

話者の個性・嗜好情報を考慮したコンピュータ会話処理
Computer Conversational Processing System Based on Individuality and Personal Preference Information

吉岡 孝治[†]
Koji Yoshioka

吉村 枝里子[†]
Eriko Yoshimura

渡部 広一[†]
Hirokazu Watabe

河岡 司[†]
Tsukasa Kawaoka

1. はじめに

現在、コンピュータが目覚ましい発展を遂げ、人間の生活に欠かすことのできないものとなってきている。コンピュータとの会話において個性・嗜好情報が重要である。

本稿では、会話文中に出現する個性・嗜好を表現する語を整理し、自動的に個性・嗜好情報を知識ベースとして構築する手法を提案している。具体的には入力文の目的語と用言の関係に注目し、必要な個性・嗜好情報の知識を獲得する手法を提案する。さらに応答処理として、個性・嗜好情報に基づいて話者の関心が高いと思われる Web ニュース記事を用いた話題の提案、さらに話者の詳しい個性・嗜好情報を獲得するための質問生成手法を提案する。

2. 関連事項

2.1 概念ベース

概念ベース^[1]は、複数の国語辞書や新聞等から機械的に構築した、語(概念)とその意味特徴を単語(属性)の集合のセットを約 12 万語構築した知識ベースである。概念 A に付与される属性 a_i には、その重要性を表す重み w_i が付与されている。

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

属性 a_i を概念 A の一次属性と呼ぶ。これに対し a_i を概念とした場合の属性を A の二次属性と呼ぶ。

2.2 関連度計算方式

関連度計算方式^[2]は、概念に定義された語と語の関連の強さを、定量的に評価する手法である。関連度の値は 0~1 の実数値をとり、値が高いほど関連が強いことを表す。

2.3 シソーラス

本研究で利用するシソーラス^[3]は、一般名詞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、意味属性であるノードと、ノードに属している単語であるリーフから構成されている。

2.4 オートフィードバック

オートフィードバック^[4](以下、AF とする)は、概念ベースに登録されていない語である未定義語 X の意味的特長を表す属性とその重要性を表す重みの組を Web を用いて自動的に構成し、概念ベースに追加する。

3. 話者理解システム

話者理解システム^[5]とは、嗜好情報を含んだ入力文の情報を人が持っている「名前」、「趣味」、「好きな食べ物」などの 47 項目のうち、どの項目に該当するかを判断するシステムである。システムの実行例を表 1 に示す。またこのシステムにより作成される情報と項目がセットになった知識ベースを話者理解知識ベースと定義する。

[†] 同志社大学大学院工学研究科

Graduate School of Engineering, Doshisha University

表 1 話者理解システムの実行例

入力文	情報	項目
私は林檎が好きだ	林檎	好きな食べ物
私は音楽鑑賞が好きだ	音楽鑑賞	趣味

4. 嗜好応答システム

4.1 嗜好応答システムの概要

本稿で提案する嗜好応答システムとは会話文中から自動的に嗜好情報を獲得し、嗜好情報を考慮した会話の展開を行う会話システムである。嗜好情報の獲得のため既存の話者理解システムを応用し、話者理解知識ベースを整理した。

応答処理として、Web から構築したニュース記事応答、詳しい嗜好情報獲得のための AF 質問語応答、そして聞き返し応答の 3 つのパターンを提案する。

嗜好応答システムは全体として、入力、用言の分類、情報の抽出、嗜好の判断、応答の選択、応答の計 6 つの部分から構成されている。次節より詳細を説明する。

4.2 用言の分類

入力文の用言の種類を 3 つのパターンに分類する。分類の種類として、好き・嫌いが明確な用言(好き・嫌いグループ)、好き・嫌いを曖昧に示唆する用言(プラス・マイナスグループ)、それ以外の用言(その他のグループ)に分類する。分類方法は人手により作成された知識ベース(以下、用言知識ベースとする)と用言の関連度の値と閾値を基に判別する。用言知識ベースの例を表 2 に示す。また格納した語数は 62 語である。

表 2 用言知識ベースの例

グループ名	用言
好き・嫌いグループ	好き, 好きだ, 嫌いだ, よい, ...
プラス・マイナスグループ	得意だ, おいしい, 苦手だ, まずい...

4.3 情報の抽出

嗜好情報においてキーワードとなる語を情報と定義する。入力文によって情報が現れる場所は異なる。そのため入力文を Who・What・When・Where・Why・How・Whom・用言のフレームに分割し情報を抽出する。Who の親ノードに「人」が含まれる場合、What を情報と見なす。また含まれない場合、Who を情報と見なす。

4.4 嗜好の判断

獲得した情報が嗜好情報に関する語かどうかを判断する。この判断では話者理解の各項目にシソーラスを用いた条件を付与することにより判断する。また趣味の項目についてシソーラスによる条件付けが困難であるため Wikipedia の「趣味の一覧」から趣味を表現する語を獲得した。付与する条件の例を表 3 に示す。

用言のグループと情報が嗜好かどうかを用いて嗜好の判断をする。用言が好き・嫌いグループに属する場合は話者理解システムにより嗜好情報を獲得する。用言がプラス・マイナスグループの場合は情報が項目条件を満たせば嗜好情報に関係のある入力文と判断する。またそれ

