

曖昧検索のための ファジィ・データベース・システム

野村 恭彦 小高 俊之 横山 光男 松下 温
慶応義塾大学 理工学部

データベースは検索効率の面においては、非常に多くの改善がなされてきた。しかし、問い合わせの手法という観点から見ると、使いやすいインタフェースを持っているといえるシステムはあまりない。そこで我々は、ユーザの問い合わせが本質的に持つ曖昧性の分析を行い、人間指向的問い合わせがどのように表現されるかということについての研究を進めている。

本論文では、曖昧問い合わせを支援するため、複数の属性から抽象化された複合ファジィ属性というものを提案し、さらにその生成方法について述べる。また、このような理論を基にして、C言語によってSun3ワークステーション上にファジィ・データベース・システムを構築した。そこで、さらにこのシステムの構成とユーザ・インタフェースについての報告を行う。

A Fuzzy Database System for Ambiguous Queries

Takahiko NOMURA, Toshiyuki ODAKA, Teruo YOKOYAMA,
and Yutaka MATSUSHITA

Faculty of Science and Technology, KEIO UNIVERSITY
3-14-1, Hiyoshi, Kohoku-ku, Yokohama, 223 JAPAN

From the view point of retrieving efficiency, database management systems have improved very much. However, with regard to access interface to database management systems it must be said that there are few systems with easy-to-use interface. Therefore, we analyze the ambiguity which queries issued by users essentially have and study in what kinds of styles human-oriented queries should be expressed.

This paper proposes Compound Fuzzy Attributes derived from multiple attributes, and describes the generating scheme of such compound attributes. Such a fuzzy database system is implemented on Sun3 workstation using C language. Finally, the configuration and the user interface of this system is reported.

1 はじめに

計算機の高機能・低価格化によるデータベースの普及に伴い、データベースに対する知識の無いユーザがデータ検索を行う機会というものが、これからも益々増加していくと考えられる。これにあわせて、よりユーザ・インタフェースの優れたデータベース検索ツールが必要となるであろう。そこで、これからのデータベース検索インタフェースとして重要となると考えられるのは、次の4点である。

1. 検索要求の入力が、マニュアルレスに容易な手順で行えること。
2. 入力・検索・ブラウジングをモードレスに行えること。
3. 人間の思考に合った検索属性によって、要求が行えること。
4. 曖昧な検索要求がそのまま入力できること。

今までの検索言語の代表格のSQL言語などは、綿密な条件入力ができるかわりに、ユーザにかなりの知識が要求された。そこで、このユーザの負担を減らすためには、言語を覚える必要の無いグラフィカルなユーザ・インタフェースが必要になってくる。しかしその一方で、グラフィカル・インタフェースでは入力条件が限られてくるため、ユーザから受け渡される情報量が大きく減少してしまうことになる。これを補助するために、データベースにある程度の知識というものを与えてやるが必要になってくるのである。一言で知識と言っても色々なものが考えられるが、当システムに於いては、人間の感覚をコンピュータが理解するための関数・パラメータ等がこれにあたる。

また検索には、ある特定のデータを探す場合と、要求にかなう不特定のデータを選ぶ場合がある。ここでは、曖昧性がより要求されると考えられる後者の場合について、モデリングを行っていく。具体的なシステムとしては、人間の購買要求を例にとり、データベースを検索していくことによって意志決定支援をするツールを考える。

ところで、データベースに曖昧な問い合わせを行うための手法としてファジィ理論を応用したものは数々提案されてきている [1][2]。そして、これらは一般に次の2つに大別することができる [3]。

1. データはクリस्पで、問い合わせのみが曖昧であるもの。
2. 問い合わせだけでなく、データにも曖昧性を持たせてあるもの。

当然1はインプリメントは楽だが、現実世界の曖昧性を表現するには弱く、一方2は曖昧性は高いが実現が難しい。こういった流れから、データベースに知識を加えてエキスパートシステム化させたもの [4] や、マルチメディアを扱うシステム [5] の研究も活発に行われている。

