

平成 26 年 6 月 14 日現在

機関番号：34315

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700183

研究課題名(和文)複数人ユーザ対応の会話エージェント対話管理機構の研究

研究課題名(英文)Development of Dialogue Management Mechanism for Embodied Conversational Agent in Engaging with Multiple Users

研究代表者

黄 宏軒 (HUANG, HUNG-HSUAN)

立命館大学・情報理工学部・准教授

研究者番号：00572950

交付決定額(研究期間全体)：(直接経費) 3,400,000円、(間接経費) 1,020,000円

研究成果の概要(和文)：CGで表現される擬人化会話エージェントシステムは、一般的ユーザ向けの新しいインタフェースとして期待される。展示会などで複数の来訪者と会話可能なエージェントが望ましいが、会話状態が複雑になり、ユーザ状態の検知とエージェントの行動制御が格段に困難になる。本研究は、まずインタラクションコーパスの収集実験を行い、ユーザの言語・非言語行動を考察し、分析結果に基づいたユーザ発話の受話者推定手法とエージェントの会話管理機構、そしてそれらを利用した完全自律の案内エージェントを開発した。開発したエージェントは被験者実験によって有効性を検証した。

研究成果の概要(英文)：There are more and more real-world deployed applications of embodied conversational agents (ECA's). These agent systems are usually used by groups of visitors rather than individuals. This situation is more complex than single user one, specific features are required. This project aims to build an information providing agent for collaborative decision making tasks. A Wizard-of-Oz (WOZ) experiment was conducted for collecting human/agent interaction data. By analyzing it, the model for identifying the addressee of user utterances have been developed with machine learning techniques on visual (face direction and movements) and acoustic (intensity, pitch, and speed) data of the corpus. We then incorporated it in to a fully autonomous multiparty ECA. The system identifies the addressee and uses this information in language understanding and dialogue management. Finally, an evaluation experiment shows that the proposed addressee identification mechanism works well in a real-time system.

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：ユーザ・インターフェース 人工知能 音声対話システム

1. 研究開始当初の背景

会話エージェント (Embodied Conversational Agent, ECA) は、90年代末期 MIT メディアラボでの研究が発端であり、ロボットの物理的制限を無くし、表情や身体の動きに表現の自由度が高い CG 描画による擬人化キャラクターをヒューマン・コンピュータインターフェースに用いるものである。会話エージェントは、一般的ユーザが特に操作の訓練を受けなくても、日常生活で使い慣れた人間同士のコミュニケーション手段で操作できるため、次世代のコンピュータインターフェースのひとつの候補として注目されている。近年、計算機パワーとグラフィクス分野の目まぐるしい進歩、そして、情報工学にとらわれず人間指向の学際的研究趨勢にも伴い、世界各国で盛んに研究されるようになった。会話エージェントを専門的に扱う国際会議である IVA (Intelligent Virtual Agent) が今年で第 10 回を迎え、発表論文も参加者も年々増えている。研究の方向性も、従来の見かけ上のリアルさとコミュニケーション能力の基礎研究から、教育・訓練の現場や公開展示など、実世界での応用へ移行しつつある。

多数の一般来訪者を迎える展示会や博物館で、疲労することなく的確な情報を伝える説明員に会話エージェントの応用が期待されるが、グループでの来訪者がエージェントと会話する場面は必然的に起きる。ユーザが一人のみの場合では、ユーザの発話対象が必ずエージェント自身であることを推定でき、エージェントは会話を進めていくことが比較的容易にできる。しかし、ユーザが2人以上いた場合、ユーザの発話対象が、エージェントを含む会話参与者全員か、他のユーザか、それともエージェントかの可能性が出てくる。エージェントが複数人ユーザの会話に参加し、適切に応答するために、まず各ユーザ、そしてエージェント自身がその会話における役割(話し手、聞き手、傍参与者)を特定する必要がある。さらに、発話ターンの受け渡しの相手、適切な取得・譲渡の方法、タイミングを選択する必要もある。

