

## 新聞記事等の文書を用いた概念自動学習による概念ベース構築方式

橋本 隆志

渡部 広一

河岡 司

同志社大学大学院工学研究科  
〒610-0394 京都府京田辺市多々羅都谷 1-3

本研究では人間らしい知的な処理を実現するための常識判断メカニズムの開発を目的としている。常識判断メカニズムでは、人間の「常識」の代用として語の概念を属性（その語と関連の強い語）の集合で定義、格納した概念ベースを想定している。既存の概念ベースは国語辞書から構築されているが、日常で使用される語の概念からは差異が存在するため、他のデータソースにより補強される必要がある。本稿では、新聞記事から自動構築した概念ベースを用いて、国語辞書概念ベースの補完、精度向上を目的とする。新聞記事概念ベースを構築する上で、新聞記事から概念となる単語と、その周囲に共起する単語を属性候補として獲得、相互情報量を基にした重みを付け集計する。構築した新聞概念ベースから属性を概念間の関連規則を用いて選別し、属性追加による概念自動学習を行い精度向上を実現した。

### Method for Automatic Construcion of Concept Base with Newspaper Articles by Automatic Concept Learning

Takashi Hashimoto,

Hirokazu Watabe,

Tsukasa Kawaoka

Graduate School of Engineering, Doshisha University  
Kyotanabe, Kyoto 610-0394

The aim of this study is a development of a commonsense judgment mechanism for intelligent information processing like a human being. To develop a commonsense judgment mechanism like a human being, it is proposed that Concept Base be used, a knowledge base defining concepts by attributes. Concept Base is constructed with dictionaries. But it is insufficient to express means of word by only explanations of dictionaries. This paper describes automatic construction of Concept Base with natural documents, specified newspaper articles and defined weight of attributes as mutual information based collocation frequency. To improve accuracy of Concept Base, attributes of concepts is selected by rule of relationship between concepts. It is shown that the proposed method improve Concept Base which constructed with dictionaries by addition of concepts and attributes constructed with newspapers.

#### 1.はじめに

情報処理システムにおいて、人間のような状況に応じた柔軟な対応や、常識的・知的な判断における分野では現在も有用なシステムの発展が見られず、知的で柔軟なインターフェースを持つ情報処理システムの登場が望まれている。本研究では、知的な判断をコンピュータで実現するための知的メカニズムとして常識判断メカニズムの完成を目的としている。常識判断メカニズムの中心をなす要素として、概念ベースと呼ばれる知識ベースがある。

概念ベースでは、ある単語の意味（概念）を

その単語に関連の深い単語群（属性）で定義する。概念ベースを構築する上で、概念を手作業で入力する方法は格納する概念の数が膨大なため、大規模な概念ベースの構築は現実的に不可能である。大量の概念を格納するためには、機械的な構築が必要となる。機械的な概念ベース構築の方法として、電子化された辞書を利用することが提案されている<sup>[1]</sup>。辞書を用いた概念ベースの構築では見出し語 - 説明文といった概念の関係や、類義語、反対語などの単語間の関係が明確であるために効率よく概念知識を獲得できる。しかし、辞書の概念は定義的であり、我々が日常的に持っている概念とは多少

のずれが存在する。

本稿では、新聞記事から概念ベースを自動構築し、それを用いて国語辞書概念ベースの補完、精度向上を目指した。国語辞書による概念ベースは概念の内的定義を表す傾向が強いのにに対し、新聞記事による概念ベースは概念の意味的連想における周囲の環境を表す傾向が強い。また、新聞記事を使用することにより、日常に密着した概念が獲得できると考えられる。属性には相互情報量による重みを付与することによって、既存の概念ベースで用いられている出現頻度を用いた属性の重み付けよりも高い精度を実現した。さらに、国語辞書概念ベースの精度向上のために構築した新聞概念ベースから属性を概念間に存在する規則によって選別し、属性追加による概念自動学習を行い精度向上を実現した。本稿で用いた手法は新聞記事に限らず、他の種類の電子化文書でも概念ベースの構築が可能である。

