

## 概念連想に基づく職種判断システムの構築

## -状態文処理機能の拡張-

Constructing Occupation Judgment System based on a Word Association System  
- Extension of State Sentence Processing Function -川波 正典\*<sup>1</sup>  
Kawanami Masanori大江 奈緒子\*<sup>1</sup>  
Oe Naoko渡部 広一\*<sup>1</sup>  
Watabe Hirokazu河岡 司\*<sup>1</sup>  
Kawaoka Tsukasa\*<sup>1</sup> 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 study aims at realization of an intelligent computer gentle to man. For that purpose, it is necessary to make a smooth conversation and the commonsense and flexible judgment system which understands the intention of man on a computer. The occupation judgment system which returns a suitable occupational description to a user's demand sentence as one model of the judgment mechanism is proposed. In this research, first, the conventional occupation judgment system is explained and the problem is explored. And improvement is added to an occupation judgment system, and some techniques for considering as a more intelligent system are proposed.

## 1. はじめに

人間同士が行うコミュニケーションは極めて重要な役割を果たしており、生活の中の社会活動や知的活動など様々な活動を支えている。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事ができ、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事ができる。そして、コミュニケーションには人間の意図を理解した円滑な会話や常識的で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、ユーザの入力を満たす職種を返す『職種判断システム』が提案されている[大江 2003]。本研究では、従来の職種判断システムに改良を加え、そのシステムの精度向上を研究目的とし、より知的なシステムとする為のいくつかの手法を提案する。

## 2. 職種判断システム

職種判断システムとは、ユーザの入力に対して適切な職種を返すシステムである。このシステムに必要な不可欠な事は“本を借りるならば図書館”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、概念ベース[渡部 2001]やシソーラスより構築した連想システムを用いて、知識の連想を行い、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。

概念ベースとは語(概念)と意味(属性)のセットを約9万語分蓄積している国語辞書等から自動構築した汎用データである。概念は、ある語 A をその語と関連の強いと考えられる語(属性)  $a_i$  と重み  $w_i$  ( $>0$ ) の対の集合として定義される。属性数  $n$  は概念ごとによって異なる。

$$\text{概念 } A = \{(a_1, w_1), (a_2, w_2), \dots, (a_n, w_n)\}$$

ここで、 $a_i$  を概念 A の1次属性と呼ぶ。任意の1次属性  $a_i$  は、必ず概念ベース内に含まれている。つまり、属性を表す語もまた概念ベース上で概念として定義されている。従って、1次属性それぞれを1つの概念と見た場合、1次属性からさらにそれぞれの属性を導ける。これを2次属性と呼ぶ。概念ベースにおいて概念は  $n$  次までの属性連鎖集合により定義されている。図1に概念「自転車」の属性と重みの様子を示す。

概念	属性, 重み		
自転車	自転車, 1.000	二輪車, 0.853	… 乗る, 0.034

図1: 概念「自転車」の属性と重み

また、概念 A と概念 B の関係の深さを定量的にあらわすのが関連度計算[渡部 2001]という方法である。それぞれの概念が持っている属性と重みによって関連度計算は行われ、その結果は関連度という数値で表すことができる。関連度は0以上1以下の実数で表され、関連度が高いものが関連の深い語ということになる。図2に概念「自転車」の関連度計算の様子を示す。

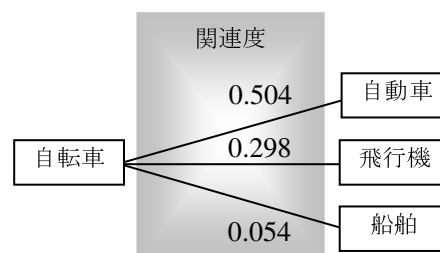


図2: 「自転車」との関連度

関連度計算を用いることにより、知識に持っていない語(未知語)を知識内にある関連の深い語に置き換えることができる。これにより未知語も扱うことができ知識の拡張が行われる。

シソーラスとは広く同義語、類義語を整理したもので、約2700の意味属性の上位下位関係、全体部分関係が木構造で示されたものである。約13万語が登録されており、親子・兄弟関係を持つ語についてはその関係も記述している。例えば、「酒」と「ビール」は親子関係、「ビール」と「ウイスキー」は兄弟関