このように複数人ユーザに対応するには、システムの複雑度が格段に高くなってしまったため、これまで会話エージェント分野の研究の殆どは、一体のエージェント対一人のユーザ、あるいは、複数のエージェント対一人のユーザという比較的シンプルな場面しか想定していなかった。複数人ユーザを想定した数少ない最近の研究でも、会話の自由度が制限された場面、例えば、宇宙空間で共同作業しているロボットに対するコマンド (NASA Research 2006)、ラウンド制のゲーム (Rehm et al. 2008)、個々の参加者が各自の発話ボタンを押して発話意思を表明するもの

(Bickmore 2010)、もしくは、会話プロセスの一部のみに注目する研究、例えば、ユーザの会話参与意思 (engagement) の有無を判断する程度 (MicroSoft Research 2009) にとどまっており、複数人ユーザに対応できる会話エージェントの研究がまだ十分されていないことが伺える。一方、人文科学の分野でも、人間同士の多人数会話において、音声言語の他に、視線、体勢、ジェスチャーなどの非言語行動の役割について詳細な分析がなされているが (坊農ら 2009)、体系的な理論にはまだ至っていない。

2. 研究の目的

本研究は、複数人ユーザと言語・非言語のマルチモーダルで自然なインタラクションができる会話エージェントの対話管理機構の実現を目指す。

任意な会話場面に対応できる会話エージェントは現段階の技術では実現するのが困難なため、タスクを会話エージェントの応用が最も自然だと考えられる案内役と設定した。複数人のユーザが、案内エージェントから情報をもらいながら、意思決定を行うものである。このような会話場面のなかで、自律した会話管理機構により、案内タスクをこなす会話エージェントの開発を行う。具体的に以下の機能を開発するものである。

- (1) ユーザの発話を、エージェントに向けたのか、別のユーザに向けたのかを区別する受話者推定手法を開発する。
- (2) ユーザ発話のたびに、受話者推定の結果を受け、エージェントに向けた発話のみに対し、音声認識と自然言語理解を通して適切な応答を行う対話管理機構を開発する。

3. 研究の方法

研究の目的を達成するために、成蹊大学情報科学科中野有紀子教授のグループと協力しながら、以下の順序で研究を進めた。

- (1) 実際にエージェントを相手とした場合のユーザの行動パターンを収集するために、WOZ (Wizard of Oz) 法による被験者対話実験を行なった。会話タスクは、観光案内、授業登録、アルバイト紹介の3つの案内タスクとした。実験協力者は2名×18ペアであった。映像、音声のデータ収録に加えて深度センサなどの計測装置を用いて、実験参加者の頭部の動きだけでなく、音声、姿勢、視線方向、立ち位置の変化など、できる限り詳細なインタラクションコーパスを記録した。
- (2) インタラクションコーパスに収録されたユーザとエージェントの会話、そして、映像データに顔認識のソフトツールを適応することで得られた顔の動き、(近似し

た)視線方向,そして,声の韻律情報に対して機械学習の手法を適用することで,受話者推定モデルを構築する.

- (3)(2)に合わせて,完全自律のガイドエージェントを構築するにあたり,これまでに黄が開発していた汎用会話エージェントプラットフォーム GECA(Generic Embodied Conversational Agent, Huang 2008)を基に,新たに Information State ベースの会話管理機構,Google エンジンによる音声認識,言語理解のコンポーネントを開発した.
- (4)(2)(3)で開発したエージェントの有効性を評価するために,被験者により実際にシステムを使った評価実験を行う.
- (5)会話状態の自動推定が完全でない場合,エージェントの行動にエラー(質問されていないのに答えてしまう,現在の話題と逸れた発言をする,ユーザ同士の会話に唐突なタイミングで割り込むなど)が必ず生じると思われる.ユーザの反応(困惑な姿勢,話し声,質問の繰り返し,笑いなど)からエージェントの行動にあったエラーを検出する手法を提案する.さらに,エラーが生じた際に会話の適切な修復手法(ユーザへの確認,話題を戻すなど)を提案する(研究進行中).
- (6)ユーザに対して積極的な案内ができるように,ユーザ間の会話に適切に介入できるタイミングを非言語情報から推定するメカニズムを開発する(研究進行中).