しかし、データベースのデータ構造などを変更し過ぎてしまうと、今までに貯めた既存のリレーショナル・データベース上のデータを利用することが難しくなってしまう。これではあまりに非能率的である。そこで考えなければならないのが、1と2の両者の利点を持ち合わせたデータの形である。つまり、既存のクリस्पなデータを用いて、抽象的なデータを導出することによって、データの持つ曖昧性というものを表現しようということである。

本論文では、まず検索における曖昧性というものを分析し、そのモデリングを行い、どのような検索条件が曖昧要求に必要なのかを分析する。そして、そのための手段として、ファジィ理論を用いた複合ファジィ属性というものを提案する。このデータモデルを用いることにより、抽象的な検索要求の入力が可能になる。そして最後に、これらの理論を基に作成したファジィ・データベース・システムのインタフェースを示すことにする。

2 曖昧問い合わせ

2.1 データ検索モデル

データ検索の流れというものを考えると、基本的には検索イメージを入力し、検索結果を見てそれによりイメージの修正を行っていくことになる。データ検索を潤滑に、というより人間的思考に合った形で行うためには、次の3点に注意する必要があるだろう。

1. 検索イメージ通りの条件設定が可能であること。
2. 曖昧なイメージに適応した検索が可能であること。
3. 誤った検索イメージを修正させるような結果表示を行えること。

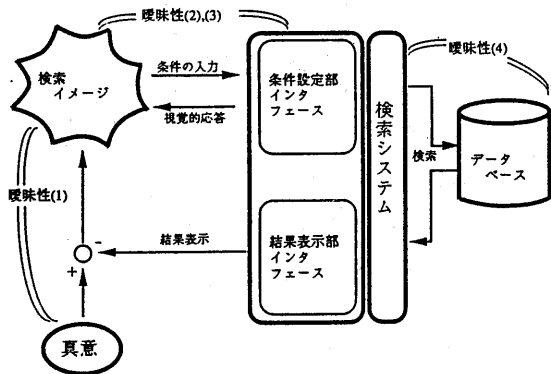


図1: データ検索のモデリングとシステム構成

これらを達成するためには、検索においてどのような曖昧性が存在するかということ进行分析が必要があるだろう。そこで、曖昧さの起こる原因について考察してみると、知識不足や不正確さに起因するもの、解釈が何通りもあるために生じるもの、主観的な基準によるため定義することができないために起こるもの等が挙げられる。

これらの曖昧性が、検索においてどのようなところに発生するのであろうか。これをはっきりさせるために、データ検索のモデリングを行い、それを図1に示す。そして、この図中で示した曖昧性の(1)~(4)について以下に述べる。

人間の検索イメージというものは、必ずしも真意と同一であるとは限らない。それは、人間が真の要求物に気付いていない場合もあるし、たとえそれに気付いていても、記憶の曖昧さによって検索イメージが狂うことは十分有り得ることである。

→曖昧性(1)

また、ある検索イメージがユーザに十分あったとしても、そのイメージが入力インタフェースにそのまま受け入れられることは少ないだろう。ユーザの所望の検索属性が無かった場合、そこでユーザが自分の検索イメージを他の検索属性に頭の中で翻訳する必要がある。そこで少なからず曖昧なものが生まれてしまうことになるであろう。

→曖昧性(2)

さらには、たとえ検索イメージ通りの入力が行えたとしても、本当にイメージ通りの検索が行われる保証はない。それは、ユーザの持つ常識とシステムの検索指標が一致しているとは限らないからである。こういった曖昧性は、SQLのような厳密な条件設定を行うシステムには余り存在しないが、ファジィ検索のような感覚的な問い合わせを行おうとすると、どうしてもユーザとシステム間の言葉の定義の違いが問題になってくるのである。

→曖昧性(3)

それともう一つ、ファジィ検索の最も得意とするところで、境界付近での曖昧性である。クリスピーな検索と違い、ファジィ検索では境界をはっきりさせずに検索を行う。そこで、各ユーザによって、というよりも各状況ごとにその限定の仕方に違いがあるはずである。そこで、過剰の限定をしすぎたり、限定の不足が起こったりといった曖昧性がついてまわるのである。

