

事例ベース推論を利用した情報検索ノウハウの蓄積と利用 — 経済統計データベースへの適用 —

鍋田 茂子 *1,*2, 寺野 隆雄 *2

*1 (株)CSK, *2 筑波大学大学院経営システム科学専攻
email: terano@gssm.otsuka.tsukuba.ac.jp

事例ベース推論を適用した経済統計データベースに関する検索支援システムを開発し、事例ベース推論の効果に関して評価を行った。経済統計データベースの検索の特徴として、次の項目が挙げられる。1 経済統計の知識や、データベースに関する知識が必要である。2 要求通りのデータが存在しない場合に代替の情報を探す必要がある。3 欠測等検索の後でなければ判らない情報がある。事例ベース推論導入の狙いは、過去の検索事例を使って、検索効率を向上させることと、従来の知識ベースでは、吸収しにくい検索上のノウハウを吸収することである。事例ベース推論の導入効果を推定するための指標として、領域分割度、類推効果、再現効果という3種類を定義し、実験によってこの有効性を検証した。

An Empirical Study on a Case-Based Reasoner for Retrieving Financial Statistical Databases

Shigeko Nabeta *1, *2, Takao Terano *2

*1 CSK Corp.,

*2 Graduate School of Systems Management, The University of Tsukuba, Tokyo
3-29-1, Otsuka, Bunkyo-ku, Tokyo 112, Japan

This paper describes the result of the development on an Intelligent Information Retriever using case-based reasoning(CBR). A case-based reasoner solves new problems by adapting relevant cases from the library(case-base). In this paper, we applied CBR method to the domain of information retrieval for financial statistics, in which the expertise on financial statistics, databases, and IR commands are required. Further, the users need semi-optimal solutions even when the databases do not generate exact solutions. This system has two components: a rule based reasoner for general problem solving and a CBR for using past cases. To evaluate the system performance, the authors carried out intensive experiments based on the observation data of end-users' works. The experiments show the effectiveness of CBR on some problems, which cannot be solved by a general problem solver. The experiments also suggest the advantages of the CBR method for improving the performance of the conventional information retrieval system.

1 はじめに

事例ベース推論とは、過去に経験した事例（問題とその解決法）を、統一した形式に整理して蓄積し、新たな問題が発生した場合に、蓄積された過去の事例を直接問題解決に利用する推論方法である [1], [2]。蓄積された事例の集まりのことを事例ベースと呼ぶ。一般的な事例ベース推論では、入力された問題に対して、次のような手順で、問題解決を行う。

1. 事例ベースのなかから適切な事例を検索する。
2. 新たな問題に事例をあてはめる。
3. 事例をあてはめた結果が成功すれば、入力された問題を事例の形式に整理して事例ベースに登録する。
4. 事例をあてはめた結果が失敗した場合は、成功するように事例に修復を加える。
5. 修復した事例を適用して成功した場合は、修復した事例を事例ベースに登録する。
6. 修復によっても成功しない場合（既存の事例では、問題解決が不可能の場合）は一般的な問題解決の手続きによって解を求め、事例ベースに登録する [1]。

事例ベース推論の提唱された背景には、過去の出来事を想起することによる問題解決が人間が行う問題解決の場面によく現れ、しかもこれらの過去の出来事に関する知識は、プロダクションルールとして整理するには、因果関係が明確でないまま使われていることが多い、という事実がある。エキスパートシステムへの適用としては、専門家が、主に過去の経験からの類推によって問題解決をおこなっているような場合に、事例ベース推論が有効であるといえる [3]。

また、事例ベース推論では、推論の過程で新たな事例が事例ベースに蓄積される仕組みになっており、知識獲得の機能を内包しているといえる。本論文では、経済統計データベースの検索支援への事例ベース推論の適用を試み、事例ベース推論の効果について考察する。DB検索に事例ベース推論を適用した研究としては、文献検索への適用 [4] などが報告されているが、統計数値データベースへの応用例は見当たらない。

本論文では、事例ベース推論を適用した経済統計データベース検索支援システムを紹介し、事例ベース推論の導入効果を推定するための指標の利用を提案する。本システムは、パソコン、PROLOG、ホスト上のデータベースをアクセスするためのフロントエンドプロセッサから構成されている。

