

個人の主観の相違を考慮したファジイデータベースの提案

大木 直人 田代 秀夫 亀倉 龍 横山 光男 松下 温
慶應義塾大学理工学部

近年、コンピュータであいまいな情報を処理する必要性が生じている。そこで、あいまいな情報をメンバシップ関数によって処理するファジイデータベースが提案されている。しかし、あいまいさは主観的なものであり、個人によって異なっているのが普通である。このため、従来のファジイデータベースのあいまい処理では、全てのユーザのあいまいさに対応してはいたわけではない。そこで、ユーザ個人のあいまいさを表すメンバシップ関数を自動的にチューニングするファジイデータベースシステムを提案する。このシステムでは、「若い」などの数値的尺度を持ったあいまいさだけでなく、「かっこよい」などの数値的な尺度を持たないあいまいさをも処理することを可能とした。

PROPOSAL OF A DATABASE SYSTEM CONSIDERING EACH PERSON'S SUBJECTIVITY

Naoto Oki Hideo Tashiro Ryu Kamekura Teruo Yokoyama Yutaka Matsushita
Keio University
3-14-1, Hiyoshi, Kohoku-Ku, Yokohama 223, Japan

Recently, computers that can handle imprecision are badly needed. As an approach, fuzzy database systems which allow imprecise data have been presented. In these systems, imprecise data are represented by means of membership functions. However, imprecise information are usually subjective, and differ among individuals. Because this subjectivity has not been considered, the membership functions installed in previous systems were not well-suited for each user. Here, we present a fuzzy database system which generate membership functions automatically for each user. This system enables users to store subjective data, and also manages each user's subjectivity. Moreover, our system allows the use of non-numerical attributes, consequently enlarging the area of application.

1 問題の提起

最近、さまざまな分野で、人間の最も重要な特徴である、あいまいさに関する研究が盛んになっている。特にデータベースの分野では、「年齢が若い」「値段が高い」などのファジイ表現と呼ばれる、主観的あいまい情報を取り扱うファジイデータベースシステム (FDBS) についての研究が盛んになっている [馬野 89] [本多 91] [大木 91] [野村 92]。しかし、FDBS については、あいまいさを扱うが故の問題点が従来より数多く指摘されている。その中でも今回、我々が着目し改善を試みたのは次のような点である。

まず第1に、従来のFDBSでは、個々のユーザによる主観的なあいまいさの違いを考慮していなかったという点である。人間は、十人十色と言われるように、ものの考え方や感じ方が個人によって異なるのが当然である。例えば20才のユーザと60才のユーザとでは、age という属性の *young* というファジイ表現を表すメンバシップ関数が異なるであろうことは容易に推測できる (図1) しかし、今までFDBSで扱ってきたあいまいさは、システムの構築者あるいは管理者が設定したものがほとんどであり、必ずしも個々のユーザの主観的なあいまいさとは一致していなかった。

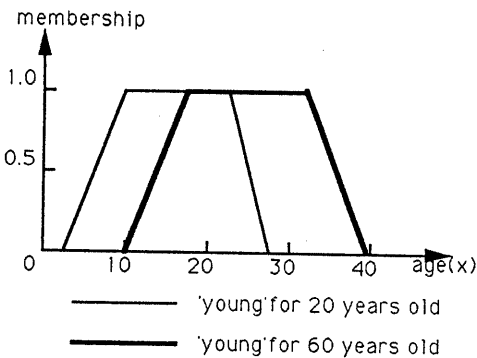


図1: ユーザ間のメンバシップ関数の相違

第2に、一般にFDBSにおいてはあいまいな情報を処理するために、Zadeh教授がファジイ集合論 [Zadeh 65] の中で提案したメンバシップ関数を用いているが、このメンバシップ関数の設定に手間がかかるということである。1つのファジイ属性には、通常3~5のファジイ表現が含まれている。例えばageというファジイ属性には *young, middle-aged, old* の3つのファジイ表現が含まれると考えられ、また、*very*

young, young, middle-aged, old, very old と5つのファジイ表現を含むこともできる。これらファジイ表現一つ一つにメンバシップ関数を設定していくのであるから、ファジイ属性が増えるにつれ、設定しなければならないメンバシップ関数の数は爆発的に増えていくのが予想できる。さらに、システム管理者以外の一般のユーザが、自分の主観的なあいまいな情報を、数多くのメンバシップ関数の形で表現するのは一段と難しい。従来のシステムでは、一般のユーザにメンバシップ関数を設定させるために、ユーザに対してアンケートを行うなどの統計的手法を用いてメンバシップ関数をチューニングしていた [Turksen 91] が、この方法も、統計的な手法を用いているため、時間をかなり費やさなければ適切なメンバシップ関数を設定することはできない。

