

WWW 感性検索システム KiSS におけるユーザモデルの構築

三浦 康史, 小澤 順, 今中 武

松下電器産業(株) 研究本部 中央研究所

「楽しい」などといった利用者の感性表現を用いて WWW データベースを検索する WWW 感性検索システム KiSS について述べる。本システムは年齢・性別に応じたユーザモデルを持ち、データベース構築者とユーザモデルとの相対的な感性の差異によって WWW データベースを補正する。これにより、個人の感性の違いを扱う。本稿では、ホームページから受ける感性に関するアンケート結果から、ユーザモデルを構築する方法を提案する。また、提案した方法の感性検索システムへの有効性を検討する。

A user model construction for WWW Kansei retrieval system KiSS

Kouji MIURA, Jun OZAWA and Takeshi IMANAKA

Central Research Laboratories
Matsushita Electric Industrial Co., Ltd.

In this paper, we explain the WWW Kansei retrieval system KiSS that enables users to retrieve by individual feeling such as "joyful". This system holds user models to adapt the difference of individualities by translating individual feeling of database constructor into that of user model. We present a method to construct user models from the result of questionnaires on WWW pages, and examine its effectivity.

1 まえがき

近年、インターネットでの WWW (World Wide Web) などに見られるように、広域型の情報サービスが急速に進歩・普及している。それにつれ情報の量が増大し、利用者にとっては欲しい情報の発見が困難となっている。

このような問題に対して、キーワードを用いた全文検索技術や意味解析・シソーラスなどを用いた情報絞り込み技術などが提案されている [3,7]。また、技術そのものの追求と共に使いやすさも重要である。

我々は、やわらかい・人にやさしいインタフェースとして、「楽しい」や「ホットな」などの感性語を用いた情報検索システムの研究を行っている。日常での会話を考えると、人は意味が曖昧な言葉を頻繁に用いており、検索においても感性語のような曖昧な言葉を用いると、手軽な検索ができると考えられる [2]。

しかし、同じ感性語に対しても受ける感覚は人によって異なる。この、人による感性の違いを扱うため、我々は年齢・性別の異なるユーザモデルをあらかじめシステム内に構築し、利用者が検

索を行なう場合にそのユーザモデルを選択する方法を用いる。つまり、利用者のステレオタイプ(代表的な集団)をあらかじめ用意しておく。これは、WWW 検索システムのような広域型システムでの利用者が不特定多数であり、一人一人の利用者に対応したモデルの構築が困難であるという理由からでもある。

感性語を用いた WWW の検索を行なうため、WWW の URL (Uniform Resource Location) が記憶されている URL 感性データベースを持つ。データベース中の各 URL は図 1 に示すように、タイトル、コメント、感性語属性を持つ。データ評価者は、データベース中のそれぞれの URL について、その URL のページから受ける感性度を評価する。検索を行なう場合は、この感性語属性を利用する。ここで、問題となるのは利用者の選択したユーザモデルと評価者との感性の違いである。我々は、この問題を解決するため、ユーザモデルに評価者との相対的な感性(相対感性)の差を持たせる。検索を行なう場合は、データ評価者の感性で評価された URL 感性データベースを、選択されたユーザモデル用のものに相対感性を用いて補正する(図 2)。

URL	タイトル	コメント	楽しさ	難易度
http://town.hi-ho.or.jp/	ワンダータウン	1.0	0.1
http://www.mei.co.jp/	ハナソニック ホームページ	0.6	0.5
	↓				

図 1: URL データ

我々は、年齢・性別の異なる 5 つのグループに対してホームページから受ける感性に関するアンケート調査を行ない、その結果から、ユーザモデルを構築する方法を提案する。また、提案した方法の感性検索システムへの有効性を検証する。

第 2 節では、我々の提案するユーザモデルと開発を進めている感性 WWW 検索システム KiSS (Kansei internet Search System) について説明する。第 3 節では、ユーザモデルの構築法について説明する。また、アンケートを用いて構