4. 研究成果

本研究で主に以下の成果を上げたと考えられる.

- (1)複数人ユーザとエージェントの会話コーパス収集:

本研究のベースとなったのは,2010年3月に成蹊大学で(21ペア),2011年8月立命館大学で(10ペア),2012年7月立命館大学で(18ペア)集めた会話コーパス.当コーパスは人間同士の会話ではなく,人間が操作するエージェントと複数人ユーザとの会話 WOZ 実験を行い,収集した物であった.実験のタスクは,典型的な会話エージェントのアプリケーションである情報提供サービスとし,複数人のユーザがエージェントから情報を提供してもらいながら,グループ内での議論を経て,単位履修と旅行先の意思決定を行う形をとった.

エージェントの役割(操作役の実験参加者)は,単にユーザの情報リクエストに答えるだけではなく,ユーザの議論に積極的に介入し,新しい情報を提供したり議論のまとめりを促進したりする.ユーザ役の実験参加者には顔認識ソフトを用いて,マイクロホンも付けてもらい,頭部運動と音声を映像データ

と共にコーパスとして収録した.このデータ数のある会話コーパスは,本研究プロジェクトのこれまでの成果に活用できたほか,さらに詳細に分析すれば,エージェントと複数人ユーザの会話研究にも適用できると考えられる.

- (2)ユーザ発話の受話者推定機構の提案:

コーパスデータで集めた被験者の非言語行動パターン(視線方向の近似として頭部の動きとパワー,基本周波数と話速の音声特徴量を用いる)を機械学習で分類可能な形にして受話者推定モデルを構築した.次に,このモデルを取り入れてリアルタイムで受話者推定可能な会話エージェントシステム用のコンポーネントを開発した.2011年8月に立命館大学でコーパス収集実験と同条件で20名10ペアのリアルタイム推定コンポーネントの精度評価実験を行った.機械学習の教師信号に用いたコーパスデータと同等の80%前後の推定精度を得ている.

- (3)複数人ユーザ対応の自律案内エージェントの開発

会話管理機構は,Information State の考え方[Larsson 99]に基づいて設計した.時間,システムの現在・以前の状態,ユーザの人数,言語理解結果,発話者,受話者,エージェントの発話といった情報の管理を行う.ユーザ発話が発検出されるか,あるいはエージェントの発話が生成されると,これらの会話状態を示す情報が更新される.言語理解結果と受話者を入力とし,状態遷移モデルによって会話の制御を行う.もし,受話者がエージェントであると推定されると,認識されたユーザ発話の意味表現に応じて状態が遷移し,遷移した状態に基づきエージェント発話が決定され,出力すべき発話の文字列がアニメーション生成部に送信される.状態遷移モデルは,黄が開発を行なっている汎用的会話エージェント開発枠組み GECA のシナリオ記述言語(GECA Scenario Markup Language, GSML)を拡張し,使用した.今回行った拡張は,新規の言語理解部,ジェネレーション部に合わせて,旧バージョンでは一意に決まっていたユーザ入力・エージェント出力の対応ペアを,実行時に動的に値を関係付けられるようにした.これによって膨大な可能性がある言語理解部の入力とジェネレーション部の出力を簡潔に記述できるようになった.

音声認識部はこれまでに利用していた Julius 音声認識エンジンからより認識精度の高い Google 社のエンジンに切り替えた.さらに,ビデオ会議用の高指向性マイクも導入することで,音声認識の成功率は飛躍的に向上した.受話者推定部で利用している発話者情報も,マイクロソフト社の Kinect センサに搭載されているマイクロホンアレイを用いて推定することで,利用者にヘッドセットを装着してもらう必要がなくなった.新規の言語理解部では,音声認識部から取得する