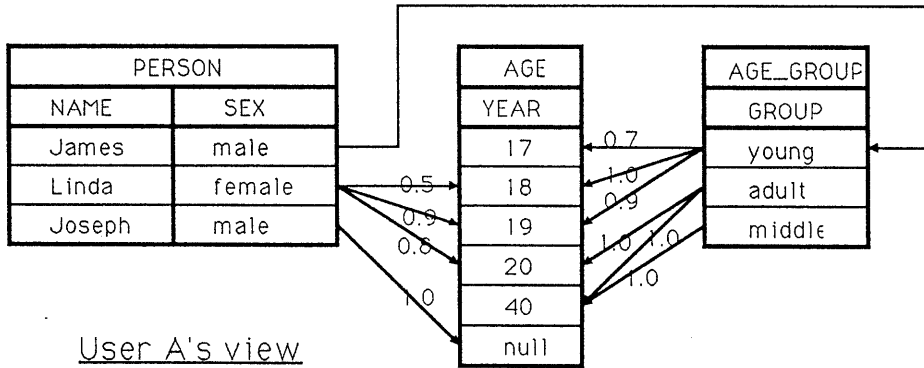
第3は、数値的な尺度に対応できない情報を扱うことが困難であるということである。従来のFDBSで扱ってきたあいまいな情報は、「年齢が若い」「値段が高い」などの数値的な尺度を持つものに限られていた。しかし、「車がかっこいい」「料理がおいしい」など、従来のFDBSでは扱えなかった、数値的な尺度を持たないあいまいな情報を処理する要求が生じてきている。

本論文では、以上の問題点を解決する手法を示し、主観的あいまい情報をより柔軟に扱えるFDBSを提案する。

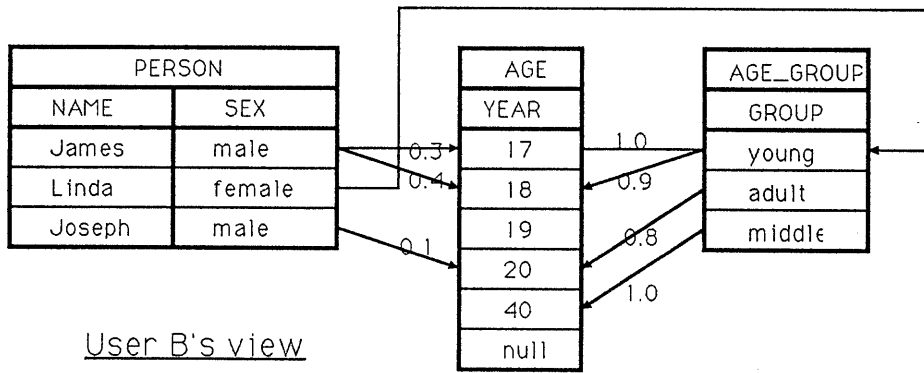
2 問題の解決

第1の問題点については、個々のユーザそれぞれが自分の主観に基づくメンバシップ関数を持つことによって解決する手法である、主観的ファジイ実現値ビューが提案されている [藤代 91]。この手法では、ユーザが自分自身の主観に基づいたあいまい属性のメンバシップ関数を、データの見方 (local view) としてFDBSに登録しておく。このとき、あいまい属性とメンバシップ値をリンクで表現し、管理するのがこのシステムの特徴であり、これによって、効率的なデータ管理を実現している。図2はユーザAとユーザBのビューの違いによって、同じデータでも違うあいまい属性値を持ち得る事がわかる。

しかしこの場合、ユーザー一人一人について、メンバシップ関数に相当するリンクをはる作業が必要となる。ユーザおよびデータの量が増えるにつれて、FDBS全体で使用されるメンバシップ関数の数が増加し、[藤代 91] でも