築したユーザモデルに対して、WWW 検索に適用する場合の考察を行なう。第 4 節では、まとめと今後の課題について述べる。

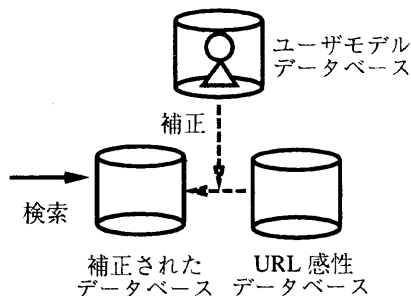


図 2: 相対感性によるデータベースの補正

2 ユーザモデル

2.1 ユーザモデル

情報検索を行なう場合、利用者の好みや性格、知識、思考法などを理解するためにユーザモデルを構築することは、欲しい情報を少ない労力で得るために必須のものと言える。我々は、利用者の感性を理解するための一手法として、感性語を用いたユーザモデルを提案する。

利用者の感性に合ったユーザモデルを構築するためには、検索前に利用者いくつかの質問を行なうことによって利用者の感性を学習する方法も考えられる [8]。しかし、この方法は利用者個人の感性を理解できる反面、利用者にかかりの負担を強いることとなり、我々の目標であるやわらかい・人にやさしい検索インタフェースとは言い難い。そこで、利用者のステレオタイプ(代表的な集団)をあらかじめ想定し、各ステレオタイプごとに代表的なホームページから受ける感性に関するアンケート調査を行ない、その結果からユーザモデルを学習する方法をとった。この方法では、利用者は検索時に質問に答える必要がなく、あらかじめ構築されたユーザモデルを選択する。これにより、ユーザの感性にあった検索結果が得られる。

我々の提案するユーザモデルを説明する。

まず URL 感性データベースには、図 1 に示すように、タイトル、コメント、感性語属性が各 URL について蓄えられている。感性語属性の値は、ページ評価者が各ページについて、そのページから受ける感性の度合で入力する。例えば、あるホームページが「楽しい」かどうかを 0 から 1 までの値で入力する。

あるページ評価者がすべての URL について感性語属性の値を入力することから、検索されるデータベースはページ評価者の主観的なデータベースであるといえる。ある利用者が、この主観的なデータベースを利用者の感性で検索しようとする、再び感性語属性の値を入力する必要がある。そこで、ページ評価者と利用者の感性の差異を表現するユーザモデルを構築し、このユーザモデルを用いて主観的なデータベースの値を補正する。本稿では、代表的なユーザ層（ステレオタイプ）とページ評価者との感性の差異を表現するモデルを構築する。

ステレオタイプには年齢・性別で分けた次の 5 タイプを定めた。

1. 老年層 (50 才以上の男女),
2. 男性中年層 (30 才以上 50 才未満の男性),
3. 女性中年層 (30 才以上 50 才未満の女性),
4. 男性若年層 (15 才以上 30 才未満の男性),
5. 女性若年層 (15 才以上 30 才未満の女性).

各層における評価値にはばらつきがあると考えられるため、ファジィ集合を用い、図 3 のような三角型のメンバシップ関数を用いて評価値を近似する。図 3 の例は、「とても楽しい」というメンバシップ関数を表現している。横軸はあるホームページの「楽しさ」に対する評価値であり、縦軸は評価値の度合である。つまり、「楽しさ」が \bar{x} という値は度合 1.0 で「とても楽しい」に属し、「楽しさ」が x' という値は度合 0.6 で「とても楽しい」に属する。