認識結果の文字列を形態素解析で単語ベクトルに分割し、事前に定義したベクトル空間とコサイン類似度を求め、最も近い結果にマッピングされる。開発終了後に、京都の観光案内を題材として、評価実験を行なった。WOZ法によるコーパス収集実験で集めたユーザ発話の韻律情報と顔向き情報に機械学習を適用することで得られたモデルを用いる提案システムに対し、評価対象となる比較システムは、発話間の注視対象(顔向き)の継続長のみで受話者を判定する。その結果、提案システムにおける対エージェント発話の判定はF値0.8であり、十分な精度が得られた。対パートナー発話については、F値が0.56であり、今後精度向上の必要があるが、顔向きの継続長のみに基づく従来の実装手法[Bohus 10]に比べて精度はよく、また、誤ってユーザ同士の会話に割り込む危険性も低いことが確認された。

(4)応用課題：

本研究の派生課題として、2012年度から複数人対話を行う際に積極的にサービスを提供するエージェントの適切な割り込むタイミングの推定にも取り組んだ。2012年7月に立命館大学で追加の22組の被験者WOZ実験を行い、これまでの顔の動き、声の韻律情報の他、Kinect センサから得られる深度情報から推定した体勢の変化も記録した。実験で集めたデータから、会話の停滞、直近のエージェントの発話に対しての更なる支援が必要、ユーザへの回想支援が必要、質問内容を議論していると4つのタイミングを自動的に推定する手法を考案した。現在では、F値0.4のパフォーマンスであるが、文脈情報をも利用することで向上できると考えられる。また、これまでに受話者推定を中心に研究を行ってきたが、その精度が100%でない以上、誤った推定結果が起きうる。現在のエージェントは、推定結果が100%正しい前提で動作しているため、誤った推定をしていることが分からなければ、そこから会話の流れが崩れ、ユーザが希望する情報を提供できなくなる。そこで、ユーザの表情の変化などから受話者推定の誤りを発見し、会話の流れを修正する手法にも取り組んだ。前記の実験で恣意的にエラーを起こし、ユーザの反応を集めた。笑い、びっくり、戸惑いの表情をよく観察したため、まず、こういった表情の認識に取り組んだ。今のところ、90%を超える精度で、この3種類の表情を無表情の状況から区別できた。しかし、表情だけでは、誤りの発生を断定できないため、今後、音声情報などを含めて、誤りの発生を判別できる手法を考案したい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計3件)

著者名：中野有紀子，馬場直哉，黄 宏軒，

林 佑樹，論文標題：非言語情報に基づく受話者推定機構を用いた多人数会話システム，雑誌名：人工知能学会論文誌，査読：有，巻：29(1)，発行年：2014，ページ：69-79，<http://dx.doi.org/10.1527/tjsai.29.69>

著者名：馬場直哉，黄 宏軒，中野有紀子，論文標題：人対会話エージェントとの多人数会話における頭部方向と音声情報を用いた受話者推定機構，雑誌名：人工知能学会論文誌，査読：有，巻：28(2)，発行年：2013，ページ：149-159

著者名：Hung-Hsuan Huang and Toyoaki Nishida，論文標題：Evaluating a Virtual Agent Who Responds Attentively to Multiple Players in a Quiz Game，雑誌名：ヒューマンインタフェース学会論文誌，査読：有，巻：14(4)，発行年：2012，ページ：351-365

[学会発表](計13件)

著者名：Hotta, R., Huang, H.H., Shochi Otogi, and Kawagoe, K., 論文標題：Detecting Address Estimation Errors from Users' Reactions in Multi-user Agent Conversation, 学会名：HCI International Conference (HCII 2014), 発表年月日：June 26, 2014 発表場所：Crete (Greece)

著者名：Otogi, S., Huang, H.H., Hotta, R., and Kawagoe, K., 論文標題：Finding the Timings for a Guide Agent to Intervene Inter-user Conversation in Considering Their Gaze Behaviors, 学会名：6th Workshop on Eye Gaze in Intelligent Human Machine Interaction, 15th International Conference on Multimodal Interaction (ICMI 2013), 発表年月日：December 13, 2013, 発表場所：Sydney, (Australia)

