

# 名詞概念の典型的特徴の分類手法の検討

原田 佳史† 森田 悠介‡ 榊井 文人‡ 井須 尚紀‡ 河合 敦夫‡  
† 三重大学工学部 ‡ 三重大学大学院工学研究科

## 1 はじめに

計算機上でことばの比喩的關係や類似關係などを把握するためのひとつの手段として、概念をその特徴(属性値)の集合で表現するモデル[10]を用いる手法がある。例えば、「リンゴ」という名詞概念は、「赤い、丸い、甘い、酸っぱい、…」といった属性値の列挙によって表現される。このうち、「赤い、丸い」など、概念の典型性を表す属性値(顕現属性値)は、その概念を把握する上で重要な役割を担う。

岩山らは、概念を属性値とその生起確率の集合で表現することによって、比喩的關係を把握する計算モデルを提案している[1]。Utsumi や今井らは、心理学実験によって人間の判断に基づく名詞概念の属性値集合を収集した知識ベースを構築し、比喩的關係を把握する実験を行っている[2][3]。榊井らは大規模テキスト集合(コーパス)を利用して同様の知識ベースを自動構築する手法を提案している[4][5]。森田らは、知識源としてWWWを利用することによって、より網羅性に優れた知識ベースが構築可能であることを実験的に示している[6]。

しかしながら、精度の良い知識ベース構築を実現するためには、自動構築による知識ベースと、人間判断による知識ベースの詳細な比較が必要である。これまでのところ、自動判定によって得られる属性値集合については、顕現属性値としては概ね人間判断に近いものが得られることは確認されているものの、属性値集合全体としての性能、さらには、顕現属性値と他の属性値の割合などについて、どの程度人間判断に近いものが得られているかという検証結果についてはほとんど報告されていない。

そこで本論文では、自動処理によって得られた知識ベースと、人間判断に基づく知識ベースとを比較し、分析評価を行った結果について考察する。

以下、2章では、心理学実験によって、人間判断に基づく複数の名詞概念に関する属性値集合を収集し、それらの属性値、顕現属性値の分布について調べる。3章では、森田らの手法を拡張した知識ベース構築手法を用いて、

複数の名詞概念に関する属性値集合を収集し、同様の分析を行う。さらに、4章で両知識ベースを比較し、両者がどの程度一致し、どのような点において違いがみられるかについて考察する。

## 2 人間判断による属性値データ

本章では、名詞概念を表現する属性値集合に関して、人間判断に基づいて収集した結果について述べる。以下、人間判断に基づいた属性値データの調査環境、方法、および調査結果について詳述する。

### 2.1 収集方法

人間判断による属性値データの収集手法について説明する。

まず連想実験を実施し、名詞概念とそれに関連する属性値データを収集する。連想実験の実施環境は以下の通りである。

- 被験者：大学生 29 名
- 提示データ：名詞概念 30 組
- 調査方法：提示概念の中からランダムに振り分けられた名詞概念 10 組を紙面で提示し、制限時間 5 分以内で想起できる特徴(属性値)を記入してもらった。

制限時間は、1 概念あたり 30 秒に設定している。これは、複雑な連想的推論を経ずに思いつく限りの属性値を答えるために必要な時間として、予備的な実験を事前に実施して決定した。

次に、収集した属性値データを概念毎に整理し、属性値として妥当性を人手で評価した。評価環境は以下の通りである。

- 被験者：大学生 9 名
- 提示データ：連想実験で収集した名詞概念 30 組とその属性値集合

- 調査方法：提示された名詞概念と属性値集合に関して,S,A,Bの3レベルで妥当性を評価した. 3レベルの評価基準は以下の通りである.

- \* S：名詞概念の典型的な特徴(顕現属性値)として妥当である.
- \* A：名詞概念の特徴(属性値)として妥当である.
- \* B：名詞概念の特徴として妥当ではない.

## 2.2 調査結果

連想実験の結果,30組の名詞概念に対して410組の属性値データが収集できた. その属性値データの妥当性評価において,各属性値で一番人数が多かった評価のレベルを付与した(同人数の評価に対してはB A Sの順に優先して付与). レベル付与の結果,Sレベルが38%,Aレベルが36%,Bレベルが26%となった. また,各概念毎にみると,レモン,チョコレートなど典型性が明確な概念ではSレベルの割合が多く,サンタクロースのような典型性が表現し難い概念ではBレベルの割合が多かった.

図1に「チョコレート」に関するS,A,B別の属性値数を表す度数分布を示す.

