

語概念連想を用いた常識的感覚・知覚想起システムの構築

芋野 美紗子 吉村 枝里子 土屋 誠司 渡部 広一

同志社大学 理工学部, 大学院 工学研究科

{mimono, eyoshimura}@indy.doshisha.ac.jp

{stsuchiy, hwatabe}@mail.doshisha.ac.jp

1 はじめに

ある事柄に対して、多くの人間が共通に持つ知識や思考を「常識」と呼ぶ。常識は種々の事物・事象に対して存在しており、「このように言われれば普通はこう感じる」のような感情の機微、「この行動は普通これぐらいの時期に行う」のような時間感覚といったように様々である。人間は自身、もしくは他者の経験により、常識に関する知識を蓄積し、その基盤を確立している。人間同士の関わり・コミュニケーションにおいて、互いがこのような種々の常識を理解することで、我々は柔軟な解釈や判断を行っていると考えられる。そしてこのような人間らしい「常識」を機械的に理解するような機構が存在すれば、人間とコンピュータとの関わりもより違和感の無いものになるのではないかと考える。

本稿では様々な「常識」のうち、感覚・知覚に着目する。国語辞書の定義によれば、感覚とは「目・耳・鼻・舌・皮膚などによって感じ取った外部の刺激」、知覚とは「外部の事物を認識することで頭の中に思い浮かぶもの」である。外部に存在する何かしらの事物・事象から、五感や心に想起されるものを感覚・知覚と考える。例えば普通、人間は「言い逃れ」という事象に対して「醜い」のようなものを想起すると考えられる。このような感覚・知覚を人間らしい常識に沿って想起する機構の構築を目指し、研究 [1][2] が行われてきた。本稿ではこれらの研究結果を元に、ある1つの事物・事象をあらゆる語から常識的な感覚・知覚を表す語を出力するシステムを提案する。

2 語概念連想システム

語概念連想システムとは、人間のように柔軟な語句の意味理解を行うための機構である。語の意味を定義した「概念ベース」および関連性の定量化を行う「関

連度計算方式」により語概念連想システムは構築される。本稿ではこの語概念連想システムの機構を用い、知識ベースに存在しない名詞の入力に対しても適した感覚・知覚語を出力することを目指す。

2.1 概念ベース

概念ベース [3] は複数の電子化国語辞書などの見出し語を概念として定義し、人間が持つ概念への常識的な知識をモデル化した知識ベースである。ある概念の意味定義は、属性と呼ばれる他の概念群と属性それぞれの重要さを表す重みによってなされる。概念ベースの具体例を表 1 に示す。

表 1: 概念ベースの具体例

概念	属性
蝸牛	(腹足類,1.62)(螺旋,0.71)(梅雨,0.35)...
梅雨	(紫陽花,1.14)(季語,0.49)(蝸牛,0.48)...

概念ベースにおいて属性に現れる語句は、全て概念として定義されている。そのため、概念「蝸牛」の意味定義を行う属性「梅雨」も概念ベースにおいて意味定義がなされている。このような意味定義の連鎖的な繋がりにより、より人間らしい語句の意味定義が可能となる。

2.2 関連度計算方式

関連度計算方式 [4] は概念と概念の関連性を関連度とよばれる数値で定量的に表現する手法であり、その有効性が示されている。関連度は 0.0 から 1.0 の値を取り、概念間の関連が強いほど大きな値を示す。例えば概念「自動車」と概念「車」の関連度は 0.912 と非常に大きな値を示す。一方、概念「自動車」と概念「学

校」の関連度は0.012となり、関連が薄い概念同士では小さな値となる。関連度は概念同士の属性の対応により算出される。互いが持つ属性の内、最も意味が近いもの同士の組を作った上でそれぞれの重みを用いて関連度を算出する。

3 提案システム

図1に本稿で提案する感覚・知覚想起システムの流れを示す。

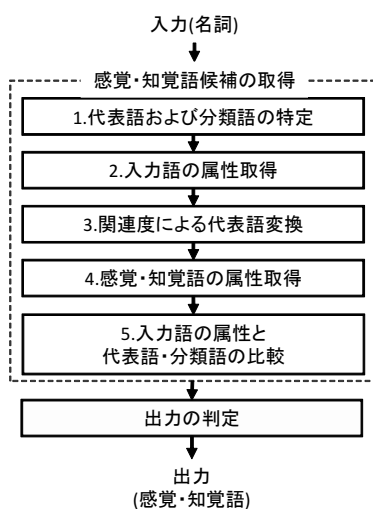


図1: 提案システムの流れ

システムでは入力名詞から、5つの処理により感覚・知覚語の候補を取得する。これらの処理で得られた候補語に対して最終出力とするかの判定を行い、感覚・知覚語を想起する。

3.1 代表語および分類語の特定

この処理では感覚・知覚語知識ベースにおいて定義されている「代表語知識」および「分類語知識」に、入力名詞と表記一致する知識が存在するか検索を行う。表2および表3にそれぞれの知識の具体例を示す。

