

ファジィ積分を用いたファジィデータ検索システム

高萩栄一郎*

1 はじめに

従来のデータ検索システム—データベースなどに格納された膨大なデータから、検索条件にあてはまるデータを検索するシステム—は、検索条件にぴったりあてはまるデータしか検索できなかった。たとえば、賃貸住宅の検索で「家賃7万円以下、2DK」としたとき、この条件に少しでも当てはまらないデータ、たとえば「家賃7万2千円、2DK」は、検索されなかった。もし人間の検索者（不動産屋の店員）であったら、家賃が7万2千円であっても「ちょっと予算オーバーだけど、こんな物件がある」として、7万2千円のデータ（物件）を提示するだろう。本システムは、このようなデータも検索できるような人間に近い検索が行えるシステムを目指している。

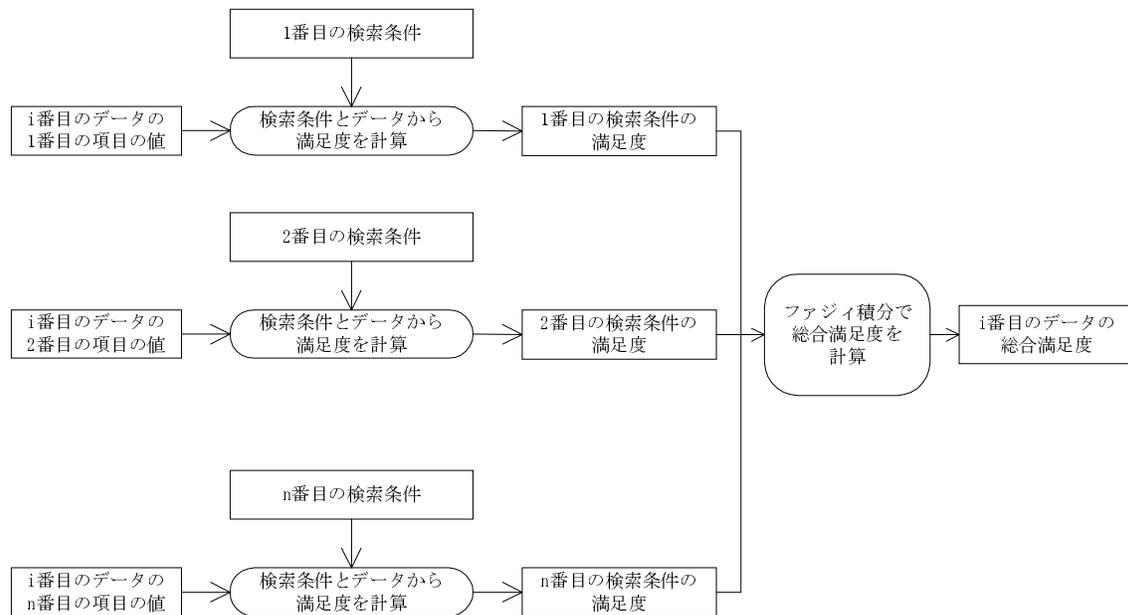


図 1: 本データ検索システムの概略 (i 番目のデータの総合満足度の計算)

本システムでは、図 1 のように、個々の項目（家賃、間取りなど）について、検索条件とそのデータの値を比較し、その満足度を求めている。次に、個々の項目の満足度を統合して、総合満足度を求める。総合満足度は、各データ毎に求め、総合満足度が高いデータを表示している。ここで、個々の満足度から総合満足度を求めるときにファジィ積分を利用する。ファジィ積分は、従来