User A's view



User B's view

図 2: Local view [藤代 91] より引用

指摘されているが、第2の問題点がますます重要になってくると思われる。このため、第2の問題を解決するためには、ユーザ毎のメンバシップ関数をユーザの手を煩わせることなく自動的に、かつ簡便にチューニングする手法が要求される。特に我々は以下の点に留意し、手法を開発した。

- ユーザはコンピュータの初心者を含む一般のユーザを想定しているため、メンバシップ関数チューニングのための特別な操作を必要としない。
- FDBSに組み込んで運用したときに、メンバシップ関数チューニングのための時間がかかりすぎ、システムの応答性が悪くなるのを避ける。

ユーザに対して特別な操作を要求しないのであるから、ユーザの主観を得ることができる情報源は、ユーザがFDBSに登録した主観的データが考えられる。そこで、この主観的データを基にして、そのユーザが「どんな傾

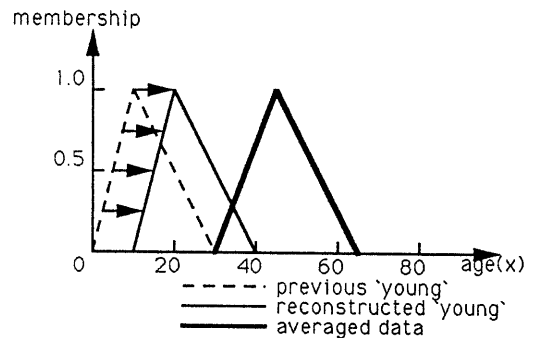


図 3: メンバシップ関数のチューニング

向を持っているか」を判断し、ファジイ表現のメンバシップ関数をチューニングすることを考えた。

例えば、個人情報のFDBSの場合、ユーザAがある人Bについての属性 *age* に対して *young* と登録したとする。このとき、A以外の大多数のユーザがBについて *middle-aged* であると登録していたならば、「Aの *young*

という表現は、大多数の平均的な意見よりも *middle-aged* の方向にずれている」という傾向のあることが判断できる。この判断された傾向をもとにして、A の *young* というファジイ表現に対応するメンバシップ関数をチューニングする。

より具体的なアルゴリズムについては後に述べる。

このように、ユーザが FDBS に登録したデータを、他のユーザの平均的なデータと比較することによって、そのユーザのメンバシップ関数をチューニングする。データとデータの比較という方法を用いているため、この方法は数値的な尺度を持たない情報にも対応できる。例えば、飲食店のデータベースで、大多数のユーザが *normal* であると登録している店について A 氏が *delicious* と登録していたならば、「A 氏の *delicious* という表現は、大多数の平均的な意見よりも *normal* の方向にずれている」という傾向のあることが、数値的尺度を持った情報と同様に判断できる。これが、第3の問題を解決する1手法である。

3 アプリケーション

3.1 システムの構成とインタフェース

以上に述べたような、ユーザによる主観の相違が重要な意味を持つ環境を実現する FDBS として、どのようなシステムが可能か考えてみた。

ユーザの主観的な情報を登録し、活用するのは、基本的にはそのユーザ専用のデータベースであると考えられる。現在の電子手帳がまさにこれにあたると思われる。このような環境ならば、ユーザ A は、「ふと入った飲食店 α 屋がおいしかったので、また今度来よう」と思ったら、データベースにその店の住所と場所と電話番号、それに、*delicious* とデータを登録するだろう。このデータは A の全くの主観的な情報であるが、この情報は A の主観の中で閉じているので、A の好みが変わらない限りは主観の相違という問題は生じない。しかし、この電子手帳がネットワークによって多くの他の人の電子手帳と情報のやりとりができて、A 以外の人がこの A が登録した α 屋のデータを見るときには、主観の相違が発生する。つまり、A 以外の人が実際に α 屋に行ったとしても *delicious* と思うかどうかはこれだけでは分からない。ここで、もし、「A はよほどうまい店でない *delicious* とは言わない」ということが分かっていると、それに基づいた

A の *delicious* というメンバシップ関数でデータが登録されていたならば、このデータは他の人からは「 α 屋は *very delicious* である」と見ることができる。