→曖昧性(4)

曖昧性(1)を埋めるには、結果表示・ブラウズ機能も強力でなければならないし、たとえユーザの真意に合ったデータが見つからなかったとしても、次の検索の指標を与えておくことの必要性は大きい。曖昧性の(2)と(3)に関しては、検索属性をユーザによって容易に作成・修正が可能である必要がある。そのためには、ユーザとシステムの会話を親密なものとするインタフェースを確立することが重要となる。曖昧性の(4)は、自然言語による限定に頼るという方法もあるが、グラフィカルなインタフェースで綿密な限定を行えるようにすることによって、より改善されると考えられる。

2.2 曖昧要求に対するインタフェース

2.2.1 属性の分類

曖昧要求を行うためのインタフェースを作成するにあたり、属性を大きく図2のように分類した。数値属性とは、数値上に全てのデータが配列できる属性のことである。この属性は、メンバシップ関数を用いてデータの検索を行うことが可能である。一般的に考えられているファジィ検索においては、検索に用いる形容詞・副詞に対応するメンバシップ関数をシステムが持っており、これを用いてユーザは問い合わせを行うわけだが、ユーザとの感覚の違いを考慮すると曖昧性の(3),(4)が

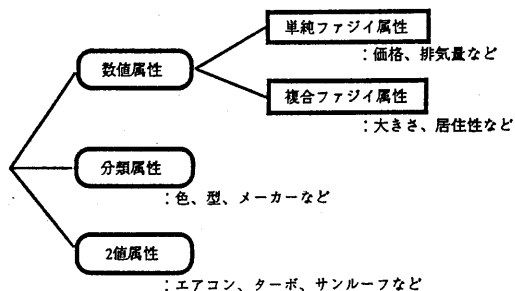


図 2: 入力インタフェースにおける属性の分類

問題になってくるので、ここではグラフィカルなインタフェースで容易に閾値を設定できるようにした。数値属性の中でも単純ファジィ属性と複合ファジィ属性に分類している。単純ファジィ属性とはそのままの数値で条件設定が可能なので、価格等の我々の数値感覚が発達した属性を扱う。一方複合ファジィ属性とは、そのままの数値では我々の感覚に合わないものを抽象的な概念に昇華させることによって、検索を行いやすくした属性である。この属性の設定や扱いについては次章で述べる。

分類属性とは、非数値属性のうちデータ全体を分類できる属性のことである。二値属性との大きな違いは、全てのデータが必ずどれかただ一つのカテゴリに属することである。それぞれの分類同士の類似度は類似行列によって与えられることになる。しかし、ただ単に類似行列を用いたのではユーザの主観の表現が弱いので、「どちらでもよい・よくわからない」といったユーザの場合にのみ類似行列が効果を発揮するようにし、それぞれの分類への綿密な条件設定をすることも可能としている。

二値属性とは *yes/no* の要求であるオプション的な属性のことである。この属性に関しては、特にアナログ的な入力インタフェースは用意していない。あまりに綿密な入力を用意してしまうと、逆に操作性を損なってしまう恐れがあるからである。

2.2.2 要求の強さ

曖昧な条件の属性の強さを次の3段階に分ける。これによって、自然言語で行う副詞的表現のニュアンスを容易に伝えることが出来る。

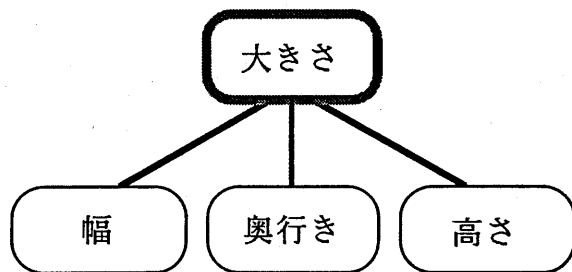


図 3: 抽象的な曖昧さ「大きさ」

1. 重要条件: 強く満たしてほしい条件
2. 通常条件: 満たしてほしい条件
3. 期待条件: できれば満たしてほしい条件

重要条件と設定した属性は、検索において通常条件と設定した属性よりも重視する。具体的には、適合度の計算時に重要条件の属性の適合度に重みをかけて、総合的な適合度を算出するのである。