*専修大学 商学部

の加重和や最小値演算を含むさまざまな演算を可能にする新しい計算方法である。

このテキストは、表計算ソフトウェア (Excel) を用いて、このファジィデータ検索システムの作成法を記述したものである。

2 ファジィ積分を使ったデータ検索システムの概略

本節では、賃貸住宅の物件の検索システムを例にして説明する。簡単化のために、検索項目は、「家賃」、「間取り」、「駅からの時間(徒歩)」の3つとする。

2.1 従来のデータ検索

従来のデータ検索システムを計算方法を図2に示す。

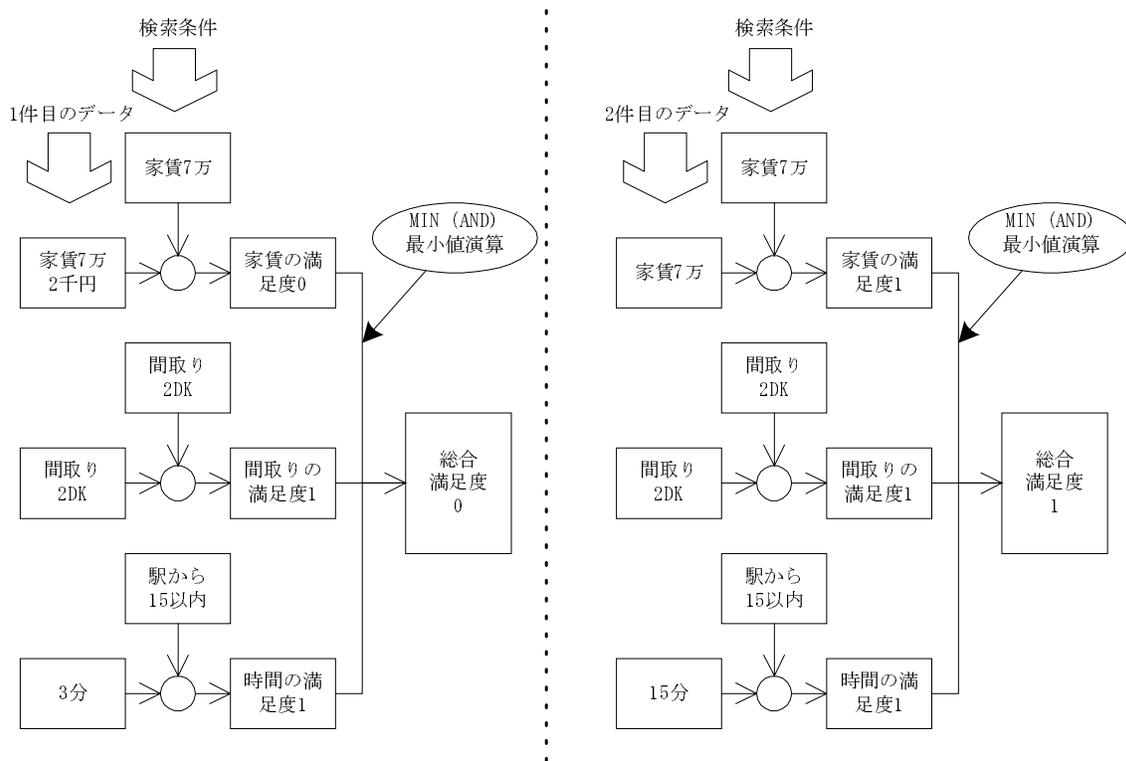


図 2: 従来のデータ検索システム

図2は「家賃:7万円以下」、「間取り:2DK」、「駅からの時間:15分以内」という条件で、データ検索を行った場合の例である。例として、2件のデータについて検索されるか否かを示したものである。

1. 1件目のデータについて、個々の検索条件があてはまるか調べる

家賃 検索条件が7万以下で、データは7万2千円なので、このデータはあてはまらない。したがって、満足度は0である。

間取り 検索条件にあてはまるので満足度は, 1.

駅からの時間 検索条件の 15 分以内にあてはまるので満足度は, 1.

2. 従来の検索システムでは, すべての条件を満足することが要求される. したがって, 3つの満足度の最小値 (AND 演算) をとり, それを総合満足度とする. したがって, 個々の満足度で1つでも0のものがあれば総合満足度は, 0となる. したがって, 1件目のデータの総合満足度は0となる.

