# 情報提供エージェントとの多人数対話における対話制御方式

武田　信也†　　中野　有紀子‡　黄　宏軒‡
東京農工大学　工学府†　　成蹊大学　理工学部‡

## 1．　はじめに

情報提供端末として利用されるエージェントは複数ユーザと応対できることが望ましいが，その設計の基礎となる理論はまだ確立していない．これは，1 対 1 の対話の場合は，ユーザの発話はすべてエージェントに向けられていると予測できるのに対し，多人数対話の場合は，誰が誰に話しかけているのかという事と，どのようなタイミングで会話へ参入するのかを判断するために，会話内容をエージェントが理解する必要があり，同時にエージェント自身も意思表示をしながら，会話の場に積極的に関与しなければならないため，その設計が非常に困難なものとなっているためである[1].

我々は，エージェントと二人のユーザとの三者対話の実現を目指し，会話コーパスの収集と分析を行ってきた[2]．そしてコーパス分析の結果，ユーザが発話を行っている場合，ユーザの会話内容とユーザペアの顔向きの間に一定のパターンが存在することを見出した．また，会話内容と顔向き方向が変化する時間との関係からも，会話内容毎に一定の傾向が見られる事が判明した．

以上の分析結果に基づき，本稿ではユーザペアの顔向きと音声入力から会話内容を推定するモデルを提案し，それを利用した対話制御方式を提案・実装する．

## 2．　会話内容とユーザペアの顔向きについてのコーパス分析

まず，我々はユーザ毎の顔向きを 1 つの組として扱い，それをユーザペアの顔向きと定義した．次に，(1) ユーザペアの顔向きが会話内容によってどのように異なるのか，(2)ユーザペアの顔向きが変化するタイミングは，会話内容によってどのように異なるのかを分析し，ユーザペアの顔向きと音声入力からの経過時間が，会話内容の推定において有用であることを示す．

### 2．1．ユーザペアの顔向きの分析

我々は既に会話コーパスを分析することにより，表 1 に示す会話内容とユーザペアの顔向きとの間に，相関関係を見出している[2]．代表的な例として，会話内容が説明の場合は，ユーザは 2 人とも正面を向いている，質問の場合は，質問を行うユーザが正面を向いたり，質問しないユーザが質問を行うユーザの方を向く，相談と雑談の場合は，ユーザ同士お互いの方を向きあう，理解の場合は，ユーザの顔向きが多様に変化するため，一定の傾向が見受けられないといった関係が見出された．

表 1　会話内容の分類

■ **説明**
案内役がユーザペアに行う，店舗についての情報提示
■ **質問**
ユーザが案内役に行う，店舗情報についての質問
■ **相談**
ユーザペアが行う，次に取るべき行動に関する議論
■ **雑談**
ユーザペアが行う，店舗とは無関係の会話
■ **理解**
ユーザによる，案内役の店舗説明に対する納得の表示

### 2．2．ユーザペアの顔向き変化の時間的分析

2.1 の分析では時間的要因を考慮していなかったため，次に，会話内容とユーザペアの顔向きとの時間的関係を分析した．その結果を表 2 に示す．

表 2　ユーザペアの顔向き変化に要する時間

| ユーザペアの顔向き | 説明 | 質問 | 相談 | 雑談 | 理解 |
|---|---|---|---|---|---|
| 正面→どちらか一方が相手の方を向く | 3.98 | 2.27 | 0.51 | 0.69 | 0.17 |
| どちらか一方が相手の方を向く→正面 | 5.16 | 1.87 | 1.00 | 1.00 | 0.56 |
| 顔合わせ状態への移行 | 4.76 | 1.61 | 0.76 | 0.59 | |
| 顔合わせ状態からの移行 | 4.65 | 2.46 | 1.07 | 1.75 | 0.23 |
| 正面→その他の場所を見る | 4.43 | 2.02 | | | |
| その他の場所を見る→正面 | 4.53 | 3.11 | 0.03 | | |
| 平均反応時間 | 4.59 | 2.22 | 0.67 | 1.01 | 0.32 |

