

事例ベース推論によるデータベースアクセス

寺本 浩史 阿江 忠

広島大学工学部

本研究の目的は、一種のデータベース検索代行システムのプロトタイプの構築である。データベースの検索はどの項目をどの程度重視するのかという主観によって異なってくる。したがってデータベース検索代行システムを開発する場合、主観を考慮する必要がある。本研究ではこの曖昧な主観を獲得するため、ユーザに直接主観を問い合わせるのではなく、ユーザが行った検索結果から定量化する方法を採る。

また、ユーザがデータベースからどのデータを選択するかという判断は非常に曖昧であり、従来のルールをベースとしたAI的な手法ですべてを記述するのは難しい。そのためニューラルネットの立場からCBRを実現するSBRによって推論を行い、逐次決定過程によるデータベースアクセスが実現される。

Database Access using Case-Based Reasoning

Koji Teramoto Tadashi Ae

Electrical Engineering, Faculty of Engineering, Hiroshima University
Kagamiyama 1-4-1, Higashi-Hiroshima, 724 Japan

We describe a method to construct a prototype of Automatic Database Searcher. The retrieved data in database depends on subjectivity. Thus we need to consider the subjectivity for constructing the system. The simplest method for acquiring the subjectivity is to ask of each user directly, but it is difficult to be implemented. In this paper, we obtain subjectivity from results that the user has retrieved. Since the judgments of what he or she chooses the data from the database include ambiguity, it is not necessarily adequate to construct a system by traditional rule-based AI technique, where we introduce SBR that realizes a Case Based Reasoning by Neural Net.

1 はじめに

エンドユーザとデータベース間のインテリジェントな仲介者として検索を支援するデータベースアシスタントシステムが広く研究されてきている [1-4]。これらの研究はデータベースに関する知識を持たないユーザに対し、リレーショナルデータベースに述語論理型の推論機能を持たせて簡単な自然言語理解機能を持つ利用者インターフェイスを実現させようと言う試みなどであり [5]、一般にデータベース検索システムと言えればこのデータベースアシスタントシステムを意味することが多い。

これに対し、本研究がめざすデータベース検索代行システムは、少なくともデータベースの扱い方を知っているユーザを対象とし、そのユーザーが日頃行うデータベース検索結果から知識を獲得し、ユーザに成り代わり検索を行うことを目的としたシステムであり、一般的なデータベースアシスタントシステムとはスタンスが異なる。前述のデータベースアシスタントシステムはデータベース研究からの要請であり、本研究のデータベース検索代行システムは、人間の行動をニューラルネットと見立て、ニューラルネットによる AI (Artificial Intelligence) 技法からのデータベースへのアプローチを試みるという位置づけにある [8]。

私たちがデータベースを検索し、その中から望ましいと考えるデータを選択するとき、どの項目をどのくらい重視するかという主観によって各項目に対する重み付けを行い、データ間に優劣の判断を下した上で、種々雑多な条件によりデータを選択するかどうかの判断を行っている。つまり、データベースからデータを選択するという行為は、主観によるデータベースの順位付けと、その上でのデータの評価という2つのフェーズで考えることができる。

そこで本論文では、主観の獲得方法と、その上でのデータを選択を行う推論方法について述べる。

しかしながら、ユーザの検索は一般に漠然としており、どういうデータを選択するかという曖昧な基準をルールベースの知識表現に形式化するの是非常に困難であろう。

そこで本研究では明確な知識表現の必要がない記憶に基づく推論 (MBR : Memory-Based Reasoning) [6, 7] にカテゴリの概念を採り入れ、処理を軽減した SBR (Small memory-Based Reasoning) [8, 9] を用いて推論を行う。

2 主観の導入

2.1 基本的考え方

中古車データベースから中古車を購入する状況を考えてみよう。私たちはまず、データベースのどの項目に着目するか考え、それをもとにデータを評価し、データに優劣を付け、そのうえで購入するかどうかの判断を行っている。つまり、データベースからデータを選択するという行為は2つのフェーズ、主観によるデータベースの並べ換えとその上での種々雑多な条件によるデータを選択に分けることができる。

