

比喩解釈を目的とする World Wide Web を利用した属性値の 適合性判定手法とそのフィードバック

梶井 文人^{†,††a)} 福本 淳一^{†††} 荒木 健治[†]

An Automatic Relevance Estimation of Property Values and Its Feedback
Based on World Wide Web for Metaphor Recognition

Fumito MASUI^{†,††a)}, Jun'ichi FUKUMOTO^{†††}, and Kenji ARAKI[†]

あらまし 本論文では、比喩解釈における評価分析作業の効率化のために、比喩解釈過程でクローズアップされた属性値の適合性判定と属性値集合への判定結果フィードバックを自動的に行う手法を提案する。提案手法は、対象概念とクローズアップ属性値を用いて生成した特定表現について、World Wide Web (WWW) 上の出現状況を調べることによって、クローズアップ属性値の適合性判定を行う。不適合と判定された場合は、WWW から取得した属性値知識に基づいて属性値集合を再ランキングすることでフィードバックを行う。実験の結果、自動判定結果と人間による判定結果の間では、約 80% の一致率が得られ、十分な判定性能が得られることが分かった。また、フィードバックにより、属性値のランキング精度を約 20% 向上させることにも成功した。

キーワード 比喩解釈, WWW, 顕現属性値, 適合性判定, フィードバック

1. ま え が き

比喩とは、ある概念を他の概念によって説明または強調する修辞的手法の一つであり、認知科学、言語学、心理学、人工知能など、様々な分野で研究対象として取り上げられている [1] ~ [5]。自然言語処理の分野においても例外ではなく、処理対象を一般的な文書や会話などへ拡大し、柔軟な処理を行うためには、比喩表現の処理は重要であり、比喩表現を検出するだけでも有益である [6] ~ [8]。例えば、機械翻訳では、比喩を直訳することによる意味的な誤訳回避や、微妙なニュアンスの違いの吸収などに結び付けることが可能であるし、対話処理でも、比喩生成による発話数の削減や柔軟な発話が期待できる [9], [10]。

比喩解釈の過程には、属性値に注目するもの、構造や関係に注目するもの、文脈やイメージスキーマ^(注1)に注目するものがある [11]。後者二つの比喩解釈のためには、概念の構造や関係の対応付け、文脈情報の利用などが必要であり、そのためには、属性値の把握がやはり重要である。そこで本論文では、比喩解釈の基本的な処理を実装することを念頭に置き、属性値に基づく比喩（属性比喩）の解釈を扱う [12]。

属性比喩を扱った研究では、概念を属性値の集合で表し、顕現性について定量的計算を行う研究が主流である [13] ~ [15]。岩山らは、プロトタイプ概念を表す属性値集合を確率的概念記述で表現する計算モデルを提案している [13]。Utsumi らや今井らは、認知心理学実験によって属性値を評定し、そこで得た評定値を用いて属性値集合を表現している [16], [17]。また、Martin による、実際の文章に現れる比喩表現を扱う研究や、梶井らによる、大規模なテキストを利用して自動構築した属性値集合を使った比喩解釈の研究もある [8], [18]。

(注1): Lakoff らによると、イメージスキーマとは、人間の身体的経験に基づいた概念組織化の中心となる概念の対応関係である [1]。例えば、「喜びは上、悲しみは下」、「未来は前、過去は後ろ」などがある。

[†] 北海道大学大学院情報科学研究科, 札幌市
Graduate School of Information Science and Technology,
Hokkaido University, Sapporo-shi, 060-8628 Japan

^{††} 三重大学工学部情報工学科, 津市
Faculty of Engineering, Mie University, 1515 Kamihama-cho,
Tsu-shi, 514-8507 Japan

^{†††} 立命館大学情報理工学部メディア情報学科, 草津市
Department of Media Technology, Ritsumeikan University,
Kusatsu-shi, 525-8577 Japan

a) E-mail: masui@ai.info.mie-u.ac.jp

属性比喩の解釈過程は、比喩表現において強調される特徴（顕現属性値）が、例える概念（source 概念）から例えられる概念（target 概念）へ重ね合わせられ、クローズアップされることによる、と説明できる [6], [8], [13], [14], [17]。例えば、「砂のような雪」という直喩の解釈で「細かい」や「さらさらした」という特徴がイメージされる過程は、source 概念「砂」の顕現属性値「細かい」が、target 概念「雪」で強調される属性値としてクローズアップされて解釈されると説明できる。よって、source 概念の顕現属性値ではない「新しい」はクローズアップされないし、クローズアップされても理解不能となる。このように、比喩解釈にとって属性値のクローズアップは非常に重要であり、概念間で正しい属性値がクローズアップされなければ、比喩解釈に悪影響を与える。

属性値のクローズアップ失敗の主な要因として、概念を表す属性値集合（のランキング）のひずみがある。例えば、今井らは、認知心理学実験の評定値を用いて、概念間でクローズアップされる顕現属性値の抽出を試みているが、評定値分布がばらつく場合の抽出精度が問題であると述べている [6], [17]。梶井らは、提案した比喩性検出手法の考察において、計算上比喩性検出には成功していても、人間の直観と異なる属性値がクローズアップされる問題があると述べている [8]。

この問題は、特に、属性値集合を自動構築する場合に顕著となる。梶井らの手法では、連体修飾関係をもつ語の共起確率が、属性値集合の生起確率分布とよく一致していることが前提となっている。この点についてももう少し詳しく吟味すると、次のようになる。連体修飾（装定）の関係にある修飾語と名詞は、主述の関係に置換可能 [19], [20] なので、連体修飾関係（形容詞・形容動詞・名詞）による語の共起関係を、概念-属性値の関係とみなして抽出する手法自体は妥当といえる。

一方、山梨によると、「連体修飾によって特徴づけられる装体表現は、一回事象的で、かつ対象指向的な機能をもつ表現 [21]」であり、必ずしも典型的特徴（顕現属性値）を表すものではない。例えば、「赤いリンゴ」とは、概念「リンゴ」の属性値ではなく、今対象としている 1 個のインスタンスとしての「そのリンゴ」の属性値を示している。ただし、概念の顕現属性値は、インスタンスの属性値としても繰り返し表現されると考えられるので、特定の名詞に対して高頻度で出現する連体修飾語は、概念の顕現属性値である可能性が高

いと推測できる。この場合も、「赤い」は、概念「リンゴ」の顕現属性値と一致している。ただし、高頻度で出現する修飾語には、「大きいリンゴ」のような表現もある。この場合は、「そのリンゴ」の属性値を示しており、概念「リンゴ」の顕現属性値とは言いがたい。よって、連体修飾語については、高頻度であるものが必ず顕現属性値であるとはいえないのである。

