

# 共起頻度と意味的特徴を用いた 五感に関する知識の一般性評価手法

三橋 奎太<sup>†</sup> ジェプカ・ラファウ<sup>†</sup> 荒木 健治<sup>†</sup>

<sup>†</sup> 北海道大学大学院情報科学研究科

{mitsuhashi, rzepka, araki}@ist.hokudai.ac.jp

**概要** 五感に関する知識は人間の意味処理において重要であるが、コンピュータにとってはそれらの知識を理解することは困難である。これまでに我々が行なった研究において五感知識データベースを自動構築する手法の提案を行なったが、それらの知識の一般性について最適な手法が確立されていなかったため、「りんごの色が赤い」と「風船の色が赤い」といった知識のように一般性のレベルに差があるような知識について、そのレベルを考慮することなく、コーパスからルールに当てはまった知識をすべてデータベースに登録していた。そこで獲得した五感に関する知識の一般性について、共起頻度と意味的特徴を用いて評価する手法の提案を行なった。その結果、正解データの値をそのまま学習させる評価手法よりも RMSE の値が 5~6 ポイント改善され、ベースラインに対する提案手法の優位性が確認された。

**キーワード** 五感, 知識獲得, 回帰問題, 評価手法

## 1 はじめに

人間は五感を通して様々な概念を理解し、言語化していくことができるため、五感によって得られる知識(以降、五感知識と呼ぶ)は人間にとって最も基本的な知識であると考えられる。近年の科学技術の発達により、様々なセンサーを用いることでロボットに擬似的に五感を与えることは可能であるが、センサーから得られたデータを言語化することは容易ではないため、ロボットは人間のように言葉を正しく理解することができない。このような問題を解決するために、五感知識データベースを構築する研究が様々行われている。

これまでの我々が行なった研究[1]では、ルールベースでブログデータから五感知識データベースを自動構築する手法の提案を行なったが、それらの知識の一般性について最適な評価手法が確立されていなかったため、一般性を考慮せずにルールに当てはまった知識をすべてデータベースに登録していた。

五感知識の一般性の評価手法を確立し、五感知識とその一般性を表す評価値を同時にデータベースに登録することでできるようになれば、アプリケーションに応じて柔軟にデータベースを使用することができるように考えると考えられる。

そこで、本研究では共起頻度と意味的特徴を用いて新たに五感知識の一般性を評価する手法を提案する。

## 2 関連研究

渡部らの研究[2]では、電子化国語辞書などから作成した概念ベースや、人手で作成した感覚判断知識ベースを用いて、名詞に対し感覚語を付加している。感覚語を付加する際には、付加方法ごとに求めた信頼度を用いて信頼度の合成を行い、閾値によって不適切な感覚語を除去している。

また林らの研究[3]では、想起関係の強さの予測を回帰問題として扱い、単語間の各種の類似度・関連度に加え、荒い意味分類、品詞、PWN (Princeton WordNet) [4]の意味ネットワーク構造から得た素性のほか、Word2Vec[5]による代表単語の分散表現の差分、および AutoExtend[6]による語義・概念の分散表現の差分を素性とする教師つき学習により、従来研究を上回る精度で想起関係の強さを予測している。

本研究では、評価値を用いて不適切な知識を省くのではなく、獲得した知識とその評価値をすべてデータベースに登録することを目的としているため、五感知識の一般性の評価値を予測することができなければならない。この点で渡部らの研究とは異なっている。また、想起ではなく、五感知識の一般性を取り扱っているという点で林らの研究とも異なっている。

## 3 システム概要

五感知識は名詞、感覚語カテゴリ、感覚語の3つの要素で表記される。感覚語カテゴリとは、sight(視覚)、touch(触覚)、hearing(聴覚)、taste(味覚)、smell(嗅覚)の5種類の感覚のいずれかであり、感覚語とは「赤い」

や「甘い」などの感覚器官から得られた情報を言語化したものである。つまり五感知識は「(りんご, sight, 赤い)」といった形で表記される。この五感知識獲得までの流れを図1に示す。この章では、これまでの研究に関する部分と新たに改善した部分についての説明を行う。