したがって、人の行動パターンを獲得するためには第一にこの主観を獲得する必要がある。しかし、主観は常に一定ではなく、データ検索とともに連続的に変化していると考えの方が自然であり、また一度にすべてのデータの順序付けをしているのではなく、部分的に半順序を付けているだけである。よって、この曖昧な主観をユーザに直接システムから入力要求するのは問題がある。

そこで本研究ではユーザに直接主観を入力してもらうのではなく、データベースを主観によって並べ換えてもらい、その結果から主観を各項目の重み係数として定量化する。

その方法として、ユーザは理想とするデータに近い順にデータベースを並べ換えると考え。したがって、理想データとユーザが並べ換えたデータとの重み付き距離を考え、この距離の大小関係がユーザが並べ換えたように成立する重み係数を見つければよい。そして、この大小関係が成立しなくなるデータにおいて主観が変化したとする。

2.2 主観の定量化

項目数 m のデータベースが与えられ、ユーザーが好みの順に並べ替えたとする。各項目は、その項目の上限と下限を用いて 0 から 1 に正規化される。

正規化された i 番目に並べられたデータを m 次元ベクトル

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$$

で表す。

次に、ユーザーが理想とする仮想的な理想データを作成するため、各項目で

- (1) 数値が大きい方がよい
- (2) 数値が小さい方がよい
- (3) 特定の数値に近い方がよい
- (4) 重視しない

と言う質問を行い、理想とする数値を (1) ならば、1、(2) ならば 0、(3) ならば理想とする数値を入力させ、それを正規化し仮想的な理想データ

$$\mathbf{r} = (r_1, r_2, \dots, r_m)$$

を決定する。(4) のときは項目自体の評価を行わない。

つぎに、各項目の重み係数を

$$\mathbf{a} = (a_1, a_2, \dots, a_k, \dots, a_m) \quad |a_k| \leq 1$$

とし、データ \mathbf{x}_i と \mathbf{r} との距離

$$d(\mathbf{x}_i, \mathbf{r}) = \sqrt{a_1(x_{i1} - r_1)^2 + a_2(x_{i2} - r_2)^2 + \dots + a_m(x_{im} - r_m)^2}$$

を考える。

この距離 d がユーザーが並べた順に小さいものから並べばよい。

$$\begin{aligned} d(\mathbf{x}_1, \mathbf{r}) &< d(\mathbf{x}_2, \mathbf{r}) \\ d(\mathbf{x}_2, \mathbf{r}) &< d(\mathbf{x}_3, \mathbf{r}) \\ &\vdots \\ d(\mathbf{x}_i, \mathbf{r}) &< d(\mathbf{x}_{i+1}, \mathbf{r}) \\ &\vdots \end{aligned}$$

つまり、 $d(\mathbf{x}_i, \mathbf{r}) < d(\mathbf{x}_{i+1}, \mathbf{r})$ が成立するような重み係数 \mathbf{a} が存在する範囲を求める。逆に、この範囲を求め、主観を求めることによって、データを並べ直すことができる。

したがって、重み係数 \mathbf{a} が作る空間つまり、主観空間の中の 1 点をとることによって重み係数が決定でき、それが主観となる。また、この空間を構成するデータの範囲(カテゴリ)を代表するデータとして、カテゴリ内のデータの平均値をとり、参照データとして記憶しておく。

$$d(\mathbf{x}_i, \mathbf{r}) < d(\mathbf{x}_{i+1}, \mathbf{r})$$

が成立しなくなれば、成立しなくなったところから新たに主観空間を構成し直す。

2.3 カテゴリズ

主観を定量化する方法を前節で述べた。この方法は連続的に変化する主観のサンプル点を求めたものである。

この主観を用いてデータベースをユーザの望むように並べ直すためには、ある主観が適応できるデータの範囲を決定し、そのデータに対して主観を用いる必要がある。

そこで本研究では、このデータベースの分割に、主観の適応範囲をカテゴリとし、そのカテゴリ内のデータの平均値を参照データとすることによって、Kohonen の自己組織化アルゴリズムの1つである LVQ アルゴリズムを用いる [10]。