著者名：堀田 怜，乙木 翔地，黄 宏軒，川越 恭二，論文標題：複数人ユーザとエージェントの会話における受話者推定の誤り時の状況の考察, 学会名：HAI シンポジウム 2013, 発表年月日：2013年12月8日，発表場所：岐阜大学岐阜駅前サテライトキャンパス (岐阜県)

著者名：Nakano, Y., Baba, N., Huang, H.H., and Hayashi, Y., 論文標題：Implementation and Evaluation of Multimodal Addressee Identification, Mechanism for Multiparty Conversation Systems, 学会名：15th International Conference on Multimodal Interaction (ICMI 2013), 発表年月日：December 10, 2013, 発表場所：Sydney, (Australia)

著者名：Otogi, S., Huang, H.H., Hotta, R., and Kawagoe, K., 論文標題：Finding the Timings for a Guide Agent to Intervene User-user Conversation to Provide Information Actively, 学会名：the 13th International Conference on

Intelligent Virtual Agents (IVA 2013),
発表年月日: August 29, 2013, 発表場所:
Edinburgh (UK)

著者名: 堀田 怜, 乙木 翔地, 黄 宏軒,
川越 恭二, 論文標題: 複数人と会話するエ
ージェントの受話者推定の誤り検知のため
の特徴量の検討, 学会名: 第 27 回人工知能
学会全国大会, 発表年月日: 2013 年 6 月 5
日, 発表場所: 富山国際会議場(富山県)

著者名: 乙木 翔地, 堀田 怜, 黄 宏軒,
川越 恭二, 論文標題: 複数人ユーザ会話に
おけるエージェントの割り込みタイミング
の推定手法の提案, 学会名: 第 27 回人工知
能学会全国大会, 発表年月日: 2013 年 6 月
5 日, 発表場所: 富山国際会議場(富山県)

著者名: 馬場直哉, 黄 宏軒, 林 佑樹, 中
野有紀子, 論文標題: 複数人ユーザに対応
可能な情報提供会話エージェント, 学会
名: 第 75 回情報処理学会全国大会, 発表年
月日: 2013 年 3 月 7 日, 発表場所: 東北大
学(宮城県)

著者名: 乙木 翔地, 堀田 怜, 黄 宏軒,
川越 恭二, 論文標題: 複数人ユーザ会話に
おけるエージェントの割り込みタイミング
の推定手法の検討, 学会名: HAI シンポジ
ウム 2012, 発表年月日: 2012 年 12 月 8 日,
発表場所: 京都工芸繊維大学(京都府)

著者名: Baba, N., Huang, H.H., Nakano,
Y., 論文標題: Addressee Identification
for Human-Human-Agent Multiparty
Conversations in Different Proxemics,
学会名: 4th Workshop on Eye Gaze in
Intelligent Human Machine Interaction:
Eye Gaze and Multimodality, 14th
International Conference on Multimodal
Interaction (ICMI 2012), 発表年月日:
October 26, 2012, 発表場所: Santa Monica
(USA)

著者名: 馬場直哉, 黄 宏軒, 中野有紀子,
論文標題: グループ会話対応型会話エー
ージェントにおける受話者推定システム, 学会
名: HAI シンポジウム 2011, 発表年月日:
2011 年 12 月 3 日, 発表場所: 京都工芸
繊維大学(京都府)

著者名: Huang, H.H., Baba, N., Nakano,
Y., 論文標題: Making Virtual
Conversational Agent Aware of the
Addressee of Users' Utterances in
Multi-user Conversation from Nonverbal
Information, 学会名: 13th International
Conference on Multimodal Interaction
(ICMI'11), 発表年月日: November 16, 2011
発表場所: Alicante (Spain)

著者名: Baba, N., Huang, H.H., Nakano,
Y., 論文標題: Identifying Utterances
Addressed to an Agent in Multiparty
Human-Agent Conversations, 学会名: the
11th International Conference on
Intelligent Virtual Agents (IVA'11), 発
表年月日: September 16, 2011, 発表場所:

Reykjavik (Iceland)

6. 研究組織

(1) 研究代表者

黄 宏軒 (HUANG HUNG-HSUAN)

立命館大学・情報理工学部・准教授

研究者番号: 00572950