

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成25年 5月 27日現在

機関番号：32689

研究種目：若手研究（B）

研究期間：2011～2012

課題番号：23700239

研究課題名（和文）

複数人会話活性化ロボットの開発と評価

研究課題名（英文）

Development and Evaluation of A Multiparty Conversation Facilitation Robot

研究代表者

松山 洋一（MATSUYAMA YOICHI）

早稲田大学・理工学術院・助手

研究者番号：90584467

研究成果の概要（和文）：

複数人が存在し進行される会話に参加し、その会話を活性化するロボットの開発を行った。従来、音声対話システムおよび会話ロボットは1対1の対話を対象にしてきたが、本研究では人同士で行われている複数人会話にロボットを参加させ、かつその会話を支援したり活性化したりすることで新たな価値を生み出すロボットシステムの開発を行った。

研究成果の概要（英文）：

We proposed a conversational robot system facilitating multiparty conversations. Although conventional spoken dialogue systems and conversational robots have focused on one-to-one dialogue, we focused on multiparty conversation situation. We developed a conversational robot that can support conversations to produce invaluable experiences.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
交付決定額	3,000,000	900,000	3,900,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報処理・知能ロボティクス

キーワード：知能ロボティクス、音声対話処理、人-ロボット・インタラクション

1. 研究開始当初の背景

我々の日常における会話は1対1の対話だけでなく、複数人で構成させる会話の場で進行される状況が多い。従来、音声対話システムおよび会話ロボットの研究では1対1の対話を対象にしてきたが、将来ロボットが我々の日常で共存する状況を考えたとき、複数人の存在する環境下で適応的に行動できる枠組みが検討される必要がある。このとき、1対1の対話では意識されなかったいくつかの難しい問題が立ち現れる。まず、複数人会話では発話権の授受が極めて複雑になる。ある発話者の発話が終了したときの次話者が誰であるかが自明ではない。この次話者推定には、言語情報だけでなく、会話の各参加者の

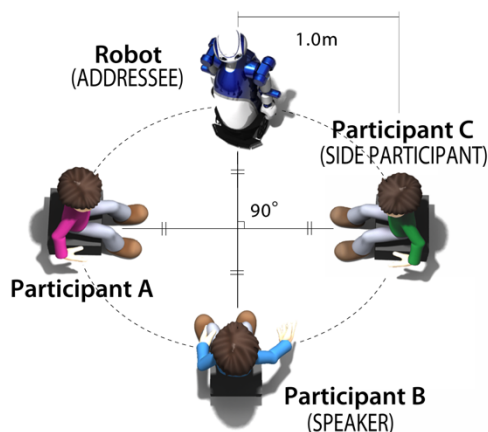


図1. 複数人会話の状況

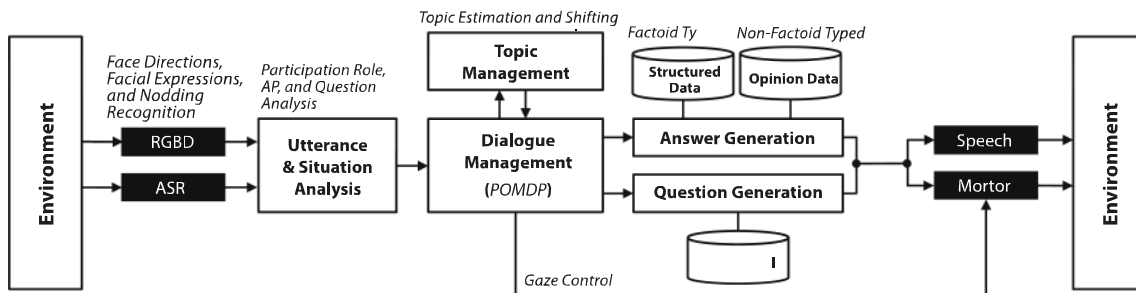


図 2 システムアーキテクチャ

視線や身体の向き等の情報も必要となる。また、会話に参加するロボットの視線の方向や身体の方向はそれぞれに重要な社会的意味を有するものであり、状況に応じて適切に制御しなければならない。故に複数人会話を実現するためには、ロボットのセンサ（音声、画像、その他）や形態、自由度といった身体性を総合的に検討する必要がある。このような複数人会話へ参加するロボットの行動選択はロボット工学や人工知能分野において近年非常に重要かつ興味深い問題として指摘され始めている。しかし、実際に複数人を対象として動作できるロボットの研究は世界的に見ても現在ほとんど存在しない。