2 経済統計データベース検索の特徴と事例ベース推論適用の狙い

経済統計分野で、複数のデータベースを利用した総合的な情報検索を行う場合に、ユーザは知りたい情報を経済統計の用語で正確に表現するとともに、種々のデータベースの中からの確なものを選択し、そのデータベースに特有の検索コマンドを使って検索をおこなわなければならない。

経済統計分野での情報検索の特徴として次のものが挙げられる [5]。

1. データベースの収録状態に関する知識、経済統計用語に関する知識、データベース検索コマンドに関する知識等が必要である。
2. 検索したい情報が直接データベースに収録されていない場合に、代替となる情報が存在する場合が多い。これらの代替情報に関する知識のなかには、常識として固定化してはいないが、代替情報としての利用例が多いために認知されているものがある。
3. データベース収録時の欠測等、検索実行後でなければわからない知識があり、これらの知識は、検索の実行によって増加し続ける。

これらの特性のため、経済統計データベースを利用して、適切なデータを検索することは、専門的な知識のないユーザにとって困難を伴う場合が多い。エキスパートシステムとして、これらの特徴を捉えると、項目2、3は、従来のプロダクションルールやフレームで

は、表現しにくい知識と考えられる。これらの知識を吸収し、効率的に問題解決を行うために、検索事例を事例ベースとして蓄え、事例ベース推論によって検索の支援を行うシステムを開発した。

3 開発方針

3.1 検索要求の表現

システムの受け付ける入力範囲を、経済統計のなかの貿易統計に限定した。ユーザの検索要求は、1 報告国、2 相手国、3 品目、4 事項という4つの属性によって表現する。例えば日本からアメリカへのコンピュータの輸出額は、報告国（日本）、相手国（アメリカ）、品目（コンピュータ）、事項（輸出額）と表現される。実際に入力する検索要求には、さらに、期間や、期種、単位、加工方法等の情報が付加されているが、データベース選択の上で、重要な要素として、事例ベースに登録する情報は、上記4項目とした。検索要求の入力は、図1に示すテンプレート型の入力画面を使って属性毎に行うが、報告国、相手国、品目に関しては、キーボード入力によってユーザの要求そのものを入力し、事項に関しては、データベースの収録上、検索可能なものをメニュー選択方式で、入力するようにした。

また、検索要求によっては、全ての属性に意味のある値が入るとは限らないことが、複数のデータベースを対象とした情報検索問題の特徴となっている。例えば、「日本からアメリカへの輸出額」の場合、品目に相当するものは、'全品目'となり、全域を示す用語になる。また「日本の海外資産」の場合、相手国、品目は、それ自体意味を持たなくなる。

①<状態表示領域>		説明	入力確定	前画面へ	終了	操作メニュー ④基本項目入力域
(報告国) _____ の (期間) _____ を		(相手国) _____ に対する (品目) _____ の (事項) _____ に関して (加工) _____ で表示してほしい				
(輸出・輸入・貿易) 額 (輸出・輸入・貿易) 依存度 輸(出・入) シェア 産業内分業度指数 輸(出・入) 価格指数 外貨準備		(輸出・輸入・貿易) 量 双曲化指数 GNP GDP 為替レート 消費者物価 海外(資産・負債) 人口	原数値 指数 前年同期比 同期比 成長率 対GNP 比 所得弾性値 実質値			
数字を入力して下さい。						

図 1: システム入力画面

3.2 一般的な問題解決方法

検索支援のための処理は、基本的には、3.1 に示した検索要求を、存在するデータベースの範囲で、検索可能なものに変換することと同義であるといえる。一般的な問題解決は、次のような手順で行う。

1. 検索可能なデータベースに対して、データベース毎に検索可能な事項、及び、属性の組合せを知識として蓄え（この知識をデータベース辞書と呼ぶ）、入力された検索要

求にマッチするデータベースを探索する。適切なデータベースが存在しない場合は、代替情報（次善情報）を、知識ベースを使って生成する。代替情報の種類と生成に必要な知識ベースの関係を表1に示す [6]。

表 1: 代替情報の種類と必要な知識ベース

代替情報の種類	生成に必要な知識ベース名
用語を同義語で代替する	類義語辞書
下位の項目の演算により導出する	導出辞書
数値的に近いもので代替する	代替データ辞書
属性構成を変更する	データベース選択辞書
検索要求の論理的誤りを訂正する	データベース辞書

