

用言に着目した話者の嗜好推測システムの検討 Consideration of Speaker's Preference Estimation System Focusing on Declinable Word

山本 若波† Wakaba Yamamoto
土屋 誠司‡ Seiji Tsutiya
渡部 広一‡ Hirokazu Watabe

1. はじめに

現在のロボットは業務用のロボットが主流であるが、今後求められるのは人間と親密なコミュニケーションをとることができ、人間のパートナーとして活躍するロボットである。

人間とのコミュニケーションにおいては、相手の趣味や好みなどを理解することで、相手に対して親しみを持つことが重要である。そのため、ロボットが話者との会話から嗜好を取得することが必要であると考えられる。

また、会話中の用言には「好き」などの嗜好を直接的に表現している語や、「美味しい」などの嗜好を曖昧に示唆する語が存在する。そこで本稿では、会話に出現する用言に着目した、話者の嗜好推測システムの検討を行う。

2. 関連技術

2.1 概念ベース

概念ベース^[1]は電子化された国語辞書などから自動的に構築された知識ベースである。ある語を概念と定義し、概念の特徴を表す語である属性と、属性の重要さを表す重みの対で構成されている。概念ベースの例を表1に示す。

表1 概念ベースの例

概念	属性
医者	(医師,0.34)(患者,0.11)(病院,0.08)・・・
病院	(医院,0.25)(手術,0.11)(施設,0.04)・・・
治す	(治療,0.43)(医療,0.21)(病気,0.13)・・・

2.2 関連度計算

関連度計算方式^[2]とは、ある二つの概念間の関連の強さを定量的に表現する手法である。関連度は0から1の間で変動し値が大きいほど関連は強くなる。

概念A, Bの持つ属性を a_i, b_j , 重みを u_i, v_j とし、属性数をL個, M個とすると、概念A, Bは次のように表される。

$$A = \{(a_1, u_1), (a_2, u_2), \dots, (a_L, u_L)\} \quad (1)$$

$$B = \{(b_{x_1}, v_{x_1}), \dots, (b_{x_M}, v_{x_M})\} \quad (2)$$

これより、概念A, Bの一致度 $DoM(A, B)$ と関連度 $DoA(A, B)$ は以下のように表される。なお、 $a_i = b_j$ は属性 a_i と b_j が一致したことを表す。

$$DoM(A, B) = \sum_{a_i=b_j} \min(u_i, v_j) \quad (3)$$

$$DoA(A, B) = \sum_i DoM(a_i, b_{x_i}) \times \frac{u_i + b_{x_i}}{2} \times \frac{\min(u_i, v_{x_i})}{\max(u_i, v_{x_i})} \quad (4)$$

† 同志社大学大学院理工学研究科

Graduate School of Science and Engineering, Doshisha University

‡ 同志社大学理工学部

Faculty of Science and Engineering, Doshisha University

2.3 MeCab

MeCab^[3]とは、入力された文を、意味を持つ最小単位である形態素に分割し、各形態素に品詞を付与する形態素解析機である。

2.4 用言知識ベース

用言知識ベース^[4]とは、好き・嫌いを明確に示唆する用言と、曖昧に示唆する用言がグループ化された知識ベースである。

2.5 シソーラス

シソーラス^[5]とは、広く同義語、類義語を木構造で整理したもので、本稿では約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NTTシソーラスを用いる。

2.6 会話システム

会話システムとは^[6]、人間と同様に自己開示をする発話を行うこと、また対話の展開の仕方をランダムにすることによってより多様性を持たせることで、人間とのより自然な雑談型対話の実現を目指したシステムである。

