2M5-OS-20b-4

確率的ルールを用いた多人数ロボット対話

Probabilistic Multiparty Dialogue Management for a Game Master Robot

○高橋 裕己[*1] Casey Kennington[*2] 船越 孝太郎[*3] 中野 幹生[*1*3] 菅野 重樹[*1]
Yuki TAKAHASHI Kotaro FUNAKOSHI Mikio NAKANO Shigeki SUGANO

[*1] 早稲田大学 Waseda University [*2] Bielefeld University [*3] ㈱ホンダ・リサーチ・インスティチュート・ジャパン Honda Research Institute Japan Co., Ltd.

We present our ongoing research on multiparty dialogue management for a game master robot that engages multiple human participants to play a quiz game. The robot invites passing people to join the game, instructs participants on the rules of the game, and leads them in the game. The robot has to manage people leaving and coming at arbitrary times. Our approach maintains a dialogue manager for each participant, and a module takes a final action with each decision cycle; responsible to decide "what/whom/when to say" in interaction. We have implemented the dialogue manager with a probabilistic rules approach [Lison 2012] and made preliminary evaluations with our multiparty human-robot game dialogue data that was collected in a Wizard-of-Oz fashion.

1. はじめに

多人数で行われる対話は，日常生活ではそれほど特別なものではない．例えば，3 人以上での雑談，学校の授業，会議など，会話する相手が 2 人以上ではあることは多い．日常生活で人間とコミュニケーションを取ることのできるロボットを作ることを考えた時に，多人数と対話できる能力が必要になる．

多人数と対話する仕組みとして考える必要があるのが，対話管理である．対話管理は盛んに研究されているが，これまで多くの研究は1対1の対話の管理に焦点を当ててきた．石崎らは1対1対話と比較して，多人数対話の特徴を探っている[Ishizaki 1998]．最近の研究ではターンテイキングの側面も含んだ仮想対話エージェントを使った多人数対話システムも存在する[Bohus 2011] [Traum 2012]．Keizerらは複数人相手のロボットバーテンダーのための多人数対話管理システムを提案している[Keizer 2013]．また，中野らは，韻律と顔向き情報から受話者推定を行い，多人数対話を行うエージェントを開発している[中野 2014]．多人数の対話管理では，「誰に対して」「何を話すか」を考えていくことになる．特に 1 対 1 の対話管理と異なるのは，話しかける事の出来る相手が複数いるので，会話の「受け手」の識別（Focus of Attention；以下 FOA）が必要となることである[Akker 2009]．

本稿では，クイズゲーム対話というドメインで，人間の参加者が何人いても対応できる多人数対話管理のモデルを提案する．このモデルは，近づいてきた人間に話しかけ，彼らをゲームに誘うという点でプロアクティブである．さらに，参加者がゲーム中に混乱しているようであればヒントを提供することで，ゲームを円滑に進めることができる．このモデルを実装したシステムには，参加者毎の個別の対話管理モジュール，それらの出力結果を時間軸上に一列に整列するモジュール，そしてそれぞれの行動の「受け手」(FOA)を予測するモジュールが含まれる．

以降では，まず2節で提案モデルを説明する．次に，3節で我々が収集したコーパスと対話行為のアノテーションについて説明する．最後に4節で，予備的な評価実験の結果を示す．

連絡先：高橋裕己，早稲田大学，
y_takahashi@sugano.mech.waseda.ac.jp

2. 対話管理モデル

2.1 多人数対話管理手法

Traum [2004]は，多人数対話管理のいくつかの問題を検討したうえで，2つのアプローチに言及している．1つ目は，多人数対話を1対1対話のペアの束として扱う方法で，2 つ目はすべての参加者を単一の統合モデルで扱う方法である．後者の方がおそらく理論的・表現能力的には優れた方法ではあるが，我々はその単純さから前者の方法を取る．本研究が対象とするクイズゲームドメインでは，全ての参加者の状態を統合してロボットの行動を決定しなければならない状況は多くないと考えられる．それ以上に，1 つ目のアプローチには，参加者が何人に増えても状態空間を指数関数的に増大させることなしに対応することができるという大きな利点がある．