3. 2件目のデータについて, 個々の検索条件があてはまるか調べる

家賃 検索条件が7万以下で, データは7万円なので, このデータはあてはまる. したがって, 満足度は1である.

間取り 検索条件にあてはまるので満足度は, 1.

駅からの時間 検索条件の 15 分以内にあてはまるので満足度は, 1.

4. 個々の満足度はすべて1なので, 3つの満足度の最小値 (AND 演算) をとると1になり, 2件目のデータの総合満足度は1となる.
5. 各データのうち, 総合満足度が1のデータが表示される. この場合, 2件目のデータのみが表示される.

2.2 ファジィデータ検索システム

§2.1 では, 1件目のデータは, 家賃の条件を満たしていないので検索されない. しかし, 不動産屋の店員に「家賃7万以下, 2DK, 駅から15分以内」という条件を出したら, たぶん, この1件目のデータの物件を「家賃は少しオーバーしますが, 駅に近くて便利な物件ですよ」と提示するだろう. これは, 1つの検索条件がほんのわずかでも満たしていなければ, その検索条件の満足度は0となり, したがって総合満足度が0となることによる.

そこで, 各検索条件の満足度について, 0か1のどちらかをとるのではなく, 0.7とかいう0と1の間のファジィ値 ($[0, 1]$ 区間の値) を取るのようする. これを「原始的なファジィデータ検索システム」と呼ぶことにする. たとえば, 各検索条件について, 次のように考えとする.

家賃:7万円以下 この条件は, 「だいたい7万円以下くらいにしたい」を意味しており, 「6万以下だったら大満足」「7万以下だったら, まあ満足」「8万以下だったら, ぎりぎり許せる」「8万をこえたら, 不満」と考えてみる. この場合, 6万以下だったら満足度は1, 7万だった0.8, 7万2千円だったら0.6, 8万をこえたら0と考える.

間取り:2DK この条件は, 2DKなら満足(満足度1)で, ちょっと外れて, 1LDKでもやや満足(満足度0.8), 2Kでは, やや不満(満足度0.4)などを表すと考えられる.

駅からの時間:15分以下 これは「駅から近いほうが良いが, 徒歩15分くらいまでだったら, 多少離れても許せる」と解釈できる. したがって, 近ければ大満足で, たとえば, 徒歩3分だったら, 満足度は0.95, 徒歩15分は, ぎりぎりまあまあ満足で, 0.7くらいとする.

これらの実際の計算方法は, 後の節()で検討する.

図3は, 個々の満足度をファジィ値で表現したものである. また, 個々の満足度の集計には, 最小値(MIN)と平均値の2つを計算してみた.