よって、共起頻度に基づくゆう度のみでは顕現属性値としての保証は得られず、属性値ランキングにひずみが生じる可能性がある。

上記のような、誤った比喩解釈を回避するためには、強調されてクローズアップされる属性値（以降、クローズアップ属性値）の選択誤りを発見する必要がある。誤った属性値については、属性値集合内のランキングを補正することによって、解釈可能な範囲を拡大することができる。しかしながら、これらの作業を人手で実施するにはコストが高く、大規模な評価実験を実施することは困難である。

そこで本論文では、比喩解釈においてクローズアップ属性値の適合性判定及び判定結果のフィードバック手法を提案する。WWW を知識源とし、対象概念とクローズアップ属性値による特定表現の生成能力を検証することによって属性値の適合性判定を行う。判定結果に応じて、WWW から得られる頻度情報に基づいて属性値集合を再構成し、フィードバックを行う。適合性判定では、定型パターンを利用した知識獲得の枠組みを利用する。定型パターンを利用して知識を獲得しようとする試みはこれまでも報告されており、有効性はある程度確認されている [22] ~ [26]^[注2]。知識源として WWW を用いる理由は 2.3 で詳しく述べるが、より人間判断に近い適合性判定を実現するための知識源として考えた場合、新聞記事新聞記事などの文書よりも有効であると判断したためである。

以下、2. で提案手法の基本的考え方を述べ、3. で提案手法の詳細について説明する。4. で提案手法の妥当性検証のための実験を行い、その結果を考察する。

2. 基本的な考え方

本章では、提案手法の基本的な考え方について述べる。

(注2): 例えば、Hearst は、「A such as B」という定型パターンを用いてテキストから共起単語を収集し、その知識獲得能力を確認している [22], [23] し、安藤らは、複数の定型パターンを利用して得られる名詞の上位下位関係について考察している [26]。

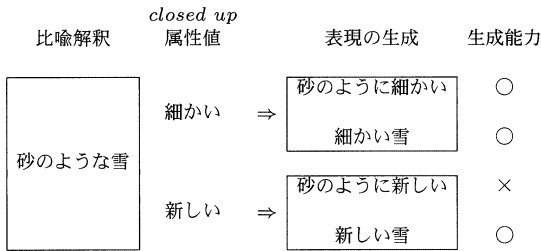


図 1 クローズアップ属性値と特定表現の関係

Fig. 1 Relationships between closed-up property values and the specific expressions.

2.1 クローズアップ属性値の適合性判定

本論文でいう「正しいクローズアップ属性値」とは、source 概念の典型性を表す顕現属性値から選択される。よって、source 概念についての適合性判定としては、選択された属性値が、概念の顕現属性値であるかどうかを検証すればよい。source 概念とその顕現属性値を用いれば、「砂のような細かい」や「ハチのように忙しい」のように、比喩または例示を意味する比較表現^(注3)が生成可能と考えられる。このことから、source 概念と、クローズアップ属性値を用いて比較表現を生成し、その比較表現が一般に利用されているかどうかを調べれば、属性値の適合性を判定できる(図 1 上段)。このとき、比較表現を構成する「~のような」、「~のように」といった定型パターンを指標表現(*ground*)と呼ぶ。

target 概念について属性値の適合性を考えると、比喩解釈においてクローズアップされる属性値は、概念の顕現属性値である必要はないので、target 概念における属性値の適合性については、概念の属性値として適しているかどうかを判定すればよい。また、連体修飾(装定)関係の機能を考慮すると、target 概念と属性値を組み合わせて典型的または一回事象的な連体修飾表現が生成できるはずである。このことから、target 概念とクローズアップ属性値を用いて連体修飾表現を生成し、その連体修飾表現が一般に利用される表現であるかどうかを調べれば、属性値の適合性を判定できる(図 1 下段)。

2.2 判定結果の属性値集合への反映

クローズアップ属性値が不適合と判定された場合、その属性値集合のランキングにははずみが生じていると考えられる。ここで、判定結果を反映して、対象属性値集合のランキングを補正すれば、同じ概念対についての属性値のクローズアップも補正することになり、

比喩解釈の性能向上に貢献できる。

source 概念については、不適合となった属性値は、顕現属性値ではないとみなせる。よって、属性値集合を該当属性値のゆう度を下げないように再構成すれば、ランキングを補正できる。target 概念については、該当属性値はノイズであり、属性値として適さないとみなせる。よって、該当属性値を概念の特徴を示す属性値集合から削除すれば、ランキングを補正できる。

以上の操作を繰り返していくと、各概念の属性値ランキングのはずみは徐々に補正されるはずである。

2.3 知識源としての WWW の利用

適合性判定には、比較表現の一般的な利用状況を把握できるだけの網羅性をもった知識源が必要となる。田添らの調査によると、新聞記事 1 カ月分のコーパスから、指標表現「のような」を用いた比較表現が 500 組得られ、そのうち名詞句同士を比較したものは、311 組である[28]。この場合、10 年分の新聞記事から得られる比較表現は 37,000 組余りにしかならない。指標表現「のように」についても同様の調査を行ったところ、10 年分の新聞記事から得られる比較表現は 20,000 組余りにとどまることが分かった。これらのことから、比較表現を用いて適合性判定を行うためには、新聞記事 10 年分程度では不十分であり、更に大規模な知識源が必要であるという結論が得られる。

関口らの報告によると、WWW データは、質に留意することで新聞記事よりも網羅性が高く、かつ分野依存性が低い知識源として利用できる[29]。本論文では、知識源の利用は、単純な定型表現の文字列照合のみに限定されるため、WWW を知識源として利用しても、質の低下等の影響は小さいと考えられる。よって本論文では、適合性判定に適した知識源として WWW を利用する。

3. 提案手法

本章では、提案手法の詳細について説明する。提案手法の概要を図 2 に、処理例を図 3 に示す。提案手法は、「クローズアップ属性値の適合性判定」と「フィードバック」の二つの処理過程からなる。以下、二つの過程ごとに詳述する。

3.1 属性値の適合性判定

適合性判定は、比喩性検出結果からの (1) 要素の取

(注3): 例えば、「ハチのように忙しい」という表現は、“as busy as a bee”という英語の比較表現に対応し、意味的な比較構造と見ることができるので、日本語においても比較表現という呼称を用いる[27]。

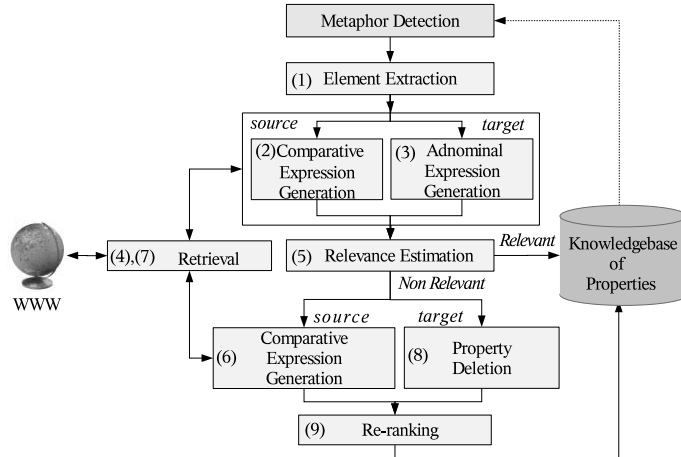


図 2 クローズアップ属性値の適合性判定及びフィードバックの概要
 Fig.2 Overview of the relevance estimation of closed-up property and its feedback.