しかし同時に、ロボットのいくつかの行動が場の活性を阻害することも明らかになってきた。これを改善しより会話を活性化させるためには、複数人会話のモデルをより精緻なものにし、その上で会話活性化のための行動戦略を検討する必要がある。よって本応募研究課題では、複数人会話にロボットを参加させ、かつ場を活性化させるための詳細な計算モデルを検討する。

2. 研究の目的

本研究の目的は、大きく分けて以下の2つである。

- (1) 会話活性化戦略
- (2) 発話文自動生成

3. 研究の方法

3.1 会話活性化戦略

3.1.2. 参与役割推定

本論文では、参与役割推定とは会話に承認されている各参加者に対して話し手、受け手、傍参与者の何れかの参与役割を正しく割り当てる問題であるにとらえ、話し手は会話内において発話権を保持している人物であると定義し、場には多くとも1名のみに存在しているとする。受け手は話し手によって発話をアドレスされている人物であり、アドレスされていない人物は傍参与者となる。実際の会話において誰が受け手であるかは、常に曖昧性なく判断できるとは限らないが、ここでは、話し手同様に受け手

も多くとも1名のみ存在するものとし、参加者全員に発話を行っている場合は特定の受け手は存在せず、話し手以外は傍参与者として扱うことにする。

話し手:発話権を保持している人物。一人のみ。

受け手:話し手にアドレスされた特定の人物。

傍参与者:話し手でも受け手でもない人物。

手順としては、まず各参加者の発話終了時の会話参加者全員の発話状態(発話中か否か)と顔向きを入力として話し手の推定を行う。次に推定された話し手顔向きの情報を用いて受け手を推定する。受け手の推定は、話し手の視線方向から決定的に決まるものとした。本論文では、ルールベースの推定器を用いた。図1のようなセッティングでロボットの位置にも人間の参加者が座って自由会話を収録したものに人手で役割をラベリングしたものをテストデータとして用いて実験した結果、話し手の推定精度は75.1%、受け手の推定精度は67.2%であった。

3.1.1 「置いてけぼり」状態の推定

参与役割の視点から、「置いてけぼり」の状態について考える。図1のような状況で、AとBが活発に会話を行い、Cに発話機会がない状態であったとする。このとき、Cは傍参与者である時間帯が結果的にしばらく続いていることになる。また、このような「置いてけぼり」状態は、話題が変わると改善される場合がある。本論文では、各参加者が当該話題区間における傍参与者にアサインされている割合を、「置いてけぼり度(=傍参与状態の深さ $Depth_{SPT}$)」と定義する。

$$Depth_{SPT} = (Duration_{SPT}) / (Duration_{topic})$$

3.1.3. 会話参加意欲度

「置いてけぼり」状態にある参加者は、さらに会話参加への意欲をもっているか否かで状態が分かれると考えられる。すなわち、発話機会が与えられていないが自分も積極的に参加しようと思っている参加者と、そもそも現在の会話にあまり興味がない場合である。我々がロボットを介して活性化させた会話(各参加者の参加意欲が高い状態

にある会話)の心理的側面について分析した結果によれば、「楽しさ」,「賑やかさ」,「快適さ」などの因子が作用していることがわかっている。客観的な指標として,そのような心理状態にあるときに各参加者の発話量,笑顔頻度,頷き頻度などが上昇傾向を示すこともわかっている。よって本研究では,参加者 i の活性度を,発話量 $SpeechAmount_i$, 笑顔量 $SmilingAmount_i$, 頷き時間 $NoddingAmount_i$ から以下のように計算されるものと定義する。

$$Activity_i = \alpha(SpeechAmount)_i + \beta(SmilingAmount)_i + \gamma(NoddingAmount)_i$$