## 2. 概念ベース

概念ベースの構造と格納される概念の基本的な定義、常識判断メカニズム中の概念ベースの役割について述べる。

### 2.1 常識判断メカニズム

知的な判断をするシステムとして、常識判断メカニズムが提案されている。常識判断メカニズムとは、会話メカニズムによって人間と会話し、各種の判断メカニズムによって判断を行うシステムである(図1)。判断メカニズムに対する入力が自分の知識ベースにないときは背後に存在する概念ベースを利用し、関連度計算などを使用して処理を行う。

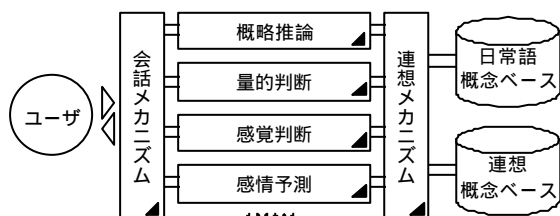


図1. 常識判断メカニズム

また、概念ベースの利用法として、概念間の関連の深さを数値化した関連度の計算への利用が挙げられる。関連度は概念間の類似性を計るために使われ、各種の判断メカニズムで使用

されている。

### 2.2 概念ベースの定義

概念ベースに格納される概念の定義、および概念ベースの構造を定める。概念ベースの単純な機能として、一つの入力単語に対して複数の単語を返すという機能がある。そのために、概念を次のように定義する。

[概念]: 単語に対し定義された、その単語の意味特徴をあらわす単語(属性)の集合。

例) 木{葉, 枝, 植物, 水, 酸素, ...}

これに加えて、概念ベースでは概念と属性の関連の深さを表す重みが付加されている。この属性と重みの集合で一つの概念を構成している。よって概念ベース  $K$  における一つの概念  $Word_i$  を、意味特徴をあらわす属性  $p_{ij}(j=1,2,...)$  と概念  $Word_i$  と属性  $p_{ij}$  の重み( $q_{ij} \geq 0$ )の対集合で表現する。

概念  $Word_i : \{(p_{i1}, q_{i1}), \dots, (p_{ij}, q_{ij}), \dots\}$

この表現により、概念ベース  $K$  を概念  $Word_i$  の集合で表す。

### 2.3 概念間関連度

関連度とは概念間の関連の深さを定量化した値である。本稿では関連度計算方式として「重み付き概念連鎖関連度計算方式」<sup>[2]</sup>を利用する。

## 3. 新聞記事からの自動構築

概念ベースを機械的に自動構築する必要性と、新聞記事からの概念ベース自動構築の手法について述べる。

### 3.1 概念ベース自動構築

概念ベースを構築する手法には大きく分けて手作業による手法と機械的に構築する手法との2つの方法が存在する。手作業による手法は、より人間から見て適切な概念を概念ベースに入力できる利点がある。しかし、入力時の人間の負担が大きいという欠点がある。実世界の概念の数が膨大であること、概念の微妙な変化や時事語への迅速な対応ができないことなどを考慮すると、概念の入力すべてを手作業で行うのは現実的ではない。そこで機械的に概念ベ

ースを構築する手法が必要となる。これにより大量の概念を機械的に獲得でき、概念獲得のルールを再現することによって異なるソースからの構築が容易である。

2. 2節で述べた通り、概念ベースは見出し語となる単語（概念）と、概念に付随する属性群より構成されている。実際は、個々のレコード「概念」「属性」という構造になっている。ここで、ある一つのレコードに注目して、その概念レコードが適切なものか否かは次のように判別できる。

概念、属性の単語自体が適切であるか。  
概念と属性の関係の深さが適切であるか。

は概念、属性それぞれの単語自体が言語的に適切かどうか、また日常的に使用する単語であるかの判断である。は概念と属性の関係の深さが人間から見て適切であるかの判断である。そこで、この2つの項目「単語」および「属性の適切さ」に対して自動構築のための検討を行った。

### 3.1.1 概念・属性となる単語

概念ベース中の概念および属性となる単語は、単独で意味を持つ単語であることが必要である。新聞記事を形態素解析したデータより、副詞を除いた自立語、つまり名詞、動詞、形容詞、形容動詞を抽出して獲得した。また、名詞中の形式名詞、固有名詞は除外した。