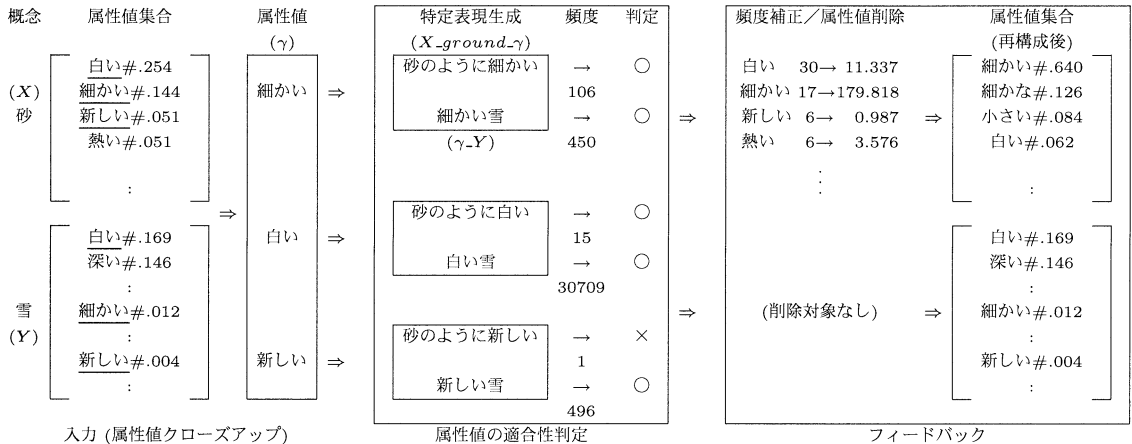


図 3 クローズアップ属性値の適合性判定及びフィードバックの実例
 Fig.3 An example of the relevance estimation and its feedback.

得, (2) 比較表現生成, (3) 連体修飾表現の生成, (4) 表現検索, (5) 適合性判定, からなる.

(1) では, 入力された情報 (比喩性検出結果) から, source 概念, target 概念, クローズアップ属性値を取り出す.

(2) は, source 概念についての操作である. まず, 「 X_ground_{γ} 」というフレーム (定型パターン) を用意しておく. このとき, $ground$ は指標表現であり, フレームの X には source 概念を, γ にはクローズアップ属性値を当てはめて比較表現を生成する.

ところで, 指標表現 ($ground$) として, 中村の調査による「比喩指標実現形」35 種類 [30] のうち, 上で述べた比較表現を生成できるもの 9 種類^{注4)}につい

て WWW 中の頻度数を調べたところ, 上位 2 組の指標表現「のように」, 「のような」だけで全表現の頻度の 70% 以上を占めた. この結果から, 本論文では, 用意するフレームとして, 「 X_{γ} 」, 「 X_{γ} 」の 2 種類のみを用いればよいと考える.

(3) は, target 概念についての操作である. クローズアップ属性値 γ と target 概念 Y を 「 γ_Y 」というパターンで組み合わせ, 連体修飾表現を生成する.

(4) では, 生成した比較表現, または連体修飾表現をクエリとして WWW 検索を行う. 検索の結果得られた件数を, 生成表現の頻度とみなして取得する.

(注4): 上位から順に, 「のように」, 「のような」, 「みたいな」, 「みたいに」, 「に似て」, 「に似た」, 「のごとく」, 「のごとき」, 「じみた」.

(5) では適合性判定を行う。表現検索で得られた頻度がしきい値 α 以上であれば、比較表現は生成可能であり、属性値は適合と判定される。頻度が α 未満であれば、比較表現は生成不能であり、属性値は不適合と判定される。このとき、しきい値 α は、対象となっている属性値がクローズアップ属性値として適合が否かを判断するための基準である。 α を大きくすると典型性重視となり、小さくすると網羅性重視となる。

[実行例]

図 3 では、「砂、雪」という概念対が処理対象であり、「砂」が source 概念、「雪」が target 概念である。適合性判定のしきい値は、 $\alpha = 15$ として考える。概念対に対する比喩解釈過程によって、属性値「細かい」、 「白い」、 「新しい」がクローズアップされている。

source 概念については、まず、フレーム「 X_{γ} 」のような γ 上の X に「砂」、 γ に「細かい」「白い」「新しい」を適用して比較表現「砂のように細かい」、「砂のように白い」「砂のように新しい」を生成する。これらの比較表現を WWW 検索し、得た検索件数は、「砂のように細かい」が 106、「砂のように白い」が 15、「砂のように新しい」が 1 である。その結果、しきい値以上の頻度が得られた属性値「細かい」、「白い」は適合、「新しい」は不適合と判定される。

target 概念については、まず、「 γ_Y 」の Y に「砂」、 γ に「細かい」、「白い」、「新しい」を適用して「細かい雪」、「白い雪」、「新しい雪」という連体修飾表現を生成する。これらの連体修飾表現を WWW 検索し、得た検索件数は、「細かい雪」が 450、「白い雪」が 30,709、「新しい雪」が 496 である。この結果、すべての属性値が適合と判定される。

3.2 判定結果のフィードバック

フィードバックは、(6) 比較表現生成、(7) 表現検索、(8) 不適合属性値の削除、(9) 再ランキング、からなる。

source 概念について、(6) は、3.1 の (2) と同じ、(7) は、3.1 の (3) と同じ処理過程であり、概念と属性値をフレームに適用して比較表現を生成し、その頻度を取得する。ただし、(6)、(7) の過程を対象とする属性値集合のすべての属性値に対して適用する。これにより、WWW における各属性値の頻度が取得できる。