連絡先: 同志社大学工学研究科 知識情報処理研究室

〒610-0394 京都府京田辺市多々羅都谷 1-3

Tel : 0774-65-6944

係である、シソーラスの親とはシソーラス上で上位(親)に当たる関係のものを言う。

## 2.1 職種ベース

職種ベースは次の 5 個のフィールドで構成されている(表 1)。  
[職種名]には 423 個の店名および職業名が登録されている。  
[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, 貨出, 鑑賞, 食事, 修理, 贈答, 売却, 建設, 余暇』の 14 個で表している。[動詞]には 214 個の特定の職種のみに関連のある動詞が登録されている。  
[商品名]は職種で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(一部)

職種名	キー動詞	動詞	商品名	
居酒屋	食事	酔う	忘年会	ビール
病院	治療		医療	熱

## 2.2 従来の職種判断システムの流れ

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』, △△に当たる用言を『要求動詞』と呼ぶ。キー動詞候補 14 個の中で、要求動詞との関連度が最も高いものを『キー動詞』とする。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要求名詞・要求動詞のセットで処理を行う。要求動詞に含まれる職種に候補を絞る。
  - ④職種ベースを参照し、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要求名詞を、そのシソーラスの親と置き換え、④の処理を行う。適切な職種がなければ⑥を行う。
  - ⑥要求名詞と[商品名]との関連度をとり、最も関連度が高く、かつその値が 0.160 以上(過去の研究により実験的に示された閾値)となる職種を検索する。適切な職種がなければ⑦を行う。
  - ⑦要求名詞・キー動詞のセットで処理を行う。キー動詞に含まれる職種に候補を絞る。処理の流れは④～⑥である。適切な職種がなければ⑧を行う。
  - ⑧要求名詞のみで処理を行う。処理の流れは④～⑥である。適切な職種がなければ、無回答となる。
- ①～⑧の順に処理を行い、④～⑥で適切な職種が見つかった時点で終了とする。

## 2.3 感覚判断システムの利用

ある入力「□□な○○を△△したい」といった入力についての処理方法を説明する。例えば、「辛い食べ物を食べたい」のような「形容詞+名詞+動詞」の入力に対して、感覚判断システム[堀口 2002]を利用する。このような入力に対して、知的に“辛い食べ物とはキムチやカレー”と想起してくれれば、「辛い食べ物が食べたい」は「キムチ, カレーが食べたい」と入力を置き換えることができ、「韓国料理店, カレー屋」という知的な出力ができる。この“辛い食べ物とはキムチやカレー”という想起を感覚判断システムにより行うことができる。感覚判断システムとは、ある名詞から想起される五感に関する語である、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。感覚判断シス

テムを利用した処理の流れは、入力「□□な○○を△△したい」と入力があったとき、形容詞□□と名詞○○を感覚判断システムにかけ、連想語を取得する。連想語を要求名詞とし、①～⑧の順に処理を行い、④～⑥で適切な職種が見つかった時点で終了とする。

## 3. 従来の職種判断システムの分析

従来の職種判断システムについて分析を行う。このシステムでの、対応可能な入力文と、その精度について調べる。また、そこから従来の職種判断システムの問題点について考える。

### 3.1 対応可能な入力文

従来の職種判断システムで対応可能な入力文は、「○○を△△したい」といったような要求文で、大きく分けて「目的語+動詞」と「動詞のみ」の 2 つに分けられる。また、「目的語+動詞」は、「修飾のない要求文」、「形容詞(修飾語)を含む要求文」、「所有格を含む要求文」の 3 つに分けられる。具体的には次に示すような入出力がある。

- 目的語+動詞
  - 普通の要求文
    - 「本を買いたい」 → 書店, 古本屋
    - 「餃子が食べたい」 → 中華料理店, ラーメン屋
  - 形容詞を含む要求文
    - 「冷たい飲み物を飲みたい」 → 喫茶店, 酒屋, 居酒屋
    - 「辛い食べ物を食べたい」 → 韓国料理店, カレー屋
  - 所有格を含む要求文
    - 「猫の病気を治したい」 → 動物病院
    - 「時計の修理がしたい」 → 時計屋
- 動詞のみ
  - 「引っ越したい」 → 運送屋
  - 「泳ぎたい」 → プール

### 3.2 要求文に対する精度

入力文によって職種の出力が望まれる様々な形態の要求文、サ変名詞文を人手で 500 個作成し、それを評価データとして、従来の職種判断システムの評価をとった。結果を図 3 に示す。○は正解を表し、△は複数ある出力のうち 1 個でも正解を含むもの、×は無回答または不正解である。精度は(○のパーセンテージ)+(△のパーセンテージ)で表す。従来の精度は 85.4% という結果になった。

