

PRINTEPS におけるロボット喫茶店の注文時ユーザモデリングの実現

User modeling for Robot Café Interaction in PRINTEPS

眞鍋 麟太郎^{*1}
Manabe Rintaro

高瀬 裕^{*2}
Takase Yutaka

西村 良太^{*3}
Nishimura Ryota

中野有紀子^{*2}
Nakano Yukiko

^{*1} 成蹊大学大学院
理工学研究科
Graduate School of Science and
Technology, Seikei University

^{*2} 成蹊大学
理工学部情報科学科
Dept. of Computer and Information
Science, Seikei University

^{*3} 慶応義塾大学大学院
理工学研究科
Keio University
Graduate School of Science and Technology

There are many situations that social robots can serve, such as shopping malls, care facilities, and museums. As one of such situations, this paper focuses on a café situation, and discusses the importance of recognizing the user's characteristics as a technical challenge in developing a social robot for the wild. First, we will discuss the user characteristics that the robots should recognize and explain how to recognize them. Then, by collecting and analyzing conversations in café situation, we demonstrate that it is necessary to recognize social signals to estimate the user models. Then, we describe our prototype system in which such user modeling functionality is implemented.

1. はじめに

近年、社会的ロボットが役に立つ状況が多数存在する。例えば、ショッピングモールや、介護施設、美術館などである[Kanda 2009] [Sabelli 2011] [Yamazaki 2008]。そのような状況において、ロボットに要求されるスキルや行動はそれぞれの場所に応じたものである必要がある。しかし、社会的ロボットにおいてより重要なのは、ロボットとインタラクションをとるユーザに応じて、行動を選ぶことである。例えば、喫茶店では、ロボットは来店者の特徴を見てテーブルに案内し、メニューを勧め、注文を取らなければならない。来店者は家族連れ、ビジネスマン、大学生など様々であり、店員は来店者に応じて、それぞれ違うテーブルに案内し、異なるメニューを勧めるなど、接客を変更していると考えられる。

本研究では、ロボットが来店者の特徴や、グループ構成に応じて、適切な接客を行うことを目的とし、来店者のグループに応じた接客を行うために、ロボットが認識すべきユーザの特徴について議論する。2章で、必要と考えられるユーザ特徴を述べた後、3章では、来店者の親密度の推定に必要な社会的シグナルの同定を行うために、接客対話データの収集と分析を行った結果を述べる。さらに4章では、ロボットが接客を行うために必要なユーザの特徴を認識するための、プロトタイプシステムを提案する。最後に、プロトタイプシステムの問題点と、今後の課題について述べる。

2. 接客ロボットが認識すべきユーザの特徴

喫茶店という状況において、ロボットの行動決定に利用するユーザモデルを作成する際に有用であると思われる情報を以下に提案する。

- 性別・・・来店者に勧めるメニューや話しかける話題の決定に役立つと考えられる。

- 年齢・・・来店者に勧めるメニューを決めるのに役立つと考えられる。例えば、小さな子供には、お子様ランチを勧める。
- 性別と年齢を組み合わせた情報・・・来店者のグループ構成を推測するのに役に立つと考えられる。例えば、来店者のグループが全員中年の男性ならば、そのグループはビジネスマンのグループかもしれない。
- 人の立ち位置・・・来店したグループのリーダーを判定するのに役に立つと考えられる。また、来店者の位置を認識し続けるために必要となる。
- 視線・・・テーブルでの会話時の視線を測定することにより、来店グループの親密度を計測することに役立つと考えられる。

これらの情報を用いれば、喫茶店での注文時において、以下のようなインタラクションを行えるだろう。

- ① ロボットが来店者を店の入り口で迎える。
- ② 来店者がテーブルに着席する。
- ③ その数分後に、ロボットは来店者が着席している席に移動し、各来店者に応じて、オススメのメニューを提示し、提案する。
- ④ ロボットは来店者からの注文を聞き、戻る。

ユーザの特徴を認識したり、来店者同士の社会的な関係を推測したりするために、システムはさまざまなセンシング技術を使うことが求められる。例えば、コンピュータビジョン技術は、来店者の年齢、性別を認識するために必要不可欠である。しかし、人の親密度などを推測するためには、会話中での社会的なシグナルを抽出する必要がある。

また、人の立ち位置や視線が、社会的関係を推測するのに役立つのか、分析する必要がある。

3. 接客対話の収集と分析

接客対話のデータを収集・分析し、社会的なシグナルを認識することにより、親密度やグループの推定が可能であるかの基礎検討を行った。収集したデータは、来客者が入店してから席



