

# 必ずヒットする「あいまいスペック検索システム」の開発

## Development of "Fuzzy spec search engine" that becomes a hit without fail

山川義介（株式会社 ALBERT）

### 1. はじめに

EC サイトでパソコンを購入する際、本体のサイズや重量、メモリーや HDD、モニターの大きさなど、希望のスペックを入力することで該当商品を絞り込み検索する「スペック検索」がよく用いられる。このような検索システムは、商品知識があまりなく、スペックなどへの理解が深くない消費者には適切に商品を選ぶことができない。もし選べたとしても、少しでも贅沢な要求をすると、「該当する商品はありません」というメッセージが出るが多かった。

実際の店舗での店員の対応はどうであろうか。不動産屋にアパートを探しに行くことを考えてみよう。駅から歩いて 5 分以内、2DK、家賃 10 万円で探しているのですが、とリクエストすると、いや 12 万円ならあるけど 10 万円じゃあないとか、7 分歩いてくれればいいアパートがあるのだけどどうするかね、と言われてたりする。人間が商品をお薦めする場合は、自動的にあいまいな検索をし、より消費者のニーズに近いものを推薦している。

これと同じことをネット上のショッピングでも実現できないだろうかという発想のもとで、本システムを開発した。

### 2. あいまい検索とは

「あいまい検索」とは、検索条件が完全一致しない対象を、一定のルールのもとで抽出する検索方法のことで、ファジィ検索ともいう。一般的にファジィ検索というと、テキスト検索において、指定した検索文字列パターンに被検索テキストが完全に一致しない場合でも検索結果が導出される方法を指すことが多い。類似検索を行なうための手法として、ファジィ理論[1]を応用した研究が多く行なわれており、例えばファジィシソーラスを用いたあいまいな検索から、データベースの属性を組み合わせて新たな属性で検索を行なう手法や、感性検索へ応用した研究などがある[2]。小澤[3]は、Fuzzy-C-means 法を用いたデータベースからの知識発見手法と、発見された知識を不動産データベース検索に応用している。これは、指定複合条件に完全一致するデータが無い場合に、条件を分解し検索要求を緩和するというものである。

### 3. 方法

本研究では、スペックで定義しやすい商品カテゴリ（電化製品、不動産、求人情報など）に関して、消費者が求める要求特性と、スペックの属性・水準から当該商品との類似度を計算し、完全一致の検索条件だけではなく、完全一致の商品がない場合には距離の近いものから順に商品を抽出し提示する方法を提案する。

#### 3.1 アイテムカテゴリ型属性の類似度計算

メーカーという属性に対し、A社、B社、C社といった水準があった場合、通常の検索ではA社を指定したらA社の商品しか抽出されない。しかし、B社はA社に極めて類似したメーカーであったとしたら検索対象にしてもよいだろう。あらかじめその水準間の類似性を定義しておき、水準が等しくなくても類似していれば検索されるシステムを作ることが可能である。

水準間の類似性は、当該水準を決定する要因に分解し、要因毎の類似性を測定し、これの総合評価で定義することができる。類似性総合評価はユークリッド距離などを用いて計算できるが、計算方式はこれに限らない。

類似性の要因には重み付けがある場合があり、検索する人によって異なった重み付けをした上で類似性の総合評価を計算することもできる。

商品が定義されるデータベースは、該当する水準に対して、等しいものには1、等しくないものには0のフラグを立てるが、類似しているものには、0より大きく1未満のフラグを立てる。類似する水準に立てるフラグには、その類似度によって変動する数値を使用する。商品を定義する属性数のうち、検索条件とどの程度が合致しているかの割合でフィット率を算出する。

フィット率は、合致の割合だけではなく、類似度（f）の程度も考慮した上で以下のように計算できる。

$$\text{フィット率} = \sqrt[n]{\frac{1}{f_1 \times f_2 \times f_3 \times \dots \times f_n}} \quad (1)$$

(ただし、1つでもフラグに0がある場合はフィット率は0とする。)

### 3.2 数量的属性の類似度計算

一方、価格や重さ、距離などの数量的スペックの場合、指定した条件の範囲に入っていないくても、若干の差異であれば許容できることが多い。希望する商品と類似した商品を推薦するため、図1に示すように、あらかじめ区間P~Qの範囲で検索した場合、この範囲をどれくらいはずれたら類似度をいくつにするという定義をしておく。