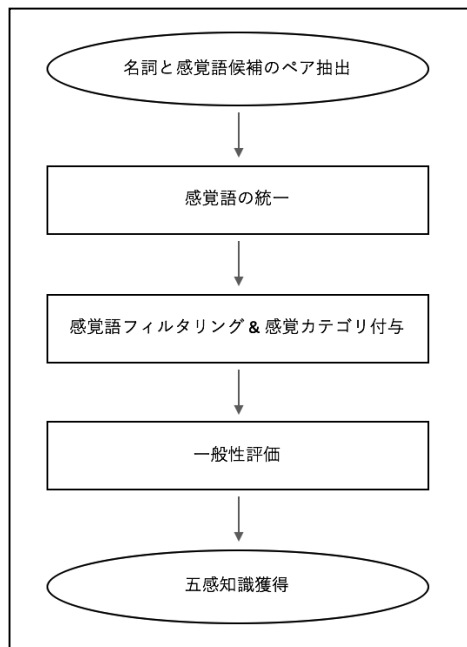


図1 システム構成

### 3.1 名詞と感覚語候補のペア抽出

名詞と感覚語候補のペアの抽出手法として、以下の二種類のルールを用いる(図2参照)。

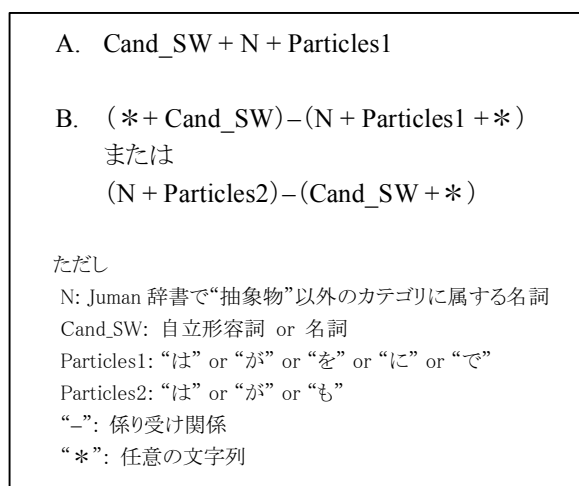


図2 名詞と感覚語候補の抽出ルール

図2でルールAは「赤いりんご」などのように感覚語が名詞の修飾語になるペアを抽出することができる。一方、ルールBはルールAの制約を強くし「赤い」と「りんご」の間に係り受け関係がある場合のみペアを抽出することができる。さらに「りんごが赤い」などのように感覚語が主格補語となるペアも抽出することができる。この二種類のルールは、抽

出数等を考慮しなかった場合、感覚語が修飾語となるペアを抽出するルール、感覚語が主格補語となるペアを抽出するルール、構文解析を用いたルールの組み合わせのうち、ルールAが最も精度が高く、ルールBが最もF値が高くなることが従来研究[1]からわかっている。

### 3.2 感覚語フィルタの作成

3.1のルールだけでは、ノイズが多く含まれるため、従来研究[1]と同様に、渡部ら[2]の作成した感覚語を使用して感覚語フィルタを作成する。

渡部らの作成した98種の感覚語には感覚カテゴリの情報も付随しているため、これを使用して3.1で得られたペアに感覚カテゴリを付与していく。このとき複数のカテゴリに属する感覚語は一意に感覚語カテゴリを決定できなくなってしまうため除外する。さらに、「痛い」や「痒い」など物体の特徴を表さない感覚語を除外し、残った83種の感覚語をフィルタとして採用する。

### 3.3 感覚語の統一

従来研究[1]より、ルールベースによってコーパスから得られた五感知識の中には、「(餅, touch, 暖かい)」などのように感覚語の漢字の表記が適切でない知識が存在することや、フィルタにひらがな表記の感覚語が登録されていないため、「(餅, touch, あたたかい)」などの知識が除外されてしまうということがわかっている。そこで、新たに感覚語のフィルタリングを行う前に感覚語の統一処理を行うことで、これらの問題に対応する。