2.2 確率的ルール

我々のモデルでは，確率的ルールを変換したベイジアンネットワーク[Lison 2012]を使用する．確率的ルールは，Lison の提案する対話管理モデルの手法で，確率モデルと手書きルールを組み合わせたものである．確率モデルには，音声におけるノイズや発話内容の不確実性などに強く，またデータからモデルのパラメータを最適化できる利点がある．しかし一方で，そのパラメータを最適化するための多量のデータが必要になる欠点を持っている．対話管理は一般にドメイン依存性が高いため，対話のドメイン毎にデータを集める必要が出てくる問題がある．確率的ルールでは，対話管理のルールを予め書き下しておくことで，パラメータの最適化を少ないデータで行う事ができる．この確率的ルールを用いて，それぞれの参加者について独立した個別対話管理モデル(IDM)を保持する．確率的ルールの具体的な実装については，4.2 節で説明する．

我々のアプローチの概略を図 1 に示す．ダイアログマネージャは，様々な情報を入力として受け取り，発話と視線（頭の向き）を出力とする．対話管理は参加者毎に個別の対話状態を保持し，様々な情報を入力として受け取る．受け取る情報は，誰が誰に向かって話しているか（Speaking）・視線（Looking）・発言内容（Speech/NLU）・対話行為（DA）・参加状態(Participating)である．

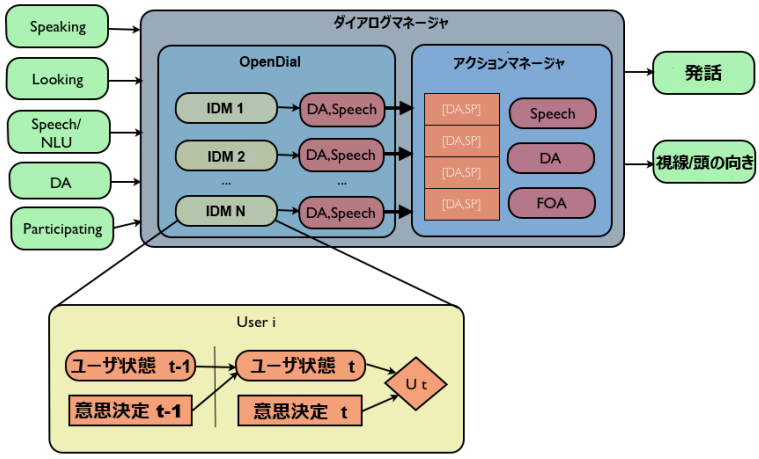

図 1: モデルの概要図

IDM の実装には[Lison 2012]実装である OpenDial[1]を用いた.

我々のアプローチのもう1つの重要な特徴は, 参加者(ユーザ)毎の IDM の決定に基づいて最終的にシステムの行う1つの行動とそのタイミングを決定するアクションマネージャの存在である. アクションマネージャは優先度付きキューとして実装されており, IDM から優先度が設定された対話行為を受け取る. 例えば, ゲームのルールを説明する場合, 5つの別々の対話行為それぞれがキューに入れられて, 順番に発話される. もし, 説明中のある時点でシステム発話の繰り返しが要求されると, 優先度の高い「発言の繰り返し」の対話行為がキューに追加され, ルールの残りの部分が完了する前に直近の発話の繰り返しが実行される.

アクションマネージャが発話行為を実行する際には FOA モジュールが呼び出される. FOA モジュールは, 単純な特徴量セットを使用して, ロボットの次の行為のすべての受け手(受話者)を推定する. 本稿ではこの受話者の推定に最大エントロピー法を用いる. IDM が反応した場合に, その IDM が対象とする参加者を受話者と決定する方法もある. 本稿では, この反応した IDM を FOA の対象とする方法を「単純ルールによる判定」と呼ぶ. しかし, この方法では FOA を誤ってしまう場合がある. 例えば参加者 A がもう一人の参加者 B と相談をしていて, A の方がロボットに質問をした場合を考えると, IDM は A に対応するもののみが反応するが, FOA としては A,B の両方を対象としなくてはならない.

研究を進める上で, 我々はいくつかの前提をおいている. システムは参加者を完全に検出することができ, また IDM のために必要なユーザの情報を入力信号から正確に分離することができるとする. そのため, 現在はユーザの情報が完全に手に入ることを前提にしたモデルとなっているが, 将来的には, 部分的な情報だけでも動くモデルに改良する予定である.

3. 多人数対話コーパス

3.1 データ収録

今回用いたデータは, 先に石川らが集めたものである[石川 2013]. 今回はそのデータにアノテーションをおこない, 実験に利用した. 以下でそのデータについて説明する.