発話継続時間 $SpeechAmount_i$ は,各参加者 i の Voice Activity Detection(VAD)の結果を用いた。発話が継続するというはその参加者の発話意欲が高いと考えられるので,定時間活性状態が続くと考えられるので,音声区間が継続する間は発話量 $SpeechAmount_i$ が線形に単調増加し,認識されていない間単調減少するものとした。笑顔・頷きも同様に笑顔・頷きであると認識されている時間は笑顔量 $SmilingAmount_i$ 頷き量 $NoddingAmount_i$ とが線形に単調増加し,認識されていない間単調減少するものとした。笑顔の推定には,会話参加者の正面に置かれた深度センサ(Kinect)を利用し,顔の表情の推定結果を用いた。頷きは,顔のピッチ角の振動の周波数を用いた。それらの線形和で表される活性度 $Activity_i$ を用いて,ある閾値以上のときに活性度 High 状態,それ以外を Low 状態とすることとした。

3.1.4 POMDP による活性化戦略最適化

会話活性化戦略の効果を最大化させるために,部分観測マルコフ決定過程 (Partially Observable Markov Decision Process: POMDP) を用いてモデル化する。

一般に,POMDP は式 (3) のように状態 S , システム行動 A , 状態遷移確率 T , 報酬 R , 観測確率 Z , 割引率 γ , 初期の信念状態 b_0 の集合として定義される。

$$\beta = (S, A, T, R, Z, \gamma, b_0)$$

信念状態は以下のように更新される。

$$b'(s') = \gamma \cdot P(o'|s',a) \sum_s P(s'|s,a)b(s)$$

ここで, S を各参加者の調和状態を S_h , 「置いてけぼり」状態の参加者の参加意欲を S_m , 参加者の行動を A_p を用いて定義する。

$$s = (s_h, s_m, a_p)$$

すると,状態の遷移確率は以下のように表現

$$P(s'|s,a) = P(s'_h, s'_m, a'_p | s_h, s_m, a_p, a_s) \\ = P(s'_h | s_h, s_m, a_p, a_s) \cdot \\ P(s'_m | s'_h, s_h, s_m, a_p, a_s) \cdot \\ P(a'_p | s'_m, s'_h, s_h, s_m, a_p, a_s)$$

できる。

ここで,簡単のため状態遷移確率に次のような仮定を置く。

$$T_{S_h} = P(s'_h | s_h, a_p, a_s) \quad (\text{調和状態モデル})$$

$$T_{S_m} = P(s'_m | a_s) \quad (\text{参加意欲モデル})$$

$$T_{A_p} = P(a'_p | s'_h, a_p, a_s) \quad (\text{参加者行動モデル})$$

よって,信念状態の更新式は以下のようなになる。

$$b'(s'_a, a'_p) = \gamma \cdot P(o' | s'_h, s'_m, a'_p, a_s) \cdot \\ \sum_{s_h} P(s'_h | s_h, a_p, a_s) \sum_{s_m} P(s'_m | a_s) \cdot \\ \sum_{a_p} P(a'_p | s'_h, a_p, a_s) \sum_{a_p} b(s_a, a_p)$$

調和状態,ユーザ行動,システム行動はそれぞれ表 1~3 のようになる。状態遷移図は図 3 のようになる。

表 1 調和状態

Harmony states	Meaning
<i>U-Harmonized</i>	the robot isn't harmonising with the current conversation.
<i>Pre-harmonized</i>	the robot is waiting for an approval to harmonize with the current conversation.
<i>Harmonized</i>	the robot is harmonising with the current conversation.

表 2 ユーザ行動

Participants' actions	Meaning
<i>first-part</i>	a participant made an adjacency part (question)
<i>second-part</i>	a participant made an second part (answer)
<i>third-part</i>	a participant made a third adjacency part
<i>other</i>	a participant asked or answered to the other participant
<i>call</i>	a participant called the robot's name

表 3 システム行動

System actions	Meaning
<i>answer</i>	answering the current speaker's question
<i>question-new-t</i>	
<i>-current-t</i>	
<i>-reaction</i>	reacting simply
<i>nod</i>	nodding to the current speaker
<i>none</i>	doing nothing