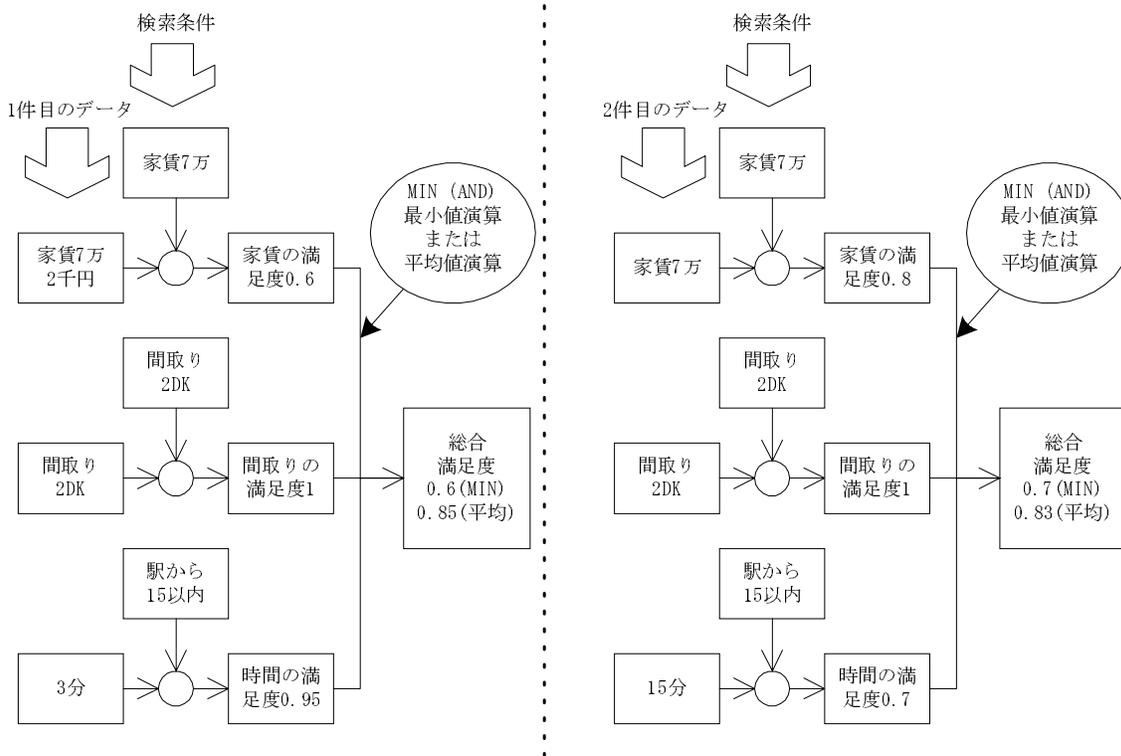


図 3: (原始的な) ファジィデータ検索システム

最小値 従来のデータ検索システムでは、1つでも満たさないとき、そのデータは検索されないという最小値演算が行われていた。これに習い、各満足度の最小値を総合満足度とする。商品選択などでは、人はもっとも悪いところ（条件に合わない点）を気にするという考え方による。図3では、1件目のデータの最小値（もっとも満足しない点）は、家賃でその満足度は0.6であるので、総合評価値は0.6である。2件目のデータの最小値（もっとも満足しない点）は、駅からの時間でその満足度は0.7であるので、総合評価値は0.7である。

最小値での総合満足度の計算では、2件目のデータがもっとも総合満足度が高い。

平均値 最小値では、最小の値（1件目では家賃、2件目では駅からの時間）のみが考慮されていて、他の項目の値は考慮されていない。最小の値のみ総合評価値に影響を与え、他の評価項目の値で、最小の値の低さを補えないので、これを専門用語では「非補償型」と呼ぶ。したがって、1件目のデータで、他は同じで駅からの時間が3分から1分になっても総合評価値は変わらない。通常、ある満足度が上がれば、総合満足度があがると考える。そこで、平均値を使ってみる。平均値は、ある項目の満足度を他の満足度で補うことができるので、「補償型」と呼ばれている。

図3では、1件目のデータの平均値は、0.85、2件目のデータの平均値は、0.83となり、1件目のデータが高くなる。これは、2件目と比べて、家賃の0.6という満足度を、駅から距離の0.95により補ったことによる。

2.3 ファジィ積分型データ検索システム

§2.2 では、総合評価値の計算に、最小値方式、平均値方式の 2 つの方式を示した。それぞれの問題点を示すと、

最小値方式 最小値の検索項目の満足度しか総合評価値に反映されない。たとえ、他の項目の満足度が上昇しても総合評価値は同じである。ある項目の満足度の上昇は、すくなくとも多少は、総合評価値の上昇につながるの自然であろう。

平均値方式 平均値での総合評価は、低い満足度の項目も高い満足度の項目も同じように平均する。商品選択では、自分の希望に 1 つでも合わないことを気にすることが多い。すなわち、すべての項目での満足度がバランスを取って満たされていることを望むことが多い。たとえば、1 件目のデータで、時間の満足度が 0.1 下がっても、家賃の満足度が 0.1 あがる物件を望む（たとえば、徒歩 5 分で、家賃 7 万 1 千円の物件）ことが多いだろう。

