

第1章 緒論

最近、コンピュータやインターネットを用いた情報検索に対する関心がとみに高まっている。使用者の側面からはより便利に利用できる知能的な質問処理システム開発に対する要求が現れている。既存の関係形データベースシステム (RDBMS) [28] からの情報検索は事前知識や経験がある人の場合は自分が望む情報をキーワードを利用してたやすく検索できる。しかし、こんな場合にも正確に match されるキーワードがないし、知らない場合は円滑な資料検索が困難な実状である。例えば‘背が高いし英語成績が良い学生たちの名前を出力しなさい’ という質問に対して現在の関係形データベースでは検索が不可能な場合がある。しかし、曖昧 (vague) な概念を使用する場合にはその関係を示す適切な演算子によってある程度の誤差を持つ情報などをデータベースから検索することが可能となる。曖昧な質問を円滑に処理できるシステムがまだ基礎的な研究に局限されて実用化されない重要な理由としては曖昧な質問を処理できるシステムの性能が低いことと、システム間の互換性の問題と使用者の認識不足の問題等が挙げられる。一方、ファジィ理論 [32] の応用の中では情報処理への応用が挙げられる。情報処理への応用の中で特に研究されている分野がファジィ・データベース [29] 又はファジィ情報検索 [15] である。ファジィ情報検索とファジィ・データベースを区別せずに広義の意味でファジィ・データベースと略称する場合もあるが、しかし、情報検索という主題は非常にその応用範囲が広いため、一旦、情報検索システムを中心にいろいろの主題に分けて見ると、

- ①情報検索システム
- ②データベース管理システム
- ③経営管理システム
- ④意思決定システム
- ⑤質問応答システム

などがある [26]。これらにはすべてファジィ理論の応用が可能である。

上に述べた状況を背景にして、本論文ではファジィマルチ集合の理論を考察し、それを使用した関係データ・ベースへの応用を行なう。そしてファジィマルチ集合についてこれまで研究されてきた既存のファジィマルチ集合の演算と関係などの理論を新たに定義し直す。そこで、ファジィ関係データ・ベースにおいて、マルチ集合の理論を利用したファジィ SQL を考える。ファジィデータ・ベースでの検索対象集合やデータ・ベースへの問い合わせの結果などにおいて、同じ記号が多数生じることは例外的ではないのでマルチ集合論は有望と思われる。大量のデータを処理するためにはデータを統一的に蓄積し取り扱う必要がある。しかし人間の直接に関係する分野では明確に定義できないこともあるし、定義する必要がないものもある。そこで、あいまいなデータとか質問を可能性分布あるいはファジィ集合によって表現することができる。そして関係データベースの理論では SQL による質問の結果としてマルチ集合が生じる場合がある。一方、バッグ (bag) とも呼ばれるマルチ集合は数学的構造としては普通の集合より弱いため理論的考察はあまりされてこなかった。しかし、最近、情報処理システムでは新しい数学的構造が要求されている。たとえば関係データベース・システムはすでに標準的なものであるが、そのデータベース・システムの問い合わせのいくつかは既存の数学的構造としては記述できない面などがある。すなわち、SQL が要求されたテーブル

では同じ tuple の多重コピー (multiple copy) が存在するが反面、古典数学に基盤をおいた関係データベースではそのような multiple copy は許されない。

マルチ集合はバッグとも呼ばれ、重複した要素を保つ集合のような構造である [1, 9, 14]。そのことによって関係データベース・システムから生じるさまざまな形態の質問を処理するための構造として適当だと見られる。マルチ集合は普通の集合に類似しているが要素の重複が許される。普通の集合からマルチ集合を形成するために、全体集合の要素を選択する場合、同一の要素を複数回とることができ。すなわち、全体集合を $X = \{a, b, c, d\}$ とすれば $A = \{a, a, b, b, c\}$ はマルチ集合である。つまり通常の集合をマルチ集合の一種とみなすことができる。Yager によって提案されたファジイマルチ集合では、

