

1 H-4 拡張演算子のあいまい要求処理への適用

平岩 真一 阿倍 博信 小高 俊之 野村 恭彦 横山 光男 松下 温
(慶應義塾大学 理工学部)

1. はじめに

現在のデータベースでは、情報が要求に対してマッチしているか否かの2値的な判断のみを行っているため、あいまいな概念(大きいとか、かっこいいなど)を含んだ要求を処理することが、非常に困難である。そのため、近年、あいまいさを含んだ検索要求が可能な、ファジ理論を用いたデータベースがいくつか提案されている[1]。しかし、何かを選んだり、買ったりするような場合には、より抽象的な概念による検索項目が必要となる。

こうしたあいまいな検索要求に応えるため、本稿では、従来のファジデータベースにおけるand、orの見直しと、検索項目の生成について提案する。

2. ファジ集合の演算

ファジ集合は、要素が集合に属するか属さないかを明確にできないときに用いられる。全体集合 U におけるファジ集合 A は、

$$\mu_A : U \rightarrow [0, 1]$$

なるメンバーシップ関数 μ_A によって特性づけられた集合であり、 $\mu_A(u)$ は要素 u のファジ集合 A におけるグレードを表す。

2つのファジ集合 A, B の和集合(AND)、共通集合(or)は次のように表される。

$$\text{和集合} : A \text{ AND } B = \min \{ \mu_A(u_A), \mu_B(u_B) \}$$

$$\text{共通集合} : A \text{ OR } B = \max \{ \mu_A(u_A), \mu_B(u_B) \}$$

これをn次元に拡張すると、 $x_i = \mu_i(u_i)$ とすれば、

$$\text{和集合} : \min \{ x_i \mid i=1, n \}$$

$$\text{共通集合} : \max \{ x_i \mid i=1, n \}$$

と表される。

これまでのファジデータベースでは、各アトリビュート間のandとorはそれぞれminとmaxで計算されてきた。しかし、これでは、質問に対する結果としてでてくる値が、一つのアトリビュートに大きく依存してしまう可能性があり、必ずしもユーザーの直感とあうとは考えにくい。たとえば、2つの条件に対する一致度が、(0.9, 0.2), (0.2, 0.2) の場合条件への一致度が同じ値になってしまう。また、検索条件の設定において、minとmaxでは、条件Aと条件Bに対し条件Aに重点を置きたいときのように、ある項目に重みをおくことが非常に困難である。これらの点から、検索方法の手法として

1) 新しいandとor

2) 条件間の重み付けの計算法

について提案する。

2.1 新しいand, orの検討

ここでは、andとorに、あいまいな意味をもたせるため、平均演算子[2]を用いることを考えた。平均演算子には、算術平均、幾何平均、調和平均といった古典的な平均があげられるが、これらの違いについて、考えてみる。いま、2つのデータの条件との一致度が x_1, x_2 だとすると、3つの平均演算子を用いたandはそれぞれ次のように表される。

$$\text{算術平均} : x_a = \frac{x_1 + x_2}{2}$$

$$\text{幾何平均} : x_g = (x_1 \cdot x_2)^{1/2}$$

$$\text{調和平均} : x_h = \frac{2}{1/x_1 + 1/x_2}$$

また、orもandの双対として次のように簡単に表すことができる。

$$\text{算術平均} : x_a' = 1 - \frac{(1-x_1) + (1-x_2)}{2}$$

$$\text{幾何平均} : x_g' = 1 - ((1-x_1) \cdot (1-x_2))^{1/2}$$

$$\text{調和平均} : x_h' = 1 - \frac{2}{1/(1-x_1) + 1/(1-x_2)}$$

これらを2次元平面上で考えると、演算結果はFig. 1に示されるように $x_1 = x_2$ 上に写影される。この結果から、算術平均は、andとorの中間の値をとり、幾何平均、調和平均となるにつれ、and、orに近い値をとるようになる。この点から考えると、アトリビュート間の関係によって、演算子を使い分けることにより、よりあいまいな表現ができる。また、多くの条件がandで結ばれているような厳しい条件の場合、平均演算子を用いることにより、条件全体に対する一致度の差を明確にすることができる。

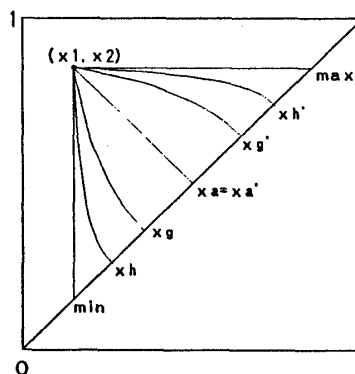


Fig.1 平均演算子

2.2 アトリビュート間の重み付け

条件を設定する場合、アトリビュート間の重み付けが必要になる場合がある。例えば、“若い and 背が高い”という条件において、“若い”より“背が高い”のほうが重要度が高いたいときなどに重み付けが必要となる。