期待条件とは、適合度の計算時にその値を含めた方が総合的な適合度が良くなる時のみ用いるということである。

3 検索用属性の生成

3.1 複合ファジィ属性の提案

複合ファジィ属性とは、いくつかのクリस्पな属性を複合することによって生成される、抽象的な概念の属性である [6]。例えば、幅・奥行き・高さが具体的な数値としてわかっているとき、「大きさ」というあいまいな属性をこれらから合成して作り出すことが可能である。複合ファジィ属性とそれを構成するクリस्प属性の関係は、一般に *is-a* 関係であると同時に全体としては *part-of* 関係となっていると言える。幅や高さなどのデータは事実の知識としてデータベースに蓄えられているわけだが、実際に「幅何 cm」といった条件は与えにくい。我々の頭の中では、「大きめの」といったような曖昧な、というより抽象的な概念としてあることが多いだろう。こういった条件設定を可能にするため、全体のデータから、この検索物ほどのくらい大ききなのかといったことをシステムが判断する

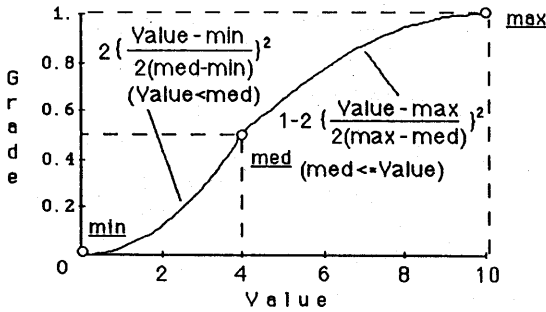


図4: 値0~10の[0,1]変換例

ことが求められるのである。その演算のために、従来の and/or 演算を拡張した形の平均演算子を用いる。

3.2 複合属性のための評価関数

3.2.1 数値データの[0,1]変換

例えば「パワー」という属性を最高出力・最大トルクから合成しようとしたとき、出力は何百馬力で表現され、トルクは何十kgである。これら二つをそのまま合成しようと思っても、値のレンジが全く違うため、そのまま式に代入したのでは片寄った演算になってしまう。そこで、それぞれの車種の最高出力と最大トルクを、0.0~1.0の値に置き換える作業が必要になってくる。

我々が目指すところは、複合ファジィ属性をできるだけ自動的に、ユーザの手を煩わせずに生成することである。そこで、その属性の最小値を0、最大値を1、中央値を0.5にとるように図4のようなS関数をつくる。こうして[0,1]変換を行うと、外れ値があった場合でもその影響を最小限に食い止めることができる。

3.2.2 平均演算子の適用

[0,1]変換を行った後、それらの数値属性をn次元座標の各軸にとると、全データがこのn次元空間のどこかにプロットされることになる。これら各々を $x_1 = x_2 = \dots = x_n$ 直線の上に評価関数を用いて落とし、その値をファジィ属性値とする。この評価関数は重み付きの平均演算子[7]であり、ここでは算術平均・幾何平均・調和平均とその双対なものを用いている。平均演算子を用いると、既存の min/max による and/or 演算よりも、かなり我々の感覚にあった結果が得られるよ