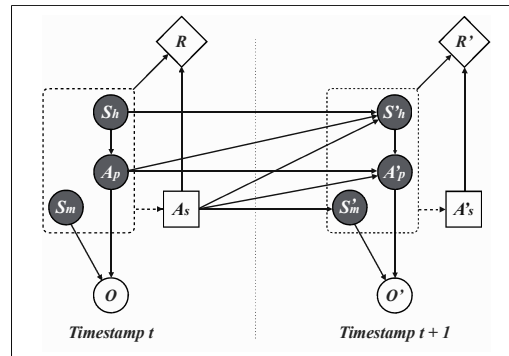


図 3 状態遷移図

3.2 発話文自動生成

発話相手として魅力的な、「意外性のある」意見・感想発話文の自動生成の方法について検討する。

例えば、「オードリー・ヘップバーンは綺麗だね」という文章は、多くの人が共感できるであろう内容である。しかし大多数の人が抱くような通り一遍の感想だけを述べるシステムは、常に楽しいシステムになるとは言えない。相手の興味を引くような意見・感想の意外性は、意見を述べる立場や、表現の多様さに現れると考えられる。大衆におもねるだけの意見や、簡潔すぎる意見を述べるだけでは、興味深い意見にはなりえない。上記のオードリー・ヘップバーンの例では、「オードリー・ヘップバーンは綺麗だね」という発話は簡潔すぎるものであるが、「オードリーは期待を裏切ることなく、プライベートの彼女もとても素敵な美しい女性だよ。」などのような文章は、より豊富な表現でオードリー・ヘップバーンについて感想を述べたものであると言える。つまり、文章がある程度長く、多様な形容詞表現を含むものがより意外性を含む意見・感想である可能性が高いと考えられる。この形容詞の多様性は、使用する形容詞自体の利用頻度にある程度反映すると考えられる。すなわち、比較的使用頻度の低い形容詞をいくつか含む文章は、大多数の意見とは少し違う視点で意見を述べている可能性が高い。しかし、あまりに頻度の低い表現だけで構成された文章は、逆にユーザの共感を得られないことも考えられる。また、冗長すぎる意見・感想文が続くことも楽しい対話にはマイナスであると考えられる。以上より、本論文では対話の楽しさのための意見・感想発話文を生成するために、形容詞表現の多様性と長さを制御しながら文章を生成する手法について検討することにする。

意見・感想発話文生成は、(1) 文書収集、(2) 評価表現抽出、(3) 文体変換、(4) ランキングの4段階のプロセスから構成される。

3.2.1 文書収集

文書収集の対象として、意見・感想の文章が多く存在する Yahoo!映画のレビューサイトを対象とした。任意の話題(映画名、俳優名で 101 話題)で検索するとレビューが表示される。このレビューサイトに対するクローラを実装し、レビュー文を取得した。レビュー文は採点順(レビューアの 5 段階評価)でソートし、上限 1000 レビューをまとめたものをレビュー文書とする。採点順でソートしたため、肯定的意見が多く含まれていると考えられる。

3.2.2 評価表現抽出

評価表現抽出(意見マイニング)のプロセスは「評価表現の抽出」、「評価極性の判定」の2段階となる。評価表現の抽出は、文書内から意見・感想の文章のみを抽出する手法である。抽出手法として、条件付き確率場(CRF)によって文中の各形態素に評価表現の開始(B)、中間(I)、評価表現以外(O)を表すタグを付与する。学習文は従来我々が手動で作成した発話文とレビューサイトからランダムに選んだレビュー文を使用し、BIOタグを付与した。素性は前後2つまでの形態素の出現形、原形、品詞大分類、品詞細分類、評価表現辞書中での極性を利用し、CRFの学習を行った。評価極性の判定は、評価文が肯定的または否定的な意見・感想なのかを判定する手法である。文中の極性反転を考慮して、条件付き確率場で極性を判定している。