具体的には形態素解析ツール MeCab[7]を用いて各感覚語候補をひらがなに変換し、同音の感覚語候補を含むペアを検索する。同音の感覚語候補が見つかった場合は、3.2で作成したフィルタに登録されている感覚語に統一し、抽出数を足し合わせる。フィルタに登録されている感覚語が複数見つかった場合には、抽出数が大きい感覚語に統一し、抽出数を足し合わせるという処理を行う。

### 3.4 感覚語のフィルタリングと感覚カテゴリの付与

3.2で作成したフィルタを使用し、感覚語のフィルタリングを行い、さらに渡部らの作成した感覚語の情報から五感カテゴリを決定する。これにより五感知識が得られる。従来研究[1]では、このままデータベースに登録していた。

## 4 提案手法

新たに五感知識の一般性を表すスコアを導入する。本章では、スコアに入れる要素として共起頻度を用いた要素と意味的特徴を用いた要素の説明と、今後の検討案として経験則を用いた要素についての説明を行う。

#### 4.1 共起頻度を用いた要素

一般性のある名詞と感覚語のペアほど、共起しやすいと考え、一般性を表すスコアの要素の一つとして共起頻度  $CP$  を利用する。名詞  $N$  が感覚語  $SW$  と共起する頻度  $CP_{N,SW}$  を(1)式で表す。ここで、 $co\_freq(N,SW)$  はルールベースにより抽出された名詞  $N$  と感覚語  $SW$  のペアの抽出数を表し、 $Cand\_SW_i$  は感覚語候補を表す。

$$CP_{N,SW} = \frac{co\_freq(N,SW)}{\sum_i co\_freq(N,Cand\_SW_i)} \quad (1)$$

#### 4.2 意味的特徴を用いた要素

「(声, sight, 黄色い)」などのように、名詞が特徴として持ち得ない(声は目に見えない)感覚語を除去するため、意味的特徴として五感特徴保有率というものを作成する。五感特徴保有率とは「どの程度 sight に属する感覚語を持ちやすいか」「どの程度 touch に属する感覚語を持ちやすいか」などの情報である。名詞ごとに五感特徴保有率を算出することが理想的であるが、名詞ごとに作成するにはあまりに膨大な量になってしまうため、同じカテゴリに属する語は同じ五感特徴保有率を持ちやすいと仮定し、名詞カテゴリごとに五感特徴保有率を作成する。

名詞カテゴリには形態素解析ツール Juman[8]辞書の 22 カテゴリを採用し、被験者7名(20代男性3名(理系学生2名, 会社員1名), 20代女性会社員4名)に対し、各名詞カテゴリからランダムに10個の単語を提示した。被験者には各単語と各感覚カテゴリのキーワード(例: 視覚カテゴリのキーワード→色, 形, 明るさ, 等)を結びつけて考えてもらい、違和感のないキーワードが一つでもあれば、その感覚カテゴリに属する特徴を持ちうるものとしてチェックしてもらった。そして全員のチェック率の平均値から名詞カテゴリごとに五感特徴保有率を算出した。

名詞カテゴリ  $NC$  における感覚カテゴリ  $SWC$  の特徴保有率は  $Prop_{NC,SWC}$  と表記する。

#### 4.3 経験則による要素

今後の検討案として経験則による要素も導入する。今回用いる図2のペア抽出手法のルールBでの抽出パターンは大きく二つに分類される。一つは「りんごは赤い」のように感覚語が主格補語になるパターンであり、もう一つは「赤いりんご」のように感覚語が名詞の修飾語になるパターンである。これら2つのパターンの両方で抽出されたペアの方が一般性があると仮定し、(2)式で表される関数  $Flag_{N,cand\_sw}$  を用意する。図2のルールAでは感覚語が主格補語になるパターンしか抽出できないため、この要素はルールBでのみ扱う。

$$Flag_{N,cand\_sw} = \begin{cases} 0 & (\text{片パターンでのみ抽出}) \\ 1 & (\text{両パターンで抽出}) \end{cases} \quad (2)$$

### 5 実験

#### 5.1 正解データの作成

五感知識の一般性スコアの正解データを作成するため、被験者11名(20代男性7名(理系学生4名, 会社員3名), 20代女性会社員4名)に五感知識の一般性の高さを1~5の数値で評価してもらった。