図 3 のような三角型のメンバシップ関数は最大値 x_{max} 、最小値 x_{min} 、および平均値 \bar{x} で

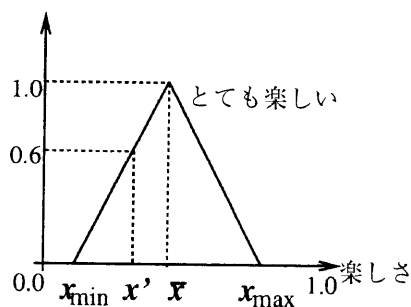


図 3: メンバシップ関数

特徴づけられる。このとき、メンバシップ関数 $\mu(x)$ は、

$$\mu(x) = \begin{cases} x/(\bar{x} - x_{min}) - (x_{min}/(\bar{x} - x_{min})) & (x_{min} \leq x \leq \bar{x}) \\ x/(\bar{x} - x_{max}) - (x_{max}/(\bar{x} - x_{max})) & (\bar{x} < x \leq x_{max}) \end{cases}$$

となる。

感性データベースに対して補正を行う場合は、データのばらつきを考慮する。つまり、感性語属性のメンバシップ関数をページ評価者の評価値と、メンバシップ関数の平均値との差だけ平行移動を行う。具体的には各感性語属性の軸を x_i 、ページ評価者の軸 x_i における感性語属性の値を p_i 、感性語属性のメンバシップ関数を $\mu(x_i)$ とすれば、補正された評価値は三角型のメンバシップ関数

$$\pi(x_i) = \mu(x_i - p_i + \bar{x}_i) \quad (1)$$

となる。2 属性の場合の補正例を図 4 に示す。図に示すように補正は点の移動ではなく、点から平面 (n 属性の場合は n 次元空間) への補正となる。

1 つのステレオタイプに対して 1 パターンの補正では表現できない場合がある。例えば、あるページに対して同じステレオタイプでも「楽しさ」に対する感性の度合が複数に分かれる場合がある。その場合、メンバシップ関数は図 5 のような形になる。つまり複数のメンバシップ関数で表される。図 5 では $\mu_1(x)$ と $\mu_2(x)$ の 2 つのメンバシップ関数からなっている。

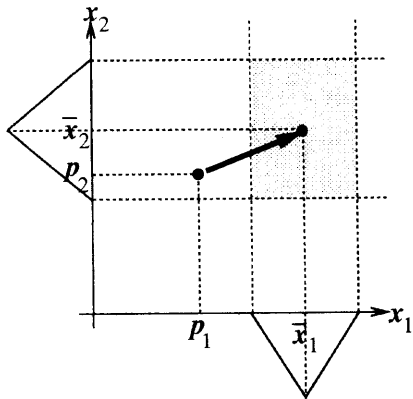


図 4: 評価値の補正

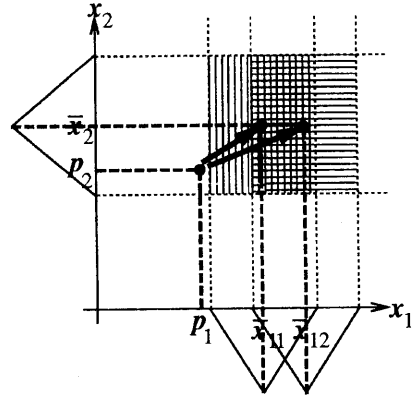


図 6: 複数パターンによる評価値の補正

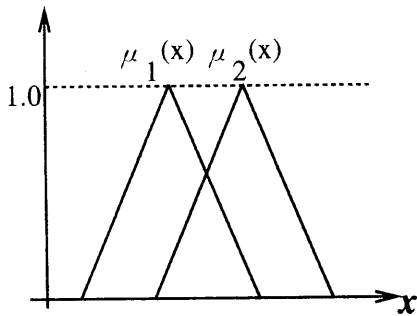


図 5: タイプが分かれた場合のメンバシップ関数

メンバシップ関数が複数存在する場合は、各メンバシップ関数について補正を行なう。それぞれの補正方法はメンバシップ関数が1つの場合と同じである。メンバシップ関数が複数の場合の補正を図に示すと図6のようになる。

以上のことから、ユーザモデルを URL 感性データベースの検索に用いられる感性語属性値の補正用メンバシップ関数集合とみなす。つまり、ユーザモデルを次のように定義する。検索に用いられる感性語属性を x_1, \dots, x_n とするとき、ユーザモデル U は m タイプのパターンに補正する場合 $\{T_1, \dots, T_m\}$ からなり、 $T_j = (\mu_j(x_1), \dots, \mu_j(x_n))$ とする。