2. 複数のデータベースが検索可能な場合は、各データベースの収録条件を示して、ユーザに、希望するデータベースを選択させる。
3. データベース毎の検索コマンド生成手順に従って検索コマンドを生成し、検索を行う
4. 既存の辞書の範囲で、検索コマンドが生成できない場合は、専門家に検索コマンドを入力させる。

3.3 事例ベースの利用

3. 2で述べた一般的な問題解決方法を行う前に、事例ベースを検索し、過去の検索事例が利用できる場合は、これを利用する仕組みを構築した。事例は、PROLOG のデータベースとして蓄えられ、1 検索要求、2 標準化された検索要求、3 データベース名、4 属性指標、5 検索コマンド、6 一般化した検索コマンド、から構成される。属性指標とは、検索要求の属性値を値によって1から3までの数値で表現したもので、事例を検索するためのインデックスとして使われる。

例 case([日本, 米国, 車, 輸出額], [日本, アメリカ, 乗用車, 輸出額], db1, [1, 1, 1, 輸出額], [\$RETR 317010503/304/80FY-82FY\$, \$DISP V3\$], [\$RETR 3/1/5\$, \$DISP V3\$]) .

下線部の意味は次のとおりである。

1. ユーザの入力した検索要求（属性毎に表現されたもの）。
2. 検索要求を類義語辞書を使って基本名詞に変換したもの。基本名詞は、国名、品目等の一般的な呼称で、データベース検索コマンド生成に必要な検索コードが基本名詞との対応表によって与えられている。但し、属性‘事項’に関しては、メニュー選択によって入力が行われるので、入力された属性と基本名詞は同じものである。
3. データベース名
4. 属性指標検索要求を示す属性を次の規則で、1から3までの数字に指標化したもの

- (a) 属性の値が全世界、全品目等の全域を示す用語になっている場合：3
- (b) 日本のGNP等の検索要求の場合の相手国や品目のように、属性自体に意味の無い場合：2
- (c) 上の1, 2以外の用語が入力されている場合：1

属性'事項'に関しては、データ構造の性質上、必ず指標は1になるので、指標の代わりに、入力された値そのものを指標として使う。

- 5. 検索コマンド1の検索要求に対応する検索コマンド(成功したもの)。
- 6. 一般化検索コマンド5の検索コマンドのうち、報告国コードを示すものを⁻1に、相手国コードを示すものを⁻2に、品目コードを示すものを⁻3に、事項を示すものを⁻4に、期間を示すものを⁻5におきかえたもの。

事例の適用は次の3段階になっている。

- 1. データベースに関する知識を参照するまえに、事例ベースを検索し、同じ検索要求が登録されていれば、事例として蓄えられた検索コマンドを利用する。
- 2. 同じ検索要求がないとき、'若干の修正'を加えれば適用できる事例を検索し、これを適用する。'若干の修正'とは、類義語辞書を使い、事例の国名や品目名を検索要求と同じものにする操作のことである。
- 3. '若干の修正'による適用が不可能な場合、部分的に別の事例を合成すれば適用できるかどうかを検証し、できればこれを適用する。

(例) 検索要求: 米国から西側先進国への一般機械の輸出額

事例1 米国から日本への電気製品の輸出額:

(検索要求と属性指標が一致する事例)

事例2 西側先進国のGNP

(西側先進国という用語を使った事例)

(類義語辞書には、'西側先進国'という用語は、存在しないが、'先進国'という用語を'西側先進国'の代わりに使って検索を成功させた事例)

本システムのモジュール構成を図2に示す。

4 事例ベース推論の導入による効果

4.1 導入効果を示す指標

事例ベース推論の導入効果に関して次の指標を定義し、実験によって妥当性を確認した。

- 1. 領域分割度: 同じインデックスを持つ問題の集合の最大のものが、問題全体に占める割合。
- 2. 類推効果: 一般的な問題解決方法に従った場合と事例からの類推(事例に修正を加えて問題解決を行うことを指す)を行った場合の処理時間の比。
- 3. 再現効果: 過去に経験した問題と同じ問題を処理する場合の処理時間の短縮比。