3.2.3 文体変換

レビュー文書の文章の文体はレビューアによって様々であるため、対話システムの発話文に一貫性を持たせるように文体変換を行う。今回文体変換で最も重要となってくる文末表現に注目した。例えば、レビュー文書においては「良いと思います」となっているが、対話システムの発話文とするときには「良いと思うんだ」と変換する。文末表現の変換はルールベースで実装した。形態素解析の結果、文末から参照する。特殊な文字(句点など)、助詞、助動詞、指示詞などは必要ないため除去する。除去した後に最後の形態素の品詞によって場合分けで変換する。例えば、「良いと思います」の最後は接尾詞「ます」であるので、動詞「思い」を原形である「思う」に変換し、直後に「んだ」を付け加えた。その結果、「良いと思います」は「良いと思うんだ」という文体に変換できる。

3.2.2 発話文ランキング

ランキングのための尺度は、情報抽出、検索検索の分野などで一般的に用いられる TF-IDF (Term Frequency-Inverse Document Frequency)、形容詞に関する TF (Term Frequency)、および形態素数の 3 要素を用いることとした。単語重要度はその文章が話題に関連する程度である。形容詞 TF は共感または意外性の尺度である。形態素数は簡潔な短文と根拠も含まれた長文を抽出するために利用する尺度である。

ランキングする際はまず話題と関連した文章をランキング上位にする必要がある。レビュー文書自体の話題と関連した名詞が含まれた文章を認識しなければならない。例えば話題「ローマの休日」のレビュー文

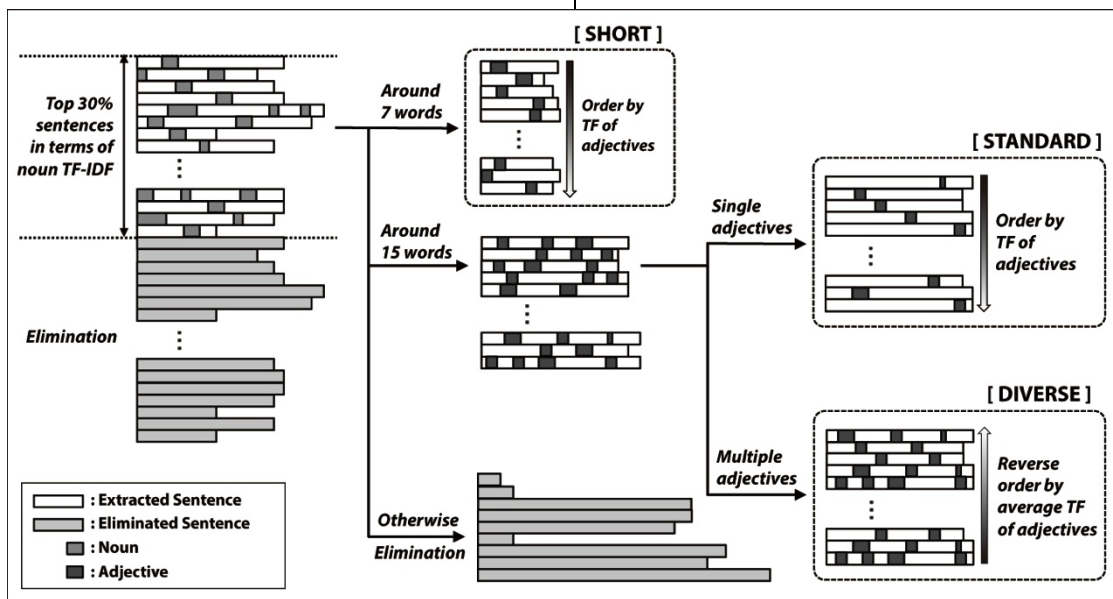


図4 ランキング手法

書においては、「王女」などの関連単語が含まれた文章であるので、TF-IDF 尺度を用いてソートする TF は任意の文書中の単語の出現頻度を示し、文書との関連度と比例する。また、IDF は任意の単語が文書(ここでは 101 文書)のうち現れる文書の数であり、低い値の単語は広く使われる単語となる。よって TF-IDF 値が高い名詞は、そのレビュー文書において重要な意味を持つ名詞となる。

意見・感想発話文の表現の豊かさの尺度として形容詞の TF(Term Frequency)を利用することにした。オードリーヘプバーンのレビュー文書の全体の中では、「綺麗だ」という形容詞は頻繁に利用されるため、形容詞の TF の値は高い。「綺麗だ」という形容詞の表現は、多くのレビュアーによって表現されていることになる。よって「綺麗だ」が含まれる意見・感想文はユーザが共感する可能性が高いと考える。一方、TF が低い形容詞は「御転婆だ」というように、レビュアー数人にしか表現されない形容詞となる。