代表語は一般的な名詞とそこから想起する感覚・知覚語のセットを人手で格納したものである。分類語はシソーラス[5]のノードを元に作成されたもので、これを用いて各代表語を分類している。また、分類語のものにも固有の感覚・知覚語が存在しており、それが分類語知識において定義されている。イメージおよび五感それぞれ、感覚語と知覚語に付随して定義されるものである。五感は感覚語がどの部位によ

表2: 代表語知識の具体例

代表語	分類語	感覚・知覚語	イメージ	五感
差別	別け隔て	つらい	-	×
優しさ	善意	親切的な	+	×
百合	花	白い	×	視覚
向日葵	花	大きい	×	視覚

表3: 分類語知識の具体例

分類語	感覚・知覚語	イメージ	五感
別け隔て	不公平	-	×
善意	有難い	+	×
花	美しい	×	視覚

て感じ取れるものかを、イメージは知覚語が与える一般的な印象の方向をプラスとマイナスのどちらかで表現している。

この処理では、まず「代表語の特定」処理として入力名詞が代表語知識において代表語として定義されているかを検索する。表記一致するものが存在するならば、対応する感覚・知覚語を候補として取得する。次に「分類語の特定」処理として、入力名詞が代表語であった場合にはその代表語が分類されている分類語を取得する。もし代表語では無い場合には、分類語知識において分類語として定義されているかを検索する。代表語でもなく、分類語にも存在しない場合には入力名詞をシソーラス上で検索し、その1つ上のノードが分類語として存在するかを検索する。以上の検索において表記が一致する分類語が得られた場合には、分類語知識を用いて対応する感覚・知覚語を取得する。

3.2 入力語の属性取得

概念ベースにおいて入力名詞が概念として定義されている場合、その属性を取得する。属性内に知識ベース内で感覚・知覚語と定義されている語が存在した場合にはそれを候補として取得する。図2に例を示す。

例では入力名詞として「恋人」を想定する。概念ベースにおいて概念「恋人」の属性には「彼氏、恋する、大切な」といった語が存在している。ここで「大切な」という属性は感覚・知覚語知識ベースにおいて知覚語として定義されているため、これを候補として取得する。

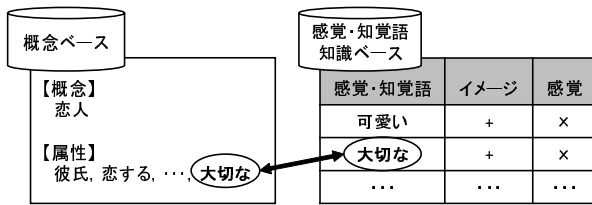


図 2: 入力語の属性取得例

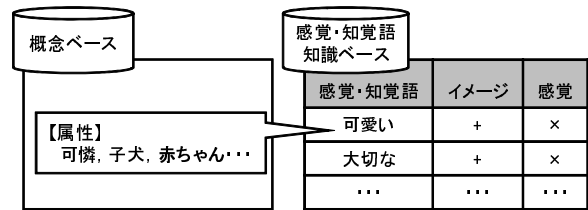


図 4: 感覚・知覚語の属性取得例

3.3 関連度による代表語変換

入力名詞と代表語との関連度を算出し、値が閾値以上の場合にはその代表語が入力名詞と変換できる語であると判断してその感覚・知覚語を取得する。図 3 に例を示す。

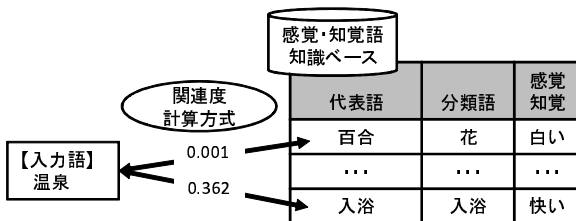


図 3: 関連度による代表語変換例

ここでは入力名詞「温泉」と全ての代表語との関連度を算出している。このとき、最も関連度が高くなる代表語として「入浴」を得る事が出来る。本稿では実験的に求めた数値として閾値 0.3 を採用し、入力名詞との関連度が最大かつ 0.3 以上の代表語が持つ感覚・知覚語を候補として取得する。例では「快い」を候補として得る事が出来る。

3.4 感覚・知覚語の属性取得

感覚・知覚語知識ベースに登録されている全ての感覚・知覚語の属性を取得し、その属性内に入力名詞があった場合には属性の取得元の感覚・知覚語を候補として取得する。例を図 4 に示す。

例では入力名詞が「赤ちゃん」であったとする。このとき、感覚・知覚語知識ベースにおいて定義されている知覚語「可愛い」の属性に「赤ちゃん」が存在している。よって属性の展開元である知覚語「可愛い」が候補として取得される。

3.5 入力語の属性と代表語・分類語の比較

概念ベースにおいて入力名詞が概念として定義されている場合、その属性を取得する。属性中に、感覚・知覚語知識ベースにおいて代表語もしくは分類語として用いられている語が存在した場合には、その代表語もしくは分類語に付随する感覚・知覚語を候補として取得する。例を図 5 に示す。