2.2 感性 WWW 検索システム KiSS

我々は、前節のユーザモデルを用いた感性 WWW 検索システム KiSS の開発を進めている。KiSS の基本構成を図7に示す。

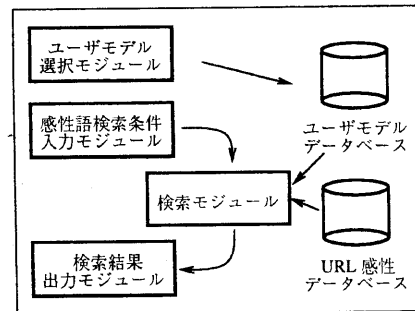


図 7: KiSS のシステム構成図

システムは WWW の URL が蓄えられている URL 感性データベースを持つ。データベース中の各 URL は図1に示したように、タイトルとコメントの属性が付加され、さらに感性語を用いて URL を検索するために感性語属性を持つ。

各ユーザモデルはユーザモデルデータベースに蓄えられている。

検索は次のような手順で行う(図8)。利用者

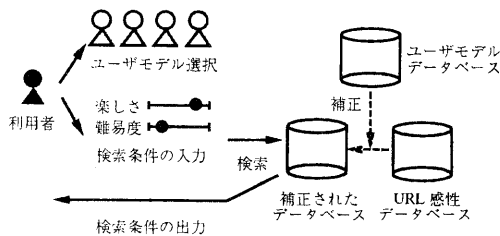


図 8: KiSS の検索手順

はまずユーザモデルを選ぶ。その後、システムの提示する各感性語に対してその度合を入力する。システムは利用者が選んだユーザモデルと各感性語の度合から、URL 感性データベースを検索し、検索結果を適合順に表示する。

URL の検索は次のように行なう。

(step1) 利用者の選んだユーザモデルをユーザモデルデータベースから取り出す。選ばれたユーザモデルを $U = \{T_1, \dots, T_m\}$ とする。

(step2) URL 感性データベースからデータを取り出す。取り出されたデータを $p = (p_1, \dots, p_n)$ とする。ここで、URL 感性データベースに蓄えられている感性語属性の数を n 、データ p の i 番目の感性語属性の値を p_i とする。

(step3) データ d をユーザモデル U の感性におけるデータに補正する。補正は各タイプ T_j ($1 \leq j \leq m$) のメンバシップ関数 ($\mu_j(x_1), \dots, \mu_j(x_n)$) に対して、2 節の式 (1) を用いる。補正後のメンバシップ関数を ($\pi_j(x_1), \dots, \pi_j(x_n)$) とする。

(step4) システムの提示する感性語 x_i ($1 \leq i \leq n$) に対して利用者の入力した度合を $w(x_i)$ とするとき、データ p の各タイプ T_j に対する適合度 $\alpha(p, T_j)$ を計算する。適合度 $\alpha(p, T_j)$ は次の計算式を用いる。

$$\alpha(p, T_j) = \frac{1}{n} \sum_{i=1}^n \pi_j(w(x_i))$$

つまり、補正が行なわれたデータのメンバシップ関数における利用者の入力した度合の値の平均を計算している。よって、適合度が高いほど $\alpha(p, T_j)$ の値は大きくなる。

データ p のユーザモデルに対する適合度 $\alpha(p)$ は各タイプに対する適合度の最大値とする。つまり、

$$\alpha(p) = \max_j \alpha(p, T_j)$$

とする。

(step5) データベース中のすべてのデータに対して (step3), (step4) を行い、適合度を計算する。

(step6) URL、タイトル、コメントを適合度の高い順にシステムで定められた個数だけ表示する。

利用者の感性語度合の入力および検索結果の出力は、HTML (Hyper Text Markup Language) などで記述されたグラフィカルなインターフェースを用いる。また、その場合の検索モジュールは CGI (Common Gateway Interface) を用いる。