(8) は、target 概念についての操作で、属性値が不適合であった場合、該当する属性値をそれが属する属性値集合から削除する。

(9) で、属性値集合の再ランキングを行う。source 概念については、新たに WWW から取得した頻度情

報を用いて、属性値の生起確率を再計算する。target 概念については、クローズアップ属性値削除後の属性値集合に対して生起確率を再計算する^(注5)。

[属性値集合再構成のための生起確率の計算]

属性値集合の再ランキングにおける、属性値の生起確率計算について説明する。

概念 X に関して、属性値を指す用言を w_i とすると、再構成の結果、 w_i の X に対する生起確率は、補正頻度 (またはゆう度) $F(w_i|X)$ によって計算される。概念 X について、 w_i が、属性値集合中の属性値としてどのくらい容易に選択されるかを示す指標 $pf(w_i|X)$ と、 w_i を属性値とする概念が X 以外にどの程度存在するか、という指標 $icf(w_i)$ の積として表し、式 (1) で求められる。

$$F(w_i|X) = pf(w_i|X) * icf(w_i) \quad (1)$$

このとき、 w_i に関して、もとの知識ベース K 中に保存されている頻度 (または評定値) を $f(w_i|X, K)$ と表し、2. で述べた、「 $X_{ground-\gamma}$ 」^(注6)を用いて WWW から得られる頻度を $f(w_i|X, W)$ と表す。

$f(w_i|X, W)$ が、適合性判定のしきい値 α 以上 ($f(w_i|X, W) \geq \alpha$) であった場合、 w_i は顕現属性値であると判断される。このとき、指標 $pf(w_i|X)$ は、WWW から得た頻度情報を用いた補正值で表し、式 (2) で求められる。

$$pf(w_i|X) = \frac{f(w_i|X, W)}{\delta(X)} \quad (2)$$

ここで、 $\delta(X)$ は、各概念ごとに決まるパラメータであり、顕現属性値中の最小頻度に対する比で表し、式 (3) で求められる。

$$\delta(X) = \min(f(w_1|X, W), \dots, f(w_n|X, W)) \geq \alpha \quad (3)$$

上の条件に適合しなかった場合、 w_i は一般属性値とみなされる。このとき、指標 $pf(w_i|X)$ は、もとの属性値集合の生起確率を用いた補正值で表し、式 (4) で求められる。

(注5): 概念の属性値にゆう度やランキングの指標に相当する数値を、人間から得る方法として、Osgood ら [31] の SD (Semantic Difference) 法がある。SD 法では、被験者による評定結果をもとに属性値のゆう度を求める。提案手法のフィードバック処理では、WWW 中の定型パターンの頻度を複数の人間による評定結果とみなして属性値のゆう度を求めていると考えることもできる。

(注6): 具体的には、「 X_{γ} 」および「 X_{γ} 」。

$$pf(w_i|X) = \frac{f(w_i|X, K)}{\delta(X)} \quad (4)$$

$\delta(X)$ は、もとの属性値中の最大頻度に対する比として表し、式 (5) で求められる。

$$\delta(X) = \max(f(w_1|X, K), \dots, f(w_n|X, K)) \quad (5)$$

$icf(w_i)$ は、知識ベースに記録された全概念数 N と、 w_i が属する概念数 $cf(w_i)$ の関係を正規化した値とし、式 (6) により求める。

$$icf(w_i) = \log \frac{N}{cf(w_i)} + 1 \quad (6)$$

[実行例]

図 3 では、source 概念について「新しい」が不適合判定であるので、source 概念についてフィードバックを行う。source 概念「砂」については、適合性判定と同様に、フレーム「 X_{γ} 」の X に「砂」、 γ に属性値集合中のすべての属性値を適用して比較表現を生成し、生成した比較表現ごとに、WWW から頻度を取得する。

もとの知識ベースを 10 年分の新聞記事から構築すると、148,733 組 (N) の概念が得られ、概念「砂」の属性値集合は、図 3 のようである。各属性値の実際の頻度は、「 $f(\text{細かい} | \text{砂}, K) = 115$, $f(\text{白い} | \text{砂}, K) = 15$, $f(\text{新しい} | \text{砂}, K) = 6$, $f(\text{熱い} | \text{砂}, K) = 6$, ...」である。ここで、「 X_{γ} 」及び「 X_{γ} 」の X , γ に「砂」と各属性値を適用して比較表現を生成し、WWW から頻度情報を取得すると、「 $f(\text{細かい} | \text{砂}, W) = 106$, $f(\text{白い} | \text{砂}, W) = 15$, $f(\text{新しい} | \text{砂}, W) = 1$, $f(\text{熱い} | \text{砂}, W) = 6$, ...」が得られる。適合性判定のしきい値を $\alpha = 15$ とすると、しきい値を満たす属性値である「白い」及び「細かい」(このとき $cf(\text{白い}) = 115$, $cf(\text{細かい}) = 44$) である) の補正頻度の計算には、式 (2), (3), (6) が適用され、以下のようになる。

$$F(\text{白い} | \text{砂}) = \frac{15}{15} \times \left(\log \frac{148733}{115} + 1 \right) = 11.337$$

$$F(\text{細かい} | \text{砂}) = \frac{106}{15} \times \left(\log \frac{148733}{44} + 1 \right) = 179.818$$

一方、しきい値未満となる属性値「新しい」及び「熱い」(このとき $cf(\text{新しい}) = 9734$, $cf(\text{熱い}) = 620$) である) の補正頻度の計算には、式 (4), (5), (6) が適用され、以下のようになる。

$$F(\text{新しい} | \text{砂}) = \frac{6}{30} \times \left(\log \frac{148733}{9734} + 1 \right) = 0.987$$

$$F(\text{熱い} | \text{砂}) = \frac{6}{30} \times \left(\log \frac{148733}{620} + 1 \right) = 3.563$$

補正によって、属性値「細かい」の頻度が高くなり、また、「新しい」「熱い」はより低い順位となる。最初にランキング 1 位にあった「白い」の順位も相対的に低下する。

以上の補正を属性値すべてに適用し、生起確率を求めると、図 3 右の属性値集合(再構成後)のような属性値ランキングが得られる。

4. 評価と考察

提案手法の効果を検証するために、一連の実験と評価を行った。以下、4.1 で、適合性判定手法が、人間判断の代替となり得るかどうかを調べるために、適合性判定能力を評価する。4.2 で、フィードバックが属性値ランキングのひずみをどの程度補正できるのかを調べるために、フィードバックによる属性値再ランキング結果を評価する。

4.1 クローズアップ属性値の適合性判定の性能評価

まず、提案する属性値の適合性判定において、顕現属性値を決定するしきい値 α の最適値を調べるための予備実験を行う。知識源として、新聞記事 10 年分^(注7)を用いて、基本知識ベースを構築した^(注8)。ベースライン(BL)として、榊井らの比喩性検出手法[8]を用いた。適合性判定における WWW 上のクエリ検索には、WWW 検索エンジン^(注9)を利用した。