領域分割度は、インデキシングによって、対象とする問題(ここでは検索要求の集まり)が、どの程度の大きさのグループに分割されるかを示す指標で、問題全体の大きさと、この指標の値によって、事例の蓄積が処理効率に及ぼす効果を推定することができる。類推効果は、初期状態における事例ベース推論の効果を示す。再現効果は、過去に経験した問題に対する事例ベース推論の効果である。

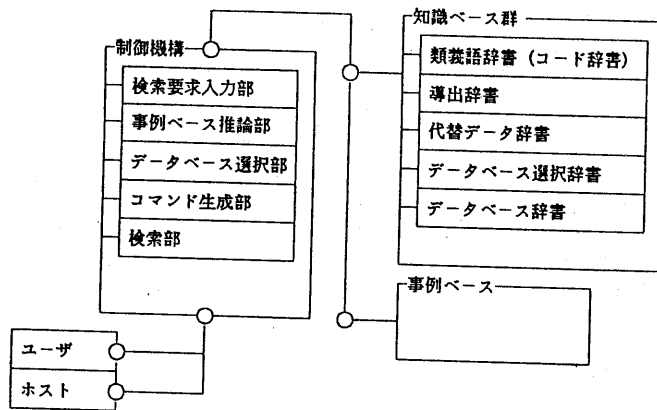


図 2: モジュール構成

4.2 実験の方法

実験データとしては、通商白書を参照して、実際のデータベースユーザを想定した60件程度の検索要求を作成した。実験には、通常の入力画面を使わず、入力ファイルからデータを入力した。検索結果の評価は、作成されたコマンドをユーザが見て評価する方法をとった。また、ホストとの通信による、処理時間の不安定性を取り除くため、ホスト側データベースとの通信は行わなかった。このような条件で、下記のテーマに従って、実験を行った。

1. 毎回異なる検索要求を入力した場合の事例ベース推論の性能について
2. 一定の検索要求の範囲で繰り返し処理を行った場合の処理時間の収束状況について
3. 事例の蓄積方法を変えた場合の処理時間の変化について
4. ユーザの独特の要求に対する事例ベース推論の反応状況について

1は、事例ベースの参照による処理効率の向上と、事例の蓄積による参照時間の増大の関係を調べるために行ったもので、60件の検索要求を順番に処理した。事例は、既に存在するものと同じでなければ必ず事例ベースに登録する方法をとり、初期状態では、事例ベースには、これらの検索要求には適用不可能な事例を2件登録した。

2は、一般的な情報検索の場面に近い状況での効果を調べるためのもので、用意した検索要求のなかから、乱数を発生させて、無作為に1件ずつ取り出して、繰り返し処理をおこなった。事例は、1と同様の規則で登録した。事例ベースの初期状態も1と同様である。合計203回の検索支援処理をおこなった。

3は、事例の蓄積方法の変化による処理効率の違いを調べる為の実験である。他の事例登録の方法とは異なり、事例は登録しない方法で、処理時間を測定した。2と同様に無作為に選択した検索要求に対して処理を行った。事例ベースの初期状態は、検索要求に適用可能なものを2件登録した場合と、適用不可能なものを2件登録した場合の2通りを用意した。それぞれ60回の検索支援処理をおこなった。

4は、ルールベースでは表現しにくい知識の吸収度合いを調べるための実験で、あまり一般的でない知識を使って、システムの回答を修正し続けた場合の結果である。事例は、1と同様の規則で登録した。事例ベースの初期状態は、検索要求に適用不可能な2件の事例を登録した状態とし、20件の検索要求の範囲で、無作為に50件の処理をおこなった。