提示した知識は3.1のAのルールを用いて抽出された五感知識である。Aのルールによって獲得された五感知識を抽出数ごとに見てみると抽出数1の知識が非常に多く、五感カテゴリごとに見ると sight に関する知識が非常に多いことがわかる(図3, 4参照)。この中からランダムに被験者に知識を提示するとデータに偏りが生じてしまうため、被験者にはランダムに選択した100件の知識とデータの偏りを抑えるために制約を加えてランダムに追加した62件の知識を提示した。なお、制約とは各感覚カテゴリで抽出数が1~10件, 10~100件, 100件以上の知識の中から、それぞれ5個ずつ出力されるようにしたものである。

これら162件のデータのうち、複数のカテゴリを持つ知識を除いた142件のデータの評価値を正規化したものを正解データとした。その結果一般性の高さを表すスコアは0.7~0.8になる知識が最も多く、その平均値は0.559となった(図5参照)。正解データの具体例として「(桜, sight, 綺麗)や(ガス, smell, 臭い), (水, touch, 冷たい)」などの知識はスコアが1に近い値となり、「(家具, taste, 渋い), (目, hearing, 不快), ((きび, touch, 重い)」などの知識はスコアが0に近い値となった。

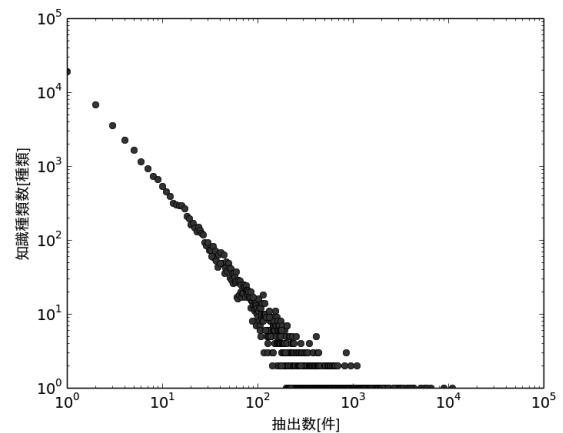


図3 抽出数と知識種類数の関係

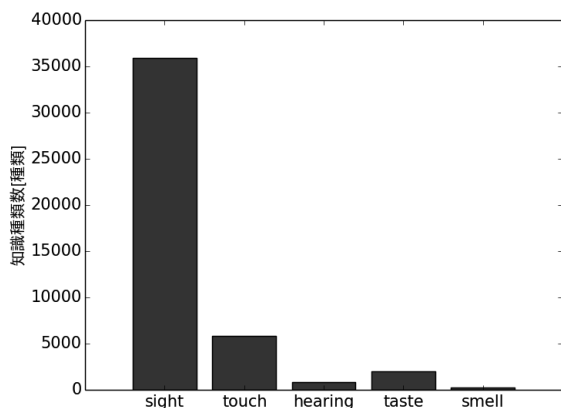


図4 獲得知識の五感カテゴリと知識種類数の関係

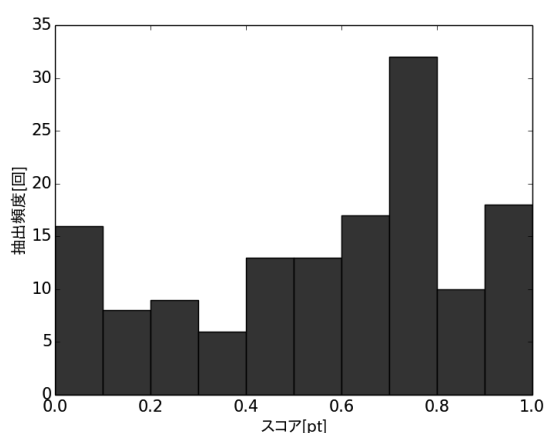


図5 正解データのスコア分布

## 5.2 比較方法

最小二乗法により学習を行い, RMSE (二乗平均平方根) で評価値の予測精度の比較を行う. RMSE は 0~1 の値を取り, 0 に近いほど誤差が小さいことを意味する. RMSE は(3)式で表される. ただし  $N$  は全予測対象数,  $y_i$  は実測値,  $\hat{y}_i$  は予測値を表す.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