また、レビュー文書内には冗長な文章、あるいは短すぎる文章が存在するため、文章の形態素数にも着目した。形態素数 7~10 である短い文章は、簡潔に意見・感想が述べられている文章である。一方、形態素数 15~20 である長い文章は、意見・感想とその根拠が述べられている可能性が高い文章となる。

発話のランキング手法として「Short」「Standard」「Diverse」の3種類提案する。ランキング手法のフローの概要を図4に示す。

Short 戦略の発話は、ユーザが共感する短文である。例えばトピック「ローマの休日」においては、「オードリーヘプバーン

の存在感が素晴らしいよね」といった文章である。簡潔に意見・感想発話をして、ユーザに発話する機会を与える。Short 戦略の場合は、形態素数が 7~10 の文章の中で、文中の名詞の TF-IDF(最大のものを採用)でソートして上位 30%を抽出する。これでトピックに関連した文章となる。次に文中に含まれる形容詞 TF 平均でソートする。形容詞が 1 語のみ含まれた文を採用し、上位から発話文を出力していく。

Standard 戦略の発話は、ユーザが共感する長文である。例えば「今回改めてスクリーンで見てオードリーの立ち姿の美しさに惚れ惚れしたんだ」といった文章である。Standard 戦略の場合は、形態素数が 15~20 の文章の中で、名詞の TF-IDF でソートし上位 30%を抽出し、形容詞が 1 語のみ含まれた文を採用し、形容詞 TF 平均でソートする。

Diverse 戦略の発話は、ユーザにとって意外性のある意見・感想発話である。多様性のある表現が含む、つまり形容詞 TF が低い形容詞が複数含まれている文章を意外性としている。例えば「キュートなオードリーとジェントルなグレゴリーが見事に織りなした、ロマンチックかつ切ないストーリーだよ」といった文章である。この戦略では長文とする。短文の場合、意見・感想の根拠が入る可能性が低いいため、「オードリーヘプバーンは切ない」のように発話してしまう。ユーザが共感する可能性が低く、その根拠が無いリスクが高いと考え、長文のみを採用した。多様性のある表現で、ユーザの興味を引き対話の楽しさを高める効果を期待する。Diverse 戦略の場合は、形態素数が 15~20 の文章の中で、名詞の TF-IDF でソートし上位 30%を抽出する。次に形容詞を複

数含む文を採用し、形容詞 TF の平均で逆ソートする。形容詞の平均の値が低いほど、希少価値の高い表現をしている文である可能性が高くなると期待される。

4. 研究成果

4.1 活性化戦略最適化 (POMDP) の評価

条件 1: 受動的な QA システム

条件 2: 「置いてけぼり」状態の参加者が検出されたときに即座に質問行動 (ルールベースによる自発的行動) を起こすシステム

条件 3: 参加者の調和状態も加味して適切な手続きを踏んで「置いてけぼり」状態の参加者にアプローチするシステム

上記の 3 条件を実装し、図 1 のような状況で会話の様子 (それぞれ約 80 秒) を撮影したものを被験者に視聴してもらい、ロボットの行動の適切性についてアンケートに回答してもらった。被験者は 24 名 (男性 11 名, 女性 13 名, 平均年齢 22.1 歳) であった。結果を図 5 に示す。

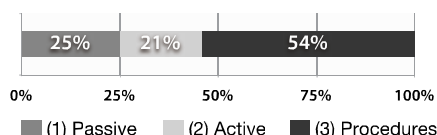


図 5 ロボットの活性化戦略の適切性

4.2 発話文の評価

条件 1: Short 手法による発話 (6 発話)

条件 2: Standard 手法による発話 (6 発話)

条件 3: Diverse 手法による発話 (6 発話)

条件 4: 上記 3 手法の混合 (2 発話 × 3)