図 1 実験風景

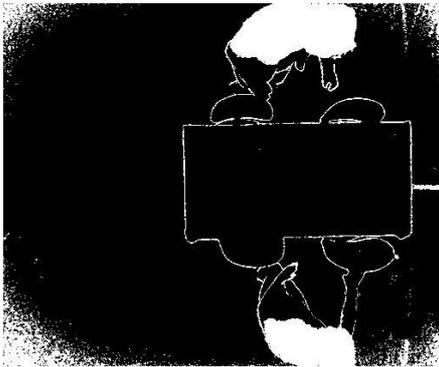


図 2 2 値画像

に座り、メニューを注文するまでである。図 1 に示す、喫茶店環境を模した空間(450cm×450cm)で実験を行い、実験の構成は以下の通りである。

- 2～3名の来店者役と店員1名を1グループとし、計11グループに実験に参加してもらった。
- 各グループの来店者役は学生もしくは社会人のみで構成される。
- 学生グループは、知り合い同士、もしくは初対面同士のグループを構成した。
- 社会人グループでは、知り合い・初対面の条件に加え、上司・部下という社会的関係の異なるグループも構成した。
- 1グループ5から10セッション行われ、総セッション数は84セッションであった。
- 収録時間は約20時間である。

実験では以下の機材を使用した。

- ヘッドセットマイク：来店者と店員の発話を収録する。
- 加速度センサ：来店者と店員の後頭部に装着し、頭部移動量を取得する。
- Microsoft Kinect センサ：実験環境の天井に設置し、来店者の移動を記録する。
- ウェアラブルカメラ：来店者と店員の側頭部に装着され、視線の方向を取得する。
- ビデオカメラ：実験の俯瞰映像を記録する。
- アンケート：グループの他者に対する印象や、自身の心境について尋ねる。アンケートは5点尺度による20項目から構成されている。

3.1 動きはじめるタイミングの違い

社会人グループのセッションにおいて、席に案内されたとき、席に向かって動き始めるタイミングの違いがあるのかを分析した。各参加者の動きを観察するために、重心移動距離を Kinect センサから得られる各フレームの深度画像を利用して求めた。そのために、まず得られた深度画像に対して閾値を床から約120cmに設定し、この閾値を超える画素には白、超えない画素は黒として、図2に示すように二値画像を生成した。この二値画像の各白領域の重心位置を各来店者の位置とし、移動が開始する時間を調べたところ、6セッション中5セッションで、上司が部下よりも平均して0.29秒先に動き出していることがわかった。

これにより、社会人のグループが来店したときには、先に動き出した人が上司であることが多いと推定できる。

3.2 同調傾向の違い

本セッションでは、着席するまでの来店者の動きに同調傾向が存在するのかが分析した。2人の来店者における加速度の差に着目し、この標準偏差を求めることで、お互いの動きに同調傾向があるかを調べることができると考えた。そこで、フレーム毎の移動距離から速度を求め、そこから加速度を計算し、2人の来店者の加速度の差の標準偏差を求めた。その結果を表1に示す。

表 1 加速度の差の標準偏差

	学生	社会人
知り合い同士	3200	3124
初対面同士	4070	2732

学生と社会人で違いがみられ、知り合い同士のセッション、初対面同士セッションどちらの場合においても社会人の方が学生よりも標準偏差の値が小さく、社会人は学生より互いの動きに同調する傾向が見てとれた。しかし、t-検定では、統計的有意差はみられなかった。

3.3 親密度

次に、親密度を測定するために、実験で取得したウェアラブルカメラの映像データと Kinect のデータ、セッション終了後に記入してもらったアンケートを分析した。アンケートは心理測定尺度集V[丹野宏昭 2008]と Funk, J. L. & Rogge, R. D. の THE COUPLES SATISFACTION INDEX(CSI)[Funk 2007]の二つを参考にした。心理測定尺度集Vからは、他者に対する印象に関する質問項目を、また、THE COUPLES SATISFACTION INDEX(CSI)は、感情についての質問項目を参考にした。このアンケートにより、初対面同士の会話では、グループの第一印象を測定することができ、知り合い同士では親密度を測定することができると考えた。

動画を観察したところ、初対面同士と知り合い同士では、相互注視行動の違いがあることに気づいた。そこで、ウェアラブルカメラで記録した動画に対し、OpenCV*1を用いて人の顔を検出し、相互に同時に相手の顔を検出できた場合、相互注視とみなした。

分析は相関分析とt-検定(分散が等しくないと仮定した2標本による検定)を行った。相互注視総時間、相互注視総回数、1回の相互注視の平均時間について、アンケートの点数との相関分析を行った。また、t-検定は、初対面同士の実験セッションと知り