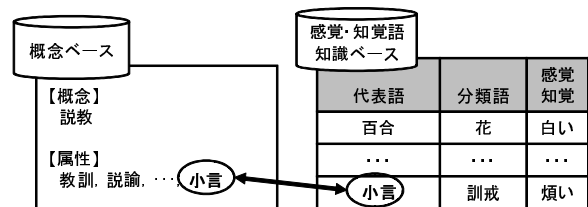


図 5: 入力語の属性と代表語・分類語の比較例

入力名詞として「説教」が与えられたとすると、その属性には「教訓, 説論, 小言」といった語が存在している。このとき、属性「小言」が感覚・知覚知識ベースにおいて代表語として用いられている。そこでこの代表語に付随する感覚語「煩い」を候補として取得する。

3.6 出力の判定

ここまでで取得された感覚・知覚語の候補について最終的に出力語とするかの判定を行う。判定基準として、まず 3.1 節, 3.2 節, 3.3 節に示した 3 つの取得処理において得られた候補語についてはそのまま出力語とする。残った 3.4 節, 3.5 節で得られた候補語については、この 2 つの処理において共通して得られた語のみを出力語と判定する。例を図 6 に示す。

例では節 3.5 の手法で得られた語のうち、「面倒な」「煩わしい」の 2 語が 3.4 節の手法では取得できていない。そのため、最終出力には双方で取得された「面白い」が残る。また、「不思議な」に関しては 3.1 節の

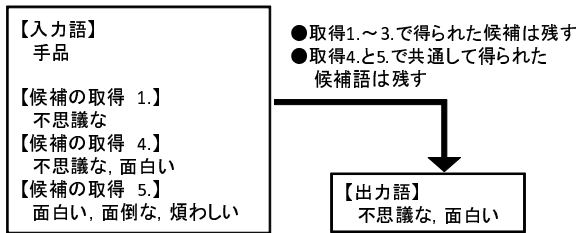


図 6: 出力の判定例

手法において取得されているため、最終出力となる。

4 評価

評価セットとして名詞 200 語をシステムへ入力し、その出力を被験者 3 名の目視により評価する。1 つの名詞から複数の感覚・知覚語が出力される場合には、それぞれについて評価を行う。それぞれの被験者は出力された語に対して「正しい」か「誤り」の二段階で評価を行う。この評価の結果を 3 名分で統合し、全員が正しいとした場合を○、2 名が正しいとした場合を△+, 1 人が正しいとした場合を△-, 全員が誤りとした場合を×として最終評価とする。表 4 に評価結果を示す。

表 4: 評価結果

○	△+	△-	×
71.0%	15.2%	9.3%	4.4%

結果として○の精度で 71.0%、多数決となる△+を合わせると 86.2%の精度となった。表 5 に入出力と評価の例を示す。

表 5: 入出力と評価例

入力語	出力語 (感覚・知覚語)	評価
温泉	快い	○
言い逃れ	醜い	○
金魚	可愛い, 美しい	○, △+
不器用	醜い, 苛立たしい	△-, ○
時間	大切な, 新しい	○, ×

入力名詞「温泉」の例では、感覚・知覚知識ベースの知識のみでは感覚・知覚を想起する事が出来ない。

語概念連想システムによって、知識として定義されている「入浴」を得る事で、適した想起を行えている例である。また「言い逃れ」のような抽象を表す名詞でも、適切な感覚・知覚語を出力することが出来ている。

入力名詞「時間」に関しては概念ベースの属性展開において「新しい」という語が取得された。これらの語が概念と属性の関係にあることは、概念ベースの定義からすれば間違いとはいえないが、「時間」という名詞 1 つから想起する感覚・知覚としては違和感があるという結果になった。

5 おわりに

本稿では感覚・知覚を人間らしい常識に沿って想起する機構の構築を目指し、ある 1 つの事物・事象をあらゆる語から常識的な感覚・知覚を表す語を出力するシステムの提案を行った。人間が人手で作成できる範囲の知識ベースを元に、語概念連想システムを用いることで様々な入力に対して柔軟に判断を行う機構を構築した。最終的な精度は 71.0%となり、人間の目視による評価で適切だと判断される「常識的な出力」を多く得ることが出来た。

謝辞

本研究の一部は、科学研究費補助金 (若手研究 (B) 24700215) の補助を受けて行った。

参考文献

- [1] 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. 人工知能学会論文誌, Vol.19, No.2, pp.73-82, 2004.
- [2] 米谷彩. 常識的知覚判断システムの構築. 第 17 回人工知能学会全国大会論文集, 2003.
- [3] 奥村紀之, 土屋誠司, 渡部広一, 河岡司. 概念間の関連度計算のための大規模概念ベースの構築. 自然言語処理, Vol.14, No.5, pp.41-64, 2007.
- [4] 渡部 広一 奥村 紀之 河岡 司. 概念の意味属性と共起情報を用いた関連度計算方式. 自然言語処理, Vol.13, No.1, pp.53-74, 2006.
- [5] NTT コミュニケーション科学研究所. 日本語語彙大系. 岩波書店, 1997.