

会話参加者の優位性推定と会話制御への応用

八城 美里¹ 池田 直弥² 林 佑樹² 中野 有紀子²

成蹊大学大学院 理工学研究科¹ 成蹊大学 理工学部²

1. はじめに

複数名のユーザとの多人数会話に参加できる会話エージェントを実現するためには、エージェントは単にユーザからの質問に答えるだけでなく、より積極的に会話に参入する能力が必要とされる。しかし、多人数会話において、エージェントが自発的に、いつ、だれに話しかけるべきかを決定する手法は確立されていない。

そこで本研究では、会話における優位性(会話を牽引する度合い)に着目し、優位性を逐次的に推定することにより、話しかける相手を決定するシステムを実装する。本システムでは、例えば、優位性の低い人に積極的に話しかけることにより、発話のチャンスを与えることができる。

2. 優位性推定モデル

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。本研究では、これを求めるために以下に示す優位性推定モデル[1]を使用する。

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

本モデルを構成するパラメータは以下の4つである。
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優位性の値は高くなる。一方、この値が低い参加者ほど、ターンをうまく取得できないことが予想される。このモデルの全てのパラメータ値は、音声処理とモーションキャプチャの計測結果から算出できるため、優位性を自動的に推定することができる。

2.1 優位性推定モデル適用例

図1は3人で訪れる観光地を決めるグループ会話のデータに優位性推定モデルを適用した結果である。図の右に示す3つのグラフは、30~40秒分のデータ

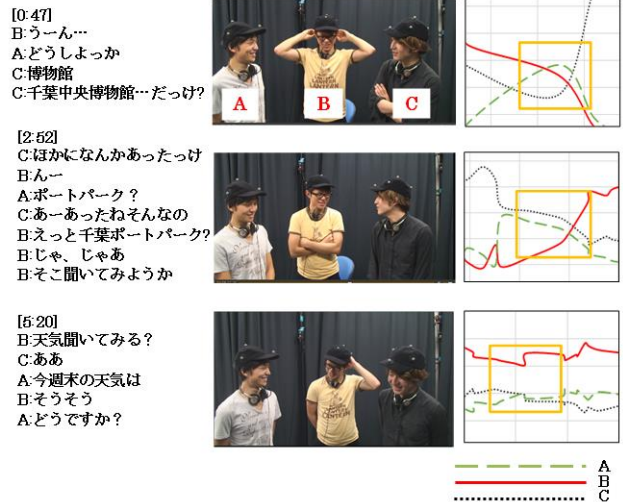


図1. 優位性推定モデルの適用例

について、優位性の値の変化をグラフ化したものである。左にある会話はグラフの黄色い枠で囲った部分の書きおこしである。

一番上の状態ではその前まであまり発言を行ってなかったCが、Aとの相互注視を維持した状態で何度か発話を行い、他の二人は発話することなくCの発話を聞いているだけだったために、Cの優位性が急上昇している。会話開始後2:52あたりでは、これまで聞き手となるが多かったBが、Cと相互注視しながら多くの発話を行ったため、Bの優位性が上昇している。5:20では、Bが会話をリードしている状態であり、AがCを注視しつつ発話を行っている。Cも発話をしているのだが、他者への注視が少なく、エージェントへの短い発話のみであったために、Aの優位性が少しずつCにせまり、その後Cを追い抜いている。

3. システム実装

2章で述べたモデルを用いて、優位性推定機構を実装する。さらに、この優位性推定機構を用いた、介入発話生成システムを実装する。

3.1 優位性推定機構

システム構成図を図2に示す。各会話参加者(User 1, 2, 3)の頭部座標、回転角度はモーションキャプチャ用PCから顔向き推定用PCへ送信される。顔向き推定用PCでは送信された頭部の動作データから顔向きの推定を行い、その結果を優位性推定サーバへ送信する。ユーザの発話が発検されると、音声データ受信/送信用PCが発話の開始・終了時間を優位性推定サーバへ送信する。また、顔向き情報から、その発話中

Estimating conversation dominance and applying it to conversation management

1 Misato Yatsushiro: Graduate School of Science and Technology, Seikei University

2 Naoya Ikeda, Yuki Hayashi and Yukiko Nakano: Faculty of Science and Technology, Seikei University

の注視時間, 相互注視時間を算出する. 以上の処理により算出したパラメータ値を2章で提案した優位性推定モデルに適用し, 優位性を推定する. 発話区間データを受信しなければ注視時間の算出を行えないことから, 本システムでは, 発話が発検出される毎に優位性推定を行う. 推定結果の値は, 会話介入システムにその都度送られる.