3 ユーザモデルの構築

3.1 アンケート調査

我々は、2 節で挙げた年齢・性別の異なる 5 つのグループのうちの 3 にあたる女性中年層 15 人に対して、ホームページから受ける感性に関するアンケート調査を行い、その結果から、ユーザモデルを構築した。その方法について述べる。

まず、ホームページから受ける感性を表現する語句を図 9 に示すようなものとして決めた。図 9 のそれぞれの語句を表現する 20 のホームページを選び、評価対象とした。

国際度	現実性	喜怒哀度
哀楽度	柔軟さ	難易度
派手さ	お色気	芸術性
どきどき度	役立ち度	トクトク度
はやり度		

図 9: ホームページに対する感性語句

アンケート回答者は、20 のホームページそれぞれについて、図 9 の語句に対して評価を行う。評価は 0 から 1 の間の 0.1 刻みの値でつける。互いに相反すると考えられる感性語句の対は、図 10 (a) に示すように一つの軸で、それ以外

の感性語句は図 10 (b) に示すような軸で評価値を決める。また必ずしもすべての感性語句が当てはまるとは限らないので、「どちらでもない」を設けた。

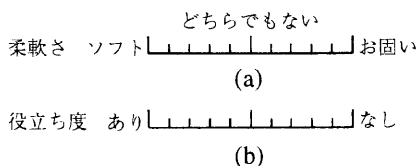


図 10: アンケート様式

ユーザモデル構築のため、次のようにして感性語属性を絞った。(1) アンケート結果から「どちらでもない」の割合が高い感性語属性を対象から外す。この場合、その感性語属性はホームページの表現には適していない。(2) アンケート結果において、ページ評価者と評価値の一致度が高い属性を対象から外す。この場合、ユーザモデルによって感性語属性の値を補正する必要がない。

このようにして、補正を行う感性語属性の絞り込みを行った。その結果、感性語属性は図 11 に示すものとなった。

喜怒哀楽	哀楽度	難易度
どきどき度	役立ち度	

図 11: 絞り込み後の感性語句

感性語句の評価値は、絞り込みを行なった感性語句で表現されるホームページに対する値を集計した。

3.2 クラスタリング

ユーザモデルの構築には、絞り込みを行った各感性語属性に対してアンケートの評価値の平均値、最大値、最小値を求め、メンバシップ関数を決定する方法が考えられる。しかし、アンケートの評価値にばらつきがあった場合、メンバシップ関数の曖昧性が高くなる。その場合、評価値の

類似度が似たもの同士をクラスタに割り当てるようにグループ分けを行なうと、曖昧性が低くなり検索精度が上がると考えられる。

例えば、図 12 のようなアンケート結果になった場合、それぞれの軸に対して 1 つずつのメンバシップ関数を割り当てるよりは、A、B の 2 つのクラスタに分割し、それぞれのクラスタに対してメンバシップ関数を割り当てる方が適切である。

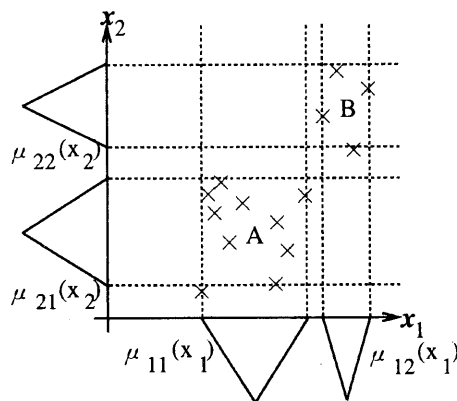


図 12: クラスタの分割

以上のことから、感性語属性間の関連性に着目して、アンケートから得られた評価値のクラスタリングを行う。またクラスタリング結果をメンバシップ関数とするため、ファジィクラスタリングを行なった。ファジィクラスタリングは、サンプル値が各クラスタに対しての帰属度を持つように行うものである。本稿では FCM (Fuzzy C-Means) 法 [1] を用いた。

