

概念連想に基づく職種判断システムの構築

Constructing Occupation Judgment System based on a Word Association System

大江 奈緒子
Oe Naoko

渡部 広一
Watabe Koichi

河岡 司
Kawaoka Tsukasa

*1 同志社大学大学院工学研究科

Department of Knowledge Engineering and Computer Sciences, Graduate School of Engineering, Doshisha University

This paper proposes a method of "Occupation Judgment System," which responds appropriate types of occupations for users' requests. Owing to the difficulty of inputting all knowledge, this system needs to expand knowledge to react to unknown words from representative words and a Word Association-System.

1. はじめに

人間同士が行うコミュニケーションをとっても、生活の中の社会活動や知的活動のような様々な活動を支えており、極めて重要な役割を果たしていると言える。コミュニケーションをコンピュータが自由自在に行えるようになれば、マンマシン・インタフェースを飛躍的に向上させることができ、人に優しく知的な活動を行えるコンピュータを実現することができると思われる。コミュニケーションには意図を理解した円滑な会話や柔軟な判断が必要であり、その判断の一つのモデルとして、ユーザーの要求を満たす職種を返すシステム(職種判断システム)を提案した。

2. 職種判断システム

職種判断システムとは、ユーザーの要求文に対して適切な職種を返すシステムである(図1)。

入力	出力
本を借りたい	→ 図書館
本を買いたい	→ 書店
ほうれん草を買いたい	→ 八百屋, スーパー
泳ぎたい	→ プール

図1: 入出力例

このシステムに必要な不可欠なのは“本を借りるなら図書館”といった人間だけが持っている知識である。従って、本システムでは人手で作成した「職種ベース」という職種に関する知識ベースを用いる(表1)。しかし、実際には知識をすべて登録することは困難であり効率が悪い。そこで、代表的な語のみを登録し、ソーラスや概念ベースにより構築した連想システムを用いて、知識の拡張を行い、職種ベースに登録されていない語に対しても対応できるようにする。

ソーラスとは一般名詞の意味的用法を表す約 2700 の意味属性の上位下位関係、全体部分の関係を木構造で示したものであり、約 13 万語が登録されている。例を示すと、“ビール”の上位(親)の位置に存在するのは“酒”である。

概念ベースとは国語辞書等から自動構築された汎用データであり、語(概念)と意味(属性)のセットを約9万語分蓄積している。また、属性にはその概念に対する重要さの度合いである、重みが付与されている。概念「コーヒー」を例にすると、概念「コーヒー」に対して属性「コーヒー」、「コーヒー豆」、「カフェ」などが存在し、それぞれに重み 0.97, 0.88, 0.45 が付いている。また、概念と概念の関連の深さを定量的に表す手法を関連度計算[渡部 2001]といい、値を関連度という。関連度は 0 から 1 までの連続値で、例えば、ビールとウイスキーの関連度は 0.351 などと表される。

2.1 職種ベース

職種ベースは次の4つのフィールドで構成されている(表1)。“職種名”には 423 の店名および職業名が登録されている。“動詞”には「現像する」のように、特定の職種のみに関連する動詞が登録されている。“キー動詞”は職種がどのような役割をしているかを“銀行, スポーツ, 勉強, 宿泊, 購入, 治療, 貸出, 鑑賞, 食事, 修理, 贈答, 売却, 建設, 余暇”の 14 語で表している。“商品名”(最大 10 個)は扱っている商品またはサービスが登録されている。

表1: 職種ベース

職種名	動詞	キー動詞	商品名 1	商品名 10
酒屋		販売	酒	おつまみ
中華料理店		食事	中華料理	餃子
ペットショップ	飼う	購入	ペット	動物

2.2 システムの流れ

入力「○○を△△したい」があったとすると、○○に当たる目的語を「要求名詞」、△△に当たる動詞、形容詞、サ変名詞を「要求動詞」と呼ぶ。“銀行, スポーツ, 勉強, 宿泊, 購入, 治療, 貸出, 鑑賞, 食事, 修理, 贈答, 売却, 建設, 余暇”の中で、要求動詞との関連度が最も高いものを「キー動詞」とする。また、キー動詞の導入により助長な語の登録を抑え、職種ベースに登録されていない動詞に対しても処理を行うことができるという。

以下に職種判断システムの処理の流れを示す。

- ① 入力文から要求名詞, 要求動詞を取り出す。
- ② 要求動詞をキー動詞に置換する。
- ③ フィールド“動詞”に動詞が含まれる職種に候補を絞る。
- ④ 職種ベースで“商品”に要求名詞が含まれる職種を検索する。適切な職種がなければ⑤を行う。
- ⑤ ソーラスを用いた処理(2.3)を行う。適切な職種がなければ⑥を行う。
- ⑥ 概念ベース(2.4)を用いた処理を行う。

①～⑥の順に行い、④～⑥で適切な職種が見つかった時点で終了する。適切な職種がない場合、動詞で候補を絞らずにキー動詞で候補を絞り再度④～⑥をたどり適切な職種を探す。それでも職種が見つからなかった場合、動詞による絞り込みを行わず要求名詞のみで④～⑥の処理を行う。