4.3 実験結果と考察

4.3.1 毎回異なる検索要求を入力した場合

毎回異なる検索要求を入力した場合の事例ベース推論の性能について実験結果を図3に示す。

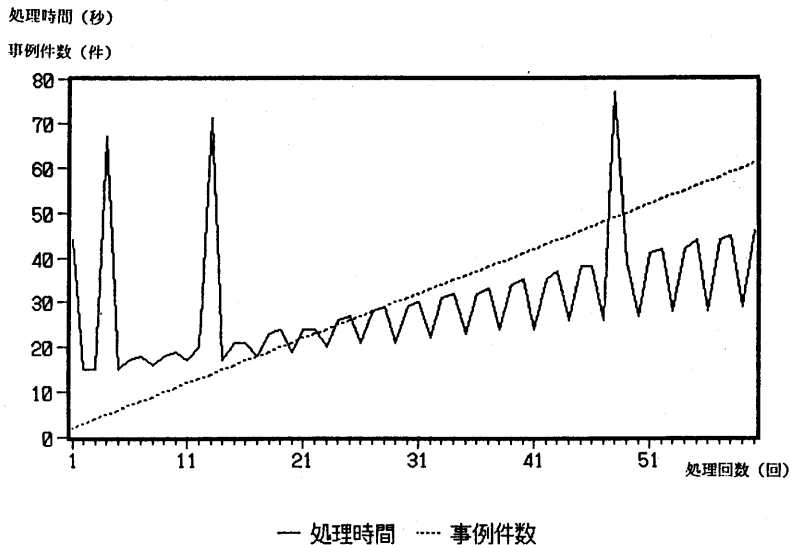


図 3: 毎回異なる検索要求を入力した場合

1 回目の処理時間 4 1 秒は、一般的な問題解決をおこなっている時間である。また、6 0 秒以上の時間が掛かる処理は、用語の学習（類義語辞書の新規登録）を行っている。グラフから、一旦利用可能な事例が登録された後、事例からの類推の時間は、2つの方向に向かって増加していることが判る。これは、問題領域が2種類の異なるインデックスを持ったグループに分割可能で、グループの大きさが2対1になっていることによる。即ち、蓄積された事例件数が増加すると、事例参照時間も増加し、参照する事例は、入力された検索要求と同じインデックスを持つものに限られる。最終時点では、事例を利用した処理時間が、一般的な問題解決の処理時間を上回る結果になっている。この問題領域に関しては、領域分割度が0.66であるが、領域分割度が低ければ、事例ベース推論は、最終時点でも、一般的な問題解決より優越することになる。また、一般的な問題解決のための処理時間が大きい場合も、事例ベース推論が有効な範囲は大きくなる。これは、類推効果によって示され、この例では、0.39 という数値が与えられている。

4.3.2 一定の検索要求の範囲で繰り返し処理を行った場合

4. 2. 1 は毎回異なる検索要求を入力した場合であるが、実際の情報検索に近い状態での実験を行うために、60件の検索要求のなかから、無作為に1件を抽出し、処理を行うことを繰り返したものが、図4である。問題領域が比較的狭い場合は、グラフの後半の部分で、定常状態となり、再現効果によって処理時間が決まる。この例では、再現効果、0.34である。一般的な情報検索では、品目、国名その他の分類項目の範囲が広く、問題領域全体と事例件数の差は大きいのが普通であるから、グラフの中間の状態が定常状態に近いといえる。この場合は、処理時間は、類推効果によって決まる。