FCM 法を簡単に説明する。

(step1) 各クラスタに対して、初期ベクトルを与え、クラスタの代表ベクトルとする。

(step2) 各クラスタの代表ベクトルとの距離を求め、代表ベクトルとの距離との逆比をクラスタに対する帰属度とする。つまり距離が近いほど帰属度が高い。これらを初期帰属度と呼ぶ。

(step3) 各クラスタに対して、クラスタの代表ベクトルの再計算を行い新しい代表ベクトルを求める。ここで、新しい代表ベクトルは帰属度

を重みとしたサンプル値の重みつき平均とする。

(step1) 再計算の結果、代表ベクトルが決められたしきい値以上移動した場合は、各サンプル値に対して各クラスとの重みつき距離を求め、その値を新たな帰属度とし (step3) を繰り返す。代表ベクトルが決められたしきい値以上移動しなかった場合は終了する。

初期ベクトルは各軸に対してサンプル値の最大値と最小値から、その存在空間を求め、それらをクラス数で等分割するものを与えた。また、サンプルは3.1節で絞り込みを行なった感性語句に対するホームページへの評価値を対象とした。

クラスタリング法では、クラスタ数を最初に与えるが、その数をいくつに設定するかという問題がある。今回はクラスタリングを行うことにより得られたメンバシップ関数の分散の大きさを見ることにより、クラスタ数を増加させていく方法をとった。

以上のようにして、アンケート結果をクラスタリングすることによって、メンバシップ関数を求め、ユーザモデルの構築を行った。クラスタ数を2として、アンケート結果からクラスタリングを行った結果を図13に示す。ただし、感性語属性空間は5次元となるため、「喜怒哀度」-「役立ち度」平面のものを図13では示した。メンバシップ関数の各軸の値はデータ評価者との相対値を用いている。

3.3 ユーザモデルの評価

中年女性層から5つの追加サンプルを取り、その評価値が構築されたユーザモデルの各タイプに対して、どの程度の割合で帰属しているかを考察する。帰属度が高いタイプが存在すれば、そのサンプルをその帰属度の高いタイプでモデリングできていると考える。逆に、どのタイプにも高い帰属度が出なければ、そのサンプルをモデリングできていないと考えられる。

帰属度の算出式は、FCM法で用いているものと同じものを用いる。つまり、ユーザモデルの各タイプについて代表ベクトルが存在し、その各代表ベクトルと評価値との距離の逆比をそれぞれのタイプに対する帰属度とする。

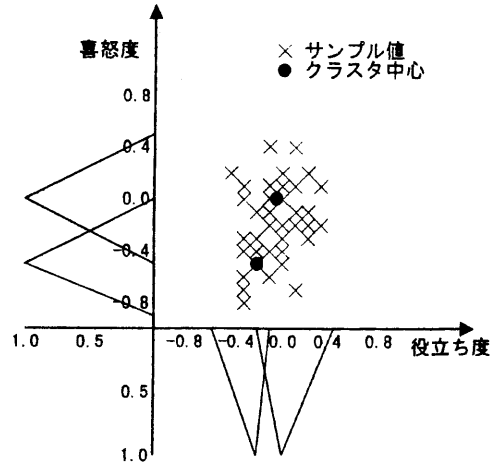


図13: クラスタリング結果

構築されたユーザモデルのタイプと追加サンプル値を図14、サンプルの各タイプへの帰属度を図15に示す。図15において、例えばサンプル1はタイプ(1)には0.78、タイプ(2)には0.22の帰属度でそれぞれ属している。

	1	2	3	4	5
タイプ(1)	-0.03	0.03	-0.07	-0.06	-0.01
タイプ(2)	-0.46	-0.46	0.11	-0.22	-0.19
サンプル1	0.0	0.0	0.0	0.3	0.0
サンプル2	-0.2	0.1	0.3	0.0	0.0
サンプル3	1.0	0.0	0.0	0.0	-0.3
サンプル4	-0.4	-0.4	0.4	-0.3	-0.2
サンプル5	0.0	0.0	-0.2	-0.3	-0.2