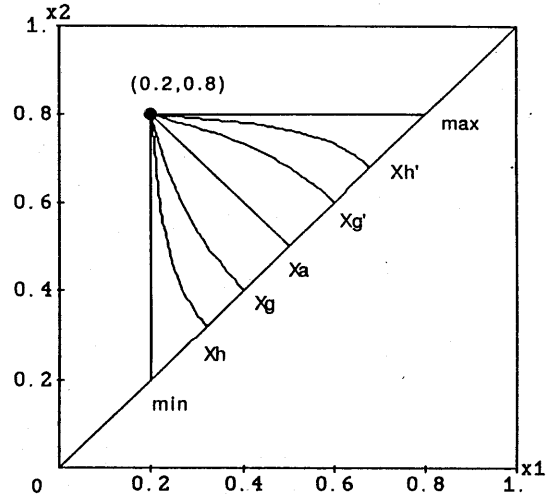


図5: 2次元平面上における平均演算子の図形的解釈

うになることが分かっている。例えば、既存の and 演算の例を示すと、 $0.1 \cdot 0.9 \cdot 0.9 = 0.1 \cdot 0.2 \cdot 0.1 = 0.1$ (・は and を示す) となり、最小値以外の値の影響が全く加わっておらず、明らかに我々の感覚にあわないことがわかる。平均演算子を用いることによって、このような問題は容易に解決する。また、属性間の重み付けも容易であるため、各属性の抽象属性に対する影響力の違いも表現できる。図5に、算術平均・幾何平均・調和平均とその双対な演算子による帰着例を示した。平均演算子によって $x_1 = x_2$ 直線上に落とされた点に対して、(0,0)を0、(1,1)を1として新たな属性値を与える。

3.2.3 評価式の曖昧さの表現

$$Size = \sqrt[3]{Length \times Width \times Height}, 0.3$$

これは、大きさという曖昧な属性が、全長・全幅・全高の幾何平均で求められ、その曖昧さが0.3程度である、ということを表している式である。この曖昧さを b とおくと、 $0 \leq b \leq 1$ の範囲で b は定義され、 $b=0$ の時は曖昧性なしを意味し、 b が1に近づけば近づくほど曖昧さが増すことになる。

この曖昧さ b は評価式の曖昧性を意味しているので、これがこの評価式から求められた結果の可能性分布の

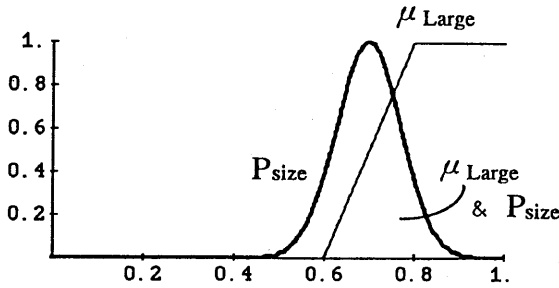


図 6: 車 i の「大きい」に対する適合度の算出

広がりとなって現れてくる。つまり、ある車 Car_i の大きさを評価式から求めた値を $Size_0$ とすると、この車の大きさ $Size$ の可能性分布は次のような式で表される (図 6)。

$$P_{Size}(Car_i) = \exp\left(-\frac{(Size - Size_0)^2}{b}\right)$$

この可能性分布の検索における役割としては、当然そのデータの適合度の演算に用いられることになる。例えば $Size = Large$ という要求があった場合、 $Large$ のメンバシップ関数を μ_{Large} とすると、この適合度 $Grade_i(Size = Large)$ は、 $P_{Size}(Car_i)$ と μ_{Large} の and の面積を計算し、これが可能性分布の何%を占めているかで適合度を計算する。こうすると、様々な特殊なメンバシップ関数 (クリस्पなものを含む) を想定した場合にも、非常に我々の感覚にあった適合度が得られる。

$$Grade_i(Size = Large) = \frac{\int \mu_{Large} \& P_{Size}(Car_i)}{\int P_{Size}(Car_i)}$$

3.3 複合ファジィ属性の例示決定法

まず、ある複合ファジィ属性がどの数値属性から成るのかを指定する必要がある。指定方法は、ある複合ファジィ属性に使用する数値属性には 1、もしくは重みを付ける場合は 2 以上の値を与える。ところが、数値属性と複合ファジィ属性が常に増加関数的関係とは限らないので、数値属性値が大きくなればなるほど複合属性値が減少する場合も考慮してやる必要がある。例えば、経済性といった複合属性を考えてみると、値