このようなアプリケーションを考えたときに、これを実現するシステムとして、図4のようなシステムを提案する。

このシステムは以下のような構成になっている。

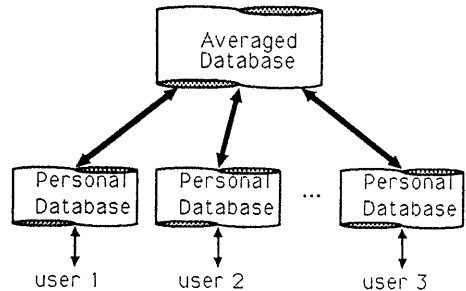


図4: システムの構成

- 個人データベース

ユーザのワークステーション上で実行され、ユーザが自分の主観で自由に情報を登録・検索できる FDBS である。この FDBS には数値データや「高い」「おいしい」などのファジイ表現データが登録できる。ファジイ表現データは、対応するメンバシップ関数が次章で述べるアルゴリズムに従ってチューニングされ、保持されている。

- 平均データベース

各個人データベースに登録されている、全レコードの全属性についてのユーザ全体の平均的な意見を蓄積しておくデータベースである。各ユーザが個人データベースにデータを登録したと同時に、そのデータは平均データベースにも送られてくる。ファジイ表現データについては、各ユーザのメンバシップ関数の形でデータが送られてくるが、複数のユーザが重複して登録している属性値があれば、それらのメンバシップ関数の平均をとって更新される。

メンバシップ関数の平均を計算する方法については、次章で述べる。

- データベース間を結ぶネットワーク

各ユーザの個人データベースは、ソケット通信 [SUN 90] を用いて平均データベースと情報の相互交換を行っている。また、個人データベース同士は平均データベースを介して相互に情報のやり取りを行うことができる。

3.2 データの登録

データの登録は、以下の手順で行われる。

1. ユーザが自分の個人データベースに情報を入力する。入力できる情報は、数値情報・文字情報・画像情報などのあいまいではない情報と、本論文の主題であるファジイ表現を用いたあいまいな情報である。

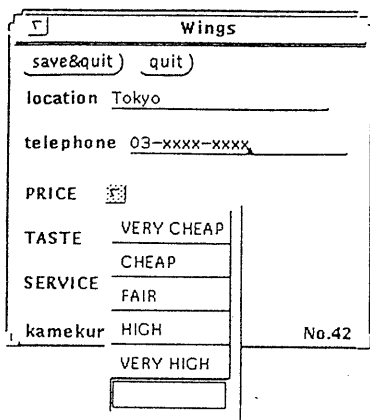


図 5: 登録インターフェース

2. 個人データベースから平均データベースに *young, delicious* に対応するこのユーザのメンバシップ関数が送られる。
3. 平均データベースでは、送られてきたメンバシップ関数を加えて平均データを更新し、その平均データを個人データベースに送り返す。
4. 個人データベースで、送られてきた平均データと、入力されたデータを基にしてメンバシップ関数を更新する。

3.3 データの検索

データの検索は、以下の手順で行われる。

1. ユーザが検索条件を入力する。あいまいな検索条件は、「*delicious* かつ *cheap* な店」のようにファジイ表現によって入力される。

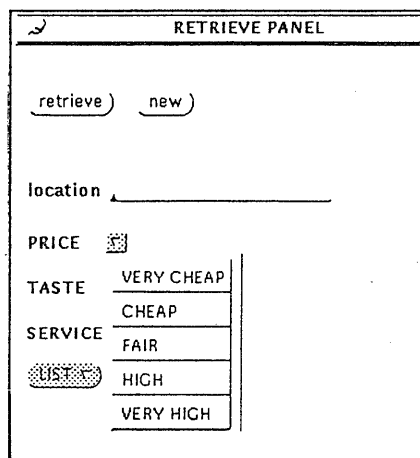


図 6: 検索インターフェース

2. まず、このユーザの個人データベースの中から、検索条件に適合するデータを検索する。
3. 次に、平均データベースに登録されているデータのなかで、個人データベースで検索されたデータ以外のデータについて同じ検索条件で検索を行う。平均データベースでは、データがメンバシップ関数の平均の形で保持されているので、検索条件に対する満足度は、一般的な手法として図 7 のように、2つのメンバシップ関数の交わりの最大値とした。