各主観ごとに LVQ の学習フェーズを用いて主観の適応範囲のデータの平均値を参照データとし、カテゴリを学習させる。こうして得られたカテゴリを用いて SBR(次節参照)を構成する。つまり、SBR には異なる主観プレーンがあり、これを切り替えることによって推論を行う。この主観を切り替えるもとなるのが状態である。

3 推論モデル

3.1 SBR

従来の AI は知識を明確な形式表現に変換し、それを用いて推論を行なうというものであった。しかし、エキスパートから得た知識からルールなどの明確な形式表現を得るのは多くの労力を必要とするボトルネックとなっている。

これに対し、MBR は、ルールを排し推論を行ないたい入力に対し、直接記憶を参照することに最近接 (nearest neighbor) データを探しだしこのデータが過去において行なわれた出力を利用することによって推論を行なう [6, 7]。これは事例ベース推論 (CBR : Case-Based Reasoning)[11, 12] の事例とデータが 1 対 1 対応する極端なものだと考えられる。しかし、統計処理を用い入力に対し MBR の記憶の中から最近接データを捜し出す MBR の基本戦略は多くの計算量を必要とし、大きなシステムになるとワークステーション向きの推論システムとはいえない。

この計算量を減少させるため、記憶を Kohonen ネットに写像し、カテゴリに分割整理して記憶しておくものが SBR であり [8, 9]、ニューラルネットによる CBR の実現と言える (図 1)。

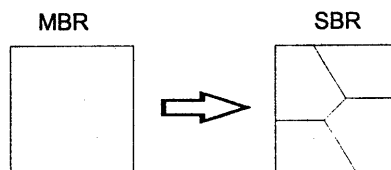


図 1: SBR

Kohonen ネットに写像する関数としてニューラルネット自己組織化アルゴリズムの1つである Kohonen の LVQ アルゴリズムを用いることができる。このとき、ビューを用いることによってよりユーザの主観に近くなるよう記憶空間をカテゴリズできる。

SBR は、入力データが与えられるとそのデータとカテゴリの代表を示す参照データとの最近接データを探し、そのカテゴリの特性を用いて推論を行う。さらに細かい推論を行いたい場合は、カテゴリ内において MBR 的に入力データとの最近接データを探し出すことによって行える。

推論 1 カテゴリレベルでの推論

例えば、SBR が中古車データベースにおける各カテゴリの参照データとして図 2 を記憶していたとする。

カテゴリ	全長	全幅	全高	HB	排気量	馬力	トルク	価格
小型車	3590	1590	1380	2405	1400	68	10.8	169.5
大衆車	4250	1670	1375	2430	1500	94	12.8	108.8
大型車	4690	1695	1405	2680	2500	180	24	244.1
高級車	4395	1695	1385	2545	2300	135	19.4	418

図 2: SBR の参照データ

このとき、図 3 がどのタイプに分類されるか推論するとしてよう。

全長	全幅	全高	HB	排気量	馬力	トルク	価格
3695	1585	1425	2360	1000	58	8.1	100

図 3: 入力データ

SBR によって、MBR のすべての記憶から最近接データ (または k-最近接) を探すまでもなく、参照データから最近接データを探すことによって、このデータは小型車に分類されることが推論できる。

推論 2 データレベルでの推論

次に、もっと細かな推論を行おう。例えば、図 3 のような中古車を購入したい顧客がいたとする。SBR は、まずこのデータがどのカテゴリに属するか推論し、そのカテゴリの中から入力データに対し最近接データを探し出し、そのデータに近いデータをいくつか顧客に提示する。

このように、推論内容がカテゴリで代表される場合 SBR は非常に高速に推論が行える。また、カテゴリの特性では表せないような推論においても、MBR 的な推論になるが、検索対象がカテゴリ内のデータだけで良く、また、MBR が統計処理を用いて行うデータの各項目に対する重み付けをビューとして与えることによって、MBR を簡略化できる。