検証方法は one-leave-out 交差検証を用いる. ベースラインには正解データの値をそのまま学習させてスコアを出力するシステムを用い, 提案手法として3章で紹介した要素を組み合わせたスコアとの比較を行う. 各スコアの要素は重みを加えて次のように表記する. ただし  $x, \alpha, w_1, w_2, w_3$  は 0~1 の変数である.

$$\cdot Base = x \quad (4)$$

$$\cdot CP = w_1 \cdot CP_{N,sw}^{\alpha} \quad (5)$$

$$\cdot Prop = w_2 \cdot Prop_{NC,swc} \quad (6)$$

$$\cdot Flag = w_3 \cdot Flag_{N,Cand_{sw}} \quad (7)$$

## 5.3 実験結果

表 1 に示すように, RMSE の値はベースラインで 0.288 となり, ルール A では CP+Prop で最小値 0.236, ルール B では CP+Prop+Flag で最小値 0.224 となった. CP, Prop は単体でベースラインよりも評価値の予測精度が高く, これらを組み合わせることで精度が 4.8~5.7 ポイント上がっていることが確認された.

ルール B にのみ適用可能な Flag は単体ではベースラインよりも予測精度が 9.1~9.5 ポイントも低く, CP, Prop のいずれかと組み合わせても, それぞれの精度向上には役に立たなかった. しかし CP, Prop, Flag 全てを用いることで予測精度は最大となり, ベースラインよりも 6.4 ポイント高い値となった.

## 6 考察

CP の結果から, ルール A とルール B を比較すると, 構文解析を用いたルール B の方が評価値の予測精度が高い結果となっていることがわかる. これは, 構文解析を用いた手法の方が名詞と感覚語候補の距離が遠く離れたペアも抽出でき, 知識の抽出漏れが少なくなるためであると考えられる. またベースラインよりも精度が上がっていることから, 一般性と共起頻度の間にはある程度の相関性があると考えられる.

また, 感覚語の統一処理がある場合とない場合を比較してみると, 全ルール, 全スコアにおいて精度が向上していることから, 感覚語の統一処理は一般性評価に有効であったと考えられる.

次にルール B で感覚語の統一処理を行なったときのシステムの出力と人間が付与した評価値の分布をしてみる. この分布図では切片 0, 傾き 1 の直線の分布に近いほど精度が高いことを意味する.

まず  $\alpha, w_1$  の値が 1 のときの CP の分布(図 6 参照)を見てみるとシステムの出力値は非常に小さいが, 人間の評価値が高くなるほどシステムの出力値も緩やかに上昇していることがわかる. ここで変数  $\alpha$  を学習させることにより, この特性を生かしたままシステムの出力値が小さい問題を解決することができ, ベースラインより精度が高い結果となっていると考えられる(図 7 参照). しかし,  $\alpha$  を設定することにより, 一般性が低い知識に関する誤差が大きくなってしまっている. この問題を解決したのが Prop である.

$w_2$  の値が 1 のときの Prop の分布(図 8 参照)を見てみると, 人間の評価値と全体的に誤差は大きい, 評価値の低い知識と高い知識に関してはある程度人間が付与した評価値に近い値を出力していることがわかる. CP+Prop では特に評価値の低い知識に対し Prop が有効に働いたことで, CP の問題を緩和することができている(図 9 参照).

表1 各スコアでの RMSE の値

ルール	A		B	
感覚語の統一処理	なし	あり	なし	あり
<i>Base</i>	0.288	0.288	0.288	0.288
<i>CP</i>	0.266	0.262	0.256	0.252
<i>Prop</i>	0.273	0.273	0.273	0.273
<i>Flag</i>	-	-	0.383	0.379
<i>CP + Prop</i>	0.240	0.236	0.236	0.231
<i>CP + Flag</i>	-	-	0.259	0.257
<i>Prop + Flag</i>	-	-	0.314	0.317
<i>CP + Prop + Flag</i>	-	-	0.226	<b>0.224</b>

※太字は最小値であることを示す.