このような場合、and・or演算に平均演算子を用いると、比較的簡単に重み付けができる。平均演算子を拡張して、n個のアトリビュートについて考えた場合、 x_1, x_2, \dots, x_n に重み w_1, w_2, \dots, w_n をつけたときの一般式は次のようになる。

$$x_a = \frac{\sum_{j=1}^n w_j \cdot x_j}{\sum_{i=1}^n w_i} = x_a'$$

$$x_g = \left(\prod_{i=1}^n x_i^{w_i} \right)^{1/\sum w_i}$$

$$x_h = \frac{\sum_{j=1}^n w_j}{\sum_{i=1}^n (w_i/x_i)}$$

$$x_g' = 1 - \left(\prod_{i=1}^n (1-x_i)^{w_i} \right)^{1/\sum w_i}$$

$$x_h' = 1 - \frac{\sum_{j=1}^n w_j}{\sum_{i=1}^n (w_i/(1-x_i))}$$

Fig.2に例として、幾何平均演算子について示した。

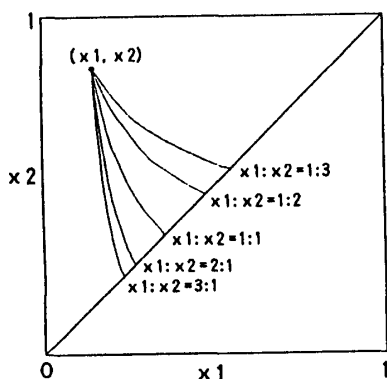


Fig.2 重み付けをしたときの幾何平均

3. あいまいな検索項目の生成について

一般にファジィデータベースでは、ユーザーがメンバーシップ関数を変更することによって、よりパーソナルなシステム環境をつくるのが可能である。従って、フロントエンドとして従来のデータベースにかぶせることができれば、どんなデータベースでもファジィ検索が可能となる。

その実現の一つの方法として、データベース中の複数のクリスパなアトリビュートから、1つのあいまいさをもつアトリビュートを作る方法について考える。新アトリビュートは、データベース内のアトリビュートからそれに関連しているものを選び、そのあいだの関係づけることによって作成される。

n個のアトリビュートから新しいアトリビュートを作成する場合、n次元空間を考え、各アトリビュートの属性を座標軸とし、すべてのデータをn次元空間内

にプロットする。それを平均演算子を用い、 $x_1 = x_2 = \dots = x_n$ 上に写影する。この直線上にプロットされた点の集合が新しいアトリビュートとなる。2次元空間での例をFig.3に示した。

自動車の“大きさ”を例にする。アトリビュートとして選ばれるものは、“全長”・“全幅”・“全高”が考えられ、これらの弱い(orを多少含んだ)andが、車の大きさと考えられる。よって、ここではandに幾何平均を用いればよい。もし“排気量”というアトリビュートも“大きさ”に関連していると思えば、上のアトリビュートに加え、その間の関係により演算子を選ばばいいのである。

これを用いれば、ユーザーは、データの更新やデータ数の変更を気にしないで、複数のアトリビュートとその間の関係を決める事により、新しいアトリビュートを作成し、あいまいな検索をすることができる。

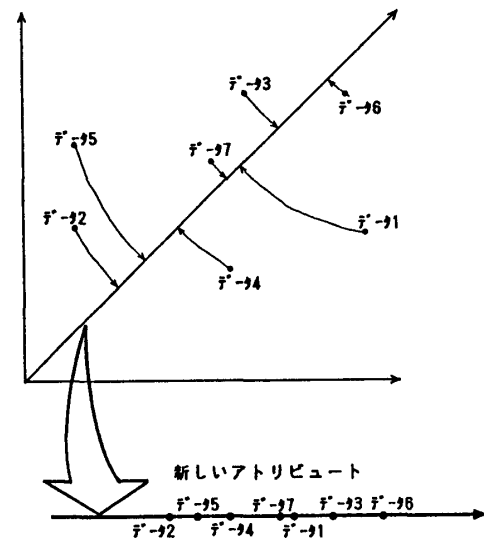


Fig.3 あいまいなアトリビュートの作成

4. おわりに

本稿では、ファジィデータベースにおいて平均演算子を用いたand、orの利用と、ユーザー定義可能なアトリビュートの生成方法についての提案をした。今後の課題として、さらにあいまいな言語への対応及びあいまいなデータの取扱いなどより一層の拡張が考えられる。また、これをもとにして検索システムを作成した[3]。

<参考文献>

- [1] Maria Zemankova, Abraham Kandel, "Implementing Imprecision in Information Systems", Information Science Vol.37, pp.107-141(1985).
- [2] D.Dubois, H.Prade, "A Review of Fuzzy Set Aggregation Connectives", Information Science Vol.36, pp.85-121(1985).
- [3] 野村, 阿倍, 小高, 平岩, 横山, 松下, "あいまい要求のための検索システム", 本大会.