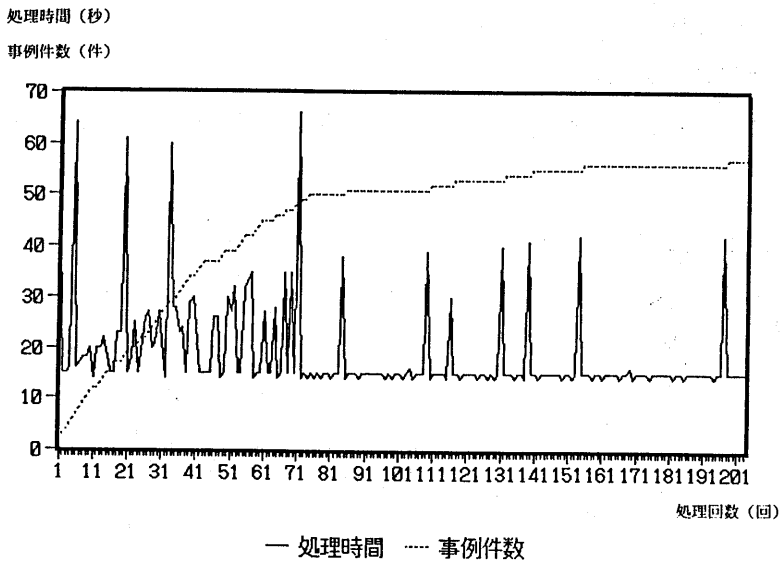


図 4: 一定の検索要求の範囲で繰り返し処理を行った場合

4.3.3 事例の蓄積方法を変えた場合の処理時間の変化について

事例ベースの増大による、参照時間の増加への対策として、事例を登録しない方法で、実験を行った。この場合、最初に参照可能な事例を登録しておく場合と、参照可能な事例がない場合の2種類の結果が得られた。図5、図6に結果を示す。参照可能な事例が存在する場合は、類推効果によって処理時間が決まる。また、参照可能な事例が無い場合は、一般的な問題解決を毎回行うことになる。このことから、予め、領域分割に関係のあるインデックスの種類だけ事例を用意しておけば、事例を登録しない方法で、効果をあげることができる。但し、類推効果と再現効果の差が大きく、再現による処理効率の向上が期待できる場合は、全ての事例を登録するほうが得策となる。

4.3.4 ユーザの独特の要求に対する事例ベース推論の反応状況について

あまり一般的でない知識（従来の知識ベースで吸収しにくい知識）として中国＝中国＋台湾ドイツ＝西ドイツ＋東ドイツ等のユーザの独自の定義を使ってシステムの回答を修正しながら処理を行った場合の結果を図7に示す。

100秒前後の時間のかかる処理は、ユーザが検索コマンドを入力しなおしている部分である。他の処理についても、他の実験と比べて、ユーザの確認操作が増えているため、時間が多く掛かっている。このように、知識ベースに収容しきれない知識の存在が認められる場合、事例ベース推論によって、それらの知識を活用することが可能になる。また、事例ベースに蓄積された内容を分析することによって、知識ベース化可能な部分を切り出して、改めて知識ベースを再編成するといった運用も可能になる。

4.4 指標の有効性について

領域分割度、類推効果、再現効果という3種類の指標と、問題領域全体の大きさを知ることにより、事例ベース推論の効果の概要を知ることができる。しかし、類推時間の増加の状態は、稼働環境によって異なるので、さらなる検討が必要である。また、本実験では、比較的単純な問題を取り上げたが、実際の問題は、さらに複雑で、一般的な問題解決に必要な

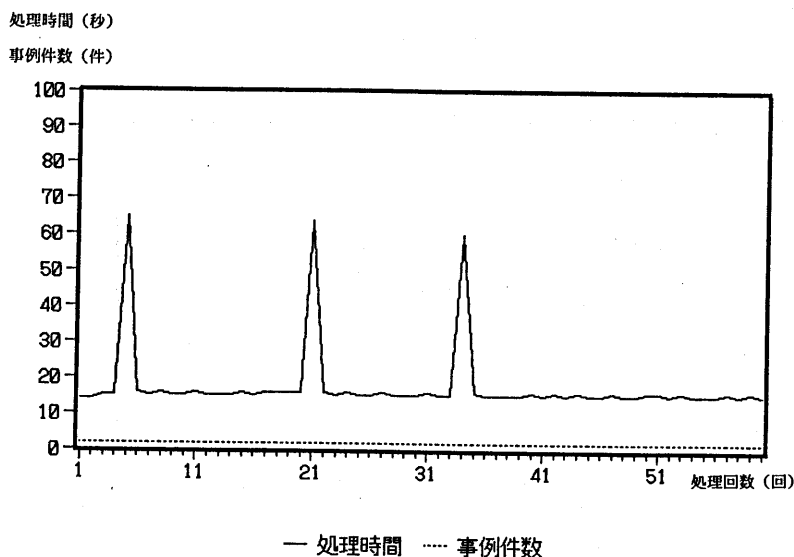


図 5: 事例を登録しない方式での結果 (参照可能な事例が存在する場合)

時間に巾があったり、類推効果に巾がある場合が考えられる。

5 おわりに

本論文では、貿易統計を中心とした経済統計データベース検索支援への事例ベース推論の適用を試みた。事例ベース推論の導入による検索効率の向上は、試作システムの評価によって実証でき、与えられた問題に対する導入効果を推定するための指標が得られた。経済統計データベースの検索に関する知識には、明文化された知識の他に、経験的に得られるものや、ある専門分野の世界では常識とされているもの、また新たな経済問題のトピックスとして追加されるもの等がある。これらの知識を運用時に吸収できるという意味でも、事例ベース推論の適用には、意味があるといえよう。今後は、事例ベース推論に蓄えられた'知識ベース化'の難しい知識に関して、分析を行うとともに、知識ベースと事例ベースを協調的に働かせる、情報検索システムのモデルを検討してゆく予定である。