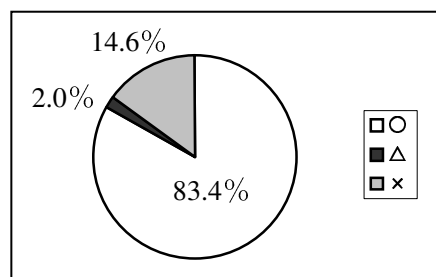


図 3: 従来の職種判断システムでの要求文に対する精度

### 3.3 従来の職種判断システムの問題点

従来の職種判断システムは、要求文に対して 85.4% という非常に高い精度で対応している。そこで、入力形態の拡張の為、状態文に対しての精度を調べる。また、状態文は「風邪をひいた」等の『職種を連想できる状態文』と、「笑った」等の『職種を連想できない状態文』の 2 種類に分けられると考え、その 2 種類について精度を調べる。

○は正解, ×は不正解とし, 精度は○のパーセンテージで表す. なお, 評価データは人手で作成した 357 個(連想できる: 200 個, 連想できない: 157 個)である. 結果を図4に示す.

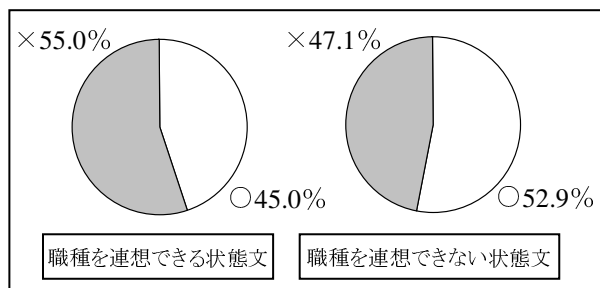


図4: 従来の職種判断システムでの状態文に対する精度

図4からわかるように, 職種を連想できる状態文の精度は, 45.0%と非常に低い. これは, 従来の職種判断システムが状態文に対応していないので当然の結果だといえる. また, 職種を連想できない状態文に対しても, 職種を出力してはならないのに, 約半分が職種を出力している. 具体的には次に示すような入出力がある.

■ 職種を連想できる状態文

正解 「風邪をひいた」 → 病院, 小児科, 内科  
不正解 「腕が折れた」 → 出力なし

■ 職種を連想できない状態文

正解 「笑った」 → 出力なし  
不正解 「目が覚めた」 → 眼科

この従来の職種判断システムが状態文に対応していないのに, これだけの出力をする理由としては, 2.2 節で説明した“[要求名詞]のみ”の処理(⑧)の為だと考えられる. 従来の職種判断システムは, “[要求名詞]+[要求動詞]”, “[要求名詞]+[キー動詞]”の処理を経て, 出力が出なかった場合, “[要求名詞]のみ”の処理により職種が出力される. つまり, ある状態文に対して, 名詞と動詞の関係を見つけないことができず, ただ単に, 名詞のみで職種を連想していることになる. これでは, 状態文の名詞と動詞の関係を考慮できないことから, コンピュータが人間の意図を知的に理解し職種を判断しているとはいえない. したがって, 状態文の名詞と動詞の関係を考慮させる為に, 状態文を対応させる新たな処理が必要であるといえる. また, “職種を連想できない状態文”を対応させるため, 職種に関係あるか, ないかの判断もさせる必要があるといえる. システムをより良いものとする為にこれら2点の問題を改善する.

4. 対策動詞

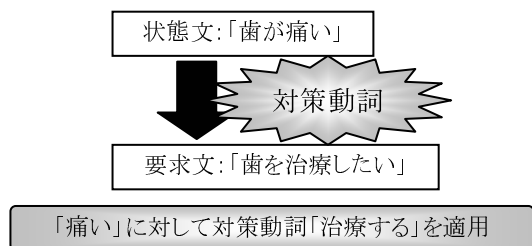


図5: 対策動詞を利用した入力文の変換

従来の職種判断システムは要求文に対して非常に高い精度で対応している. そこで, 本システムでは, 状態文を処理する際に状態文から要求文への変換を行う. この変換において, 『対

策動詞』を利用する. 例えば, 状態文「歯が痛い」に対して, “歯が痛いので, 歯を治療する”といった連想から, 「痛い」の対策動詞として「治療する」を利用する. そして, 要求文「歯を治療したい」と変換する事により, 従来の職種判断システムに対応させる事が可能となる. 図5に対策動詞を利用した入力文の変換の様子を示す.

