型、C型肝炎がどのような経過をたどって、肝硬変、肝臓癌に至るかについてはよくわかっていない。肝炎の病理像(繊維化の程度)と血液検査データとの間にある程度の相関があると推測されているが、明らかな知見に至っていない。侵襲度が高く簡単にできる検査ではない肝生検に代わって(生検)検査以外の血液データから疾患の予後が予測できれば医学的にも大きな貢献になる。共通医療データからのアクティブマイニングは、現在、新しい医療行為として注目されている Evidence Based Medicine (EBM) (「エビデンス(科学的根拠)に基づいた医療」)を実践するための有効な手段を提供するものと期待されている。本プロジェクトで、共通データとして医療データを用いた実践を行なうことにより、EBMの効果を実証する。化学薬品に関しては、毎年多数の薬品が開発されているが、既知の生理活性に対する未知の化合物構造を発見し、すでに市場に出まわっている薬品に対し必要な警告を与える可能性を検討する。

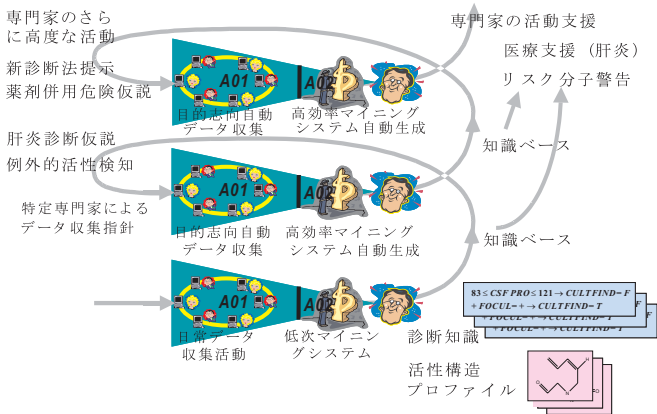


図 5 アクティブマイニングを実践した実証

3. 研究計画と現在までの主要成果

3.1 A01: アクティブ情報収集

研究項目 A01 では、必要な情報リソースをシステム自らが能動的に探しに行く仕組みを確立することを目標とし、ネットワーク上に分散したデータベース内から使用者の目的や興味に関連のありそうな情報を、積極的に探索する技術やテキスト情報処理技術、それを後段のマイニング処理に引き渡す前処理を行うことを目的とする。特に、WWW におけるメタ情報源の獲得、分散動的情報からのアクティブ情報収集、多段階学習方式によるデータ収集と前処理の自動化、という 3 つの課題に焦点をあて研究を進める。

3.1.1 WWW におけるメタ情報源の獲得

WWW などの膨大な情報空間に存在するメタ情報源（情報の所在情報 - LI 情報源、内容に関するメタ情報源 - MD 情報源）を（半）自動的に収集し、それを利用することにより、従来手法よりも質と量、そして効率とともに飛躍的に向上するような情報収集を実現する。

- 1) 関係学習による対話的文書検索 LI 情報源はハイパーリンクからなるグラフ構造をもち、またその Web ページのコンテンツにも特徴があるので、リンク構造とコンテンツからなる制約により判定できる。適合フィードバックにより LI 情報源判定のための制約を学習モデルを検討した。
- 2) WWW 上の情報収集 / 可視化のための免疫ネットワークを用いたクラスタリング MD 情報源を獲得するには、閲覧されている Web ページの系列に代表される情報ストリームから、情報の流れの抽出と流れ間の関係を捉えることが必要になる。免疫ネットワークを用いて文書集間間の時系列的関連を考慮した可塑的クラスタリングを提案した。
- 3) Web ページの部分更新のモニタリング WWW 情報は、時々刻々変化する動的なものであるため、獲得された情報源の更新を自動的に検知するシステムが必要不可欠になる。ユーザによって指定された一部分の

情報に特定の更新があった場合のみ、その更新をユーザに提示する部分更新モニタリングシステムを開発した。

3.1.2 分散動的情報からのアクティブ情報収集

頻繁に更新される情報源では、一回に更新される情報量はそれほど大きくはない。少量で速報性のある情報オブジェクトを Web 情報源から収集・統合し利用者の意思決定や問題解決を支援するアクティブ情報収集システムを開発する。