データ収集の際には互いに知人関係にある 3 名を 1 組とし, 30 組 90 名の参加者を集めて, マルチモーダル対話データを収録した. Wizard of OZ 法で, ロボット(NAO)と参加者に「20 のとびら」と呼ばれるクイズゲームを 1 組 25 分間行ってもらった. 参加者は, ゲームの監督者から無線で指示を与えられ, 下記の4つの行動のどれかを行う.

- フィールドに入って, ゲームに参加する
- フィールドに入るが, ゲームには参加せず, 傍観する
- ゲームから離脱し, フィールドから出る
- フィールドを素通りする

ロボットのオペレータは, 下記の行動指針に基づいて操作を行う.

- フィールドに新たな参加者が入場したとき, その参加者に顔を向け, ゲームへと誘う発話を行う
  - ➢ 参加者が誘いを承諾したとき, その参加者をゲーム参加者とみなす
    - ✧ 参加者が初めてゲームに参加する場合は自己紹介発話を行う
  - ➢ 参加者が誘いを拒否するか, 反応がなければその参加者を傍観者とみなす
- ゲームのシナリオに則って, ゲームに参加している参加者に向けて定型文発話をする
  - ➢ 傍観者がフィールドにいる場合, 一定時間経過するとその傍観者をゲームに誘う
  - ➢ 参加者がフィールドから去ろうとしたとき, その参加者のほうへ顔を向け, 呼び止める発話を行う

参加者の数は, 監督者の指示によって, 1 名から 3 名の間で増減する. また, 英語学習目的のゲームであるため, ロボットは基本的に英語で話すが, 参加者の方は日本語の使用も許されている. 参加者は, ロボットの発話を聞き取れなかった場合, 直前の発話を聞き返すことができる. そして, フィールドにいる参加者同士で自由に発話することができるので, ロボットの発話内容や意図について互いに相談することができる.

3.2 コーパス構築と対話行為のアノテーション

収録したデータのうち 10 セッションに対し, ELAN[Auer 2010] を用いてゲーム参加中の参加者の発話対象, 注視対象, 会話へ

[1] https://code.google.com/p/opendial/

DA の種類	DA の出現数		DA をつける状況
	ユーザ	ロボット	
Accept-Request	2	17	相手が聞き返しなどの要求をして，要求に応じた場合.
Accept-Suggestion	92	3	ゲームへの参加など，相手のした提案を承諾した場合
Agreement	187	58	相手の意見に同意した場合
Answer	63	683	相手の質問に答えた場合
Apology	6	1	謝罪した場合
Call	5	184	ゲームへの誘いなどの呼びかけをした場合
Confirm	465	3	確認をとった場合.
Congratulation	5	33	賞賛した場合
Decline-Suggestion	15	0	提案を拒否した場合
Express-Emotion	161	22	感情表現をした場合
Good-bye	15	10	さよならの挨拶
Greeting	34	125	挨拶した場合
Inform	464	116	新しい情報を話した場合
Introduce	40	35	自己紹介
Monologue	514	41	独り言
Opening	4	51	ゲーム開始の掛け声
Propositional-Question	1009	6	Yes-No で答えられる質問をした場合
Question	179	51	Propositional 以外の質問
Request	127	44	聞き返しなどの要求をした場合
Return-Greeting	67	7	挨拶を返した場合
Suggestion	43	355	提案をした場合
Time-Management	304	0	言い淀みなどをした場合
Thanking	25	0	感謝を述べた場合
Turn-Taking	37	4	発言を求める発話
Unhearable	202	0	音声の重なりなどで聞き取れないもの

表 1: DA タグ一覧と実験中の数

Nao :Please ask me an Yes-No question. (*Suggestion)*
ユーザ :なんて言った？ (*Request)*
Nao :Please ask me an Yes-No question. (*Suggestion)*
ユーザ :それは食べ物？ (*Propositional-Question)*
Nao :Yes (*Answer)*

図 2: DA 付与した対話例

の参加状態，対話行為に関する注釈層をそれぞれ設定し，タグを付与した．DA タグの一覧を表 1 に示す．

Speaking 注釈層は，参加者の発話対象を表す．参加者の発話 1 つにつき，1 つ付与される．参加者 A が参加者 B に話した場合は toB がつき，B と C の 2 人に話した時は，toBC となる．ひとりごとの場合は Monologue，笑いの場合は Laughter タグが付与される．