#### 形態素解析

前述の通り、本稿では新聞記事文を形態素解析して構築に利用した。形態素解析データには解析に失敗している部分がある。形態素解析の失敗として、「死亡者」「死」/「亡者」などの単語の切れ目の誤りがある。しかし、本稿では形態素解析のこのような失敗についての対策は考慮しないとされた。例外として、語尾が「～する」で終わるサ変動詞については処理を行った。例えば「走行」と「走行する」のように、品詞は異なるが概念としては同一と考えて差し支えないからである。これらのサ変名詞+「する」については、「～する」の部分を除いた名詞と同一の単語とみなして処理した。

### 3.1.2 属性の獲得と属性領域

属性の獲得を考える条件として、概念となる

単語に関連を持つ単語を獲得しなければ意味がない。新聞記事中の単語間の関係の深さについて、近接する単語は遠く離れた単語に比べて、単語間の関係が深い可能性が高いと考えた。そこで、概念となる記事中の単語から一定の領域（これを属性領域と呼ぶ）を設定する。その内部の自立語から前述の条件に合う単語を抽出し、その概念の属性として獲得する。

属性領域については、本稿では領域の区切りとして、以下の4つの領域で構築を行った。

#### 隣り合う単語

読点（“、”）で区切られる領域

句点（“。”）で区切られる領域

隣り合う2文

### 3.2 自動構築の手順

以上をまとめた自動構築の手順を示す（図2）。

#### 属性領域を選択

大学が国立研究所など外部の研究機関に大学院の研究室を置く、「連携大学院」が拡大している。新年度から佐賀大学などが加わり実施校が六大学から八大学になるほか、神奈川県が設立した神奈川科学技術アカデミーという異色の...

#### 形態素解析

1. 大学(1100)/が(7410)
2. 国立(1100)/研究所(1100)/など(7510)
3. 外部(1100)/の(7410)
4. 研究(1220)/機関(1100)/に(7430)
5. 大学院(1100)/の(7410)
6. 研究室(1100)/を(7430)
7. 置く(2317)

#### 自立語を抽出

大学 国立 研究所 外部 研究 機関 大学院 研究室 置く

#### 概念と属性として格納

概念	属性1	属性2	属性3	属性4	属性5	属性6	属性7	属性8
大学	国立	研究所	外部	研究	機関	大学院	研究室	置く
国立	研究所	外部	研究	機関	大学院	研究室	置く	大学
研究所	外部	研究	機関	大学院	研究室	置く	大学	国立
外部	研究	機関	大学院	研究室	置く	大学	国立	研究所
研究	機関	大学院	研究室	置く	大学	国立	研究所	外部
機関	大学院	研究室	置く	大学	国立	研究所	外部	研究
大学院	研究室	置く	大学	国立	研究所	外部	研究	機関
研究室	置く	大学	国立	研究所	外部	研究	機関	大学院
置く	大学	国立	研究所	外部	研究	機関	大学院	研究室

図2. 概念ベース構築の流れ

新聞記事から本文部分を形態素解析し、新聞記事を属性領域に分割する。続いて領域内から自立語を抽出し、任意の単語を概念、領域内のその他の自立語をその概念の属性とする。さら

に領域中のすべての単語について、同様にして概念を獲得する。このため、すべての概念と属性の関係は互いに対称な形で格納される。つまり概念  $C$  中に属性  $A$  が出現頻度  $q$  で存在したとすれば、概念  $A$  中にも同様に属性  $C$  が出現頻度  $q$  で存在する。また、概念ごとに記事中の属性の出現頻度を求め、属性の付随情報として格納しておく。このように概念を獲得する。

### 3.3 概念ベースの特性と問題点

以上で述べた手法によって構築された概念ベースについて、各概念の属性数を調べた。図3は属性数ごとに、概念ベース内でのその属性数の属性を持つ概念の個数を示したものである。これらより、属性が非常に多い概念が少数と、属性が10個以下の少ない概念が多数存在することが分かる。