参考文献

- [1] Riesbeck, Christopher K.: Inside Case-Based Reasoning. Lawewnce-Erlbaum- Associates, 1989.
- [2] (財) 日本情報処理開発協会: 知的情報処理システムに関する調査研究報告書 第1分冊—計画・設計型知識システムの構築方法論. 63-A001, 1989.
- [3] 中村孝太郎, 小林重信: 事例ベース推論の対話型モデルとその機械調整支援への適用. 人工知能学会, Vol.4 No.6 pp.704-713, 1989.
- [4] Ase, H., Kobayashi, A.: A Case-Based Reasoning System for Intelligent Information Retrieval. Proc. PRECAI'90, pp.583-588, 1990.
- [5] (財) 日本情報処理開発協会: 情報処理技術の応用に関する調査研究報告書—人工知能技術を応用したデータベース利用技術に関する調査—. 01-R007, 1990.
- [6] 寺野隆雄, 鍋田茂子: 事例ベース推論のデータベース検索への適用と考察. 電子情報通信学会, AI 90-79, 1990.

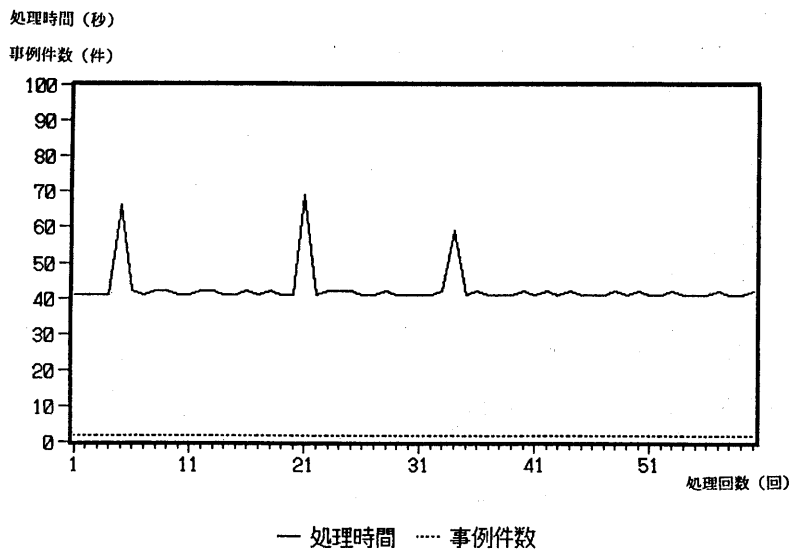


図 6: 事例を登録しない方式での結果 (参照可能な事例が存在しない場合)

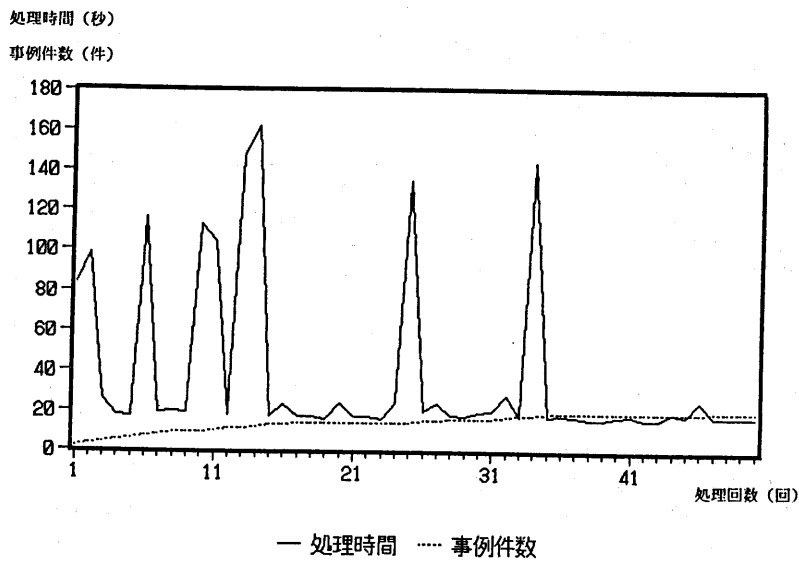


図 7: ユーザの独特の検索要求に対する反応