本研究が対象とする問題は、ユーザが過去において選択していたデータをデータベースから探し出し、それを出力しようと言う問題であり、推論 2 に近いものである。具体的には SBR の記憶として、ユーザが過去に選択したデータをカテゴリ化し記憶しておく。SBR はデータベースからデータを読み込むと推論 2 のように最近接データを探し出し、この最近接データが過去において選択されていたものならば選択可能データとしマークしておく。そして、何らかのアルゴリズム (nearest neighbor アルゴリズムなど) によりマークされた選択可能データの中からデータを選択し出力する。

3.2 状態に依存する SBR

私たちはデータベースを検索するとき、すべての項目を等しい価値としてとらえ検索しているのではなく、主観によって項目に対し重み付けを行い検索をしている [13]。またその主観は常に一定ではなく、必要としているデータや、そのときの状況 (状態) によって変化していると考えるのが自然であろう。例えば車を購入するとき、欲しい車のタイプや予算などで主観は変わってくる。

したがって、データベースから複数のデータのある予算内で選択すると言うような問題を考えた場合、SBR に何らかの主観を変化させる手段が必要となる。これが状態に依存する SBR である。つまり、現在の状態から次の状態が規定され、SBR の主観プレーンが変化する (図 4)。

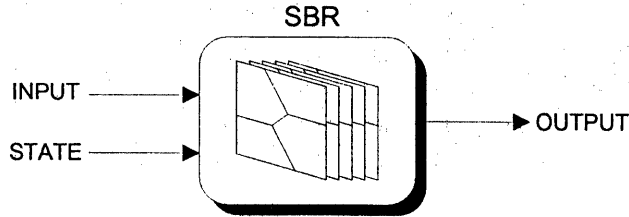


図 4: 状態に依存する SBR

4 データベースへの逐次アクセス

4.1 状態に依存する SBR と逐次決定処理

データベースからのデータ選択は、主観によるデータの優劣の判断と、その上での何らかの制約によるデータ選択の 2つのフェーズに分けて考えることができると述べた。したがって、選択可能なデータからどのデータを選択するかという逐次決定処理を行う必要がある。そこで、モデルに DP (Decision Process) を付加することによってデータベースの逐次アクセスを実現する (図 5)。

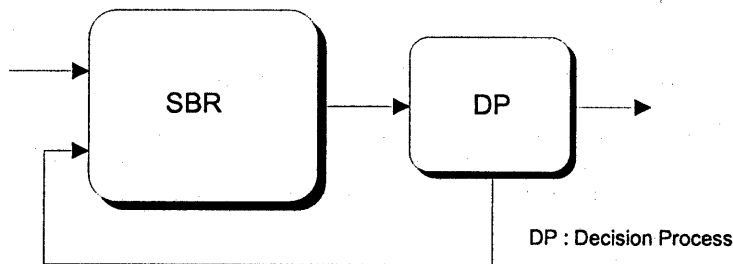


図 5: 状態に依存する SBR をベースとする逐次決定処理

4.1.1 SBR

SBR は、ユーザが過去に行ったデータ選択結果を記憶しておき、状態により主観プレーンを変えながら推論を行う。その基本的な考え方は、「読み込んだデータの最近接データがその状態下で過去において選択されていたならば、読み込んだデータも選択される可能性がある」とするということである。逆に言えば、最近接データがそのときの状態下で選択されていないならば今回も選択しないとする。したがって、図 5 の SBR は、データベースからそのときの状態下で選択可能なデータを出力するという一種のフィルターのよう働きを担う。

4.1.2 逐次決定処理 (DP)

DP は、SBR が出力した選択可能データを主観によって並べることによりデータの順序づけを行い、上位に並べられたデータから選択可能な範囲にあるデータを選択する。その選択されたデータをもとに DP 内部のプログラムにより、次の状態を決定する。

ケーススタディ:野球選手データベースへの適用例

具体的に DP を説明するためにプロ野球選手データベースでの例を示す。

ユーザが行うデータ選択として、一定の予算が与えられその予算の中で野球チームを作ること考える。ユーザはシステムへの入力として (1) データベースの並べ換え (2) データベースからのデータ選択の 2 つの試行を行う。