以外の入力文は嗜好に関係ない文となる。

表3 項目条件の例

項目	設定条件
食べ物	情報がノード「食料」に存在し、「飲料」に存在しない
スポーツ	情報がノード「スポーツ」に存在

4.5 応答の選択

話者の嗜好情報に関係のある語が入力された場合は、話の展開を促す必要がある。そのため嗜好に関するニュース記事を用いた応答を出力する。入力文に対しニュース記事が得られない場合に AF 質問語応答に移る。そしてこの応答文も得られない場合に聞き返し応答へと移る。

獲得した情報がすでに嗜好情報の知識として存在するかどうかは話者理解を参照してデータの比較を行う。

4.6 応答

4.6.1 ニュース記事応答の判断とその応答

情報より連想されるニュース記事があるかの判断を行う。ニュース記事は Web 上のニュースサイトの見出しを獲得し知識ベース化 (以下、ニュース記事知識ベースとする) する。

情報からの連想の流れとして、ニュース記事知識ベースよりニュース記事の自立語を獲得し、情報と自立語の関連度の値を求める。閾値以上の記事を応答文に用いる。

4.6.2 AF 質問語応答の判断とその応答

詳しい嗜好情報を獲得するために質問固有の語 (以下、質問語とする) を用いた質問文を生成する。質問語とは情報から連想される詳しい嗜好情報の項目と定義する。例えばサッカーの質問語として「選手・チーム・リーグ」が考えられる。AF 質問語応答作成の流れを図1に示す。応答文に用いる質問語を獲得する段階で、質問語の候補となる語を候補語として獲得する。候補語の獲得法については以下で説明する。また質問語を連想するため代表語から連想される質問語を格納した代表質問語知識ベース (249 語) と、一般的に質問語となる語を格納した一般質問語知識ベース (19 語) の例を表4、表5に示す。

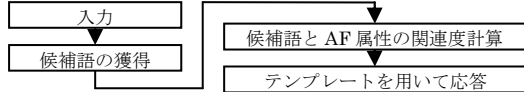


図1 AF 質問語応答作成の流れ

表4 代表質問語 KB の例

カテゴリ	代表質問語
競馬	レース, 騎手
林檎	産地, 味

表5 一般質問語 KB の例

一般質問語
産地, 種類, チーム, リーグ, 試合, 種類...

候補語の作成には以下の2つのパターンで行う。

- A) 情報と代表質問語 KB のカテゴリの関連度が閾値以上
候補語 = 閾値以上の語の質問語 + 一般質問語
- B) 情報と代表質問語 KB のカテゴリの関連度が閾値以下
候補語 = 一般質問語

質問語を連想し、以下の2つのテンプレートを用いる。

- a) 質問語の親ノードに「人」が存在する場合
好きな[質問語]はいますか?
- b) 質問語の親ノードに「人」が存在しない場合
好きな[質問語]はありますか?

4.6.3 聞き返し応答

聞き返し応答では以下の2つのパターンを用いる。

① 選択型質問

情報がシソーラスのリーフに存在する場合に生成する。

テンプレートとしては「情報[情報]は好きですか?」を用いる。
② 疑問詞型質問

情報がシソーラスのノードに存在する場合に生成する。テンプレートとしては「どんな情報[情報]が好きですか?」を用いる。

5. 実験評価

話者の嗜好情報を与えたデータを用意し、テストセット 200 文を用いて評価を行った。システムは入力文に対し、妥当であると考えられることのできるニュース記事の見出し、AF 質問語応答、聞き返し応答を出力する。評価方法は3人の被験者に応答文が適切かどうかを○, △, ×の3段階で評価した。最終的な評価として3人の評価の多数決の結果を出力した。3人とも異なる場合は△とした。システムの結果を図2に、成功例を表6に示す。

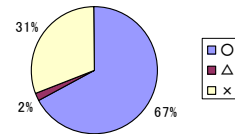


図2 嗜好応答システムの評価結果

表6 成功例

入力文	出力結果
私は肉が好きだ	吉野家: ステーキ最王手「どん」 買取交渉 牛肉調達で相乗効果
スケートは楽しい	フィギュア: 浅田真央ら公式練習… GP ファイナル
私はメロンが好きだ	好きな産地はありますか?

6. 考察

成功例より各応答文で適切な応答が選択されていることがわかる。ニュース記事選択ではサッカーとラグビーとの関連度の値で誤りが起きた。スポーツの観点で2つのスポーツは意味的に近いのだが、嗜好情報という観点では近くない。またニュース記事には固有名詞を多く含むため、固有名詞を有しない概念ベースを用いた関連度計算で誤りが生じていた。

7. おわりに

本稿では、個性・嗜好情報の整理方法、話者の個性・嗜好情報を考慮した応答を提案した。常識知識を用いた応答に加え、新たに会話に必要な個性・嗜好情報を整理したことにより応答の幅が広がったと考えられる。

このシステムを構築したことにより人間と円滑にコミュニケーションをとることのできるコンピュータ会話の実現に近づいたと考えられる。

参考文献

- [1] 奥村紀之, 北川晋也, 渡部広一, 河岡司, “概念ベースの分析と構築”, 同志社大学理工学研究所報告, Vol.46, No.3, pp.133-141, 2005.
- [2] 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.
- [3] NTT コミュニケーション科学研究所監修, 「日本語語彙体系」, 岩波書店, 東京, 1997.
- [4] 辻泰希, 渡部広一, 河岡司, “www を用いた概念ベースにない概念の属性候補自動獲得”, 第18回人工知能学会全国大会, 2D1-01, 2004.
- [5] 八尾旬扇, 渡部広一, 河岡司, “連想システムを用いた話者理解”, 言語処理学会第11回年次大会発表論文集, D1-6, pp.125-128, 2005.