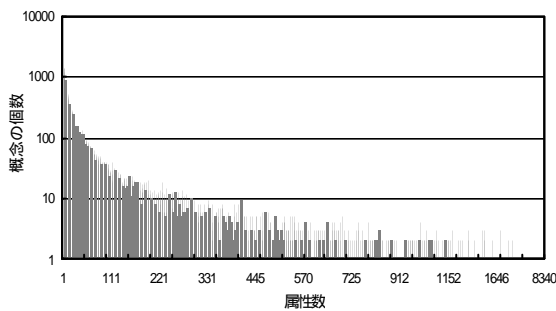


図3. 構築した概念ベースの属性数

構築された概念ベースは、属性を選択する必要がある。構築された概念ベースは属性数にばらつきが激しく、このままの利用は困難なためである。また、適切でない属性も多数含まれている。よって、属性数を揃えるための属性の選択方法について考える。属性数の上限については、概念ベースの精度のためには多過ぎる属性数は問題である。概念ベース利用の面から見ると少数過ぎてもうまく利用できない。そのために属性数の検討が必要である。これは3.5節で検討を行う。

### 3.4 属性の重み

属性の重みを以下のように定める。出現頻度による属性の選択は、普遍的な概念が上位にカウントされる。つまりどの概念においても頻出単語である「会社、企業、見る、もの」などの出現頻度が高くなるが、概念に対する意味的な重要度は低い。反対に少数の概念だけに含まれ

る属性は他と比べて特に重要視すべき属性であると考えられる。また、属性数の少ない概念では出現回数の少ない属性が大半を占めるため、属性の出現頻度によって各属性間に有意な差を見出すことができない。このような出現頻度では解決できない問題に対処するために、別の属性選択の基準を考える。

本稿では概念ベースが現在持っている情報、つまり概念と属性の組および属性の持つ出現頻度のみから選択基準を考えた。概念ベースの属性において、属性を概念としたときの属性（属性の属性）を二次属性と呼ぶ。属性の出現頻度が同じでも、二次属性全体の数に差があった場合、属性は概念にとっての重要度は異なると考えられる。つまり、二次属性の総数によってその属性の重要性を判断できる。次の図4はこの考え方を概念からの観点と概念ベースからの観点で示したものである。

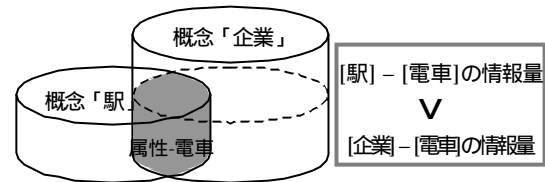


図4. 相互情報量

概念  $i$  中の属性  $j$  に対して、属性の重み  $q_{ij}$  を、概念  $i$ -属性  $j$  の共起頻度  $f_{ij}$  と概念  $i$  の頻度合計（各属性における共起頻度  $f_{in}$  の和  $f_{ik}$ ）と概念  $j$  の頻度合計（同  $f_{jk}$ ）を使い、相互情報量<sup>[3]</sup>を用いて次式のように定義する。

$$q_{ij} = \log_2 \frac{N_A \cdot f_{ij}}{\sum_k f_{ik} + \sum_k f_{jk} - f_{ij}}$$

ただし  $N_A$  は全概念の共起頻度合計である。ここで概念  $j$  の頻度合計に加えて、概念  $i$  の頻度合計を使用することによって重み  $q_{ij}$  と重み  $q_{ji}$  に対称性が維持される。得られた重み  $q_{ij}$  を共起頻度  $f_{ij}$  に置き換わる属性の適切さの値として用い、属性の選択に利用する。

### 3.5 構築パラメータ

属性獲得における属性領域の違いによって、構築された概念ベースの属性が適切である割合の変化を評価した。

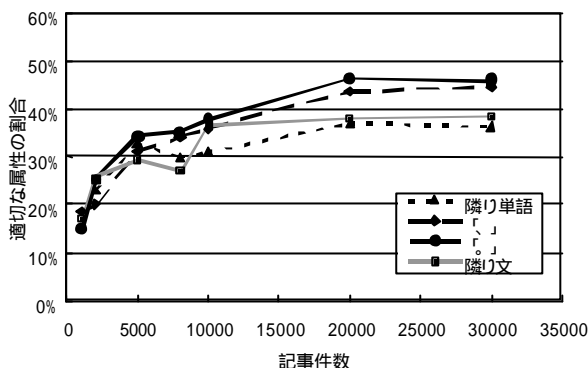


図5.属性領域と適切な属性の割合

図5より,記事件数の増加による精度の向上は,記事20000件から30000件の辺りでほぼ収束している。また,属性領域による変化では,読点(「,」),句点(「。」)を境界にした概念ベースが精度がよく,隣り単語・隣り文にした場合には精度が低くなっていることが分かる。このことから,最適な属性領域は発散せず,一定の値を取ることが推測される。

