1J4-OS-22b-4

会話参加者の優位性を考慮した会話エージェントによる多人数会話への介入

Intervening in Multiparty Conversations by Conversational Agents based on Conversational Dominance

八城 美里[*1] 林 佑樹[*2] 中野 有紀子[*2]
Misato Yatsushiro Yuki Hayashi Yukiko Nakano

[*1] 成蹊大学大学院 理工学研究科 [*2] 成蹊大学理工学部
Seikei University Seikei University

It is important for conversational humanoids that manage multiparty conversations to recognize the group dynamics existing among the users. This paper proposes a prototype system that intervenes multiparty conversations among the users. The main characteristic of our intervening strategy is that the system estimates the conversational dominance of each participant in group interactions and uses the information to determine the system's intervening utterance. First, we conducted a Wizard-of-Oz experiment to collect conversational speech, and motion data. Then, using the collected data, we examine the validity of our dominance estimation model, and discuss proper intervening timing. Based on the discussion, we implement a prototype system that generates intervening utterances towards the most dominant participant or most submissive participant.

1. はじめに

人と共生するロボットやアニメーションエージェント(本研究では，これらを会話エージェントと呼ぶ)を実現するうえで，このような人工物が人同士の会話に適切に参加・参入できる必要がある．会話エージェントと複数のユーザによる多人数会話に関する研究として，受付 (receptionist) のエージェントや [Bohus 2009]，クイズを出題するエージェント [Huang 2010] 等の研究がおこなわれているが，会話エージェントがより積極的に人同士の会話に参入するには，会話内容の理解のみならず，会話参加者間の関係なども考慮して，誰に対して，どのような発話を行うことが有効であるかを判断する必要がある．

一方，会話参加者を特徴づける 1 つの側面として，「優位性」がある．優位性とは，会話参加者が会話をけん引する度合いである．会話において，中心的な役割を担い，会話をリードする人は優位性の高い参加者であり，従属的であり発言の機会の少ない人は優位性の低い参加者であるといえる．このような，会話参加者の会話における優位性は，社会的な立場，役割，あるいは性格等様々な要因によって規定されているが，背後にある人間関係を知らずとも，発言の内容や非言語行動を見ることで，優位性の高い人とそうでない人を推し量ることができる．

そこで本研究では，適切なタイミングで多人数会話に介入するエージェントシステムの実現を目指し，会話における優位性に着目し，①会話の非言語行動から優位性を推定し，②それに基づいて，エージェントが誰にどのように介入するかを決定する方法を提案する．また，エージェントの発話内容の決定では，ユーザ発話の音声認識結果からキーワードを抽出し，それに特定の地名などが現れた場合はそれが現在の話題であると考え，それに関連する発話を行う．本論文では，この方式の実現可能性，有用性を確認するために，Wizard-of-Oz 実験で収集した多人数会話の音声，頭部姿勢のオフラインデータを入力としたシミュレーションシステムを実装する．

連絡先：八城美里，成蹊大学大学院 理工学研究科，
dm136215(at)cc.seikei.ac.jp

図 1：WOZ 実験の様子

2. データ収集実験

多人数会話における優位性順位の予測，また介入タイミングの決定方式を検討するための有用な言語，非言語データを収集するために，Wizard-of-Oz (WOZ) エージェントシステムを用いた会話実験を行った．被験者は 3 人 1 組のグループとし，週末の外出先を決める会話を行ってもらった．被験者には，エージェントに質問することにより，場所や施設についての情報を得ることができると指示した．また，情報提供を行うエージェントとして，ヴィストン社製の Robovie-R Ver.3 を使用した．実験の様子を図 1 に示す．

2.1 実験環境

実験環境を図 2 に示す．エージェントの発話には音声合成による合成音を用い，WOZ システムを操作することにより，ロボット内部に搭載したスピーカーから音声が出力される環境を実装した．WOZ 操作者の姿が被験者から見えないようスクリーンで実験スペースを区切り，お互いの姿は見えないようになっている．3 名の被験者には図 2 のように被験者前方のカメラを中心とした視点で，右側に立っている被験者を right，真ん中に立っている