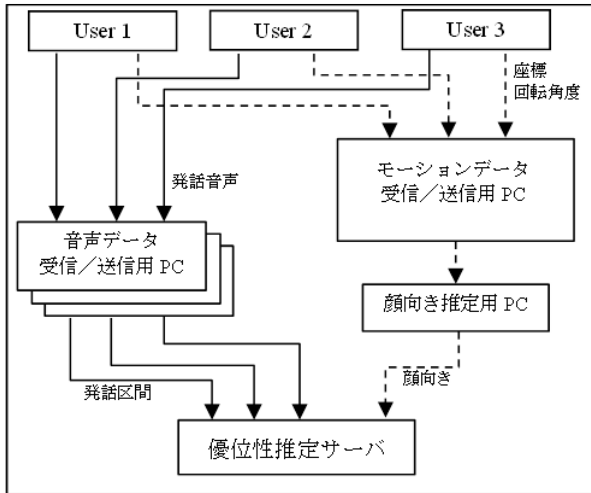


図2. 優位性推定機構のシステム構成図

3.2 介入発話生成システム

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)のどちらかに話しかけるかを, 会話介入システムにあらかじめ介入対象として設定しておく. 介入すべきタイミングであると判定されると, システムは, 指定された優位性の参加者の発話に関連した発話を生成する.

システム構成を図3に示す. マイクを通した音声データとモーションキャプチャにより取得された頭部動作データは優位性推定機構に送られ, 推定結果は優位性スコアとして介入決定モジュールに送られる. 音声データは同時に音声認識器にも送られ, 音声認識結果は介入決定モジュールに送られる. ここでは, まず, 音声認識結果にキーワードルールを適用し, 認識結果からキーワードを抽出する. 抽出されたキーワードは, 現在のトピックを判定するために用いられる. そして, キーワード, 介入対象, 現在のトピックの情報を用いて介入ルールを参照し, 介入内容を決定する.

さらにキーワードは介入タイミング決定部にも送られ, **Argument** 状態[2]を判断するためにも使用される. タイミング決定部では, キーワードの中の特定の言葉が1分間に3回以上出てきた場合, **Argument** 状態であると判断し, 介入を実行する. また, 音声データの音量に基づき沈黙状態を判定し, 3秒以上の沈黙が続いた場合に **Stagnation** 状態[2]と判断し, 介入決定モジュールで決定された発話が出力される.

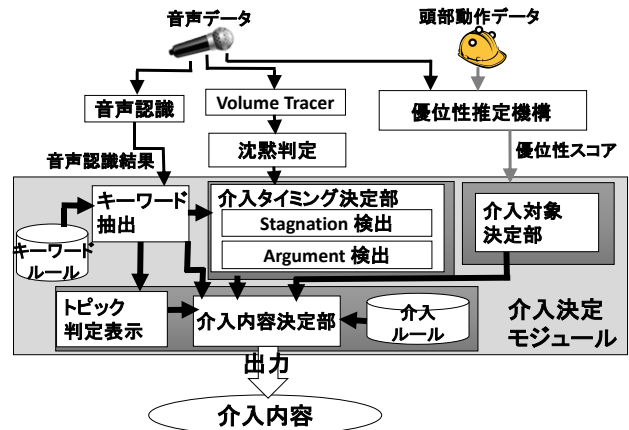


図3. 介入発話生成システム

4. 介入発話生成システムの動作例

介入対象を「優位性が高いユーザ」と設定した状況を想定する. ここで, 現在のトピックが「東京タワー」であるときに, ユーザから「イベント何やってる?」という発話があり, 音声認識結果が「イベント なに やる」であった場合, まず介入決定モジュールでは, 音声認識結果からキーワード「イベント」が抽出される. 次に, このキーワードが介入内容決定部に送られると, 現在のキーワード, 介入対象, トピックから介入内容が決定される. この場合, キーワードが「イベント」, 介入対象が「優位性が高いユーザ」, トピックが「東京タワー」なので, 「東京タワーでは現在お笑いライブ 333 が行われています. 行き先にしますか?」という介入内容が決定される.

介入対象が「優位性が低いユーザ」であった場合は, 介入内容が「東京タワーでは現在お笑いライブ 333 が行われています. お笑いは好きですか?」のように介入内容が変更される. 介入内容は介入タイミングであると判断された時点で出力される.

5. おわりに

多人数会話に積極的に参入するエージェントを目指し, 本稿では, 会話における優位性に基づき, 誰に話しかけるかを決定し, さらに介入発話を生成するシステムを実装した. 今後は, 本稿で提案した介入発話生成システムを全自動の会話エージェントに組み込むことにより, 多人数会話エージェントの実現に向けた実装を進める. また, 現在は介入タイミングが2種類でのみであるが, 介入タイミングのバリエーションを増やす必要がある. また介入内容の決定方法にも改良が必要である.

参考文献

[1] Nakano, Y.I. and Fukuhara, Y.: Estimating Conversational Dominance in Multiparty Interaction, Proc. of ICMI2012, pp.77-84 (2012).
 [2] 乙木翔地 他: 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, 2B-1 (2012).