がある。したがって、人間は、個々の検索項目の満足度は総合評価値に影響を与えるが、低い満足度の項目を重要視すると言えるだろう。したがって、平均値と最小値の評価の中間の評価 – 低い満足度の項目の重視度を大きくした加重平均 – を行えばよいことがわかる。このような総合評価を可能にしたのがファジィ積分である。

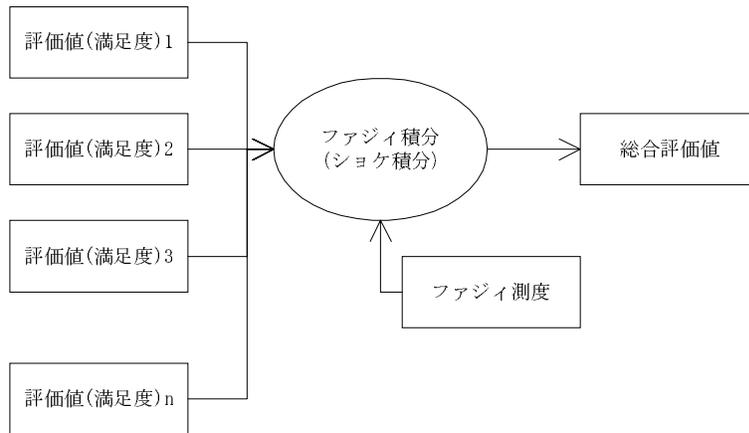


図 4: ファジィ積分のイメージ

図 4 のように、ファジィ積分は、 n 個の評価値から 1 つの総合評価値を求める計算を行う。このような総合評価値を求める計算方法には、§2.2 で示した最小値による方法や平均値による方法など多数ある。ファジィ積分は、平均値や最小値などとは異なり 1 つの計算方法ではなく、ファジィ測度をとりかえることにより、さまざまな総合評価を行うことができる。実際、平均値をとるファジィ測度を使えば、平均値による総合評価になり、別のファジィ測度を使えば、最小値、また別のファジィ測度を使えば、最大値となる。そして、うまくファジィ測度を与えると、先ほどの述べた、「平均値と最小値の評価の中間の評価」が可能になる。表 1 は、平均値と最小値の評価の中間の評価が行えるようにファジィ測度を作り、そのファジィ測度によって総合評価したものである。ファジィ積分の場合、1 件目のデータは、家賃の満足度が低いことを駅からの時間で補えず、2 件目より低い総合満足度である¹。ファジィ積分には、菅野積分、シヨケ積分など多数あるが、本稿では、

¹この程度の差なら、1 件目の方が高い総合評価値がよいと考えることも可能である。1 件目の総合評価値を高くするには、さまざまなパラメータの設定（チューニング）により可能である。

シヨケ積分を用いることにする。

表 1: ファジィ積分による評価例

評価法	1 件目のデータ	2 件目のデータ
平均値による評価	0.85	0.83
ファジィ積分	0.73	0.76
最小値	0.60	0.70

3 ファジィ積分

前節 (§2.3) で、ファジィ測度をとりにかえることによりさまざまな評価が可能であるあると書いた。本節では、ファジィ測度の作り方を説明する。ただし、ファジィ測度の説明は、他書に譲ることとし、ここでは、パラメータ ξ (グザイ) と各評価項目の重要度により、ファジィ測度を作る方法に説明を絞ることとする。