$$\{(a, 0.5), (a, 0.6), (b, 1.0), (b, 1.0), (c, 0.0)\}$$

のように同じ要素が同じグレードあるいは異なるグレードとともに複数個存在することができる。Yager [30] は包含、相等関係、和集合、積集合などの基本的関係と演算を定義しているが、しかし、その定義では通常のファジイ集合の基本的関係・演算とは整合性がない部分があるため不適切な点がある。そこで本論文では Yager の提案したファジイマルチ集合の基本演算などを新しい 'grade - sequence (グレード列)' という概念を使用して基本演算と関係を新たに定義し直し、通常のファジイ集合演算の一般化としての包含・相等関係や和・積 ($A \cup B$, $A \cap B$) 演算を定義する。この演算や包含関係の定義の要点は各要素に対するグレードの集合を順序付け一例に並べておくことである。グレード列によって先のファジイマルチ集合は、

$$\{(0.6, 0.5)/a, (1.0, 1.0)/b, (0, 0)/c\}$$

と表される。また、ファジイマルチ集合からクリスプマルチ集合への α -カットを定義し、 α -カットと集合演算の可換性を用いてファジイマルチ集合演算が、交換、結合、分配法則を満たすことを示す。またそれらがどのような意味で、Yager

の提案よりも適切な演算であるかを示す。マルチ集合に特有の演算として、直和演算 $A \oplus B$ が考察されている。これについては Yager によるファジィマルチ集合の直和演算の定義が適切であると思われる。本論文ではファジィデータベースに関する応用的な部分として、クリस्पマルチ集合を用いた通常の SQL を考察する。そのため、 α -カットと交換可能なファジィマルチ集合の新しい像の演算について提案する。この部分ではマルチ集合が $f[A]$ と $f(A)$ すなわち、

$$f[A] = \bigoplus \{f(x) : x \in A\}$$

$$f(A) = \bigcup \{f(x) : x \in A\}$$

で表される 2 種類の像を持つことを示す。マルチ集合の最も顕著な特徴は和演算 $A_1 \oplus A_2$ をもつことである。ファジィマルチ集合に関する $f[A]$ は和演算と可換であるが、合併演算とは可換でない。このように像 $f[A]$ の意義は、和演算から生じてくると考えられる。ファジィマルチ集合に関する 2 種類の像のうち像 $f[A]$ の応用上の意義を明らかにする。一方、情報処理の観点から見ると、ファジィとは若干違うが知識表現、知識獲得などに応用されるラフ集合も重要である。ラフ集合の概念を初めて提案したのは Pawlak [22] であるが、最初のラフ集合の基本概念は classification と categories というある知識の概念に対する理論だったが、近年、ファジィ理論とラフ理論の両分野の関連についての研究が盛んになされており、またこの理論を利用した応用分野すなわち、エキスパートシステム、機械学習、パターン認識、決定支援システムなどのさまざまな人工知能分野にまたがって多様な分野で活発な研究がされている。ラフ集合の基本概念は分類を用いた集合の上近似と下近似である。本論文ではファジィマルチ集合に関する像 $f[A]$ を適用してみることを考える。

そこで、像 $f[A]$ に基づく単純な計算法によって集合 A のマルチ集合による上近似 $R^*[A]$ を定義する。これを用いて通常の上近似を表現することができる。マルチ集合とファジィマルチ集合の理論が詳細に論じる価値をもつ理由は、マルチ

集合は計算アルゴリズムに直接関連しているからである。他の応用対象としては、先に述べたファジィ・データベースが挙げられる。一般に、与えられた検索対象集合において、ある属性のみを見れば、同一の属性値が複数個認められるのが一般的である。この場合クリस्पあるいはファジィマルチ集合対象集合に対する自然な記述手段となる。Yager [30] はクリस्पマルチ集合をクリस्पな関係データベースの射影演算に応用することを提案している。Yager の射影演算をファジィ関係データベースにおける射影演算に拡張すれば、ファジィマルチ集合が得られる。また、これらの射影演算は上に述べた像 $f[A]$ に対応している。このことを含めて、いろいろな方法でファジィマルチ集合をファジィ・データベースに応用し、その有用性について考察することができる。要約すると、本論文ではファジィマルチ集合がファジィ集合論の中でも基本的な概念に属し、かつ、重複する特徴記述のための自然な手段であることを示し、その理論を確立するとともに、ラフ集合への応用とデータベースへの応用について考察している。