表2 一般性の誤差が大きい知識上位 5 件

knowledge	system	human	diff
(夜更け, hearing, 静か)	0.345	0.910	-0.565
(末っ子, sight, 新しい)	0.600	0.0900	+0.510
(獵師, sight, 深い)	0.571	0.0675	+0.504
(子供, sight, 鋭い)	0.564	0.0675	+0.497
(にきび, touch, 重い)	0.495	0.0225	+0.473

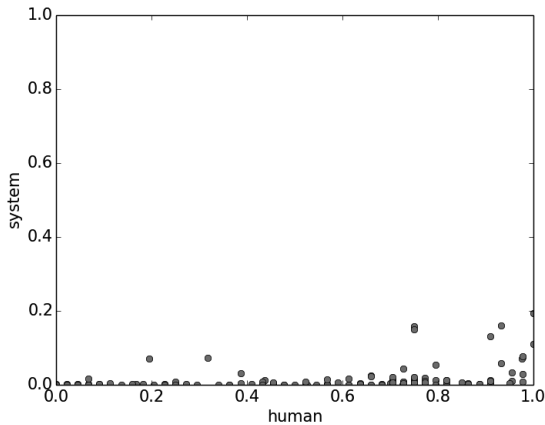


図6  $w_1, \alpha = 1$  のときの CP

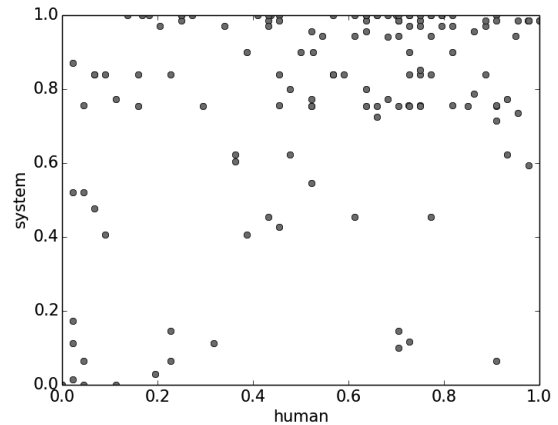


図8  $w_2 = 1$  のときの Prop

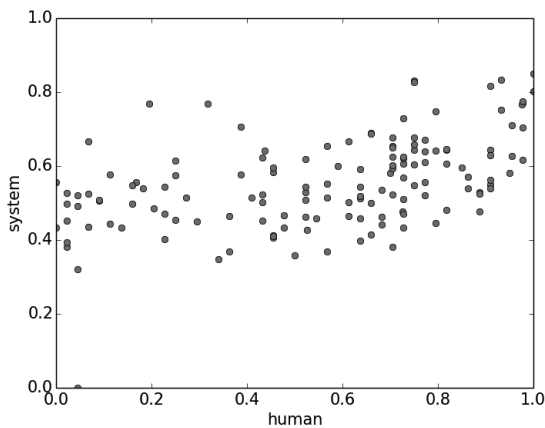


図7 重み学習後の CP

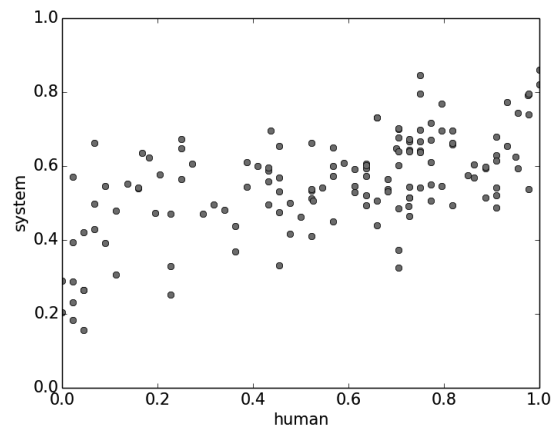


図9 重み学習後の CP+Prop

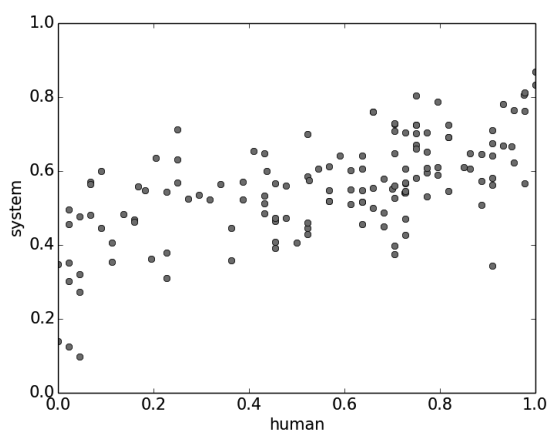


図10 重み学習後の CP+Prop+Flag

Flag は 0 か 1 のみしか出力しないため、単体で用いると精度は低いが、CP+Prop+Flag で RMSE が最小となっている。特に評価値の低い知識に対し Flag が有効に働いていると考えられる(図 10 参照)。このときの重みの平均値は  $w_1 = 0.515$ ,  $w_2 = 0.355$ ,  $w_3 = 0.130$  となっている。このことから一般性の評価値予測には CP と Prop が大きく貢献していることがわかる。

最後に RMSE が最小となったルール B で感覚語の統一処理を行なったときの CP+Prop+Flag について、人間の評価値とシステムの出力値の差が大きくなった知識を見てみる(表 2 参照)。

「(夜更け, hearing, 静か)」は「夜更け」が hearing に関する特徴を持ちにくい「時間カテゴリ」に属しているため、Prop によって大きくスコアが下がっている。「(末っ子, sight, 新しい), (猟師, sight, 深い), (子供, sight, 鋭い)」は sight に関する特徴を持ちやすい「人カテゴリ」に属しているため、Prop によってスコアを落とすことができなくなっている。この問題は五感特徴保有率作成時の粒度を「sight-新しさ, sight-深さ, など」のようにすることで改善されると考えられる。「(にきび, touch, 重い)」も同様であるが、「にきび」が「動物カテゴリ」に属しているということも問題である。このようなカテゴリの間違いは「にきび」の類似単語が属するカテゴリから訂正することができると考えられる。