実験に用いる概念対には、榊井らの実験で用いた概念対 70 組[8]を用いた。まず、概念対 70 組について、BL による比喩性検出処理を行い、クローズアップ属性値を得、次に、クローズアップ属性値に対して、提案手法による適合性判定と、人手による適合性判定を行った。適合性判定のしきい値 α を、 $\alpha = 1, 5, 10, 15, 20, 25, 30$ に変更し、7 種類の判定結

(注7): 毎日新聞 CD-ROM 版 1994~2003 年、約 111 万記事(約 1,046 万文)を用いた。

(注8): WWW から直接知識を抽出して基本知識ベース(属性値集合)を構築することも考えられる。しかし、WWW 文書は新聞記事と比較して表記や文体の自由度が非常に高く、精度の良い知識抽出を実現するためには、形態素解析や係り受け解析の解析誤りの影響を回避する手段が必須となり、かえって高コストであるため、基本知識ベース構築については新聞記事を用いた。また、たとえ WWW から知識ベースを構築したとしても、1. で述べたように、属性値ランキングの揺れは解消できないため、適合性判定とフィードバックは必要である。

(注9): Yahoo! Search を利用した。

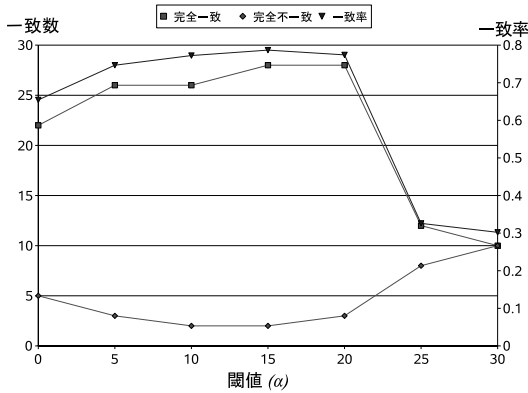


図 4 適合性判定のしきい値の変化とクローズアップ属性値の適合性判定の性能

Fig. 4 Thresholds for the relevance estimation and performance of the estimating closed-up property values.

果を得た。

提案した適合性判定（以下，適合性判定）の結果と，人手による適合性判定（以下，人手判定）の結果を比較し，一致率を求めた。

[実験結果]

実験の結果を図 4 に示す．図には，両判定結果の一致率と適合性判定のしきい値 α の関係（ ），適合性判定結果が人手判定結果と完全に一致した数（ ），全く一致しなかった数（ ）が示されている．

両手判定結果の一致率の推移を見ると，しきい値 $\alpha = 15$ のときに，0.79 と，非常に高い値が得られた． $\alpha \leq 20$ においては，0.65~0.79 と，おおむねよい一致を見せており，本手法が有効であることを示している．一方， $\alpha \geq 25$ になると一致率は 0.35 未満と急激に低下した．

[考察]

適合性判定結果と人手判定結果の一致率が $\alpha \geq 25$ で急激に低下する理由について考察する．一致率の低下傾向は，適合性判定結果と人手判定結果が完全に一致した数の推移と非常に類似しており， $\alpha \geq 25$ の条件で急激に減少している．しかし，全く一致しなかった数の推移はこれと相反する傾向を示している．このことから， $\alpha \geq 25$ の条件では，過剰判定によって有効な属性値が不適合と評価されたため，一致率が低下したと考えられる．

表 1 に，適合性判定結果 ($\alpha = 15$) と人手判定結果の例を示す．左から，対象概念対，BL によってクローズアップ属性値，適合性判定結果，人手判定結果，

表 1 クローズアップ属性値の適合性判定結果の例
Table 1 Some examples of results from the relevance estimation of closed-up property values.

概念対	BL によるクローズアップ属性値	適合性判定 ($\alpha = 15$)	人手判定	一致率
人形，表情	美しい			1.000
	可愛い			
	かわいい			
	いい	×	×	
	不思議な	×	×	
	ユーモラスな	×	×	
	素朴な	×	×	
	グロテスクな	×	×	
神様，人	新しい	×	×	.500
	新しい	×	×	
子供，心	いい		×	.500
	やすい	×	×	
	悪い	×	×	
	健康な	×		
	高い		×	
	弱い		×	
夢，プロジェクト	不自由な	×	×	.250
	幼い			
	大きい	×		

両判定結果の一致率を示している．「人形，表情」の概念対に対しては，「美しい」「可愛い」という典型的特徴を示す属性値が適合と判断され，「不思議な」「グロテスクな」「新しい」などの一回記述的な特徴となり得る属性値は不適合と判断されており，人間の判断に近い結果が得られていることが分かる．

一方，人手判断と一致しなかった結果のほとんどは，以下のようなケースであった．例えば，「神様，人」では，「偉い」という属性値がクローズアップされており，人手判定では適合とされている．適合性判定でも，WWW で実際に「神様のように偉い」と記述されているケースがあるにもかかわらず，取得した頻度がしきい値より小さかったため，不適合となってしまった．また，「枝，柱」では，「太い」という属性値がクローズアップされており，人手判定では不適合であるが，適合性判定では適合となっている．これは，比較表現「枝のように太い」の WWW 検索において，「...，木の枝のように太い根から細い根が枝分かれしています。」や，「緑化樹剪定枝のように太い枝が入っていたり，...」のように，意味の異なる表現の一部にも一致してしまい，取得される頻度が実際より大きな値になることが主な原因となった．

これらの問題については、適合性判定の基準として静的なしきい値によって一元的に区別してしまう点に工夫の余地がある。検索された文を統語構造解析し、統語範囲を把握するなどの対応が考えられる。

一致率が 0.5 を下回ったケースが 4 組あったが、すべて「夢」を source 概念とする概念対であった。これは、田添らの分析にあるように、「夢」という概念そのものが比喩を生成する機能をもち、顕現属性値を明示した比較表現が記述され難いためと考えられる [28]。

4.2 フィードバックの評価

4.1 の実験で用いた概念対の各概念から、抽象度の高い概念を取り除いた 66 組と、今井らが心理学実験（連想実験及び SD 法）により構成した概念 10 組 [17] を合わせた計 76 組の概念を用意した。これらの概念に対して、全属性値に対する再構成を行い、フィードバックが完全に適用された知識ベースを擬似的に構築した。属性値集合の再構成では、4.1 の実験結果から、適合性判定のしきい値を $\alpha = 15$ とし、補正処理を行わないもの（*type I*）と、人手で作成した不要語リスト^(注10)により不要語を除いたもの（*type II*）を試した。