3.1 ファジィ積分の概念

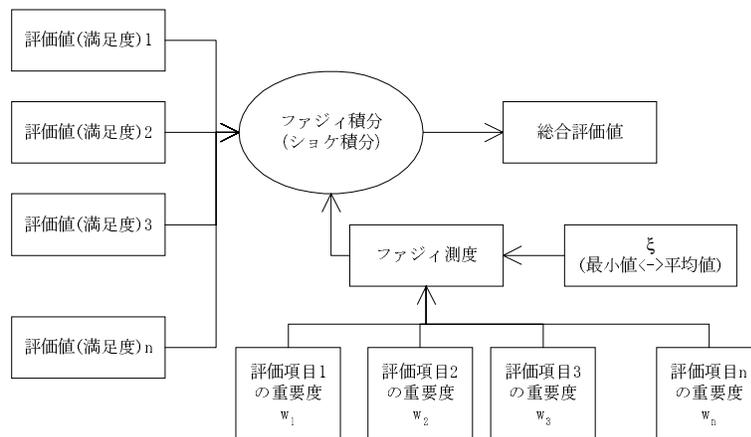


図 5: 本稿で用いるファジィ積分

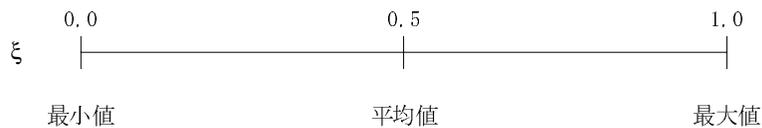


図 6: ξ の意味

図 5 は、ファジィ積分の概念図である。ファジィ測度は、各評価項目の重要度と相互作用の指標（最小値に近いのか・平均値に近いのか）で作られる。

重要度は、各検索項目をどれくらい重要視するのかを表す。たとえば、家賃を 5, 間取りを 3, 駅からの時間 2 とすると、その重要度は、 $w_{家賃} = 0.5, w_{間取り} = 0.3, w_{時間} = 0.2$ となる（合計が 1 になるように調節する）。

相互作用の度合い ξ は、最小値に近い評価をするのか、平均値に近い評価をするのかを表す値である。図 6 で示すように、0 で最小値, 0.5 で平均値, 1 で最大値になり、それぞれの値に近づけるとその評価に近くなる²。たとえば、 $\xi = 0.1$ とすると、かなり最小値に近い評価になる。ただし、 $\xi = 0, \xi = 1$ では、重要度は無視され、それぞれ最小値, 最大値になる。

3.2 感度分析

表 2: 例題

評価項目	重要度	1 のデータ	2 のデータ
家賃 (rent)	0.5	0.6	0.8
間取り (layout)	0.3	1.0	1.0
駅からの時間 (time)	0.2	0.95	0.7

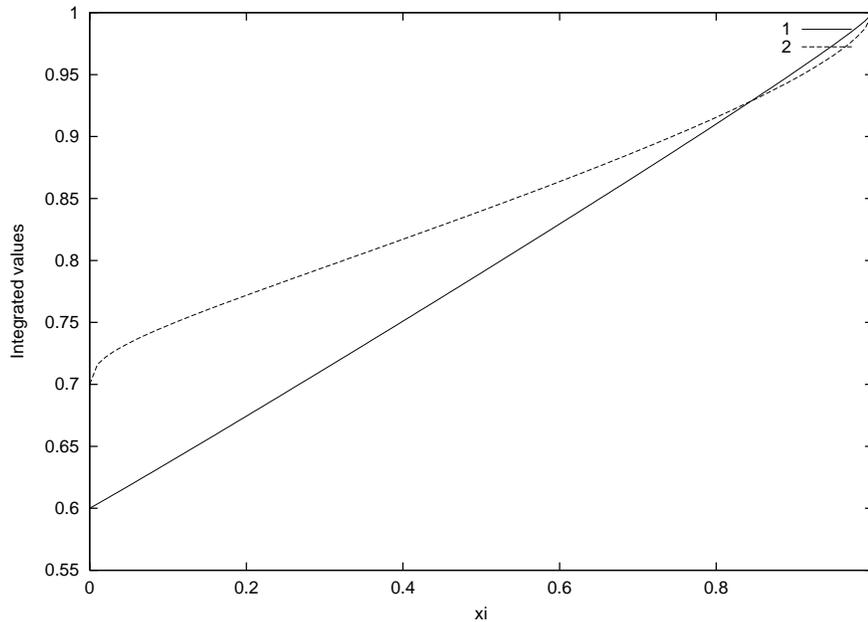


図 7: 例題の感度分析 (横軸: ξ , 縦軸: ショケ積分値 (総合評価値))