Looking は参加者が注視している対象を示す．参加者がロボットを見ている場合は toNAO，参加者やロボット以外を見ている場合は toOthers，視線が確認できない場合は，Invalid タグを付与する．

Participating は，参加者がゲームに参加しているかどうかのタグである．ロボットと会話してゲームに参加している場合は Participating，ゲームに参加せず見ているだけの時は Observing，フィールドを通り過ぎるだけのときは Passing，ゲームから離脱しようとしている場合は Leaving タグが付与されている．

対話行為のタグについては，アノテーションスキーマの 1 つである DIT++[Bunt 2009]を参考に，Suggestion(提案)や，Request(要求)などの 25 種類の DA タグを付与した．DA は，その発話の意図を考え，付加している．例えば，実際の対話への DA タグ付与の例である図 2 の対話では，参加者の「なんて言った？」という発話は Question（質問）ではなく，繰り返しの Request（要求）と解釈して，DA を付与している．

また，ゲームの進行状態について，Phase 注釈層と X_situasion 注釈層の 2 つのアノテーションをおこなった．

Phase 注釈層はゲーム全体の進行に関するもので，下記の 2 つの状態がある．人の出入りなどによって，これらの状態を移行しながらゲームが進行する．

- Engagement
- Game

Engagement は，自己紹介，ルール説明を行なっている状態で，原則としてゲーム参加者全員の自己紹介が終わり，ルールを理解している場合に Game へと移行する．ただし，この判断は WOZ に任されているために絶対ではない．

Game はゲームをしている時の状態で，このときに人の出入りがあると同じく WOZ の判断で Engagement に移行する．

X_situation 注釈層は，各参加者の状態についての注釈で，下記の 3 種類が存在する．

- NameGiven=Y :参加者が自分のニックネーム(Y)を名乗った直後の状態
- IntroCompleted :自己紹介が終わり，ゲームの説明に入る前の状態
- RuleInstructed :ルールの説明を終えた状態

これらを各参加者の状態の段階に沿って付与した．

4. 予備的評価実験

4.1 FOA

アノテーションを付与した 10 セッションのうち，01-09 番の 9 つのセッションを用いて FOA の最大エントロピー分類器を訓練し，10 番のセッション 1 つを用いて FOA 判定精度のテストを行った．正解ラベルは，1 人から 3 人の参加者のラベルセットとして用意されている．例えば正解ラベルが，2 人の参加者の集合{A,B}であったときに，FOA モジュールの出力が{A}であった場合，このときの正答率は 50%とする．テストの結果，提案手法における FOA 正解率は 71%であった．さらに 10 番のセッションのデータも含めて，事例単位での leave-one-out 交差検定を行った結果は 78%であった（ただしここでも 01-09 番のデータは訓練にしか用いていない）．

マジョリティベースラインの正解率は 33%，2.2 節で説明した単純ルールで求めたベースラインの正解率は 57%であった．

```
  <case>
   <condition>
    <if var="Participating" value="Participating" />
    <if var="answer" value="correct" />
    <if var="DA" value="Propositional-Question" />
    <if var="Speaking" value="toNAO" />
   </condition>
   <effect util="11.8555">
    <set var="dialogue-act" value="Answer" />
    <set var="ds-speech" value="you right! congratulations!" />
    <set var="correctAnswer" value="true" />
   </effect>
  </case>
<case>
```

図 3: 学習されたルール

ルールに使われる変数の種類	変数の説明
Answer	ユーザの答えがあっているかどうか
IntroCompleted	自己紹介が終わっているか
RuleInstructed	ルール説明が終わっているか
dialogue-act	ロボットが行う DA
ds-speech	ロボットの発話内容

表 2: ルールに使われる変数の例

4.2 DM

訓練データから OpenDial が用いるルールを抽出し，同じデータを用いてそのルールの確率パラメータの学習を行った．抽出したルールの一部を図 3 に示す．ルールに用いる変数には，注釈したユーザ状態（Speaking など）の他に，IntroCompleted などのゲーム進行に関連する変数が存在する．変数は全部で 13 個あり，その一部を表 3 に示す FOA の評価と同様に 10 番のセッションを用いて，OpenDial による対話行為選択の評価を行った．この評価には全ての参加者の個別対話管理モデルに同じモデルを使用した．合計で 19 個の対話行為がある中で，52%の正解率を得た．マジョリティベースラインは 37%，バイグラムのマジョリティベースラインは 51%であった．