図1の例では、 $(X_n/P)$ 、 $(Q/X_m)$ の3乗とあるが、これをn乗 ( $n=1, 2, 3, \dots$ ) とすることもできる。

たとえば、モニターを15~19インチで指定したとしよう。P=15、Q=19となるので、スペックXがこの範囲であれば、もちろん検索条件に当てはまるので、類似度は1となる。

ところが商品nが14インチだった場合、指定のスペックからははずれているので、通常ではこの商品は抽出されない。しかし、15インチ以上という指定に対し14インチだった場合は、たとえばその類似度を81%と定義することで、本来求めている商品の仕様を100%は満足していないが、満足していない商品の中では最も近い商品として検索結果に出すことができる。

また、スペック検索で商品の絞込みをする際に、複数の条件を指定することが多い。n個のスペックに対して、n個の類似度が与えられた時、例えば図4の式でフィット率を求め、フィット率の大ききで検索結果を順位づけて提示することができる。

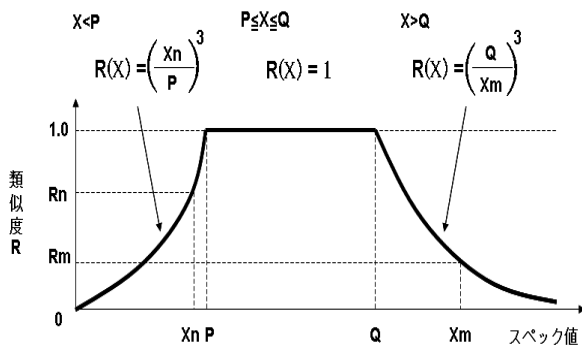


図1 比尺度スペックと類似度の関係

f: 類似度

	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	....	Spec n-1	Spec n
商品	f1	f2	f3	f4	f5	.....	fn-1	fn

$$\text{フィット率} = \sqrt[n]{f_1 \times f_2 \times f_3 \times \dots \times f_n}$$

図2 フィット率の計算方法例

価格や重さなどの比尺度ではなく、温度の様な間隔尺度の場合は、例えば図3に示すような、計算式を用いることができる。居住環境を選ぶ時に、P (0度) からQ (30度) までは快適なので、この範囲で選択するとする。たまにはXn (-2度) やXm (35度) でもいいだろうと考えた時、温度は比尺度ではないので、20度は10度の2倍暑くはない。そこで、温度として考えられる最低気温 (Xmin) と最高気温 (Xmax) をあらかじめ決めておき、これらを元に類似度Rを計算することができる。

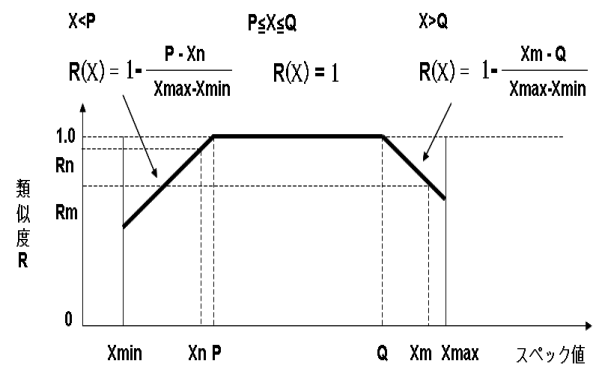


図3 間隔尺度スペックと類似度の関係

### 4. まとめ

本システムは、東芝、リコーなどのECサイトに導入され順調にアクセス数を伸ばしている。必ずヒットすることでサイト離脱率を抑え、機会損失を減らすことができていると思われる。

### 参考文献

- [1] L. A. Zader ; Fuzzy Sets, Inform. Control, Vol. 8, pp. 338-353, 1965
- [2] 小澤 順 ; ファジィ理論に基づく情報検索, 日本ファジィ学会誌, Vol. 13, No. 5, pp. 454-462, 2001
- [3] 小澤 順 ; ファジィクラスタリングとデータベースからの知識発見, 日本ファジィ学会誌, Vol. 8, No. 3, pp. 456-462, 1996