- 1) 知的情報収集アルゴリズムの開発 アクティブ情報収集システムで利用する知的情報収集アルゴリズムを開発した。本アルゴリズムは収集すべき情報を質（良い解を構成できるかどうか）と信頼度（新しい情報であるかどうか）の観点から評価し、それに応じて効率よく情報源から情報収集を行う。今後、Medline 等の医療情報源からの情報収集に応用する予定である。

3.1.3 多段階学習方式によるデータ収集と前処理の自動化

データベース間の通信ネットワークに学習能力を持たせ、関連する情報を収集し、それらの関係を自動的に求める、トポロジや情報伝達の優先順位を動的に変化させることができる Global Intelligence Associating Network (GIANT) の構成法を明らかにし、現実のデータの収集と前処理に適用する。

- 1) 前処理支援システムの構築 蓄積されたデータに対して前処理を施し、解析アルゴリズムが直接扱えるデータ（現状では XML 形式を仮定）に変換する過程を明確に定義し、その過程に含まれる処理を効率的に行えるデータ構造、及び自動化アルゴリズムを提案し、システムを構築した。
- 2) 伝言ゲーム型の情報収集と前処理 情報収集および前処理はルーチンワークではなく、従来のグループウェアの適用は困難である。情報提供者、ドメイン専門家、マイニング専門家の共同作業を支援する、伝言ゲーム型の情報収集および前処理結果の交換を提案し、システムを実装し実験で評価した（図 6）。

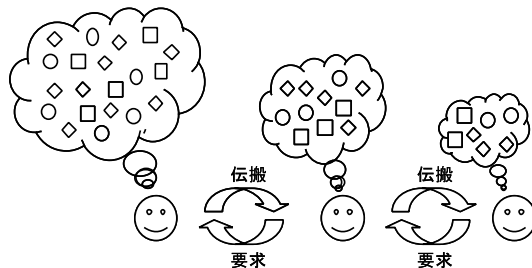


図 6 伝言ゲーム型の情報収集と前処理

- 3) HTML のリンク構造と構文的特徴に基づく知識獲得 バックリンクページがリンク集である頻度が高いこと、データ集については予め XML の雛型を準備できることを利用して、検索エンジンにより構文的な特徴

を共有する WWW ページを収集するシステムを実装し、有効性を確かめた。

- 4) 高次元インデックス技術を用いた検索処理性能向上について
 ヒトの思考は空間的な直感に基づいていることが多い。対象データとして位置情報を取り上げ、高次元インデックス技術を開発した。これにより、空間属性と非空間属性を併せもつ大量の時系列高次元データの処理に関し展望を得た。

3.2 A02: ユーザ指向アクティブマイニング

研究項目 A02 では、アクティブ情報収集機構で収集・前処理された情報から、ユーザの目的や興味に照らして重要・有用と思われる知識を発掘することを目標とし、多様な表現形式あるいは多種の情報源に対応できる「汎用性」、並びに、ユーザを含めた状況の変化に対応できる「柔軟性」の両者を兼ね備えた、大規模データからのマイニング手法の開発を目的とする。特に、構造化データからのマイニング、マイニングアプリケーション自動構築、例外性の発見、テキストデータからのマイニング、の 4 つの研究課題に焦点をあて研究を進める。

3.2.1 構造データからのアクティブマイニング

大規模なグラフ構造データ、空間分布構造データ、時系列構造データ、半構造データ、制約構造データなどの構造データを対象とし、ユーザの価値観を反映した重要なあるいは興味深い部分構造ならびにその特徴を知識として、ユーザの許容時間内に発掘するために必要な基礎技術を開発する。