また、職種ベースに存在しない語を「未知語」と呼び、処理⑤、処理⑥を「未知語処理」と呼ぶ。

2.3 ソーラスを用いた処理

処理内容は要求名詞を、そのソーラスの親と置き換え、④の処理を行う。具体的に入力文「焼酎を買いたい」を例に説明する。

要求名詞は「焼酎」、キー動詞は「購入」である。今、キー動詞が「販売」である職種に商品「焼酎」は存在しないので、処理④では職種が得られず、処理⑤であるシソーラスを用いた処理が行われる。要求名詞「焼酎」のシソーラスの親は図2より「酒」であるので、要求名詞を「酒」として④の処理を行う。その結果、職種「酒屋」を出力として得ることができる。

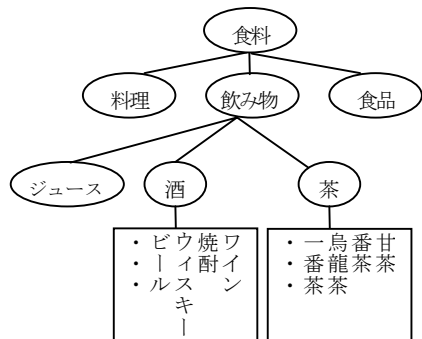


図2：シソーラス（部分）

2.4 概念ベースを用いた処理

全職種中の“商品”（代表語も含む）の中で、要求名詞との関連度が最大であり、かつ値が 0.16 以上（過去の研究により実験的に示された閾値）となる商品を持つ職種を出力する。

例を図3に示す。入力が「月刊誌を買いたい」であった場合、「買う」のキー動詞は「購入」なので、職種ベースのキー動詞“購入”と一致するものに絞る。絞った全職種で「月刊誌」と商品との関連度の最大値を出し、その値が最大のもを出力する。この場合は、出力は「書店」である。

職種名「書店」

商品名	「月刊誌」との関連度
本	0.1644
コミック	0.0021
書籍	0.3133
洋書	0.1589
雑誌	0.6800
漫画	0.1001

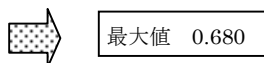


図3：「書店」の最大値

3. 入力形態

ここまで、入力形態が要求名詞と要求動詞（もしくは要求名詞のみ）に形態素解析される要求文に限られていたが、“犬の病気を治したい”、“辛い食べ物が食べたい”といった複合系の入力も考えられる。これらの入力は特によく行われるので拡張を行った。

(1) 名詞+名詞+動詞(ex.犬の病気を治したい)

「犬の病気を治したい」という入力例では、「犬の」の部分で使用されず、「病気を治したい」と同様の入力になるため「病院」という出力になる。この問題を解決するために、「犬」、「病気」それぞれの属性、重みを合わせて使用し、関連度計算を行うことにより「動物病院」との関連度が高くなる手法を用いた。

(2) 形容詞+名詞+動詞(ex.辛い食べ物が食べたい)

すでに、常識判断システムという人間の常識を持たせたシステムが構築されており、そのひとつに感覚判断システム[堀口2002]がある。感覚判断システムはある名詞から想起される五感に関する語（形容詞、形容動詞）を出力するシステムである。ま

た、このシステムは形容詞、形容動詞から名詞（連想語）を想起することもできる。このシステムを用いて「辛い食べ物」→「カレー、キムチ」と連想できるので、構築した本システムに要求名詞として使用することができる。

4. 評価

人手で、要求名詞と要求動詞に構文解析できる要求文を500データ作成して評価を行った。評価は人が判断し適当だと思われるものを正解、正解を含まないものを不正解、正解を含むものと3段階で行う。

各処理の出力の割合と各処理で出力されたものを図5、図6に示す。

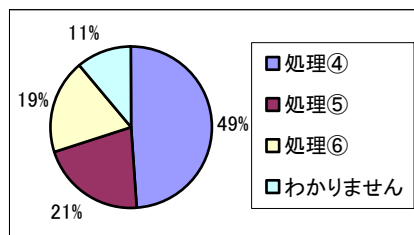


図5：各処理の出力された割合

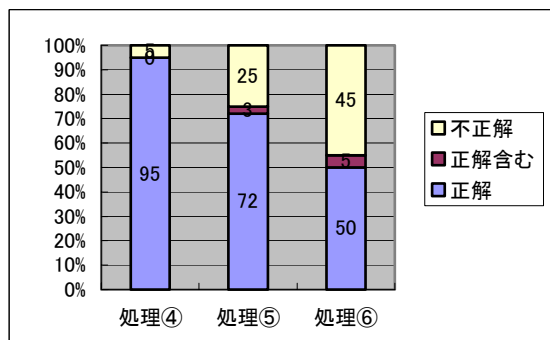


図6：各処理で得られた出力の精度

職種判断システムの精度（正解と正解を含むの合計）は、73%となる。また、概念ベースに登録されていない未定義語の45個を評価からのぞくと精度は80%となる。

5. おわりに

要求に対して適切な職種を返すことができる職種判断システムを構築することができた。しかし、今後の課題として未知語処理である処理⑥の精度向上がさらに望まれる。また、現段階では人間の要求を「～したい」というものに限定して研究を進めているが、様々な入力にも対応することが必要であると考えられる。例をあげると、「お腹が痛い」といった状態文などのような入力文形態や「モンブラン」、「ブッチャプリン」といった概念ベースに登録されていない未定義語である一般名詞、固有名詞も扱えるようになることが今後の課題である。

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における研究の一環として行った。

参考文献

[渡部 2001] 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, Vol.8, No2, pp.39-54, 2001.
 [堀口 2002] 堀口敦史, 渡部広一, 河岡司: 常識的感覚判断システムの構築, 情報処理学会研究報, 2002-ICS-130, pp.31-36, 2002.