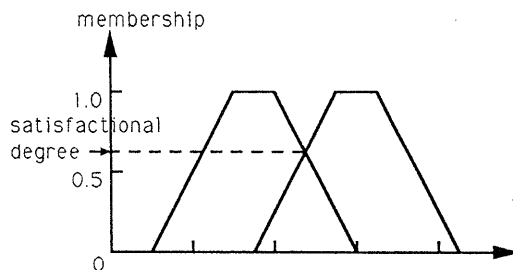


図 7: 満足度の計算

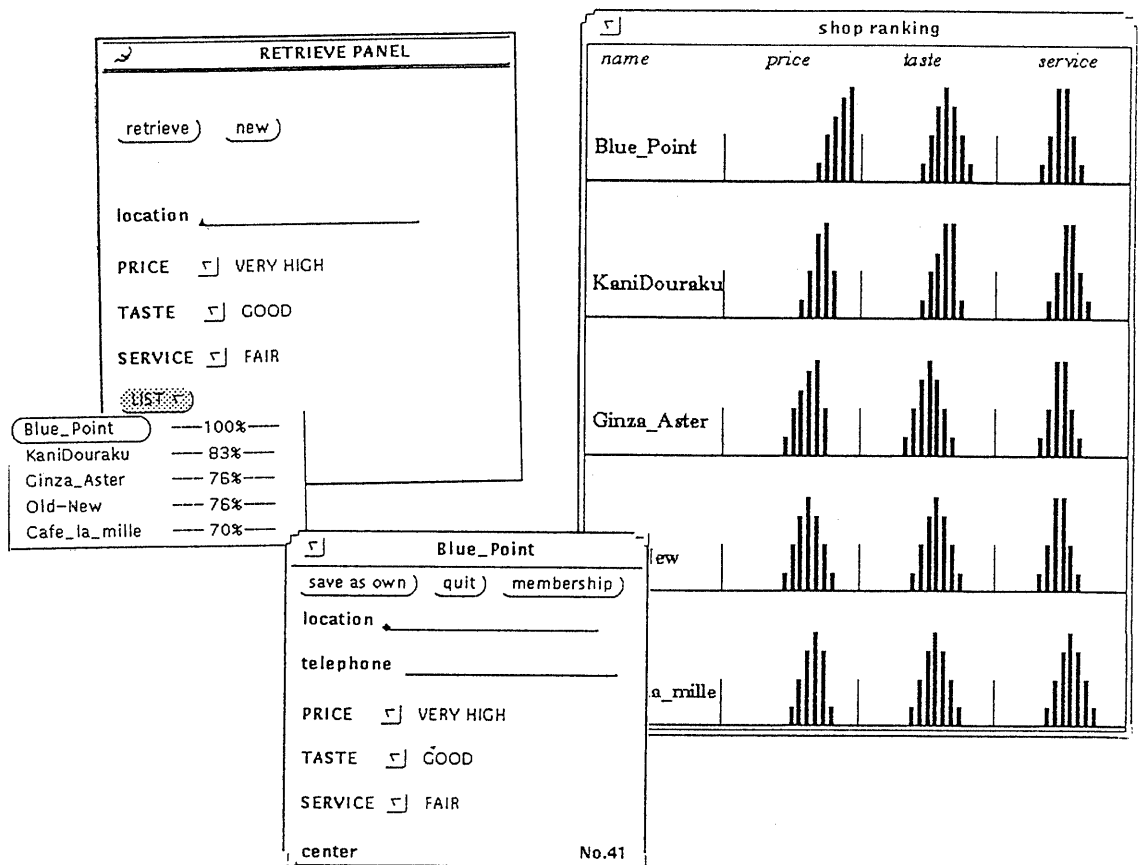


図 8: 検索結果インタフェース

4. 個人データベースと平均データベースの検索結果のなかで、満足度の高いデータが検索結果として満足度順に表示される。
5. 検索結果のデータの詳細な情報を得ることもできる。

4 メンバシップ関数チューニングのアルゴリズム

この章では、平均データベースでのメンバシップ関数の平均演算、および、個人データベースでのメンバシップ関数のチューニングについて述べる。

4.1 平均データベースでの演算

平均データベースでの平均演算は、複数のユーザの平均値を求めるのが目的なので、複数のユーザのメンバシップ関数をファジイ加

算 [Zemankova 87] することにより平均する。2つのメンバシップ関数 $\mu_1(i)$ 、 $\mu_2(j)$ のファジイ加算 $\mu(k) = \mu_1(i_1) \oplus \mu_2(i_2)$ は、以下のように定義される。

$$\mu(k) = \sup\{\min(\mu_1(i_1), \mu_2(i_2))$$

$$; i_1, i_2 \text{ where } i_1 + i_2 = k\}$$

従って、n人のユーザのメンバシップ関数の平均 $M(k)$ は、以下のように定義される。

$$M(k) = \frac{1}{n} \{\mu_1(i_1) \oplus \mu_2(i_2) \oplus \dots \oplus \mu_n(i_n)\}$$