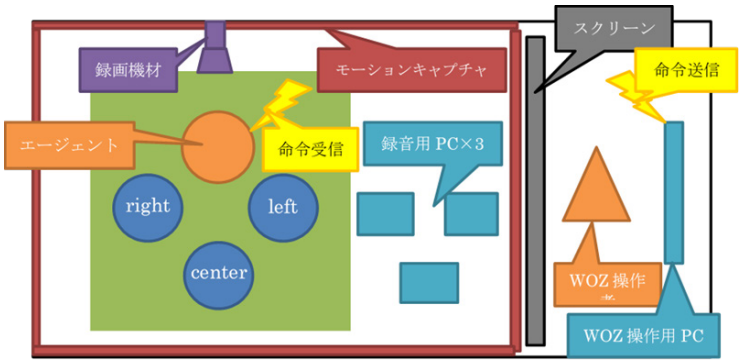

図 2.実験環境

被験者を center，左側に立っている被験者を left と呼び，それら位置にそれぞれ立ってもらった．

エージェントの後方にビデオカメラを設置し，被験者らの会話の映像データを収録した．また，被験者には複数のマーカーのついた帽子をかぶってもらい，モーションキャプチャシステム OptiTrack を用い，頭部のモーションデータを取得した．さらに，ヘッドセットマイクを装着してもらい，それぞれをノート PC につないで，音声データを収録した．

2.2 実験手続き

被験者として，20 代の男性 7 名，女性 5 名の計 12 人の 4 グループの大学生に実験に参加いただいた．社会的立場等により会話における優位性が影響されないよう，各被験者グループは同学年の親しい者同士で構成されるようにした．

被験者らは，案内員（エージェント）に質問して情報収集しながら，週末に皆で一緒に出掛ける場所を 2 か所決めることが課題として与えられた．その際，各被験者には，景観が楽しめるスポット，文化に触れられるスポット，ショッピングができるスポットのいずれかを場所選択の選好条件として与えた．従って，被験者らは話し合いにより，可能な限り多くの人の希望を満たす外出場所を 2 か所決める会話を行うことになる．

各セッション終了時にアンケートを行い，会話の中心となった度合いにより，被験者の順位づけを行ってもらった．

2.3 収集データ

以上の実験で，セッションごとに被験者がエージェントと会話する様子を収録したビデオデータ，会話の言語情報を収録した会話コーパス，モーションキャプチャによる頭部のモーションデータが得られた．また，セッション毎にアンケートを行い，各被験者の主観による優位性の順位を得た．

4 グループ×3 セッションで 12 回分のデータが得られ，そのうち 1 セッションが録音ミス，また別のセッションでは録画ミスがあったため 10 回分，計 30 人分のデータを使用し，分析を行った．

以下の章では，これらの収集データを用いて，優位性の自動推定と，音声認識結果からの介入内容の決定の可能性を検討する．

3. 優位性の推定

会話進行中に，各会話参加者の優位性を逐次推定することができれば，会話エージェントが誰に対してどのように働きかけるかを決定するための有益な情報になることが期待できる．例えば，優位性のもっとも高い参加者の発話をきっかけにエージェントが介入し，それに関連する発話を行うことにより，グループの意思決定を促進し，意思決定までの時間が短縮される効果が期待される．一方，優位性のもっとも低い参加者の発話をきっかけに介入した場合，これまで意見をなかなか取り入れてもらえなかった優位性の低い参加者の意見について議論されるきっかけを作り，優位性向上の可能性を高めるだろう．実際にこのような効果が生じるか否かについては，本研究の範囲では検証できないが，ここでは，その前段階として，優位性を逐次的に推定できるか否かを，2 章で収集したデータを用いて検討する．

優位性推定モデル[Nakano 2012]では，2 章で示した実験とほぼ同じ設定で，3 人の被験者による多人数会話を収集し（ただし，情報提供役のエージェントには，スクリーンに投影されたアニメーションキャラクタを用いた），優位性の違いにより，注視，相互注視の量や，会話における発話量，ターン譲渡成功率などが異なることを明らかにしている．さらに，これらのデータ分析結果に基づき，以下に示す，重回帰式による優位性の推定モデルを提案している．