表 2 のデータを使って、ショケ積分を行うと図 7 のようになる。このように ξ を特定の値に定めず、 ξ を変化させてみた結果を考察する方法を感度分析という。

$\xi = 0$ 最小値での評価になっている。1 は家賃の 0.6, 2 は駅からの時間の 0.7 になる。

²上記の説明では、最小値と平均値の中間の評価しか用いなかったが、最大値と平均値の中間の評価も可能である

$\xi = 0.5$ それぞれ重要度での加重平均になっている．たとえば，1 は，

$$0.5 \times 0.6 + 0.3 \times 1.0 + 0.2 \times 0.95 = 0.79$$

になっている．

$0 < \xi < 0.5$ 最小値と加重平均値の間になるように補間されている．

$\xi = 1$ 1,2とも最大値は，間取りの 1.0 であるので，1 である．

この場合， $\xi = 0.8$ 付近で逆転しているが，これは，良い値，すなわち，1 の「駅からの時間」の 0.95 を重視したためである．

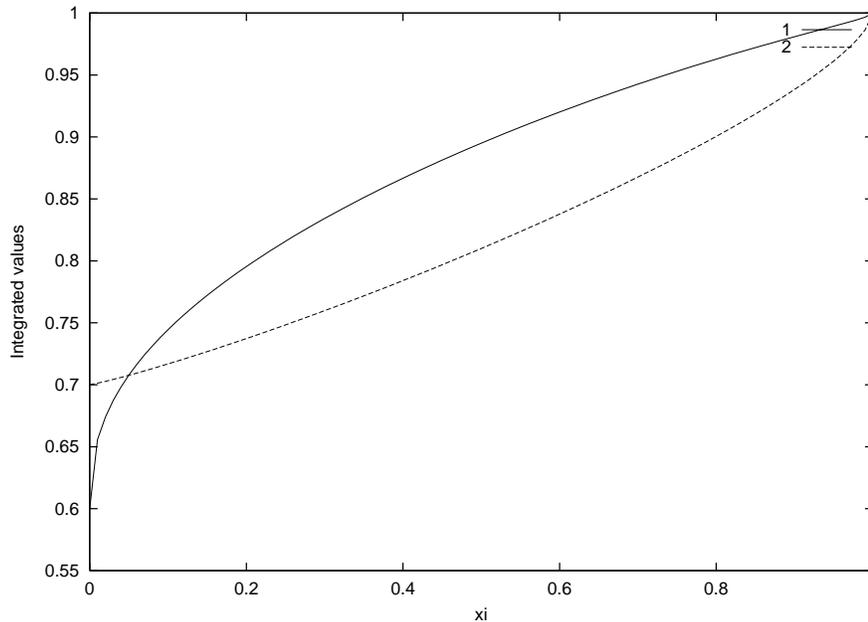


図 8: 重要度を変更して感度分析 (横軸: ξ , 縦軸:シヨケ積分値 (総合評価値))

次に，重みを変更して，駅からの時間を重要視した場合「駅からの時間が重要で，徒歩 15 分がぎりぎり，近ければ近いほどよい」という場合を考え，時間の重みを 0.05，家賃を 0.2，間取りを 0.3 としてみた．その感度分析の結果を図 8 に示す．多くの ξ で，時間の重要度が高いために，1 が上位に来ている．しかし， $\xi = 0.5$ 付近より小さい場合，逆転し，2 が上位にくる．これは， ξ が 0 付近の場合，もっとも低い評価値 (1 は家賃の 0.6，2 は時間の 0.7) を中心に評価するためである．