5. まとめ

本稿では，確率的ルールを用いて多人数と対話を行うロボットの対話管理モデルの提案を行なった．そして，予備的実験として，FOA の推定と，DM による DA の推定を行なった．結果として，人間とシステムとの多人数対話インタラクションの実現に，提案した対話管理モデルが有効であることが示唆された．FOA の推定では，最大エントロピー法を用いることによりベースラインの正解率を超え，さらにデータを増やすことによってさらに正解率が上がる可能性が示唆された．DA の推定では，バイグラムのマジョリティベースラインとの正解率の差はわずかであるものの，改善が見られた．こちらもデータを増やすことで改善できる可能性がある．さらに，DA が正解でない場合にも，実際の会話では問題がない場合もある．こちらの方は，システムと実際に会話をしてもらい，アンケートを取ることで評価を行う予定である．

今後は，より詳しい評価のために，人間とのリアルタイムのインタラクションを行う必要がある．そのために，ユーザ状態を読み取る部分・ロボットに実際の動作を行わせる部分の実装が必要になる．さらに，リアルタイムで動作し，入力されたイベントに基づき意思を決定するだけでなく，簡単な説明リクエストの処理，合成発話の生成をする必要がある．我々は，さらに多くのデータを使い，IDM の対話行為の予測の改善と，発話のタイミングを含むより複雑な状況に対処するためにアクションマネージャの改善をしていく．

謝辞

本研究成果の一部は，科学研究費補助金基盤研究(S)25220005 の助成を受けたものであり，また，早稲田大学理工研プロジェクト研究「自然と共生する知能情報機械系に関する基盤研究」の一環として行われたものである．ここに謝意を表す．

参考文献

[Bohus 2011] D. Bohus and E. Horvitz. Multiparty Turn Taking in Situated Dialog : Study , Lessons , and Directions. In Proceedings of SIGDIAL 2011, , pages 98–109. , 2011.

[Ishizaki 1998] M. Ishizaki and T. Kato. Exploring the characteristics of multi-party dialogues. In Proceedings of the 17th international conference on Computational linguistics Volume 1, pages 583–589. , 1998.

[Keizer 2013] S. Keizer, M. E. Foster, O. Lemon, A. Gaschler, and M. Giuliani. Training and evaluation of an MDP model for social multi-user human-robot interaction. In Proceedings of SIGDIAL 2013, pages 223–232, Metz, France, Aug. 2013.

[Akker 2009] R. op den Akker and D. R. Traum. A comparison of addressee detection methods for multiparty conversations. In Proceedings of the 13th Semdial Workshop on the Semantics and Pragmatics of Dialogue, pages 99–106, Stockholm, Sweden, 2009.

[Traum 2004] D. Traum. Issues in multiparty dialogues. In Advances in agent communication, pages 201–211. Springer, 2004.

[Traum 2012] D. Traum, D. DeVault, J. Lee, Z. Wang, and S. Marsella. Incremental dialogue understanding and feedback for multiparty, multimodal conversation. In Intelligent Virtual Agents, pages 275–288. Springer, 2012.

[Lison 2012] P. Lison. Probabilistic Dialogue Models with Prior Domain Knowledge. In Proceedings of SIGDIAL 2013, pages 179–188, Seoul, South Korea, July 2012.

[Auer 2010] E. Auer, A. Russel., H. Sloetjes, P. Wittenburg, O. Schreer, S. Masnieri, D. Schneider, and S. Tschpel :ELAN as Flexible Annotation Framework for Sound and Image Processing Detectors, in LREC'10, Valletta, Malta (2010), European Language Resources Association(ELRA).

[中野 2014] 中野有紀子，馬場直哉，黄宏軒，林佑樹：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，人工知能学会論文誌，Vol.29, No.1, pp.69-79，2014.

[石川 2013] 石川真也，船越孝太郎，篠田浩一，中野幹生：多人数対話ロボットの実現に向けたマルチモーダル対話データの収集と分析，人工知能学会全国大会，2013.

[Bunt 2009] Harry Bunt: The DIT++ taxonomy for functional dialogue markup. In Proceedings of EDAML 2009.