## 7 まとめと今後の課題

共起頻度と意味的特徴を用いて五感知識の一般性を評価する手法を提案した。その結果ベースラインより約 5 ポイント上回る精度で一般性の評価値を予測することができた。さらに経験則による要素を加えることで、最大でベースラインより 6.4 ポイント精度を改善することができた。

今回の実験で意味的特徴として用いた五感特徴保有率の有効性が確認されたことから、このような情報を

自動で構築することが今後の課題として挙げられる。

また経験則をスコアに換算することは難しいが、精度向上にわずかながら貢献していることから、このような要素についても今後検討していきたい。

最後に今回の手法ではルールベースでの抽出を前提しているため、ルールベース以外の方法で知識を抽出し、より多くの知識とその一般性を評価できるような手法を確立することが今後の課題として挙げられる。

## 参考文献

- [1] 三橋奎太, ジェプ・ラファウ, 荒木健治: 構文解析を用いた常識的感覚データベース構築のための知識獲得, 第 16 回情報科学技術フォーラム (FIT2017) 講演論文集, 第 2 分冊, pp.185-186, 2017.
- [2] 渡部広一, 堀口敦史, 河岡司: 常識的感覚判断システムにおける名詞からの感覚想起手法, 人工知学会論文誌, No.19, pp.73-82, 2004.
- [3] 林良彦: 概念間の想起の強さと方向性の予測, 2016 年度人工知能学会全国大会 (JSAI2016), 3O2-OS-04a-5, 2016.
- [4] Miller, G.A and Fellbaum, C.: WordNet then and now. Language Resources and Evaluation, Vol.41, No.2, pp.209-214, 2007.
- [5] Mikolov, T., et al.: Distributed Representations of Words and Phrases and their Compositionality. Proc. of Neural and Information Processing System (NIPS2013), pp.3111-3119, 2013.
- [6] Rothe, S., and Schutze, H.: AutoExtend: Extending word embeddings to embeddings from synsets and lezemes. Proc. of Association for Computational Linguistics (ACL2015), pp.1793-1803, 2015.
- [7] 工藤拓: MeCab: Yet another part-of-speech and morphological analyzer. 2006.
- [8] 黒橋・河原研究室: Juman++, <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>