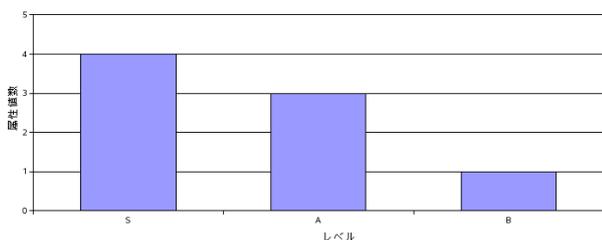


図1: 「チョコレート」に関するS,A,B別の属性値数

## 3 自動処理による属性値データ

本章では,名詞概念を表現する属性値集合に関して,自動処理に基づいて収集した結果について述べる. 以下,属性値データの収集・レベル付与手法,およびその結果について詳述する.

### 3.1 属性値収集手法

属性値は,森田ら[6]の手法を拡張し,テキスト中の共起情報と定型パターンを利用して獲得,識別する. 以下,属性値獲得手法について詳述する.

1. 対象コーパスから連体修飾関係にある形容詞・形容動詞-名詞の共起  $(X, i)$  を抽出する.
2.  $(X, i)$  を「Xは  $i$  です」というパターンに適用し,断定表現  $S_1$  を生成する. さらに,WWW上の  $S_1$  を検索して頻度情報を得る.
3. あらかじめ設定した閾値  $\theta$  により適合性を判定し,適合  $S_1$  を list1 に保存する. ただし,得られる  $S_1$  は非常に大きいため,頻度に基づく順位尺度としてRR(ReciprocalRank)を用いる.
4.  $(X, i)$  を「Xのように  $i$ 」「Xのような  $i$ 」というパターンに適用し,比喩的關係を示す表現  $S_2$  を生成する. さらに,WWW上の  $S_2$  を検索して頻度情報を得る.
5. あらかじめ設定した閾値  $\theta$  により適合性を判定し,適合  $S_2$  を list2 に保存する.
6. list1 と list2 を比較し,属性値を2段階のレベル a,b に識別する. 識別基準は,
  - (a)  $i \in \text{list2} \rightarrow a$
  - (b)  $i \in \text{list1} \rightarrow i \in \text{list2} - i \in \text{list1} \rightarrow b$
 とする.

閾値  $\theta$ ,  $\theta'$  は,事前に最適値を調査して設定した ( $\theta = 0.1$ ,  $\theta' = 15$ ).

### 3.2 獲得結果

7組の名詞概念について,前節で説明した処理を適用し,属性値データを取得し,レベルを付与した. その結果,aレベルは全体の16%,bレベルは84%であった. また,各名詞概念別にみると,「レモン」など典型性の明確な概念では,aレベルの割合が多く,「人参」などのような典型性が表現し難い概念ではbレベルの割合が多かった.

図2に「レモン」に関するレベル a,bの度数分布の結果を示す.

## 4 比較分析と考察

本章では,名詞概念に対する属性値について,2章で行った連想実験および妥当性評価によってレベル分類

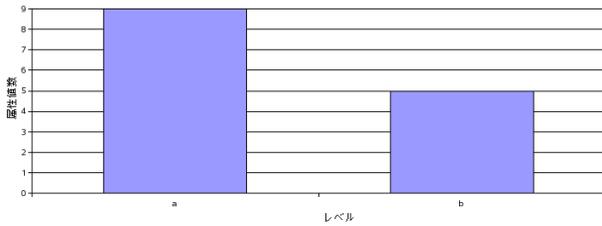


図 2: 「レモン」に関する S,A,B 別の属性値数

された属性値データ (人間判定データ) と、3章で行った自動処理により収集によってレベル付与された属性値 (自動判定データ) を比較し、その一致性について考察する。

#### 4.1 比較分析

2章では、連想実験によって得た人間判断データについて妥当性を評価することによってそれぞれ S,A,B 3つのレベルに分割した。3章では、自動処理によって属性値データを収集し、a,b 2つのレベルに分割した。ただし、2つのデータのレベルは単に大雑把な順位尺度として対応することのみを意味し、同定義であることを保証するものではない。

そこで、人間判断データの各レベルと自動判定データ間に生じるレベル対応のズレに注目する。ここでいうレベル対応のズレは、両データ間のレベル判定誤差とみなすことができる。もしレベル対応がよく一致していれば、誤差は0点 (誤差なし) を平均とした正規分布に近づくと考えられる。レベル対応に偏りがある場合は、平均値が偏った分布となり、対応がよくない場合は、正規分布から外れることになる。