本モデルを構成する 4 つのパラメータについて以下に説明する．

注視時間(s): 発話中に他の参加者を見ていた時間の累積
相互注視時間(s): 発話中に他の参加者と相互注視を行っていた時間の累積
発話時間(s): 発話時間の累積
発話権取得: 前話者の発話終了後 2 秒以上の沈黙が続いた後，ターンを取得した回数

より会話をリードしている参加者ほどこの優位性の値は高くなる．一方，この値が低い参加者ほど優位性が低く，ターンをうまく取得できない等が予想される．

3.1 優位性モデルの評価

$$\text{優位性} = (0.80)*\text{注視時間} + (0.162)*\text{相互注視時間} + (0.94)*\text{発話時間} + (0.256)*\text{発話権取得} + (-0.25)$$

2 章で収集したデータを用いて，前節で述べた優位性推定モデルの妥当性と有用性を検討する．

(1) 優位性推定のためのデータ作成

収集データから，発話単位で，優位性推定モデルのパラメータ値を算出した．

大規模音声認識システム Julius の adintool を用いて，各セッションの個人ごとの音声データを，発話単位に分割し，発話開始時間と発話終了時間を得た．さらに，発話開始時間と発話終了時間の差から，発話時間のパラメータ値を算出した．

次に，発話権取得の回数をカウントするために，先行発話の終了時間後の沈黙時間を以下の式により算出した．

現在の発話の開始時間(s)－直前の発話の終了時間(s)

この沈黙時間が，2 秒以上であり，かつ直前の音声区間データにおける発話者と異なる話者であれば，その参加者の発話権取得回数を1つカウントアップした．

注視行動に関するパラメータ値は次の方法で算出した．まず，各参加者が誰の方向を見ているかを自動的に認識するためのモデルを作成した．実験の様子を撮影したビデオデータを用いて，各被験者の顔向きのアノテーションを行い，これを教師信号として決定木学習を行った．right, center, left の立ち位置それぞれについて，モデルを作成した結果，3 種類のモデルすべてに

おいて，F-measure が 0.8 以上であったため，これらのモデルを用いて，モーションキャプチャデータから顔向きの自動推定を行った．

この方法により各被験者について顔向き情報を得ると同時に，顔向きが変化した時間の取得を行った．さらにこれらの情報から，各被験者が発話中に他者を見ていた時間（注視時間）を算出した．さらに，相互注視の区間と対象を，区間と対象のオーバーラップに基づき算出した．例えば，A が B を注視している時間と B が A を注視している時間の重なりを相互注視がおこっている時間とした．

(2) 優位性の算出

(1)で得た値を用い，発話区間ごとに優位性を計算するプログラムを作成した．このプログラムは，(1)で作成したデータを読み込み，被験者ごとにデータを保持し，各被験者の顔向きデータ，発話データから，発話や注視に関するパラメータ値の計算を行う．これにより優位性推定モデルの計算に必要な値がそろうので，3 章で説明した重回帰式を適用して，発話区間ごとに優位性の値を計算する．

その結果，会話時間の経過と優位性の値の変化をとらえるとともに，各会話参加者の最終的な優位性の順位を得ることができる．あるセッションにおける 3 人の会話参加者の優位性の推移の様子をグラフにしたものを図 3 に示す．現在の優位性推定モデルでは，会話開始からの累積データを用いているため，優位性の値が減少することはない．したがって，ある時間における 3 者間の優位性の値の差が優位性の順位を表し，値の差が大きいほど，会話における役割の違いが顕著になると考えられる．