- 1) グラフ構造データからの特徴的な部分グラフの発見手法 GBI の性能向上 構造の大きさに線形な処理時間で多頻度連結部分グラフを近似抽出する GBI アルゴリズムを、分類性能に基づく指標を用いて部分構造を並列抽出することにより、分類問題により柔軟に対応可能とした。発癌性化合物や変異源性化合物の分類に適用し良好な結果を得た。
- 2) GBI 法の属性構築への適用 決定木のノード生成時に分離能力最大のペアを選定し、チャンクして 1 つのグラフノードとし、各分岐ごとに同じ操作を再帰的に繰り返すことにより、分類に効果的な属性を逐次的に構築しながら利用するグラフ構造データ向きの決定木生成法を提案した。
- 3) 多頻度誘導部分グラフ抽出手法 AGM の連結部分グラフマイニングへの拡張 グラフ構造データからすべての多頻度誘導部分グラフパターンを高速導出する AGM 手法を、すべての多頻度連結部分グラフパターンを高速導出する AcGM 手法に拡張した。これにより、実規模の問題にも適用可能となった。
- 4) 文書クラスタリング 対称性と反射律のみを許すラフ集合モデル TRSM を用いて文書を表現し、階層的・非階層的クラスタリングアルゴリズムを開発し、情報

検索に適用した。

- 5) マイニング過程の視覚化 規則やデータの階層構造を視覚化し、ユーザがマイニングプロセスに積極的に関与し必要なモデルを選択できるマイニング環境を構築した。
- 6) 時系列データからのマイニング手法の開発 経営データを対象に、複数時点における複数ブランド間の購買関連性の分析を可能にした。その結果、時系列データとして蓄積されている POS データの特徴を活かした、時間の経過にそった状態の遷移を表現することが可能になった。

3.2.2 メタ学習機構に基づくアクティブマイニング

マイニングシステムの開発プロセスで重要な「データ加工」、「データマイニング」プロセス「ユーザの主観的基準」をリポジトリ化し、メタ学習機構に基づいて、これらのリポジトリを統合しながら、ユーザの使用目的に合致したマイニングアプリケーションを半自動合成するツールを開発する。

- 1) リポジトリに基づく帰納アプリケーション構築支援環境 CAMLET StatLog プロジェクトの 8 種類のデータセットを用いて、CAMLET を評価した。合成した帰納アプリケーションの平均正解率は、同プロジェクトで評価した 24 種類の代表的な帰納アルゴリズムのどの平均正解率よりも高い値を示した。さらに、効率的な仕様探索実現のために、仕様書換ルール (メタルール) の学習を試みた。
- 2) 慢性肝炎データセットのデータ前処理とルール発見 慢性肝炎データセットを分析した。データ前処理の後、時系列データをクラスタリングし、決定木学習により、検査項目値から GPT の 1 年間の変化を予測するルールを学習した。専門医より、乳びと GPT の関連性、GPT の周期性の予測結果などが興味深いというコメントもいくつかもらえた。
- 3) 多粒度の問題解決メソッドの検討 要求レベル - 問題解決レベル - 実装レベルと対応づけた 3 層のメソッドリポジトリを検討し、ビジネスドメインにおいて、プロトタイプを試作し、その有用性を確認した。
- 4) 領域オントロジー構築支援ツール ユーザにとって重要・有用な知識を発見するには、マイニングだけでなく、モデリング機能も必要である。ドメインモデルを構成する概念要素の仕様を与えるドメインオントロジーを構築するツールの開発を手がけ、手始めに、計算機可読型辞書とテキストコーパスから、ツールを試作した。

3.2.3 例外性発見に基づくスパイラル的アクティブマイニング

例外性は、大多数の傾向とは異なる性質を表し、それ自体が興味深いことに加えて新しい知識発見の糸口となることが多い。データ、知識、および環境などに関する例外性を連鎖的に発見するアクティブマイニング手法を