表1に、人間判断データと自動判定データ間のレベル対応の組み合わせを示す。各レベル対応には組み合わせ数に差があるので、各対応毎の度数を組み合わせ数で平均した値を評価対象として誤差分布を検証する。

表 1

誤差	人間判定	自動判定
+2	B	対応なし
+1	A a	B b
0	S a	A b
-1	S b	A 対応なし
-2	S	対応なし

上で述べた手順に従って得た誤差分布を図3に示す。全体として、誤差0 (ズレなし) であったものは74%、

±1 (ズレが1) であったものは23%、±2 (ズレが2) であったものは3%となった。このことから、自動判定結果が人間判定結果とよく一致していることがわかる。

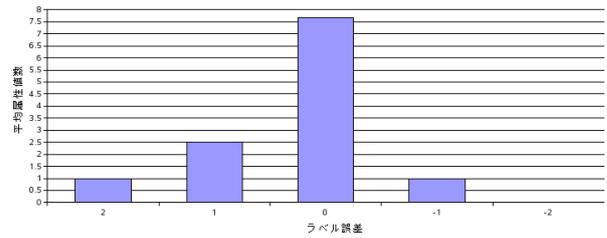


図 3: 両分類の誤差分布

#### 4.2 考察

前節の誤差分布の正規性を、正規確率表を用いて確認した。得られたレベル対応の度数を大小の順に整列させた順序統計量を作成し、各度数に対して累積相対度数を正規確率表にプロットした。その結果、プロットデータに直線性が見られ、誤差分布の正規性が確認できた。

このことから、自動判定によって付与されるレベルは人間判定によって付与されるレベルと概ね一致しており、自動判定によって高い精度で属性値を分類可能であることが示唆された。

また、誤差分布が正の方向にやや偏っていることから、自動判定の過大評価傾向がやや強いと言える。この原因としては、自動判定基準が易しすぎる可能性と、人間判断データの規模が十分でない可能性が考えられる。この原因について詳細に考察するためには、人間判断データを大規模化する必要がある。

次に判定誤差が大きかったケースについて考察する。

自動判定では a が付与されたが、人間判定では属性値として収集されなかったものが82組あった。この原因としては、上記と同様に、自動判定基準が厳しすぎる可能性と、人間判断データの規模が十分でないために、見かけ上表出したものである可能性が考えられる。しかし、前者の可能性では、上述した誤差分布の偏りの原因としてあげた可能性と矛盾してしまう。よって、後者の可能性が高いと考えられる。

人間判定では S であるにもかかわらず、自動判定では収集できなかった属性値が12組あった。この原因として、該当する属性値がいわゆる「常識」に相当するものであり、テキスト上に記述され難い可能性が考えられる。このケースに対しては、自動判定による収集は不可能で

ある。自動判定の処理過程中に、人間による適合フィードバックなどを取り組む工夫が必要であろう。

[10]Rosche,E.and Mervis,C.B,"Family resemblance:studies in internal structure of categories,"Cognitive Psychology,3,pp.382-439.

## 5 おわりに

本論文では、自動判定によって得られる属性値集合について、属性値集合全体としての性能、顕現属性値と他の属性値の割合などについて、どの程度人間判断に近いものが得られているかを検証した。心理学実験によって、人間判断に基づく複数の名詞概念に関する属性値集合を収集し、それらの属性値、顕現属性値の分布について調べた。森田らの手法 [6] を拡張した知識ベース構築手法を用いて、複数の名詞概念に関する属性値集合を収集し、同様の分析を行った。さらに、両知識ベースを比較し、両者は概ね一致していることがわかった。

今後は人間判定データの規模を拡大することで今回得た知見を確認し、知識ベース構築手法の精緻化を進める予定である。

## 文献

[1] 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.

[2] 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.

[3] 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. [4] 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.

[5] 榎井文人, 福本淳一, 荒木健治," 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," 電子情報通信学会論文誌, vol. J89-D, no.4, pp.860-870, 2006.

[6] 森田悠介, 榎井文人, 河合敦夫, 井須尚紀," WWW からの概念属性値獲得と属性値集合の洗練," 第 5 回情報科学技術フォーラム講演論文集, E-005, pp.145-148, 2006.9.

[7] 平井文人, いろのえほん, 江口克彦, PHP 研究所, 2001.

[8] 岡本健, あか あお きいろ, 小林芳郎 (監修), ひかりのくに株式会社, 1997.

[9] 丹後俊郎, 古川俊之 (監修)," 新版 医学への統計学," 朝倉書店, 1993.