4.1 対策動詞ベース

対策動詞ベースは, 職種判断システムで状態文から要求文へと変換する際に用いる知識ベースであり, 『特殊対策動詞ベース』, 『一般対策動詞ベース』の2つの知識ベースから構成される.

特殊対策動詞ベースは, 要求文から状態文へと変換する際, 対策動詞のみでは変換が不可能な特殊な文に対して利用する. 例えば, 状態文「お腹が空く」に対して, 動詞のみではなく, 目的語も変化させ, 要求文「食べ物を食べたい」と変換する. このような状態文に対して, この知識ベースを利用する. [特殊代表名詞]には入力目的語にあたる語, [状態詞]には入力の用言にあたる語が登録されている. [特殊対策名詞]には目的語と変換する語, [対策動詞]には用言と変換する語が登録されている. 知識数は36個である(表2).

表2: 特殊対策動詞ベース(一部)

特殊代表名詞	状態詞	特殊対策名詞	対策動詞
汗	かく	風呂	入る
腹	空く	食べ物	食べる

一般対策動詞ベースは, 特殊対策動詞ベースで対応できなかった入力に対して利用する. 例えば, 状態文「テレビが壊れた」に対して, 動詞のみを変換して, 要求文「テレビを修理したい」と変換する. [一般代表名詞]には入力目的語にあたる代表的な語, [状態詞]には入力の用言にあたる語が登録されている. [一般対策名詞]には目的語を変換する語, [対策動詞]には用言を変換する語が登録されている. [親]には[一般代表名詞]のシソーラスの親が登録されている. 知識数は104個である(表3).

表3: 一般対策動詞ベース(一部)

一般代表名詞	親	状態詞	一般対策名詞	対策動詞
物	人工物	壊す		修理する
人間	人間	失神する	怪我	治療する

5. 入力判断メカニズム

対策動詞により状態文を要求文へと変換し, 従来の職種判断システムで状態文を判断させることが可能となった. そこで, 入力文が要求文か状態文か, さらに, 状態文ならばそれが職種を連想できる文かできない文かを判断させる入力判断メカニズムを提案する. 処理の流れを以下に示す.

- ①入力文から要求名詞, 要求動詞を取り出す.
- ②文末の品詞が助動詞で「～たい」ならば, その入力文を要求文と判断し, 従来の職種判断システムの処理へと進む. できなければ, その入力文は状態文と判断し, ③へと進む.
- ③特殊対策動詞ベースを参照する. ここで状態文から要求文への変換ができれば, 従来の職種判断システムの処理へと進む. できなければ, ④へと進む.
- ④一般対策動詞ベースを参照するここで状態文から要求文への変換ができれば, 従来の職種判断システムの処理へと進む. できなければ, 「職種に関係ない入力です」と出力する.

③, ④の処理で, 名詞と名詞, 動詞と動詞の関連度を利用するが, それぞれの閾値は 0.160 以上, 0.140 以上(研究により実験的に示された閾値)である。

## 6. 評価

対策動詞, 入力判断メカニズムを利用した本職種判断システムと従来の職種判断システムの状態文に対する精度を比較する。評価データは 3.3 節で使用したものを利用する。職種を連想できる状態文についての結果を図6, 職種を連想できない状態文についての結果を図7に示す。本職種判断システムにおいて, 職種を連想できる状態文については, 正しい職種を出力した場合のみを○(正解)とし, 職種を連想できない状態文については, 「職種を連想できない入力文です」と出力した場合のみを○(正解)とする。精度は正解のパーセンテージで表す。