構築し、医学・商業・科学データなどに適用して領域専門家による評価で有効性を実証する。

- 1) スパイラル的例外性発見手法 与えられたデータ、既に発見された知識、および妥当性・有用性・新規性・意外性に関するユーザの評価値に照らしあわせ、既発見知識の近傍を、異なる離散化手法と確率的規準に基づいて探索する新しい例外知識の発見手法を開発した。
- 2) 状況変化の検知手法 ルール発見に重要な複数の指標に関する誤差が指定値以内である確率が指定確率以上となるために必要な事例数を与える最悪解析法 PAGA を提案し、発見された例外性を用いて、データ、知識、および環境に関する状況変化を検知する手法を開発中である。
- 3) スパイラル的特異ルール発見手法 対象データを変更し、興味深い特異データを抽出することにより、特異ルールを発見する手法を開発し、相関ルールや例外ルールと形式的に比較・分析し、特異指向マイニングの理論的根拠を確立した。現在、マイニングプロセスのメタレベルの制御メカニズム、複数のマイニングエージェント、マルチデータソースからのマイニング手法を開発中である。

3.2.4 利用者からの要求を考慮したテキストデータからの知識抽出

ある特定の意味クラスに属する用語の発見手法を、論文要旨に出現する名詞句がそのクラスの使用語であるかどうかを同定する問題を例に開発する。とくに、学習事例が少ないという現実的な想定に基づき、用語が出現するテキスト中の文脈が意味クラス推定にどの程度有用であるかを明らかにする。

- 1) 未知語を含む英文文書内の単語の品詞推定 前後の文脈、単語の綴りを手がかりとして、未知の単語の品詞を決定し、それによって、専門用語と考えられる名詞句の同定を柔軟に行うことを試み、実用的な解析時間で未知語を含む専門文書中の単語の品詞推定を約 97% という高い精度で達成した。
- 2) 文書中の基本句の自動抽出 品詞付与が行われた文書に対し、名詞句や動詞句などの基本句を精度よく同定する手法を提案した、各単語へのラベル付与問題として見直し、SVM を用いた学習システムを複数混合することにより、従来法を上回る精度を達成した。
- 3) テキストからの専門用語の抽出と分類 Medline アブストラクトを題材とし、テキスト中に現れる名詞句を対象に、その意味クラスを推定する実験を行った。病名のみ注目し、未知の名詞句に対し、それが病名であるか否か自動推定可能かどうかを確認した。文脈情報のみを用いた場合は、精度が約 70%、再現率が約 60% で病名かどうかを判定することができた。また、用語内の文字情報まで利用する場合には、精度が約 90%、再現率が約 80% で判定が可能との知見を得た。

3.3 A03: アクティブユーザリアクション

研究項目 A03 では、具体的な問題領域（医療、化学薬品）を対象にデータマイニングシステムを構築し、アクティブマイニングの結果得られた知識を、適用領域のユーザにとって有用なものとするための仕組みを具体化することを目標とし、診療情報生成システムの開発、化学物質群からのリスク分子発見およびヒューマン・システム・インタラクションに基づく知識評価と選択という 3 つの研究課題に焦点をあて、ユーザのアクティブなリアクションを容易にし、“科学的発見のらせんモデル”を実現する一般的な枠組みの構築に取り組む。

3.3.1 ラフ集合に基づくアクティブマイニングによる診療情報生成システムの開発

専門医のあいまいな知識をラフ集合に基づき定量的に取り扱うことにより、専門家の知識との総合作用による知識の発掘の促進および医療現場に有用な診療情報の生成を目指す。とくに、共通医療データが時系列データであることに着目し、時系列データからの診療情報生成に関する基礎的技術に焦点を当てる。

- 1) 多重スケールマッチングによる時系列データのセグメント化 同一患者の継続的な時系列検査データは、数日を単位とする短期間の推移のみならず、年単位の長期にわたる検査値推移パターンと疾患との対応関係を示しており、その解析には、データのセグメント化とセグメント間のマッチングが重要となる。これを効果的に実現する多重スケールマッチングを提案した。
- 2) ラフクラスタリングによる系列データの分類 ラフ集合論により対象のまとまり具合を識別不能度として表現し、分散、重心等を定義することが困難な相対的類似度で表現されたデータに対して可読性の高いクラスタを生成する手法を開発し、多重スケールマッチングでセグメント化した共通医療時系列データの分類に適用した。