ここで構築された新聞概念ベースの詳細を示す。国語辞書概念ベース(基本CB)に比べ,属性数が大きく増加していることが分かる。

表1.新聞概念ベース詳細

	新聞CB	基本CB
概念数	34998	33699
属性数合計	5126270	1503925
属性を持つ概念	34683	33542
平均属性数	147.8035	44.83707

### 3.6 新聞概念ベースの評価

構築された新聞概念ベースに対して,属性が適切かどうかの評価と,関連度による評価用データの評価を行った。

#### 3.6.1 概念の評価

構築された概念ベース中の概念と属性に使われた単語(前述のように,概念と属性の単語は同じ基準で獲得されている)を500語選出し,概念として適切であるかを調査した(表2)。格納される単語のうち,日本語として存在しない単語は明らかに不適切である。また,日本語としては存在するが,日常では使われないような単語も適切ではない。日本語として不適切な例としては,主に“権数”,“工特”,“賀涛”など日本語として存在しない単語や,“本県”,“同紙”など接頭辞が混入している単語である。

表2.概念の適切さの割合(概念500語)

適切な概念(語数)	不適切な概念(語数)	
	日本語として	日常として
91.8% (459)	4.8% (24)	3.4% (17)

これは主に形態素解析に原因があると思われる。日常に使用する語として不適切な例としては,“涵養”,“沈金”,“授權”,“ファース”などの意味は持っているが日常的にまず使用されない単語である。

#### 3.6.2 属性の評価

サンプル概念として30概念を抽出し,属性がその概念と関連が深いかを基準に人手で評価した(図6)。

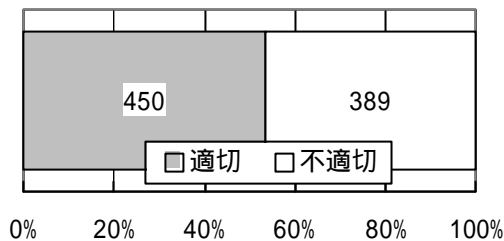


図6.適切な属性の割合

約54%が適切な属性であった。国語辞書概念ベースの適切な属性の割合は約50%であるから,適切な属性の割合は同程度といえる。

#### 3.6.3 関連度評価

590組の評価データで解釈成功率を評価した。ただし,評価データは基準概念  $M_X$ , 同義・類義の概念  $M_A$ , 関係のある概念  $M_B$ , 関係のない概念  $M_C$  で構成されており(表3),  $M_X$  との関連度  $R_A, R_B, R_C$  が  $R_A > R_B > R_C$  かつ各々の差が  $R_C$  以上であるとき, ABC が正しく解釈されたとする。AB, BC, AC も同様である。