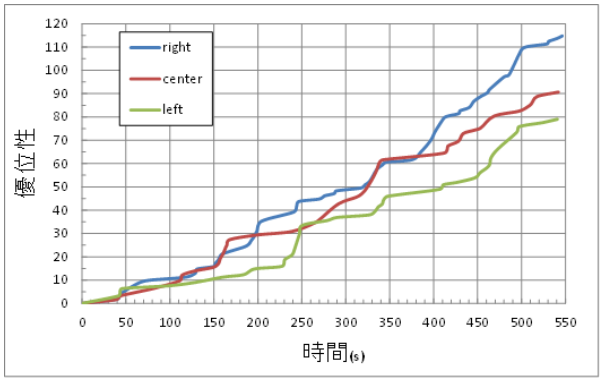

図 3：優位性の時間変化

(3) 優位性推定モデルの評価

各セッションで被験者らに行ってもらったアンケート結果から得た優位性順位と(2)のプログラムから得た最終的な優位性の順位との比較を行った．その結果，優位性順位 1 位で 6 割，2 位で 6 割，3 位では 7 割の一致が見られた為，優位性推定モデルはシステム実装に利用可能であると考えた．

4. システム実装

3 章で検討した優位性推定方式に基づき，2 章で収集したデータを用いて，エージェントによる介入発話を生成するシステムを作成した．このようなオフラインでの方式の検討を行うことにより，提案手法の妥当性を検証することができる．

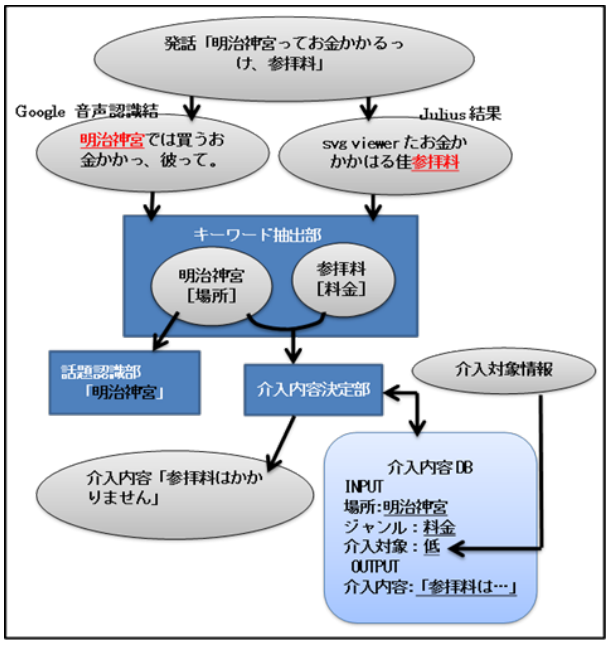

図 4：介入内容決定部の処理例

システムには事前に，優位性の最も高い人（会話の主導権を握る人），もしくは優位性の最も低い人（会話にあまり参加できていない人）のどちらにアプローチするかを指定しておき，介入すべきタイミングであると判定されたときに，システムは，指定の優位性順位の参加者の発話に関連した発話を生成する．以下に，介入内容と介入タイミングの決定について詳細を説明する．

4.1 介入内容決定部

Google 音声認識と大語彙連続音声認識システム Julius を用いて，収集した音声データに対して，音声認識処理を行った．実験で検出された発話音声データに対する両音声認識システムからの認識結果を，人が行った書きおこしと比較した結果，検出される単語は Google 音声認識のほうがやや多かったが，Google 音声認識で認識されない単語が Julius で認識される場合もあったことから，本研究では，どちらかの音声認識で得られたキーワードが書きおこしに基づいて選択したキーワードと一致していれば，それを採用することにした．将来的には両方の音声認識システムを組み合わせたキーワード抽出方法を検討する必要がある．

次に，発話ごとに介入内容を決定する．処理の流れを図 4 に示す．2 種類の音声認識エンジンからの認識結果がキーワード抽出部に送られると，キーワード抽出部で，キーワードを抽出する．次に，話題認識部で保持されている現在話題となっている場所と，キーワード抽出部で得られたキーワード，さらに介入対象者（優位性高／低）の 3 種類の情報が介入内容決定部に送られる．介入内容決定部では，これらの情報を用いて介入内容 DB を参照し，介入内容を得る．もし，キーワードが[場所]のキーワードであり，かつ現在話題名として保持している場所名と異なる場所である場合は，話題名を更新する．

4.2 介入タイミング決定部

エージェントと複数ユーザが会話を行っている際に，エージェントが会話に介入できそうなタイミングについて，[乙木 2012]は，表 1 に示す 5 種類の介入タイミングを提案している．