3.3.2 アクティブマイニングによる化学物質群からのリスク分子発見

薬品類とその活性のデータベースから、1) それぞれの生理活性に対して特徴的な部分構造や物理化学的性質を見出し、2) 新規の化学物質群から未知の生理活性を予測して、危険を回避する手法を確立することを目的とする。

- 1) 変異原性に対する検討 芳香族ニトロ化合物の変異原性問題に AGM(3.2.1) およびカスケードモデルを適用し、「芳香族ニトロ化合物で ortho 位置換基の有無が重要な因子となる」という有意義な結果を得た。
- 2) 発がん性の解析と予測 有機塩素化合物の発がん性問題にカスケードモデルを適用し、「有機塩素化合物の活性と水素結合受容体の有無および分子の柔軟性との間の高い相関」を見出すことができた。解析結果は、国際ワークショップ Predictive Toxicology Challenge 2001 にて、モデルの理解容易性において参加 14 グループ中第 1 位の評価を受け、また female rat に対

する予測精度においても、ROC 解析の結果が非常に良いとの評価を得た。

- 3) ドーパミン活性に関する検討 各種の性質にもとづくトポロジカルフラグメントスペクトルを利用して化合物構造間の距離を定義し、事例ベース推論により、化学者が見て納得できる類似構造の分子を選択することができた。ドーパミン活性については、類似 20 分子中で 3 分子が活性を示した。

3.3.3 ヒューマン・システム・インタラクションに基づく知識評価と選択

「知識が有用である」とは、利用者にとって理解が容易であり目的に応じた的確に使用できること、また、利用者の創造性を刺激しうる機能を備えることを意味する。その実現に向けて、知識の需給関係に注目し、知識の候補を供給するシステムとそれを解釈・選択・利用する専門家とのインタラクションを通じて、知識を評価・選択できるような方式を確立する。とくに、評価に関しては、利用者個人あるいはグループの主観までを評価尺度に含み、従来研究されてきた客観的な基準での知識評価方法を超えるものを目指す。

- 1) 客観的意志決定の支援: 病態モデルの構築 血液中の複数の酵素データを用いて病態モデルを臨床検査の実データから因子分析を用いて構築し、因子スコアの視覚的表現を試みた。視覚化によって、臨床家の検査データ評価に関わる負担を軽減し、さらに、重要な病的異常値の見逃しを防止する効果が得られた。
- 2) 客観的意志決定の支援: 病態パターンのクラスタリング 血清蛋白分画データの病態パターンを自己組織化マップを用いクラスタリングし、得られたクラスターを病態と対応づけるために、分類クラスと他の臨床検査データを属性として決定木分析を行った。意味づけされた分類クラスには、臨床診断につながる有用な付加価値情報が得られた。
- 3) 主観的発見: チャンス発見プロセスの実現 チャンスを評価するメタな指標として、P: 行動の提案可能性, U: 気付かれにくさ, G: 成長性の 3 つを提案した。P, G を具体的に把握するために、グループディスカッションにおける人と人のインタラクションを通してチャンスを選択してゆくプロセスを実現した。意思決定環境についてのデータ(販売者における購買データなど)に適用し、視覚的データマイニングによって刺激してチャンスを発見する過程を促進する結果を得た。
- 4) 主観的発見: チャンス発見二重らせんモデルの実現 新たなチャンスを発見するモデルとして二重らせんモデルを提案し、グループインタビューにおける各参加者のメモ内の単語の相関関係を図示する視覚的なテキストマイニング(KeyGraph)の出力図を通じた議論からチャンス表出化を促進するシステムに適用した。このシステムで、顧客の購買(POS)データからスー

パーマーケットの購入金額増加の鍵となる商品や、その店の経営状態のおおまかな変化を示す予兆を発見することができた。さらに、社会調査を行った社会学者の「関心」を起点として二重らせんモデルを実行した結果、社会的に新規性と説明能力の高い仮説が獲得された。

4. ま と め

文部科学省特定領域研究(B)「アクティブマイニング」プロジェクトの構想と現在までの主要成果を報告した。本プロジェクトは開始してまだ日が浅い。本稿で報告した成果は第一歩に過ぎず、新しい成果は逐次ホームページ[AMh]に掲載して行く所存である。これからの成果に期待して頂きたい。