発話開始からユーザペアの顔向きが変化するまでの平均時間は，会話内容が説明の場合は **4.59** 秒，質問で **2.22** 秒，雑談で **1.01** 秒，相談で **0.67** 秒，理解では **0.32** 秒であった．また，顔向き変化に要する時間の長さは，どの顔向きでも会話内容が説明の場合が一番長く，次に質問，相談または雑談，そして理解の様になる傾向があることが判明した．このことから，会話内容が 2.1 で示した 5 種類のどれに分類されるのかを，発話開始から顔向き変化に要する時間によりある程度予測できると考えられる．ただし，会話内容が理解の場合に関しては，他の 4 種類の会話内容に比べて出現頻度が極端に少なかったため信頼度に欠け，今回の予測対象からは除外した．

## 3．　会話内容推定モデルの構築

ユーザペアが現在行っている会話内容が，質問・相談・雑談・理解の 4 種類のうちいずれであるかをリアルタイムで判定するために，コーパスの分析結果を用いて決定木学習を行った．なお，会話内容が説明の場合は，必ず案内役の発話であり予測する必要が無いため，予測対象から除外した．用いた特徴量としては，ユーザペアの顔向き・ユーザペアの 1 つ前の顔向き・発話者・発話前の無言状態の長さ・発話開始からの経過時間・7 つ前までの会話内容のタイプである．会話内容推定モデルは，データマイニングツール Weka の J48 に AdaBoostM1 という集団学習アルゴリズムを適用することで決定木を構

**Conversation management in multiparty conversations with information kiosk agent**
† Shinya Takeda: Tokyo University of Agriculture and Technology
‡ Yukiko Nakano, Hung-Hsuan HUANG: Faculty of Science and Technology, Seikei University

築した[3]. AdaBoostM1 は boosting と呼ばれる集団学習アルゴリズムであり，与えた学習データを用いて学習を行い，その学習結果を踏まえて逐次に重みの調整を繰り返すことで複数の学習結果を求め，その結果を統合し組み合わせて分類精度を向上させる方法である[4].

決定木学習の結果，10 回の交差検定における分類精度は，相談が 74.5%，質問が 68.8%，雑談が 66.7%，全体での分類精度は 70.5%であった.

## 4． 提案するシステムアーキテクチャ

会話内容推定モデルを組み込んだ，情報提供端末との多人数対話における対話制御方式についてのアーキテクチャを図 1 に示す.

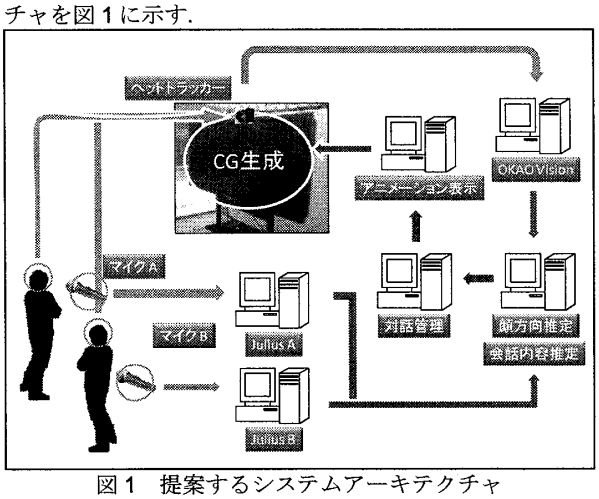


図 1 提案するシステムアーキテクチャ

**A)入力部**

入力部のデータとしては，ユーザペアの顔情報とユーザの発話音声情報である．ユーザペアの顔情報は，ヘッドトラッカーを通じて OKAO Vision に送信され，顔の位置や視線方向などが検出される．また，ユーザの発話音声情報は，マイクを通じて Julius[5]に送信され，入力された音声情報からユーザが現在発話しているのか否を判定し，発話していると判定した場合，会話のキーワードを抽出する.