属性値を生起確率順に上位 5 位までを取り出し、それらが顕現属性値であるかどうかを人手で評価した。この結果を用いて、各モデルの MRR (Mean Reciprocal Rank) を計算して比較した [32]。MRR は、ある基準によって順位付けされた情報に対し、順位の逆数をスコアとして付与し、それらの平均値として求めることができる。MRR により、フィードバックによる再ランキング性能が把握できる。

次に、フィードバック結果と、前述の、今井らによる心理学実験の結果を比較した。今井らによる心理学実験に基づく 10 組の概念に関する顕現属性値を正解データとみなし、提案手法による顕現属性値の *F* 値を計算し、平均 *F* 値を求めた [32]。属性値集合内の異表記は、1 個の属性値として数え、評価において、意味的に同じ属性値は同じ属性値とみなした。平均 *F* 値を利用することによって、顕現属性値抽出の網羅性と精度を把握できる。

[実験結果]

MRR の比較結果を表 2 に、各ランクと正解数の関係を図 5 に示す。MRR スコアは、*type I* が 0.727、*type II* が 0.757 であり、ベースライン (BL) のスコア (0.562) に対して 16~19% の向上が見られた。図 5 は、顕現属性値をどの順位で示せたかを数えた結果で

ある。これによると、*type I*、*type II* とともに、ベースラインと比較して、明らかに 1 位での正解数が多く、顕現属性値の抽出失敗数が減っている。また、*type II* は、1 位数、抽出失敗数とも最良の結果であった。

1 位にランキングされた属性値について、WWW からの頻度情報の取得と反映状況及び顕現属性値抽出性能についてまとめたものを表 3 に、WWW による検証を行った知識とコーパスから抽出した知識の例（属性値頻度順に上位 5 位まで）を表 4 に示す。76 組のうち、正しく顕現属性値を取り出せたものは全体の 69% であった。また、WWW から頻度情報を取得して反映できたものは全体の 46%、WWW から頻度情報は取得したがしきい値に満たないために反映できな

表 2 MRR 値による属性値集合のランキング精度の比較
Table 2 MRR values of property value rankings for *type I*, *type II* and BL.

	<i>type I</i>	<i>type II</i>	BL
MRR	.727	.757	.562

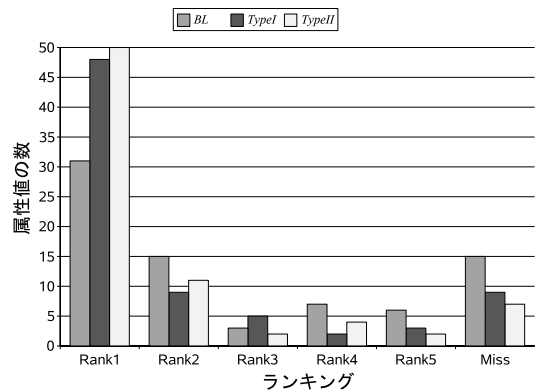


図 5 上位 5 位までの正解属性値数の比較

Fig. 5 Numbers of in top 5th ranked correct property values in results by *type I*, *type II*, and BL.

表 3 フィードバック結果におけるランキング 1 位属性値の抽出性能

Table 3 Performance of top ranked properties in the feedback results.

	(A)	(B)	(C)	計
正しく抽出	45%	21%	3%	69%
誤って抽出	1%	21%	9%	31%
計	46%	42%	12%	100%

- (A) : WWW から頻度を取得して反映
- (B) : WWW から頻度を取得したが反映せず
- (C) : WWW から頻度を取得できず

(注10) : 属性値として無意味な語、例えば、「さまざま」や「新たな」などを登録した。

かったものは 42%，WWW から頻度情報が取得できなかったものは 12%であった。誤った抽出結果のうち、ほとんどは、WWW からの頻度情報を反映できなかった場合であり、WWW からの頻度情報を反映して誤った

場合はわずか 1%であった。このことから、WWW からの頻度情報をフィードバックできれば、高精度に顕現属性値の判断が可能となることが分かる。更に、表 2、図 5 の結果から、フィードバックが、属性値ランキングのひずみを補正できることが確認できる。また、*type II* は、1 位数、抽出失敗数とも最良の結果であったことから、不要語処理を併用することがより効果的である。

表 4 再ランキング結果の例(上位 5 位)

Table 4 Some examples of re-ranked property values by feedback (in the top 5th).

	属性値	BL	属性値	type I	type II
氷	厚い	.147	冷たい	.957	.944
	薄い	.059	鋭い	.008	.008
	白い	.059	固い	.008	.007
	たかき	.049	硬い	.004	.007
	新しい	.039	美しい	.004	.004
炎	急性すい	.148	赤い	.432	.432
	赤い	.144	熱い	.288	.288
	真っ赤な	.120	激しい	.137	.137
	激しい	.076	真っ赤な	.066	.066
	慢性すい	.068	美しい	.025	.025
冷蔵庫	古い	.206	古い	.056	.067
	不思議な	.147	不思議な	.056	.067
	新しい	.118	新しい	.056	.067
	巨大な	.088	巨大な	.056	.067
	白い	.059	白い	.056	.067

：顕現属性値と判断できる最上位の属性値

今井らが心理学実験(連想実験及び SD 法)によって抽出した顕現属性値と、提案手法によって抽出した顕現属性値の比較結果を表 5 に示す。今井らの実験は人間判断に基づくので、これを正解データ集合と仮定し、提案手法の適合率、再現率、*F* 値を計算している。下線部は、両手法間で一致して抽出した顕現属性値を示す。括弧は、提案手法の結果で異表記とみなして数えなかったものを示す。

比較の結果、適合率平均は 0.440、再現率平均は 0.362、平均 *F* 値は 0.330 であった。フィードバック処理が人間の判断を代替するためには、もう少し改良の余地が残る数値である。

表 5 人手による抽出結果を正解としたときの提案手法の顕現属性値抽出性能

Table 5 Performance of the typical property value extraction against human handring results as correct answer sets.