謝 辞

本解説の執筆に当っては、山田誠二(情報研)、北村泰彦(大阪市立大)、鷲尾隆(大阪大)、鈴木英之進(横国大)、松本裕治(奈良先端大)、岡田孝(関西学院大)、大澤幸生(筑波大)、寺野隆雄(筑波大)他、アクティブマイニングプロジェクトメンバー各位の多大な協力を得た。ここに謝意を表します。

◇ 参 考 文 献 ◇

- [AMh] <http://www.ar.sanken.osaka-u.ac.jp/activemining/>.
 [Fayyad 96] Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P.: From Data Mining to Knowledge Discovery: An Overview, in Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. eds., *Advances in Knowledge Discovery and Data Mining*, pp. 1-34, AAAI Press (1996).
 [Mannila 96] Mannila, H.: Data mining: machine learning, statistics, and databases, in *Proc. of The Eight Int. Conf. on Scientific and Statistical Database Management*, pp. 1-34 (1996).
 [Motoda 02] Motoda, H. ed.: *Active Mining - New directions of Data Mining*, IOS Press (2002).
 [元田 01] 元田浩: 情報洪水時代におけるアクティブマイニングの実現, 平成 13 年度科学研究費補助金 特定領域 (B) 研究成果報告書(本プロジェクトのホームページから入手可能) (2001).

[担当委員: × ×]

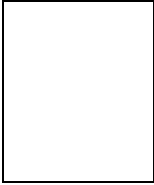
19YY 年 MM 月 DD 日 受理

著 者 紹 介

元田 洵(正会員)

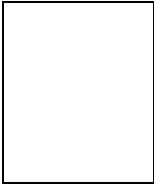
1965 年東京大学工学部原子力工学科卒業。1967 年同大学院原子力工学専攻修士課程終了。同年、日立製作所に入社。同社中央研究所、原子力研究所、エネルギー研究所、基礎研究所を経て平成 7 年退社。現在、大阪大学産業科学研究所教授(知能システム科学研究部門、高次推論研究分野)。原子力システムの設計、運用、制御に関する研究、診断型エキスパート・システムの研究を経て、現在は人工知能の

基礎研究，とくに機械学習，知識獲得，知識発見，データマイニングなどの研究に従事．工学博士．日本ソフトウェア科学会理事，人工知能学会理事，同編集委員会委員，日本認知科学会編集委員会委員，Knowledge Acquisition (Academic Press) 編集委員，IEEE Expert 編集委員を歴任．Artificial Intelligence in Engineering (Elsevier Applied Science) 編集委員，International Journal of Human-Computer Studies (Academic Press) 編集委員，Knowledge and Information Systems: An International Journal (Springer-Verlag) 編集委員．1975 年日本原子力学会奨励賞，1977，1984 年日本原子力学会論文賞，1989，1992 年人工知能学会論文賞受賞．1997 年人工知能学会研究奨励賞受賞，1997，1998 年人工知能学会全国大会優秀論文賞受賞．人工知能学会，情報処理学会，日本ソフトウェア科学会，日本認知科学会，AAAI，IEEE Computer Society，各会員．



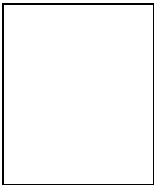
沼尾 正行 (正会員)

1982 年東京工業大学工学部電気電子工学科卒業．1987 年同大学院情報工学専攻博士課程終了．工学博士．東京工業大学工学部情報工学科講師を経て，現在同大学院情報理工学研究科計算工学専攻助教授．1989 ~ 90 年スタンフォード大学 CSLI 客員研究員．人工知能，機械学習，関数型言語などの研究に従事．情報処理学会，人工知能学会，日本認知科学会，日本ソフトウェア科学会，電子情報通信学会，AAAI 各会員．



山口 高平 (正会員)

abc



津本 周作 (正会員)

abc

Abstract:

和文アブストラクト 900 字程度 (25 字 × 36 行) *元が英文の論文には不要です．