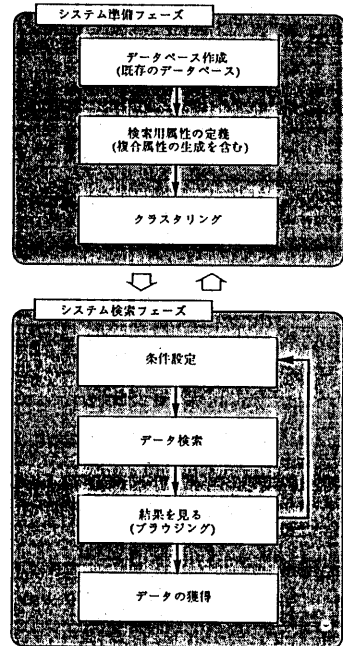


図 7: システムの処理の流れ

段が高くなればなるほど、当然経済性は下がることになる。こういった数値属性には -1 もしくは重みを付けて -2 以下の値も付けられるようになっている。

前述したように、評価関数として用いる平均演算子には算術平均・and/or幾何平均・and/or調和平均の 5 種類を用意しているのだが、そのうちのどの関数を用いるかを決定する指針を与える必要がある。具体的には、システム側が 10 種のデータを与え、人間の方でそのデータの各複合ファジィ属性に対する評価値を主観により 0.0 ~ 1.0 の値で与える。この例示に従って決定した評価関数を用いて、他の全てのデータについての計算も行うことになる。例示を行うデータ数はここでは 10 種としたが、サンプルデータの特性によってこの数は変化させる必要があるだろう。

次に、評価関数の決定方法について述べる。まずシステムが各複合ファジィ属性についての例示を得ると、それらのデータについての評価値をその数値属性を用いて各平均演算子を用いて計算してみる。そして得られた値と例示された値との自乗誤差の和が最も小さい演算子を評価関数に採用する。そして、その自乗誤差の平均を曖昧さとして、その評価式で得られた評価

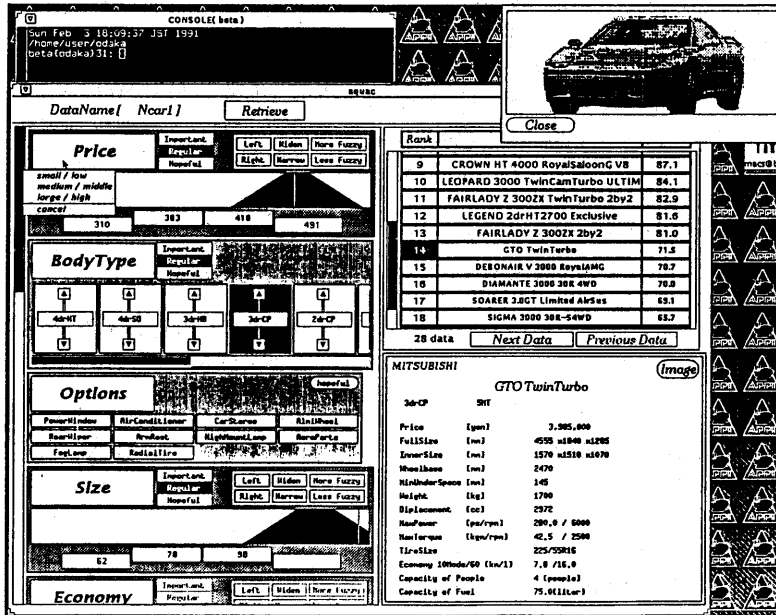


図 8: 画面出力例

値の可能性分布の広がりとした。こうすることにより、人間の与える例示と選ばれた評価式とのズレを可能性分布の広がりでもカバーする事ができると考えられる。曖昧さ b は、 n 個の例示データ ex と平均演算子から得られた評価値 $mean$ とで、次式のように表される。

$$b = \frac{\sum_{i=1}^n (ex_i - mean_i)^2}{n}$$

4 システムへの適用

4.1 システム構成

システムの行う処理の流れを図7に示すが、大きく分けて準備フェーズと検索フェーズから成る。準備フェーズは主にシステム管理者が使用することになるわけだが、検索属性やその複合ファジィ属性はユーザによってまちまちの感性を持っている。そこで、システム管理者が各層のユーザに対しての検索属性を用意しておくか、もしくは各ユーザが独自に作成する必要がある。ここでは、ユーザが容易に検索属性を定義できることを目標にシステム開発を行ってきたが、これに対するインタフェースは未だに管理者レベルの知識を必要とするものの域を出ていない。またファジィ検索には、検索空間を広げることによる検索時間の増加というもの避けられない問題であるが、それを回避するために