概念	抽出された顕現属性値		適合率	再現率	<i>F</i> 値
	人手による抽出	提案手法による抽出			
部屋	<u>広い</u> , <u>狭い</u> , <u>汚い</u> , <u>きれいな</u> , <u>暖かい</u>	<u>狭い</u> , <u>広い</u>	1.000	.400	.571
番犬	恐い, <u>うるさい</u> , <u>噛む</u> , <u>吠える</u> , <u>強い</u> , <u>大きい</u>	白い	.000	.000	.000
風船	ふわふわした, <u>軽い</u> , <u>色とりどり</u> , <u>割れる</u> , <u>飛ぶ</u> , <u>赤い</u> , <u>子供っぽい</u> , <u>やわらかい</u>	<u>軽い</u> , <u>赤い</u>	1.000	.250	.400
チーター	速い, <u>スマート</u> , <u>美しい</u> , <u>鋭い</u>	<u>優美な</u> , <u>珍しい</u> , <u>強い</u>	.333	.250	.286
湖	<u>静かな</u> , <u>きれいな</u> , <u>深い</u> , <u>冷たい</u> , <u>広い</u>	<u>大きい</u> , <u>静かな</u> , <u>深い</u> , <u>青い</u> , <u>広い</u> , <u>美しい</u>	.667	.800	.727
鏡	映る, <u>丸い</u> , <u>反射する</u> , <u>割れる</u> , <u>神秘的</u> , <u>光る</u>	<u>美しい</u> , <u>青い</u> , (<u>きれいな</u>), <u>丸い</u>	.333	.167	.222
冷蔵庫	<u>冷たい</u> , <u>大きい</u> , <u>四角い</u> , <u>白い</u> , <u>保存する</u> , <u>重い</u>	<u>重い</u> , <u>恐い</u> , <u>無害な</u> , <u>不思議な</u> , <u>最低限必要な</u> , (<u>巨大な</u>), <u>鈍い</u> , <u>大きい</u> , <u>古い</u> , <u>必要な</u> , <u>白い</u> , <u>小さい</u> , <u>廉価な</u> , <u>あいまいな</u> , <u>粗末な</u>	.214	.500	.300
鬼	恐い, <u>強い</u> , <u>赤い</u> , <u>でかい</u> , <u>悪い</u> , <u>角がある</u>	<u>怖い</u> , (<u>恐ろしい</u>), <u>悪い</u> , <u>優しい</u> , <u>赤い</u>	.750	.500	.600
風	<u>強い</u> , <u>冷たい</u>	<u>爽やかな</u> , <u>優しい</u> , (<u>さわやかな</u>), <u>自由な</u> , <u>心地よい</u> , <u>目まぐるしい</u> , <u>冷たい</u> , <u>爽快な</u> , <u>柔らかい</u> , (<u>やさしい</u>), <u>清々しい</u> , <u>暖かい</u> , <u>かすかな</u> , <u>気まぐれな</u> , (<u>やわらかい</u>), (<u>すがすがしい</u>), <u>涼しい</u> , <u>美しい</u> , <u>自然な</u> , <u>不思議な</u> , <u>遠い</u> , <u>激しい</u> , <u>強い</u>	.105	1.000	.190
流れ星	速い, <u>ロマンチック</u> , <u>願い事</u> , <u>美しい</u>	明るい	.000	.000	.000
平均 <i>F</i> 値					.330

下線：一致した属性値 括弧：異表記とした属性値

[考察]

表 4 を見ると、「氷」については、提案手法では、*BL* では抽出できなかった顕現属性値「冷たい」を第 1 位にランキングすることに成功している。「炎」については、*BL* では、「急性すい」とそれに類する語が上位の属性値として混入している。これは、「急性すい炎」が形態素解析における未知語処理の問題に由来する結果である。これに対し、提案手法は、これらをうまく除去し、的確に顕現属性値を抽出している。

一方、「冷蔵庫」の場合は、すべての属性値が不適合と判定されてしまった結果である。提案するフィードバック処理では、属性値頻度がしきい値未満の場合、式 (6) によって、実質的に属性値の統計的優位性が失われる。その結果、かえって属性値ランキングを劣化させてしまった。対策としては、もとの属性値ランキングをより強く反映させるような式 (5) の修正が考えられる。

F 値の低下の原因は、再現率低下の影響と、適合率低下の影響がある。適合率低下は、抽出された顕現属性値の数が多い場合や正解が少ない場合に起こる。前者は、属性値の頻度が発散している場合である。「冷蔵庫」の例がそうであり、すべての属性値を抽出してしまっている。この問題については、属性値の生起確率計算法の拡張が必要である。

後者は、二つの原因が考えられる。一つは、顕現属性値として動詞や名詞が含まれている場合である。例えば、「鏡」については、連想実験データには、動詞が多く含まれていたため、再現率が低下し、*F* 値を下げた。人間の判断のよりの確かな再現には、属性値の対象を拡張する必要がある。もう一つは、属性値集合のノイズの影響である。例えば、「風」については、顕現属性値「強い」「冷たい」はすべて網羅できたが、抽出した顕現属性値数が多すぎるため、適合率が低下した。しかし、不正解となった「爽やかな」「自由な」なども「風」の顕現属性値として全く不適とはいえない。連想実験では、より一般性の高い顕現属性値が得られたものと考えられる。この差異を埋めるためには、顕現属性値の一般性を判断する尺度が必要となる。

再現率低下は、抽出された顕現属性値の数が少ない場合に起こる。これは、属性値集合の網羅性に問題がある。「チーター」「流れ星」については、新聞記事から獲得したもとの知識が少なかつたため、WWW を用いた効果が得られなかった。この問題に対応するためには、もとの属性値集合の網羅性を確保する必要が

あり、更に網羅性の高い知識抽出手法や、WWW から基本知識を精度良く取り出す手法を考えなければならない。

5. む す び

本論文では、比喩解釈において、比較概念間でクローズアップ属性値の適合性を自動判定し、その結果をフィードバックして知識を自動更新する手法を提案した。実験による検証の結果、自動評価では、人間の判断を 79%シミュレートでき、人間判断の代替としての利用も考慮可能であることが分かった。フィードバックについては、提案手法により、56.2%の属性値ランキング精度が、75.7%に向上したが、人間による判断の代替とするには更に改良が必要であることも分かった。今後は、WWW の知識を直接用いた属性値集合構成手法や動詞を属性とする手法について、その実現可能性も含めて検討していく。

謝辞 本論文を執筆するにあたり、有益な議論をして頂いた、北海道大学大学院情報科学研究科の Rafal RZEPKA 助手、大学院博士後期課程の Marcin SKOWRON 氏に感謝致します。