表 1: 介入タイミング

Stagnation	グループ内で会話が停滞している
Check	質問する際に話者らで行う確認
Argument	エージェントの発話内容を議論
Arrangement	会話内容の整理を行っている
Decision	意思決定への議論

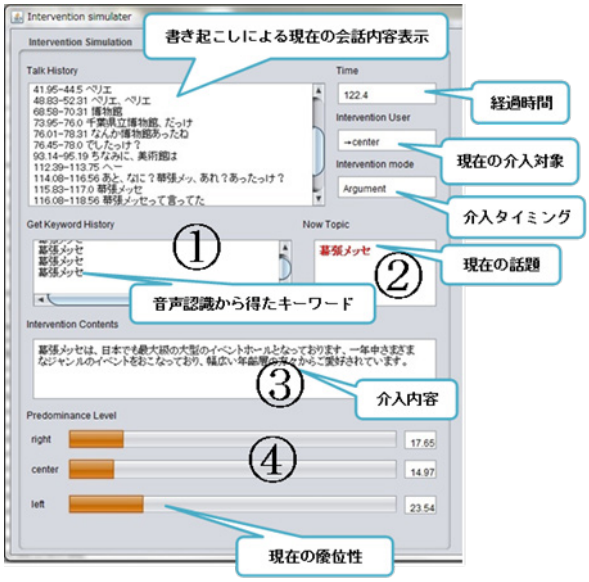

図 5：介入発話生成システムの動作例

本研究では，これらの中から Stagnation と Argument の場合を参考に介入タイミングの決定行う．具体的には，収集したコーパスの分析に基づき，沈黙が 3 秒以上あった場合を Stagnation 状態であるとした．また 1 分間で同じ場所名が 3 回以上出てきた場合，その話題について議論が行われていると判断し，Argument 状態とみなした．

これらの状態が検出された場合は会話介入が必要な状態であると考え，介入内容決定部で決定した介入発話を出力する．

5. 介入発話生成システムの動作例

実装したシステムの動作画面を図 5 に示す．図中①に音声認識結果，②には現在の話題，③に介入の発話内容，④には優位性推定結果が示されている．

沈黙時間の付与された会話ログ，紹介する地区，優位性の高い人，もしくは低い人のどちらを介入対象者にするかを選択し，スタートボタンを押すと会話のシミュレーションが開始される．①の音声認識結果から得たキーワードが介入内容決定部に送られ，②で話題の更新をし，介入内容の決定を行う，その際 Argument と Stagnation が検出されたら③に介入内容が表示される．発話終了時間になると，3 章で提案した優位性推定手法を適用して④の優位性の推定値を更新するとともに，次に介入すべき参加者(介入対象者)が決定される．

6. おわりに

本研究では，会話エージェントによる多人数会話への介入システム の実現に向けて，介入発話生成システムを提案した．今後は，リアルタイムでの介入システムの実装に向け，音声認識部の出力からリアルタイムにキーワードを抽出するとともに，モーションキャプチャデータからリアルタイムに優位性を推定するシステムの組み込み，介入内容決定方式の改良，適切な介入タイミングの更なる調査を行う予定である．

参考文献

[Bohus 2009] Bohus, D., Horvitz, E.: Open-world dialog: Challenges, directions, and prototype, In: IJCAI2009 Workshop on Knowledge and Reasoning in Practical Dialogue Systems (2009).

[Huang 2010] Huang, H.H., Furukawa, T., Ohashi, H., Nishida, T., Cerekovic, A., Pandzic, I.S., Nakano, Y.I.: How multiple concurrent users react to a quiz agent attentive to the dynamics of their game participation, In: AAMAS, pp.1281–1288 (2010).

[Nakano 2012] Yukiko I. Nakano, Yuki Fukuhara , Estimating Conversational Dominance in Multiparty Interaction, 14th ACM International Conference on Multimodal Interaction (ICMI2012), pp.77-84 (2012).

[乙木 2012] 乙木翔地, 堀田怜, 黄宏軒, 馬場直哉, 中野有紀子, 川越恭二: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, HAIシンポジウム(2012).