4.2 個人データベースでの演算

個人データベースでは、ファジイ表現のためのメンバシップ関数の更新はデータを登録したときに、登録したデータのメンバシップ関数と、平均データベースに既に登録されているメンバシップ関数との演算によって行われる。

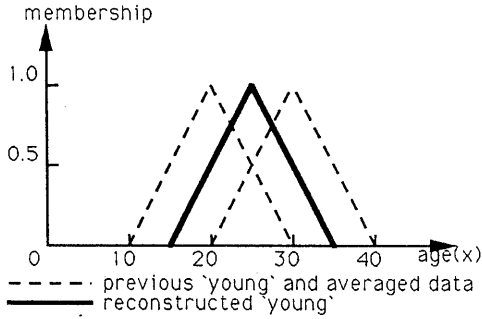


図 9: ファジィ加算法

メンバーシップ関数の更新のための演算方法として、ファジィ加算平均法の他に、和演算法・統計的平均法の手法を提案する。

• 和演算法

メンバーシップ関数と可能性分布の和集合を新たなメンバーシップ関数とするのが、この方法である。例えば、あるユーザが、平均データベースで *middle-aged* ぐらい (平均化された意見であるため *middle-aged* そのものではない) と登録されている A の年齢について、*young* と登録したとする。このユーザの *young* という概念は、平均的な *middle-aged* という概念をも含むと考えられる。和演算法はこのような考えに基づいており、以下のように定義される。

$$\mu_r(i) = \max\{M(i), \mu(i)\}$$

ここで、 $\mu_r(i)$, $M(i)$, $\mu(i)$ はそれぞれ更新後の個人データベースのメンバーシップ関数、平均データベースに登録されているユーザの平均的なデータ、更新前の個人データベースのメンバーシップ関数である。

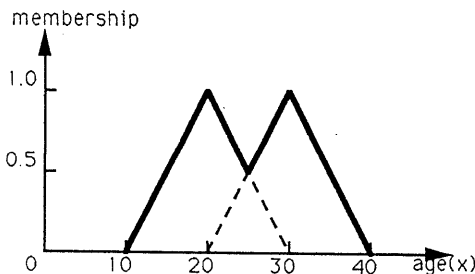


図 10: 和演算法

• 統計的平均法

ファジィ加算平均法をメンバーシップ関数の横軸方向に加算した平均と考えると、この統計的平均法はグレードを表す縦軸方向に加算して平均する方式である。

$$\mu_r(i) = \frac{1}{2}\{M(i) + \mu(i)\}$$

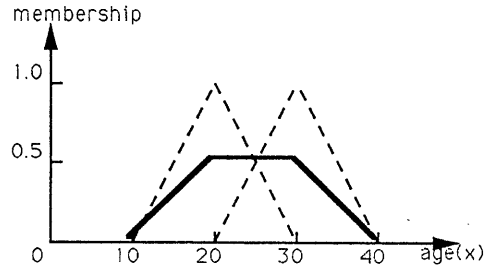


図 11: 統計的平均法

4.3 メンバシップ関数の評価

前節で提案したメンバーシップ関数のチューニングアルゴリズムについて比較検討を行った。比較検討にあたっては、数値的な尺度を持つあいまい属性と、このシステムの特徴である数値的な尺度を持たない属性について行った。

検討にあたって用いたデータは、中学生、高校生、大学生、大学院生約 50 人にアンケートを行い、集めたものである。アンケートの内容は、スポーツに関するものであり、テニス、ゴルフ、野球など 10 種類の競技に対して、楽しさ、難易度、人気など 5 種類の項目について 6 段階評価をしてもらった。

このデータを用いて、それぞれのアルゴリズムでそれぞれのユーザのメンバーシップ関数をチューニングする。さらに、ある任意の検索条件で検索を繰り返し、検索条件に該当するデータを検索結果とする割合を計算した。