3. システム概要

システムのフローを図1に示す。以下でそれぞれの処理について述べる。

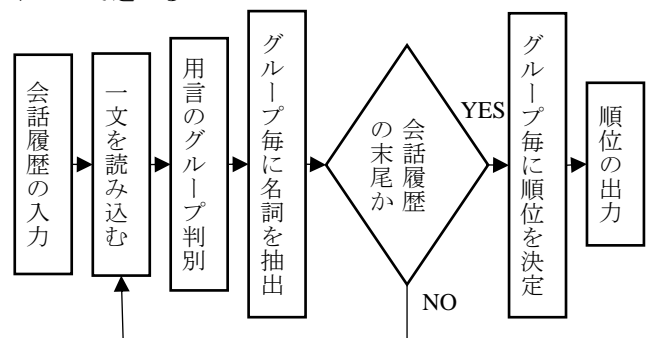


図1 システムのフロー

3.1 会話履歴の入力

図2のような話者の会話履歴を入力する。

User: こんにちは
User: 週末に焼肉に行きます
User: 焼肉を食べに行きます
User: 焼肉が好きです
User: 家族です

図2 話者の会話履歴

3.2 一文を読み込む

会話履歴の上から一文ずつ読み込む。

図 2 の場合はまず「こんにちは」を読み込む。次の一文を読み込む際には「週末に焼肉にいきます」を読み込む。

3.3 用言のグループ判別

入力文を形態素解析し、その結果と用言知識ベースを用いて入力文に嗜好情報が含まれているか、またどのグループに属するかを判別する。

図 2 の会話履歴が入力されれば、4 行目の「焼肉が好きです」という文に嗜好情報が含まれていると判断し、好きグループに分類する。

3.4 グループ毎に名詞を抽出

3.3 節で判別されたグループ毎に入力文から形態素解析で一般名詞となった名詞を抽出する。

例として、「焼肉が好きです」という文から「焼肉」が好きグループの名詞群として抽出される。

3.5 グループ毎に順位を決定

3.4 節で抽出された名詞と会話履歴のすべての自立語を用いて関連度計算を行い、総和を求める。総和を用いて名詞をソートし、総和の大きい順に順位を付ける。その後、シソーラスを用いて親ノード名を取得し、親ノード名によって 4 つの項目に分類する。分類後、項目毎の順位を決定する。4 つの項目とそれに対応する親ノード名を表 2 に示す。

表 2 項目と親ノードの対応

項目	親ノード
食べ物	食料
スポーツ	スポーツ
動物	動物
娯楽	娯楽, 楽器

「焼肉」であれば、親ノードに「食料」が含まれているので、「食べ物」に分類される。

3.6 順位の出力

出力は項目分類前の順位と、4 つの項目毎の好みを上位 3 位までのランキング形式で出力する。

4. 評価

本システムはコンピュータが人間と同じように話者の嗜好を推測するシステムであるため、システムが出力した嗜好と被験者の嗜好が合致するか否かで評価する。被験者として 20 代男性 5 人に対して評価を行った。

4.1 評価手法

被験者にアンケートを行うことで、被験者の「食べ物・スポーツ・動物・娯楽」の嗜好を決定する。そして被験者は会話システムを用いて、それぞれの嗜好について 10 回入力するまで会話する。得られた 40 回分の入力を会話履歴としてシステムに入力する。被験者の嗜好がシステムの出力の何位になるかによって評価を行う。被験者の嗜好が出力の 1 位であれば◎、2 位であれば○、3 位であれば△、出力されなければ×とする。

4.2 評価結果

図 3 にシステムの評価結果を示す。

◎が 50%、○が 10%、△が 5%、×が 35% となった。

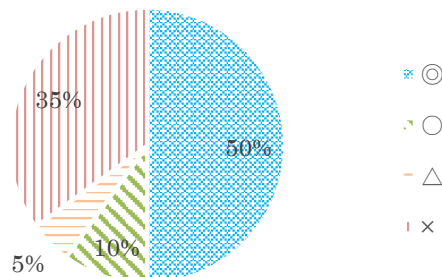


図 3 評価結果

5. 考察

システムの出力として被験者の嗜好が出力された割合は 65% となった。しかし、今回の評価では、被験者の嗜好が多義語となる場合が多く見受けられた。このような場合、シソーラスによる項目分類が正確に行えておらず、×の割合が増えてしまっていると考えられる。

また、会話履歴とシステムの出力を解析すると、会話の流れや質問に対する返答で、会話に嗜好物が出現した場合に、その嗜好物を抽出できていないことが分かった。そのため会話の前後関係なども考慮すると、より良いシステムを構築できるのではないかと考える。

6. おわりに

本稿では、会話の用言に着目した嗜好推測手法の検討を行ったが、問題点が多数見つかり、あまり満足できるものとはならなかった。しかし本システムは、用言知識ベースや関連度計算を基に嗜好推測した効果として、嗜好として取得できる情報の正確さや整合性が高いと考えられ、問題点を改善することで、より正確な嗜好推測が可能であると考えられる。

謝辞

本研究の一部は、JSPS 科研費 16K00311 の助成を受けて行ったものです。

参考文献

- [1] 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.
- [2] 井筒大志, 渡部広一, 河岡司, “概念ベースを用いた連想機能実現のための関連度計算方式”, 情報科学技術フォーラム FIT2002, pp.159?160, 2002.
- [3] “MeCab -- 形態素解析器”, <http://taku910.github.io/mecab/>, 京都大学情報学研究所-日本電信電話株式会社コミュニケーション科学基礎研究所, 2019-2-11 参照.
- [4] 吉岡孝治, “嗜好情報を考慮したコンピュータ会話処理”, 同志社大学理工学部卒業論文, 2008.
- [5] シソーラス NTT コミュニケーション科学研究所, “日本語語彙体系”, 岩波書店, 1997.
- [6] 金子駿, 吉村枝里子, 土屋誠司, 渡部広一, “話題を考慮した自然な会話システムの構築”, 研究報告知能システム, 183, No2, pp1-8, 2016.