3.3 デフォルトの ξ

ξ を $0 < \xi < 0.5$ にすると，最小値と平均値の間の中間の総合評価になることがわかった．しかし，実際にデータ検索システムを作る場合， ξ の値を固定し，それぞれのデータに対して 1 つの総合評価値を求めなくてはならない．そこで，経験的な値ではあるが， $\xi = 0.2$ を提案する．この $\xi = 0.2$ で，データ検索システムを構築し，後に，検索結果などみて， ξ の値を調整する．

3.4 感度分析の実行

Fuzzy web を使って感度分析を行ってみます。データは、表 2 のものを利用します。

<http://www.senshu-u.ac.jp/thc0456/fuzzyweb.html>

(高萩のホームページ → ファジィのページ → ファジィ積分、計算 Web)

感度分析は、次の手順で行います。

1. 「 ファジィ測度 (相互作用の度合いを指定した) Choquet 積分の計算 」をクリック
2. 評価項目数, データ数を入力
 - 評価項目数:3 (「家賃」, 「間取り」, 「駅からの時間」の3つ)
 - データ数:2 (データ 1 と 2)
3. 評価項目名の入力
「rent」, 「layout」, 「time」と入力 (日本語は使用できない)
4. 評価項目の重要度の決定方法
「手入力」を選択
5. 重要度の入力
「0.5」, 「0.3」, 「0.2」と入力
6. 相互作用の度合いの決定
「 ξ を手入力」を選択
7. 相互作用の入力
感度分析を行うので, 適当な値 (たとえば 0.2) を入力
8. ファジィ測度の同定方法の選択
「入力の数基準」を選択
9. ファジィ測度の表示
そのままクリック
10. 入力値の入力
表 2 のように入力 (行と列が入れ替わっているのに注意)
11. Choquet 積分結果の表示そのままクリック
12. 感度分析の設定
感度分析を行うデータの番号: 「1」と「2」を設定
13. 結果が表示される。
修正は, 戻るボタンで

練習問題

1. 重要度は, 自分で一対比較をし, データに rent が 0.72, layout が 0.8, 時間が 0.75 を付け加えて感度分析してみよう。
2. AHP で作った例を, ファジィ積分型にして, 感度分析を行ってみよう。

4 データベースの作成

4.1 検索対象の設定

まず、何の検索システムを作るかの選定をする。中島 [?] の p.156 で述べているように、趣味性の高い商品のデータ検索システムは、難しい。たとえば、「かわいい洋服を検索してほしい」といわれても、人によって「かわいい」の基準が大幅に異なる。したがって、人によって、評価の基準が異なる検索項目が要求されるデータの検索システムは難しい。しかし、後で述べるように、検索時のキーワードを、人によって基準があまり異ならないものに絞ることによりある程度可能になるだろう。

対象の例：中古車の検索、贈答品の検索、プレゼント用の絵本の検索、パソコンの検索（含む増設・オプション）、賃貸住宅の検索、中古住宅の検索宿の検索、旅行プラン（ツアー）の検索、ペット種類の検索、Free GIF ファイルの検索など

4.2 検索項目，入力項目の決定

4.2.1 検索項目

検索項目は、できるだけ人間の店員に聞くような項目であったほうがよい。

難しい専門用語は使わない。まず利用者を想定する。多くの場合、専門知識を持たない一般の人を想定する。その商品に詳しくない一般の人がわかる言葉を使う。

目的で検索 目的を入力させることにより、必要とする商品の機能が特定されることが多い。たとえば、パソコンの検索で「インターネットを楽しむ」だったら「モデムが必要」とか「学校の宿題をやる」だったら「Office ソフトが必要」となる。

利用者の立場で 利用者の立場で検索できるようにする。たとえば、飲食店の検索で、定休日、営業時間で検索するのではなく「10月20日の夜」というような検索をできるようにする。

4.2.2 入力項目

入力する項目は、データを特定する項目（データの名前、電話番号など）と §4.2.1 の検索項目に対応する項目である。