ここではクラスタリングというものをを用いている。クラスタリングとは、 n 次元の属性から成る検索空間を考えたときに、そのユークリッド距離の近いものどうしは似たデータであるという見地から、全データを複数個のクラスタにグループ分けすることである。全データの階層的なクラスタリングを予め行っておき、それを用いて検索することにより、検索空間を狭めて検索時間を非常に小さくすることが可能になる。なお当システムはC言語によって記述されており、X-Windowを用いてSun-WS上で動作する。

4.2 インタフェース

条件設定部インタフェースは、図2で表した様に大きく3つに分類される。システムの画面表示の様子は図8に示しており、左半分が条件設定部で右半分が結果表示部である。ここでは数値属性の単純ファジィ属性である価格、複合ファジィ属性であるスポーティ度、そして分類属性であるボディタイプ等の条件設定が行われている。数値属性はメンバシップ関数を簡単なアイコン操作によって直接入力するインタフェースを持っており、曖昧性の幅なども綿密な設定が可能である。任意の条件設定をアイコン操作で行い検索ボタンをクリックすると、直ちに検索結果が右の結果表示部に出

力される。そして、その上部の総合適合度順に並んだ表のデータ名をクリックすれば、その詳細データが下部に出力され、それと同時にそのデータがどういった条件にあたるかが左の条件設定部に指示される。具体的には、数値属性にはメンバーシップ関数上に縦線が引かれ、非数値属性においては、そのデータに当てはまる属性が反転することによって示される。こうすることによって、システムの検索に於ける判断基準というものの認識がユーザに促され、次の検索を行うときの強力な指針となり得るだろう。

5 まとめ

データベースの検索環境というものをシステム側からではなく人間指向的な観点から見た分析を行い、それに適応したインタフェースというものを考察した。そのために検索に於ける曖昧性というものを分析し、インタフェースに合わせての検索属性の分類を行った。そして、人間の曖昧指向に合った検索を行うための複合ファジィ属性というものの生成法を提案した。これにより、既存のDBMS上の曖昧性のないデータを利用して、かなり曖昧な問い合わせが可能となった。

しかし一方では、当方法は属性の大部分が数値属性からなるデータの検索には威力を発揮するが、文書検索やマルチメディア検索などの数値で表しにくいものには向かないということがいえる。現在のインタフェースには改善の余地が少なくはないが、今後はそういった改良を行うだけでなく、全く性質の異なるメディアの検索への応用を模索していく必要があると考えている。

参考文献

- [1] H.Prade and C.Testemale, "Generalizing Database Relational Algebra for the Treatment of Incomplete or Uncertain Information and Vague Queries", Information Sciences, Vol.34, pp115-143, 1984.
- [2] M.Zemankova and A.Kandel, "Implementing Imprecision in Information Systems", Information Science, Vol.37, pp107-141, 1985.
- [3] 馬野, "ファジィ・データベースの現状と動向", Advanced Database System Symposium '90, pp207-214, Dec.1989.
- [4] K.S.Leung, M.H.Wong and W.Lam, "A fuzzy expert database system", Data & Knowledge Engineering 4, pp287-304, 1989.
- [5] 佐藤, 田坂, 山本, 浪岡, 茂木, "ファジィ情報検索システム INDAS/f の試作", 情報処理学会研究報告, 90-DBS-78, 5, 1990.
- [6] 小高, 野村, 横山, 松下, "抽象的なあいまいさの表現", 第4回人工知能学会全国大会, pp219-222, 1990.
- [7] 野村, 小高, 横山, 松下, "曖昧検索のためのファジィアトリビュートの生成", 第41回情報処理学会全国大会, pp135-136, Sep.1990.