1: 喜怒哀度 2: 哀楽度 3: 難易度 4: ときどき度 5: 役立ち度

図14: タイプの代表ベクトルとサンプル値

	タイプ(1)	タイプ(2)
サンプル1	0.78	0.22
サンプル2	0.68	0.32
サンプル3	0.76	0.24
サンプル4	0.30	0.70
サンプル5	0.81	0.19

図15: サンプル値のタイプへの帰属度

図14の代表ベクトルの値から、クラスタリングによって、中年女性層は(1)補正のほとんどいらないタイプと(2)「喜怒哀度」と「哀楽度」に関して補正の必要なタイプに分かれたことがわかる。また、図15の結果から、各サンプルは

いずれかのタイプにほぼ 0.7 以上の帰属度で属することがわかる。

4 あとがき

本稿では、感性表現を用いて WWW を検索するためのユーザモデルを提案した。提案したユーザモデルは、相対感性を表現するメンバシップ関数集合からなる。また、中年女性層に対してアンケート調査を行ない、その結果からユーザモデルを構築する方法を説明した。さらに、構築されたユーザモデルに対して、そのモデルに対する帰属度を用いた評価を行い、ほぼ 0.7 以上の帰属度で中年女性層が属することが言えた。

現在、本稿で提案した考え方の一部に基づいて、感性 WWW 検索システムを試験的に運用している。¹

本稿で述べた感性 WWW 検索システム KiSS では、URL 感性データベースに記憶されている感性語属性のすべてについて、利用者はその度合を入力する。今後の課題として、入力に用いる感性語属性とその度合の組を表すような要約語を用いた入力方法を挙げる。

例えば、利用者は「役立ち度 = 0.9」, 「現実性 = 0.7」などと入力する代わりに、それらを要約した「なるほどページ」を選んで検索を行なえるようにする。

このような検索を可能にするために、感性語を基本感性語と複合感性語に分類したモデルを用意する。基本感性語は、複合感性語に対する寄与度を持ち、その構造は木構造となる。「なるほど」という言葉から受ける感性は人によって異なるため、ユーザモデルを構築し、モデルによって木構造を変える。また、寄与度もユーザモデルでばらつきがあると考え、メンバシップ関数を用いたファジィ処理を行なう。現在、このような方針で入力インタフェースを検討している。

参考文献

- [1] Bezdek J.C.: "Pattern Recognition with Fuzzy Objective Function Algorithms",

¹<http://town.hi-ho.or.jp/kansei/>

Plenum Press (1981).

- [2] 森山, 斎藤, 小沢: "音声における感情表現語と感情表現パラメータの対応付け", 信学技報, SP95-67 (1995-10).
- [3] 西村, 河野, 長谷川: "WWWデータ資源検索システムの実装と評価", 情報処理学会, データベースシステム, 109-44, pp.263-268 (1996-07).
- [4] 野村, 大木, 小高, 横山, 松下: "ファジィ検索のための平均演算子による曖昧属性の生成", 日本ファジィ学会誌, 4, 2, pp.334-360 (1992).
- [5] 奥野, 久米, 芳賀, 吉澤: "多変量解析法", 日科技連 (1981).
- [6] Sugeno M. and Yasukawa T.: "A Fuzzy-Logic-Based Approach to Qualitative Modeling", IEEE Trans. on Fuzzy Systems, 1, 1, pp.7-31 (Feb 1993).
- [7] 鳥居: "意味解析機能を備えたWWW検索システム", 情報処理学会, データベースシステム, 109-46, pp.275-280 (1996-07).
- [8] Wakami N., Naito E., Ozawa J., and Hayashi I.: "Fuzzy Retrieval System Employing Fuzzy Connectives with Learning Functions and Query Networks", Proc. of the International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and the Second International Fuzzy Engineering Symposium, pp.57-64 (March 1995).