*1 <http://opencv.jp/>

合い同士の実験セッションにおける、相互注視総時間、相互注視総回数、1回の相互注視平均時間に対して行った。

初対面の場合、相互注視回数はアンケートの「自分に好意的に評価してくれる」という項目と相関係数が 0.299 で相関があり、相互注視継続時間はアンケートの「良い印象である」という項目と相関係数が 0.309 で相関があった。知り合い同士の場合、相互注視総時間はアンケートの「とても愉快的気分になることが多い」という項目と相関係数が 0.420 であり、相互注視継続時間はアンケートの「いざという時に力になってくれる」という項目との相関係数が 0.487 であった。

この相関分析の結果より、初対面の場合、相互注視総時間が長ければ長いほど、相手に対して頼りになると感じ、注視回数が増えると相手が自分を好意的に評価してくれていると感じる。そして、相互注視の平均時間が長くなれば長くなるほど、印象が良くなることが分かった。

また、初対面同士と知り合い同士の相互注視の平均値を比較した t-検定の結果は表 2 の通りであり、初対面同士のほうが、知り合い同士より相互注視総時間と相互注視平均時間が長くなり、相互注視回数も増えることが分かった。

表 2 相互注視データの t-検定

相互注視	条件	平均	t 値	p 値
総時間	初対面	107.02	2.24	p<0.05
	知り合い	42.25		
回数	初対面	309.27	1.9	p<0.05
	知り合い	182.37		
継続時間	初対面	0.3039	1.9	p<0.05
	知り合い	0.226		

以上、人が着席するまでの行動に出現する社会的シグナルを分析した結果、着席する際の動き始めのタイミングの違いが認識できれば、来店者が社会人である時に、上司と部下の判定ができる可能性を見出せた。また、来店者が着席するまでの同調傾向を認識できれば、そのグループが社会人か学生か、あるいは互いに相手に配慮する関係であるのかがわかるだろう。そして、視線という社会的シグナルを用いて、相互注視を認識できれば、ロボットは、親密度が低い来店者グループには、その場を盛り上げる為に、冗談を交えながら接客を行うほうが望ましいかもしれない。

4. プロトタイプシステム

3 章で社会的な関係を推定するために、来店者が着席し、注文するまでの行動を分析した。本章では、ロボットが適した来店者に応じた接客を行うため、2 章で述べた来店者の特徴の一部である、性別、年齢、人の立ち位置を認識し、来店者グループを推定する機能を実装し、グループの特徴に応じてロボットが発話する内容を変化させるプロトタイプシステムの提案を行う。認識するタイミングは、来店者の入店時の行動である。図 3 に、プロトタイプシステムのシステムアーキテクチャを示す。図 4 には、現在の実験環境を示す。使用するロボットは、ヒューマノイド型ロボットの Pepper である。

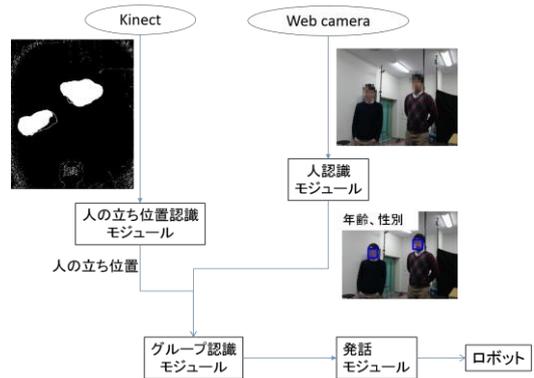


図 3 システムアーキテクチャ

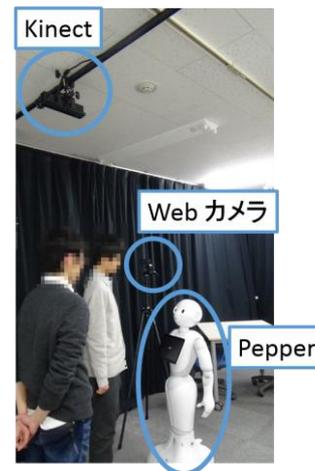


図 4 実験環境



図 5 Kinect からの深度情報



図 6 顔の位置推定

各モジュールの処理を以下に述べる。

- 立ち位置認識モジュール

図 4 にあるように天井に Microsoft Kinect センサを設置し、これにより、深度情報を取得している。このモジュールは

Kinect センサから深度情報を受け取ると、人の領域を検出する。図 5 が Kinect センサからの深度情報を二値化した結果であり、白い領域が人のいる場所を表している。各白い領域の重心を計算することでシステムが人の立ち位置を認識する。