文 献

- [1] G. Lakoff and M. Johnson, *Metaphors We Live by*, The University of Chicago Press, Chicago, IL, 1980.
- [2] 山梨正明, 比喩と理解, 東京大学出版会, 1988.
- [3] 芳賀 純, 子安増生, メタファーの心理学, 誠信書房, 1990.
- [4] 諏訪正樹, 岩山 真, “比喩の計算モデル,” 情報処理, vol.34, no.5, pp.566–575, 1993.
- [5] K. Shinohara and B. Indurkha, “Braiding the multiple threads of interdisciplinary research on metaphor,” 日本認知科学会第 17 回大会発表論文集, pp.292–297, 2000.
- [6] 今井 豊, 石崎 俊, “比喩理解のための顕現属性の幾何学的抽出法,” 人工知能誌, vol.15, no.2, pp.309–320, 2000.
- [7] 内山将夫, 村田真樹, 馬 青, 小作浩美, “統計的手法による換喩の理解,” 自然言語処理, vol.7, no.2, pp.91–116, 2000.
- [8] 榎井文人, 福本淳一, 椎野 努, 河合敦夫, “確率的尺度を用いた比喩性判定手法,” 自然言語処理, vol.9, no.5, pp.71–92, 2002.
- [9] 佐々木淳一, 荒木健治, 桃内佳雄, “名詞述語文からの概念学習システム:比喩の理解と学習,” 情処学研報, 92-NL-89-3, pp.7–12, 1992.
- [10] 池ヶ谷有希, 野口靖浩, 鈴木夕紀子, 伊藤敏彦, 小西達裕, 近藤 真, 高木 朗, 中島秀之, 伊東幸宏, “対話文脈を利用した構文・意味解析手法の検討,” 人工知能学会第 18 回全国大会講演論文集, pp.3E2–10, 2004.
- [11] 楠見 孝, “メタファー研究の総括, 21 世紀に向けて: 認知心

- 理学の立場から”日本認知言語学会論文集 2, pp.268-271, 2002.
- [12] D. Gentner, “The mechanisms of analogical learning,” in *Similarity and Analogical Reasoning*, ed. S. Vosniadou and A. Ortony, chapter 7, pp.199-241, Cambridge University Press, Cambridge, MA, 1989.
- [13] 岩山 真, 徳永健伸, 田中穂積, “比喩を含む言語理解における顕現性の役割,” *人工知能誌*, vol.6, no.5, pp.674-681, 1991.
- [14] 内海 彰, “比喩理解モデルの諸相,” *信学技報*, 79-TL-97-5, 1997.
- [15] 内山将夫, 板橋秀一, “被喩詞の意味と比喩表現の意味との違いを示す指標,” *自然言語処理*, vol.3, no.3, pp.31-51, 1996.
- [16] A. Utsumi, K. Hori, and S. Ohsuga, “An affective-similarity-based method for comprehending attributional metaphors,” *J. Natural Language Processing*, vol.5, no.3, pp.3-31, 1998.
- [17] 今井 豊, 石崎 俊, “比喩理解における顕著な属性の発見手法,” *自然言語処理*, vol.6, no.5, pp.27-42, 1999.
- [18] J.H. Martin, “Metabank: A knowledge-base of metaphoric language conventons,” *Computational Intelligence*, vol.10, no.2, pp.134-149, 1994.
- [19] 佐久間鼎, 修飾の機能, 表現文法 (日本文法講座 5), pp.23-55, 明治書院, 1958.
- [20] O. Jespersen, *The Philosophy of Grammar*, George Allen & Unwin, 1924.
- [21] 山梨正明, 認知文法論, 言語の経験値基盤とイメージスキーマ, vol.4, ひつじ書房, 1995.
- [22] M.A. Hearst, “Automated acquisition of hyponyms from large text corpora,” 14th International Conference on Computational Linguistic (COLING92), 1998.
- [23] M.A. Hearst, “Automated discovery of wordnet relations,” in *WordNet: An Electronic Lexical Database* (Chapter 5), MIT Press, 1998.
- [24] M. Berland and E. Charniak, “Finding parts in very large corpora,” *International Conference on Association of Computer Linguistic (ACL-1999)*, ACL, 1998.
- [25] 山田一郎, 住吉英樹, 柴田正啓, “ニュース記事に出現する用語と説明文の意味関係自動獲得,” *情処学研報*, NL152-21, pp.145-152, 2002.
- [26] 安藤まや, 関根 聡, 石崎 俊, “定型表現を利用した新聞記事からの下位概念単語の自動抽出,” *情処学研報*, NL157-11, pp.77-157, 2003.
- [27] 榊井文人, 網島督之, 杉尾俊之, 田添丈博, 椎野 努, “英語長文における比較構造のモデル構築と解析,” *信学論 (D-II)*, vol.J78-D-II, no.12, pp.1849-1858, Dec. 1995.
- [28] 田添丈博, 椎野 努, 榊井文人, 河合敦夫, “‘名詞 A のような名詞 B’ 表現の比喩性判定モデル,” *自然言語処理*, vol.10, no.2, pp.43-58, 2003.
- [29] 関口洋一, 山本和英, “Web コーパスの提案,” *情処学研報*, IPSJ-SIG-FI, pp.123-130, 2003.
- [30] 中村 明, 比喩表現の理論と分類, 共立出版, 1977.
- [31] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum, *The Measurement of Meaning*, University of Illinois Press, Urbana, IL, 1957.
- [32] 福本淳一, 榊井文人, “質問応答技術—大量のデータをもとに任意の質問に答える,” *情報処理*, vol.45, no.6, pp.580-585, 2004.
(平成 17 年 2 月 14 日受付, 8 月 26 日再受付)



榊井 文人 (正員)

平 2 岡山大・理・地学卒。同年沖電気工業(株)入社。平 12 三重大・工・情報工助手。平 16~17 北大・情報科・メディアネットワーク専攻客員研究員。自然言語処理, 特に, 質問応答, 情報抽出, 比較構造のモデル化の研究に従事。言語処理学会, 人工知能学会, 日本設備管理学会各会員。



福本 淳一 (正員)

昭 59 広島大・工・第 2 類卒。昭 61 同大学院工学研究科博士前期課程了。同年沖電気工業(株)入社。平 4~5 英国マンチェスター科学技術大学言語学部 Ph.D. コース在学。平 12 立命館大・理工・助教授。平 16 米国 USC/ISI 客員研究員。Ph.D. 自然言語処理, 特に, 質問応答, 情報抽出, 談話構造解析の研究などに従事。情報処理学会, 言語処理学会, 人工知能学会, ACL 各会員。



荒木 健治 (正員)

昭 57 北大・工・電子卒。昭 63 同大学院博士課程了。工博。同年, 北海学園大学工学部電子情報工学科助手。平元同講師。平 3 同助教授。平 10 同教授, 平 10 北大・工・電子情報工学専攻助教授。平 16 北大・情報科学・メディアネットワーク専攻教授。自然言語処理, 特に, 機械翻訳, 音声対話処理などの研究に従事。情報処理学会, 人工知能学会, 言語処理学会, 日本認知科学会, ACL, IEEE 各会員。