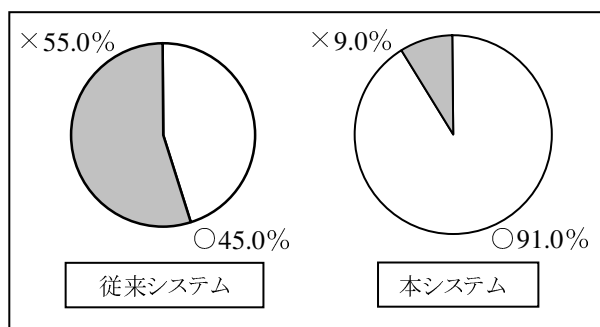


図6: 職種を連想できる状態文

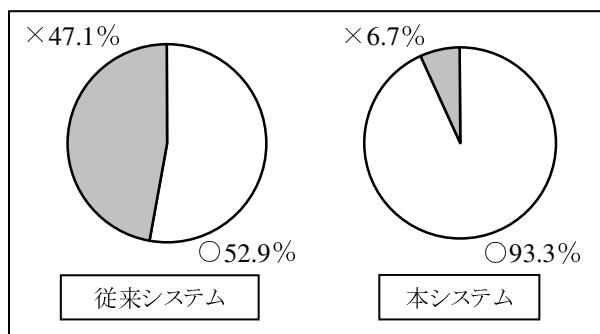


図7: 職種を連想できない状態文

図6より, 職種を連想できる状態文に対して, 本システムの精度が従来のシステムの精度よりも 46.0% 向上し, 精度からも, 本システムが状態文に対して常識的な判断がほぼ可能になったといえる。また, 図7より, 職種を連想できない状態文に対して, 本システムの精度が従来のシステムの精度よりも 40.4% 向上し, 精度からも, その入力が職種に関係あるか, ないかという常識的な判断がほぼ可能になったといえる。さらに, 職種を連想できる状態文は, 知識を多く登録すれば精度が向上するのは当然だが, 職種を連想できない状態文は, 知識が多くなれば精度は下がる。しかし, 共に高い精度を得ていることから, 対策動詞ベースにより良い知識を登録できたといえる。この結果, 今回提案した対策動詞, 入力判断メカニズムが職種判断システムの知的化に大いに貢献できたといえる。成功例と失敗例を以下に示す。

### ■ 成功例

#### ● 「腕が折れた」

従来のシステム : 出力なし

本システム : 病院, 外科, 整形外科, 整骨院

#### ● 「交通事故に遭った」

従来のシステム : 脳神経外科(目的語のみの判断)

本システム : 病院, 外科

### ■ 失敗例

#### ● 「寝坊した」 : 病院, 外科

(「失神する」と関連度が 0.564 の為)

正解 → 「職種に関係のない入力です」

#### ● 「耳が聞こえない」 : 職種とは関係のない入力(※)

正解 → 耳鼻咽喉科

※本職種判断システムの問題点として, 「耳が聞こえない」といったような否定形の状態文に対応できないという点がある。この入力の場合であると, 形態素解析によって目的語を「耳」, 動詞を「ない」と取得する。その結果, この入力は「耳がない」といった文となり, 職種に関係のない入力だと判断されてしまう。そこで, 形態素解析を行った際の動詞「ない」の手前の動詞を取得するようにすると, この入力の場合であれば, 「聞こえる」が取得され, 目的語「耳」, 動詞「聞こえる」として処理を行うことができる。しかし, それでは「耳が聞こえる」といった文となり, これもまた職種に関係のない入力だと判断されてしまう。結果, 本職種判断システムは否定形の状態文に対応できないという問題点がある。

## 7. おわりに

本稿では, 知的な判断メカニズムを用いて, コンピュータに常識的かつ柔軟で適切な会話を実現させる為の一つのモデルとして, ユーザの入力に対して適切な職種を返す職種判断システムの研究報告を行った。

従来の職種判断システムは要求文のみに対応しており, 「腕を折った」などの状態文には対応していなかった。そこで, 状態文を要求文へと変換する手法を考え, 対策動詞を提案した。これにより, 「腕を折った」という状態文から, 「骨折を治療したい」という要求文へと変換でき, 従来の職種判断システムで状態文を処理できるようにした。また, 入力判断メカニズムを提案し, 職種判断システムに関係のある入力かどうかの常識的な判断もほぼ可能にすることができた。このように, 対策動詞, 入力判断メカニズムを提案することにより, 職種判断システムをより知的化することに成功したといえる。

今後の課題としては, 研究の過程において, 入力に対する適切な職種が職種ベースにないものも多く存在しており, 職種ベース自体の拡張の必要性があると言える。また, 人間と会話を行う上で, 会話の中に多く固有名詞が使われることから, 固有名詞に対応したシステムを構築する必要があると考えられる。さらに, 職種を連想できない要求文や否定形の状態文に対応できないという問題点の改善も必要だと考える。0

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

- [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, 自然言語処理, Vol.8, No.2, pp.39-54, 2001.
- [大江 2003] 大江奈緒子, 渡部広一, 河岡司: 概念理解に基づく職種判断システムの構築, 第17回人工知能学会, 2003.
- [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, 情報処理学会研究会資料, 2002-ICS-130, pp.31-36, 2002.