●人認識モジュール

このモジュールはウェブカメラからのデータを用いて、年齢と性別、人の顔のある位置の座標を推定する。図 4 にあるように、来店者の顔が映るように、ウェブカメラを来店者の正面に設置している。また、年齢と性別、人の顔の位置を推定するために、Microsoft oxford computer vision API^{*2}を使用した。これによって図 6 に示すような顔の位置の推定に加えて、各人の年齢と性別を取得することができる。

●グループ認識モジュール

このモジュールは、立ち位置認識モジュールからのデータと人認識モジュールからのデータを統合する。例えば、図 5 と図 6 にあるように、それぞれのモジュールから認識した結果を、このモジュールが受け取ったとする。この場合は、図 5 の左の領域と図 6 の左にいる人の情報(性別、年齢)を組み合わせ、図 5 の右の領域と図 6 の右にいる人の情報を組み合わせる。また、二人の推定された年齢の差が 3 歳以下の場合、来店したグループは友達であると推定する。よって、この場合、二人の年齢が近いので、このグループの関係は、友達グループであると推定される。

●発話モジュール

このモジュールは、グループ認識モジュールで推定されたグループによって話す内容を変更する。例えば、学生グループと推定された場合、ロボットは「今日、クラブ活動はありましたか?」と尋ねる。また、現在のプロトタイプシステムでは、ロボットから距離の近い人に尋ねるようになっている。

5. プロトタイプシステムの問題点

現在の環境では、ウェブカメラは図 4 にあるように、ロボットの横に配置している。ロボットの横に配置すると、大勢の来店があった場合、後ろにいる来店者が前にいる来店者の影に隠れてしまい、性別と年齢が推定されず、グループを推定することができない。グループ推定についても、現状のシステムでは判定ルールを手で作成し、グループを推定しており、推定できるグループは 2 通りしかない。来店者には様々なグループが存在するはずであるので、2 通りでは不十分である。

6. まとめと今後の課題

本稿では、喫茶店という状況に焦点を当て、ユーザモデルに基づきコミュニケーションを変更する社会的ロボットを実装するために、ロボットが認識すべきユーザ情報について考察した。また、接客対話の収集と分析を行い、ロボットが社会的シグナルである視線や着席行動を認識できれば、社会的関係や、親密度を推定することが可能であることが分かった。そして、現在実装中のプロトタイプシステムについて説明し、システムの問題点について述べた。

今後、3 章で述べた分析の結果から、学生と社会人の推定、社会人が来店してきた場合、上司と部下の推定、着席時の視線の情報からグループの親密度推定、これらを用いて来店者の注文時、ロボットが来店者にメニューを勧める行動の実装を進めていきたい。

また、物体認識を使い、来店者が持っているものを認識し、話題を提供できる機能についても実現していきたい。例えば、来店者がスーツケースを持っている場合、ロボットはそのスーツケースを認識し、「旅行に行かれていたのですか?」と問いかける。さらには、対話機能も高度化し、ロボットが認識した人数が正しいかどうかの確認、注文時に来店者が頼んだものや、質問されたものに対して返答ができるようにしていきたいと考えている。

7. 謝辞

本研究は、科学技術振興機構(JST) 戦略的想像研究推進事業(CREST)「実践知能アプリケーション構築フレームワーク PRINTEPS の開発と社会実践」の支援によって実施した。

また、本研究にあたり、データ収集と分析に協力して頂いた、小川俊樹氏、清田晃平氏に感謝致します。

参考文献

- [Kanda 2009] Kanda, T., Shimoi, M., Miyashita, Z., Ishiguro, H., and Hagita, N.: An Affective Guide Robot in a Shopping Mall. in Proceedings of HRI' 2009. pp.173-180, 2009.
- [丹野 2008] 丹野宏昭: 大学生の内的適応に果たす友人関係機能 青年心理学研究, pp55-69, 2008 年
- [Funk 2007] Funk, J. L. & Rogge, R. D.: THE COUPLES SATISFACTION INDEX (CSI), Journal of Family Psychology, 21, pp.572-583, 2007 年
- [Sabelli 2011] Sabelli A.M., Kanda, T., and Hagita, N.: A Conversational Robot in an Elderly Care Center: an Ethnographic Study. in Proceedings of HRI 2011, 2011.
- [Yamazaki 2008] Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M., and Kuzuoka, H. Precision Timing in Human-Robot Interaction: Coordination of Head Movement and Utterance. in Proceedings of CHI 2008. pp131-140, 2008.

*2 <https://www.projectoxford.ai/>