結果としては、統計的平均法が数値あいまい属性、非数値あいまい属性ともに比較的高い評価が得られている。また、数値的な尺度を持つあいまい属性に比べると、非数値あいまい属性については、どのアルゴリズムもあまり高い評価は得られていない。これは、非数値あいまい属性ではユーザの嗜好が深く関与するためであると考えられる。例えば、ユーザ A は α 屋を *delicious* とし、ユーザ B は α 屋を *bad* とするような、ユーザの意見が 2 分されるような場合が生じる可能性がある。こ

表 1: Evaluation

	数値あいまい属性	非数値あいまい属性
和演算法	31%	19%
統計的平均法	66%	33%
ファジイ加算平均法	50%	29%

のような場合、平均データベースにおける平均演算を和演算で行うことによって処理することは可能だが、平均データベースの α 屋のデータは、「*delicious*であり、かつ*bad*である」という一見矛盾したものとなる。さらに極端な場合、 α 屋のデータが「*very delicious*かつ*delicious*かつ*normal*かつ*bad*である」となりかねない。このデータは、どんな検索要求にも適合して検索結果となり得るわけで、このようなデータが増えていくと、検索の意味がなくなってしまう。

各アルゴリズムはそれぞれに一長一短があり、それぞれの良いところを生かした折衷案的なアルゴリズムも試してみたが、アルゴリズムが複雑になってしまい、また、期待したほどの改善は見られず、これは今後の研究課題となるであろう。

5 結論

人間の知識表現を目標としている人工知能の分野において、人間の最も重要な特徴である曖昧性を処理する必要があるが、今回提案したシステムによって、個人によって異なる主観に基づいた曖昧な情報の登録・検索に必要な不可欠なメンバシップ関数を自動的にチューニングする事が可能となった。また、従来不可能であった数値的尺度を持たない曖昧な情報もある程度扱えるようになった。現在、飲食店をデータ例とした試作システムを SUN Sparc Station 上に構築し、これを用いてメンバシップ関数チューニングアルゴリズム、およびシステム全体の評価を行っている。

今後の大きな課題としては、メンバシップ関数チューニングアルゴリズムの改良がある。ファジイ理論とニューロネットワークの組み合わせは、最近盛んに研究されているが、現在の段階では、ニューロネットワークの学習・チューニングには多くの教師信号を必要とし、時間と手間がかかるのが難点となっている。しかしながら、個人の主観の学習にニューロネットワークを応用することは有効であると

考えられ、今後の研究課題の一つである。

また、現在のシステムは、個人のデータベースを基にしたあいまい検索システムだが、他のシステムにおいても、個人の主観の相違が重要となる場合があると思われ、それに対応するシステム・アプリケーションを模索中である。

参考文献

- [馬野 89] 馬野:“ファジイデータベースの現状と動向”, ADSS'89 Kyoto, pp.207-214, 1989
- [本多 91] 本多, 小高, 野村, 横山, 松下:“曖昧な問い合わせによる検索環境(1) — モデル化とインタフェース —”, 情報処理学会第42回全国大会, 4-103 ~ 104, 1992
- [大木 91] 大木, 小高, 野村, 横山, 松下:“曖昧な問い合わせによる検索環境(2) — 曖昧検索に適した LBG クラスタリング —”, 情報処理学会第42回全国大会, 4-105 ~ 106, 1992
- [野村 92] 野村, 大木, 小高, 横山, 松下:“ファジイ検索のための平均演算子による曖昧属性の生成”, 日本ファジイ学会誌, Vol.4, No.2, pp.126-142, 1992
- [Zadeh 65] L.A.Zadeh: “Fuzzy Sets”, Information Control, Vol.8, pp.338-353, 1965.
- [Turksen 91] I.B.Turksen: “Measurement of membership functions and their acquisition”, Fuzzy Sets and Systems 40, pp.5-38, 1991
- [藤代 91] 藤代, 白井, 池辺, 國井:“個別化されたファジイデータ管理のためのグラフ指向スキーマ設計”, 日本ファジイ学会誌, Vol.3, No.1, pp.86-96, 1991

[SUN 90] Sun microsystems: "A Socket-Based Interprocess Communications Tutorial", Network Programming Guide, pp.251-278, 1990

[Zemankova 87] M.Zemankova-Leech, A.Kandel, 向殿 訳: "ファジイ・リレーシヨナル・データベース", pp.37-57, 1987