表3.関連度評価データの例

基準概念 $M_X$	高関連 $M_A$	中関連 $M_B$	無関連 $M_C$
貨幣	通貨	値段	宇宙
会社	企業	仕事	草
海	海洋	塩	車

評価結果を図7に示す。国語辞書概念ベース(基本CB)に比べ新聞概念ベース(新聞CB)はBC解釈成功率がやや高いが,総じて成功率は低い結果となった。

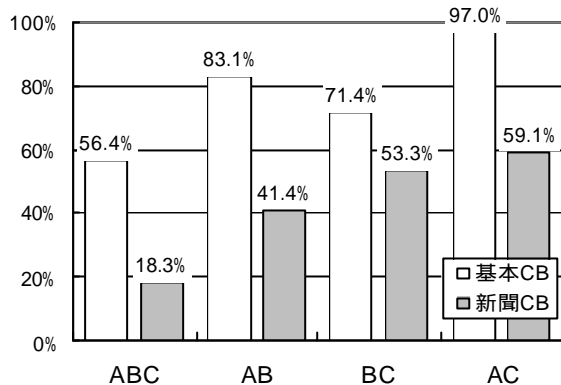


図 7. 関連度による評価

#### 4. 関連規則による属性選択

適切な属性を選別するために、以下に示す概念間の関連規則によって選別を行った(表4)。

表 4. 概念間の関連規則

	関連規則	内容
R0	共起	新聞記事内で共起している。
R1	関係語	関係語辞書に登録されている(同義・類義・上位・反対)
R2	人間の知識	各種判断 KB に登録
R3	表記特徴	漢字関連度がきわめて高い
R4	複合語	隣接して共起している
R5	関連度	概念間関連度が極めて高い
R6	情報量	属性の重みがきわめて高い

判断 KB は常識判断メカニズムにおいて使用される各種の判断知識ベースを利用した。

##### 4.1 関連規則の詳細

###### R0: 共起

新聞記事内で共起が起こった語。本稿では共起が起こった単語対から概念構築を行っているので、この規則に適合する属性は属性全体の集合に等しい。

###### R1: 関係語

国語辞書の語義文により機械構築された同義語・類義語・反対/対義語・上位の約 46 万対の語を使用し、概念-属性間に関係が登録されている属性を採用した。

###### R2: 判断知識ベース

本稿では常識判断メカニズムにおいて提案されている判断知識ベース<sup>[4]</sup>の内から感情 KB、職種 KB、場所 KB を利用した。以下に詳細を

示す。

###### R2-1: 感情知識ベース

感情判断に利用される知識ベース<sup>[4]</sup>より、同一分類の変化語、象徴語-分類語、代表語-分類語の約 12 万対の語を使用した。

表 5. 感情 KB の適用例

新聞 CB		感情 KB	
概念	属性	変化語	分類
治療	怪我	治療する	回復
治療	完治	退院する	回復
治療	直る	完治する	回復

表5のように、感情 KB の変化語テーブルにおいて、「治療」と「完治」は同じ分類語「回復」に登録されている。よって、新聞 CB 中の概念「治療」の属性「完治」がこの関連規則に適合し、選択される。

###### R2-2: 職種知識ベース

職種判断に利用される知識ベースより、商品名と表記特徴の一致のある語、および職種分類と表記特徴の一致する語とその職種に含まれる全ての商品名を使用した。

表 6. 職種 KB の適用例

新聞 CB		職種 KB	
概念	属性	職種	商品
運ぶ	配達	運送屋	配達, 引越, 宅配, 運輸, 輸送, 商品配達,
運ぶ	宅配		
運ぶ	持つ	銀行	お金, 預金, 融資, 払戻し, 残高照会, ATM, 口座
預ける	預金		

表6のように、概念「運ぶ」に対して、職種 KB の職種テーブルにおいて、職種名「運送屋」と表記特徴が一致する。そこで、新聞 CB の概念「運ぶ」の属性に対して「運送屋」の商品項目内で一致する「配達」「宅配」をこの関連規則に適合させる。概念「預ける」に対しては表記特徴が一致する職種名が登録されていないので、全ての商品から表記特徴の一致するものを探し、結果「預金」が関連規則に適合される。

###### R2-3: 場所知識ベース

場所判断に利用される知識ベースより、場所分類の代表語と目的名から約 3000 対の語を使用した。表7のように、概念「大学」を場所 KB の代表語テーブルにおいて検索すると、目的から「教育」が属性と一致し、関連規則に適合される。同様に概念「山」の目的から「ハイキング」が適合される。

表 7. 場所 KB の適用例

新聞 CB		場所 KB	
概念	属性	代表語	目的
大学	教育	大学	勉強, 学習, 学ぶ, 教育,
大学	受験		授業, クラブ活動
山	山岳	山	登る, 登山, 修行,
山	ハイキング		ハイキング, ピクニック

**R3：表記特徴**

漢字関連度を用いた表記特徴量<sup>[5]</sup>により閾値以上の値を持つ属性を採用する。本稿では閾値として漢字関連度 0.25 以上を有意な表記特徴として採用した。