上記の 4 条件を実装し、ロボットと人の 1 対 1 の会話の様子を撮影したものを被験者に視聴してもらい、アンケートに回答してもらった。被験者は 38 名 (男性 21 名, 女性 17 名, 平均年齢 22.0 歳) であった。結果を図 6 に示す。

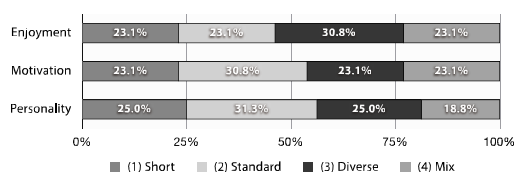


図 6 発話文の印象評価 (楽しさ, 主体的な参加動機, ロボットのパーソナリティへの好感度)

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 1 件)

(1) 藤江真也, 松山洋一, 谷山輝, 小林哲則, 人同士のコミュニケーションに参加し活性化する会話ロボット, 電子情報通信学会論文誌 A, Vol. J95-A No. 1, pp37-45, 2011. (<http://ci.nii.ac.jp/naid/110008898123>) (査読有り)

[学会発表] (計 11 件)

(1) Yoichi Matsuyama, SCHEMA: A Framework of Embodied Conversational Robots Facilitating Small Groups, Embodied Situated & Language Processing, Potsdam, Germany, July 2013. (査読無し)

(2) Yoichi Matsuyama, Yushi Xu, Akihiro Saito, Shinya Fujie and Tetsunori Kobayashi, "Multiparty Conversation Facilitation Strategy Using Combination of Question Answering and Spontaneous Utterances," IWSDS2011 Workshop on Paralinguistic Information and its Integration in Spoken Dialogue Systems, pp.99-107, Granada, Spain, Sep.2011. (査読有り)

(3) Yoichi Matsuyama, Conversation Robot Participating in and Promoting Multiparty Conversation, Workshop on Social Robots For Assisted Living, University of Aalborg, Denmark, Nov.2011. (招待講演)

(4) 秋葉巖, 松山洋一, 小林哲則, 多人数会話ファシリテーションロボットの主導権奪取手続き, 情報処理学会 SIG-SLP, July 2013. (査読無し)

(5) 松山洋一, 齋藤彰弘, 伊東篤史, 秋葉巖, 渡邊萌実, 小林哲則, 多人数会話活性化のための自発的行動タイミング検出と発話行動戦略, SIG-SLUD 2012 (人工知能学会 2012 年度研究会優秀賞), Feb. 2013. (査読無し)

(6) 齋藤彰弘, 松山洋一, 轟あずさ, 小林哲則, 意外性のある質問応答システムのための発話文生成手法, SIG-SLUD 2012, Feb. 2013. (査読無し)

(7) 松山洋一, 齋藤彰弘, 秋葉巖, 渡邊萌実, 小林哲則, 「置いてけぼり」を救う多人数会話活性化ロボット, HAI シンポジウム 2012, Dec. 2012. (査読無し)

(8) 松山洋一, 藤江真也, 齋藤彰弘, 小林哲則, 多人数会話活性化システムの発話戦略パターン, 人工知能学会第 26 回全国大会, 102-0S18-9, June 2012. (査読無し)

(9) 齋藤彰弘, 松山洋一, 藤江真也, 小林哲則, 会話ロボットの多人数会話活性化戦略とその評価, 信学技報, vol. 111, no. 225, SP2011-53, pp. 7-12, Oct. 2011. (査読無し)

(10) 藤江真也, 松山洋一, 齋藤彰弘, 小林哲則, 多人数会話に参加しコミュニケーションを活性化する会話ロボットの開発, 日本音響学会 2011 年秋季研究発表会, NO.3-10-6, Sep. 2011. (査読無し)

(11) 齋藤彰弘, 松山洋一, 藤江真也, 小林哲則, 質問応答と自発的発話の組み合わせによる複数人会話活性化戦略, 人工知能学会第 25 回全国大会, 3C2-0S19-10, June 2011. (査読無し)

[その他]

ホームページ等

www.pcl.cs.waseda.ac.jp

6. 研究組織

(1) 研究代表者: 松山洋一 (MATSUYAMA YOICHI)
早稲田大学情報理工学科・助手
研究者番号: 90584467