**B)推定部**

OKAO Vision により検出されたユーザペアの顔情報は，顔方向推定モジュールに送信され，ユーザペアの顔方向の推定を行う．ユーザペアの顔向き方向の推定精度は 93.5%であり，十分な精度であるといえる[2]．その後，ユーザが発話を行っている場合は，ユーザペアの顔方向が会話内容推定モジュールに送信され，第 3 節で示した特徴量を使用して，現在行われているユーザペアの会話内容を推定する.

**C)対話管理部**

推定されたユーザペアの会話内容は，対話管理部に送信される．対話管理部では，まずプランナーにより対話のゴールが決定され，次に推定された会話内容とエージェントが現在行っている発話内容を元に，エージェントの発話内容と次動作の決定を行う.

**D)アニメーション表示部**

対話管理部で決定されたエージェントの動作にしたがって，アニメーションの表示を行う.

## 5． エージェントとの対話例

以上の機能を実装した会話エージェントとの対話の流れを図 2 に示す.

＊[A]：エージェントの発話，[U1，U2]：ユーザの発話
**(OKAO Vision でユーザ 2 人の顔を認識)**
[A] いらっしゃいませ.
・・・
**(ア) ユーザが相談し合っている**
[U2] 旅行先なかなか決まらないな.
[A] 何かお困りであれば，こちらから情報提供をさせていただきますが？
**(イ) ユーザがエージェントに話しかける**
[U1] 私達は沖縄に旅行に行きたいと思います.
[A] 分かりました．さっそくチケットの手配をさせていただきます.
**(ウ) エージェントが会話の主導権を握る**
[A] 宿泊先はもう決まっていますか？
私共の方で格安の宿泊先の斡旋も行っていますが，ご予算はどのくらいでしょうか？
・・・
**(行き先と宿泊先が決定する)**
[A] ご利用ありがとうございました.



図 2 エージェントとの対話例

図中(ア)，(イ)，(ウ)でのシステム動作について以下に説明する.

(ア) ユーザペアの会話内容履歴が，4 回以上連続して相談(雑談)であると判定されていた場合，現在ユーザ同士は相談(雑談)し合っていると判定し，エージェント側から情報提示を行う.

(イ) ユーザの発話内容が質問で，かつユーザの発話内容に現会話固有のキーワードが検出された場合，ユーザがエージェントに話しかけていると判定し，エージェントはユーザに問いかけに対する回答を行う.

(ウ) ユーザの意思決定が行われた後は，エージェントが会話の主導権を取り，ユーザから会話のゴールを設定するために必要な情報を得る.

## 6． おわりに

本稿では，ユーザペアの顔向きと音声入力から会話内容を推定するモデルを提案し，これを利用して 2 人のユーザに応対する情報提供エージェントを実装した．今後，本システムに対し評価実験を行うことで，本アーキテクチャの有用性を確かめる.

謝辞:本研究における頭部姿勢情報の収集には，オムロン株式会社の OKAO Vision 技術を利用しています.
本研究の一部は科研費基盤(S)(課題番号:19100001)の助成による.

## 7． 参考文献

[1] 小林哲則，白井克彦: ヒューマノイドロボットにおけるマルチモーダル会話インタフェース，情報処理学会研究報告 SLP 音声言語情報処理，1999
[2] 武田信也，中野有紀子: Wizard-of-Oz による情報提供エージェントとの多人数対話における言語・非言語行動の分析，人工知能学会全国大会，2009
[3] Weka: http://www.cs.waikato.ac.nz/ml/weka/
[4] Freund, Y, and Schapier, R. E: A decision-theoretic generalization of on-line learning and application to boosting, Journal of Computer and System Sciences, 1997
[5] Julius: http://julius.sourceforge.jp/index.php