**R4：複合語**

新聞記事、国語辞書等から得られた複合語 46 万語を使用し、概念-属性が複合語の構成要素である属性を採用した。ただし、複数の自立語形態素が隣接して出現する単語を複合語の定義とした。

**R5：関連度**

2. 3 節で述べた関連度を用いた概念-属性の関連度で、閾値以上の関連度を持つ属性を採用した(図 8)。

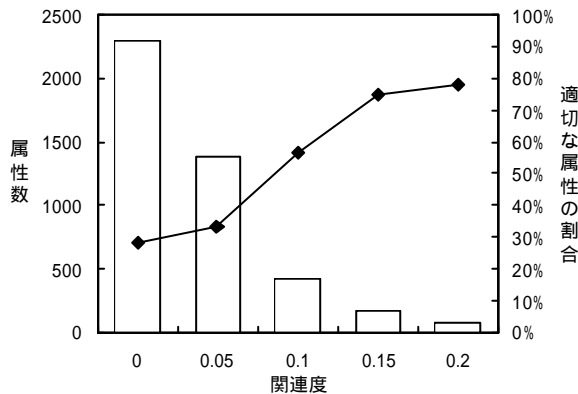


図 8. 関連度と属性の適切さの分布

図 8 のように、閾値を高くすると適切な属性の割合が増える。ここでは適切な属性が 7 割を越える関連度 0.15 を閾値として設定した。

**R6：情報量**

構築時に得られた相互情報量を元にした属性重みにおいて、閾値以上の値を持つ属性を採用した(図 9)。関連度による属性の分布と同様に閾値を高くすると適切な属性が得られる。ここでは、適切な属性の割合が 7 割を越える情報量 14 以上を閾値に設定した。

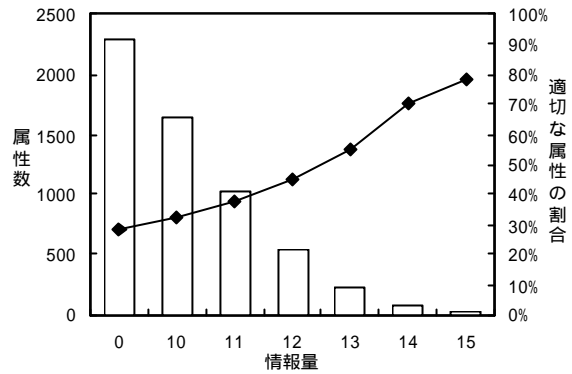


図 9. 情報量と属性の適切さの分布

4. 2 関連規則の適用

関連規則の内、R1-2 は正確な知識が記述されているのに比べ、R3-6 は不確実で雑音を多く含む属性が多い。よって R1, R2 で選択された属性はそのまま採用し、R3, R4 で選択された属性は R5, R6 の関連度又は属性の重みで閾値を設け、一定の値以上の属性を採用した(図 10)。閾値は実験の結果、 $T_5=0.15$ ,  $T_6=13$  が最適であった。

```

if (R1 or R2) then
  True
else
  if (R3 or R4) then
    if (R5 > T5 or R6 > T6) then
      True
    else
      False
  
```

図 10. 関連規則の適用

5. 国語辞書概念ベースへの概念学習

5. 1 対象概念

概念追加は動詞概念(サ変名詞概念を含む)を対象にして行った。

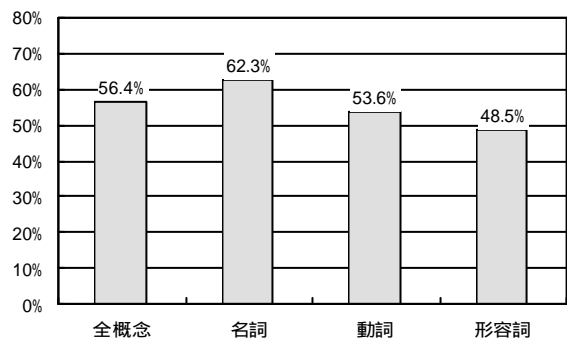


図 11. 品詞種別毎の評価 (基本 CB)