ユーザがシステムに入力したデータベースの並べ換えから主観とカテゴリを獲得し、データの選択結果と照らし合わせることによって、予算残高とそのときに選択された選手の関係から予算と主観の関係を求める。また、予算残高の推移チャートを作成することにより、ある予算残高での選択可能な年俸の範囲を得る。これを元にして、SBR の出力した選択可能データを主観によって並べ換え、上位に並んだデータから選択可能な予算内であれば、それを出力する。

選手を選択すると、それによる予算の減少から次に選択する状態を決め、SBR に出力する。

5 システム概略

検索代行システムの処理は大きく次の 2 つの処理に大別される (詳細は紙面の都合上省略)。

(1) 検索パターン獲得部

主観獲得支援ツールとデータ選択インターフェースから構成され、ユーザの主観とデータ選択結果を獲得する。

(2) データ選択処理部

検索パターン獲得部によって得られた知識から、モデルを構成しデータベースからデータを選択する処理を行う。

6 おわりに

本研究で用いた主観の獲得方法は、例外的なノイズデータに対する対策は特に行っていないが、マニュアルでは、主観の修正を行うことができる。

さらに拡張すべき機能としては、まず、DP に Join のような集合演算機能を取り込むことが考えられる [14]。このとき曖昧な処理にも対応できるようにすることが望まれ [15]、ニューラルネットベースのシステムの実現はリーズナブルな実現を提供するものと考えている [8, 9]。

また、現在システムを Sun SPARC station 上で C 言語を用い作成中であり、検索パターン獲得部まではすでに完成している。

参考文献

- [1] G. Jakobson et al. : " An Intelligent Database Assistant , " IEEE Expert , Summer , 1986.
- [2] 加納康男, 岸野文朗 : " 文献データベースにおける対話型問い合わせ理解とその評価 ", 信学技報, DE90-1, 1990.
- [3] X. Wu, T. Ichikawa : " KDA : A Knowledge-Based Database Assistant with a Query Guiding Facility," IEEE Trans. on Knowledge and Data Engineering, Vol.4, No.5, Oct., 1992.
- [4] X. Wu : " A Knowledge-Based Database Assistant with a Menu-Based Language User-Interface." IEICE Trans. Inf. & Syst., Vol. E76-D, No. 10, Oct., 1993.
- [5] 上野晴樹, 知識工学入門, オーム社, 1985.
- [6] C. Stanfill and D. Waltz : " Toward Memory-Based Reasoning, " CACM, 29, 12 (Dec. 1986)
- [7] R. Creedy et al. : " A Trading Mips and Memory for Knowledge Engineering , " CACM, August, Vol.38, No.8, 1992.
- [8] T. Ae, Y. Chikamatsu et al. : " Database Access using Self-organizing Neural Nets" , Tech. Report IPS.J, DBS-94-12, 1993.
- [9] T. Ae, S. Fujita, Y. Fujihara et al. : "A Multiprocessor for Neural Artificial Intelligence", ICASSE Proc., 1st Int. Conference on Applied Synergetic and Synergetic Engineering Erlangen. June, 1994.
- [10] T. Kohonen : " The Self-Organizing Map, " Proc. IEEE, Vol.78, No.9, 1990.
- [11] 奥田健三 : " 事例ベース形推論とその応用例 ", 情報処理, Vol.31, No.2, pp244-254, Feb. 1990.
- [12] E. Simoudis : " Using Case-Based Retrieval for Customer Technical Support," IEEE Expert, pp7-12, Oct. 1992.
- [13] 長島秀世 : " 人間の主観を重視した類似商標図形の検索の基礎的検討 ", 信学論 D-II, Vol.J74-D-II, No.3, pp.311-320, 1991.
- [14] P. Mishra, M. H. Eich : " Join Processing in Relation Database ," ACM Computing Survey, Vol. 24, No. 1, March 1992.
- [15] T. Ichikawa , M. Hirakawa : " ARES : Relational Database with the Capability of Performing Flexible Interpretation of Queries, " IEEE Trans. on Software Eng., Vol. SE-12, No. 5, May 1986.