これは基本 CB において名詞概念に比べ動詞・形容詞概念の正解率が低いため、これを補完することにより、より効果的な精度向上を実現できると考えたためである(図11)。

### 5.2 関連規則により選択された属性

4章により選別された属性を、国語辞書概念ベース(基本CB)に追加し、学習CBを作成した。追加の際、属性の重みに整合性を持たせるために基本CB内の重みを一旦削除し、追加後に関連度によってすべての属性の重みを付与し直した。追加された属性を人手により評価した結果を示す(図12)。

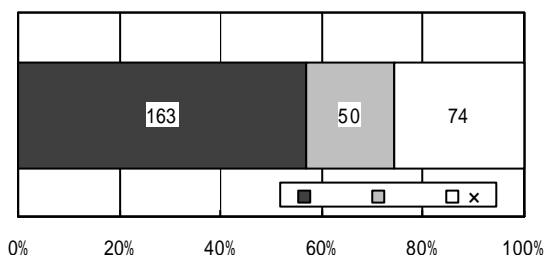


図12. 選択された属性の評価

### 5.3 評価尺度による評価

属性追加を行った概念ベースに対して、3.6.3節と同様の評価データによる評価を行った。ここで動詞評価は評価尺度590組中、基準概念が動詞またはサ変名詞であるもの270組である。

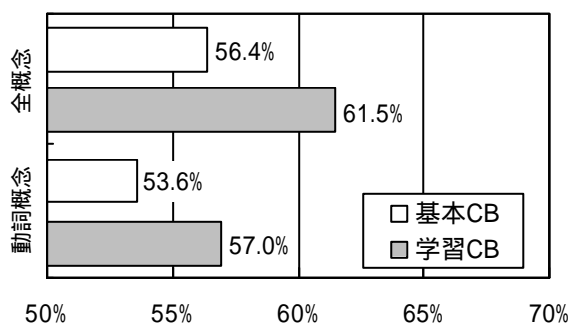


図13. 評価尺度による評価(属性追加)

図13のように、追加前の概念ベースである基本CBに比べて動詞概念で3.4%、全概念では5.1%の正解率の向上となった。属性追加を行った概念は動詞概念のみであるが、精度向上は概念全体に波及し、動詞概念の向上幅を上回る精度の向上が得られた。これは各概念に含ま

れる属性の動詞概念が改善されたため、概念全体の関連度計算に影響を与えたと考えられる。

## 6. おわりに

概念ベースを利用する上で、日々新たに出現する概念や意味属性の変化に対応するためには概念ベースを自動的に改善、学習を行う手法が必要である。本稿では、新聞記事を用いて国語辞書概念ベースを補完することにより、概念学習を実現する手法を提案した。

本稿の手法では、まず新聞記事から概念ベースの自動構築を行い、属性の重みを相互情報量によって定めた。次に獲得された概念ベースの概念属性に関連規則を適用することにより、適切な属性を選択した。さらに、選択された属性を国語辞書概念ベースに追加する形で概念学習を行い、精度向上を実現することができた。

本稿では関連規則を適用した動詞概念に限定しているため、全ての品詞に拡張を行うことによってさらなる精度の向上が期待できる。そのためには関連規則の追加、修正が必要と考えられる。

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクト「知能情報科学とその応用」における研究の一環として行った。

## 参考文献

- [1] 笠原要, 松澤和光, 石川勉: 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌, Vol.38, No.7, pp.1272-1283(1997)
- [2] 渡部広一, 河岡司: 常識的判断のための概念間の関連度評価モデル, 自然言語処理, Vol.8, No.2, pp.39-54(2000)
- [3] 杉浦正利: 英語学習者コーパスと母語話者コーパスにおける相互情報量を用いた共起関係の比較分析, 名古屋大学 教育研究改革・改善プロジェクト報告論文集(2000)
- [4] 土屋誠司, 馬場秀樹, 渡部広一, 河岡司: 入力文から感情を判断するシステムにおける未知語の処理, 電子情報通信学会 信学技報, AI99-107(2000)
- [5] 東村貴裕, 小島一秀, 渡部広一, 河岡司: 意味属性と漢字属性を用いた概念間の関連性評価法, 情報処理学会「知能と複雑系」研究会, 2001-1(2001)