文脈を考慮したロボットの対話についての研究

松元 崇裕† 大村 廉† 今井 倫太†
慶應義塾大学 理工学部 慶應義塾大学 理工学部 慶應義塾大学 理工学部

matumoto@ayu.ics.keio.ac.jp, ren@ayu.ics.keio.ac.jp, michita@ayu.ics.keio.ac.jp

1 はじめに

本稿ではロボットと人とのジェスチャを用いた音声対話を扱う．実世界で生活支援を行うロボットを実現するには，人との自然な対話が可能であることが望まれる．自然な対話を実現するための問題の1つとして，対話における単語と実世界対象の参照関係の問題が挙げられる．ロボットが対話においてユーザの発話を解釈をするためにはユーザ発話内の個々の単語の意味を特定しなければならない．単語の意味は，その単語自身と実世界における物体・概念・知識との間で1対1の参照関係を取ることで初めて特定することが出来る．そのため，単語の対応先が一意に決められない場合はユーザ発話の解釈をすることが出来なくなってしまう．

ユーザの発話における単語の参照関係の問題を扱った研究として，ユーザのジェスチャと対話時の環境情報を考慮する手法が提案されている [1]．ユーザ発話をジェスチャ情報・環境情報も含めて解釈を行うことで，発話のみを扱う手法では解決できない単語の参照関係の特定を可能にしている．

しかしながら，この手法はジェスチャや環境情報を用いても単語の対象が特定できない場合に単語の参照関係を解決することは出来ない．単語の意味はジェスチャや環境情報だけでなく対話の文脈においても変化するため，単語の参照関係を解決するためには文脈自体を取り扱う必要がある．

自然言語対話において文脈を扱う手法としては，文脈情報によりユーザ発話の解釈の精度を向上させる研究 [3] などこれまでも多く行われてきた．しかしながら，単語の参照関係に注目した研究は少なく，対話時の環境情報やジェスチャ情報も扱っていない．

文脈に基づき，ユーザ発話における単語の参照関係の問題を解決しようとする際の難しい問題としては文脈と単語の参照関係の相互依存関係がある．個々の単語の参照関係は文脈情報に照らし合わせることで初めて決めることが出来る．しかしながら，文脈はユーザ発話の解釈結果から構成されるため，正確な文脈を得るためにはユーザ発話における単語の参照関係が正しく与えられ，単語の意味を特定してなくてはならない．

そこで，本研究では文脈と単語の参照関係における相互依存関係を解決するモデル SIAC(Simultaneous Interpretation And Contextualizing) を提案する．また，環境情報・ジェスチャに加え文脈を考慮する対話システムの構築をSIACを適応することで行い，単語の参照関係の問題が解決されることを示す．

Study of Human-Robot Interaction based on Context
†Takahiro Matsumoto
†Ren Ohmura
†Michita Imai
Faculty of Science and Technology, Keio University

2 SLAMアルゴリズム

同様の相互依存関係を扱った問題として，地図と位置の関係における相互依存関係を解決するSLAMアルゴリズムが提案されている．SLAM(Simultaneous Localization and Mapping) 問題は自律移動ロボットが自己位置を知るために出てきた問題であり，SLAMアルゴリズムはロボットが移動しながら環境地図を作ると同時にロボット自体の位置を推定する．この問題の本質は，次の相互依存関係による制約を解決することである．

- ロボットの位置を知るためには地図が必要
- 地図を作るためにはロボットの位置情報が必要

ロボットの自己位置を推定するためには，この双方の制約を同時に満たす形で処理をする必要がある．SLAMアルゴリズムではSLAM問題に対し行動モデル (車輪のスリップの確立モデル等) と計測モデル (距離センサの計測の誤差モデル) を利用することで自己位置推定と地図作成の2つの相互依存関係を解決している．

そこで本稿ではSLAMアルゴリズムが行動・計測モデルから地図作成と自己位置推定における相互依存関係を同時に満たせる解を取得できる点に着目し，特にロボットが行動しながら動的に相互依存制約を解消していく事に注目する．

3 SIAC

SIACはSLAMアルゴリズムを自然言語対話に応用したモデルである．SLAMが地図と位置の相互依存関係を解決することに対し，SIACでは文脈と単語の参照関係の相互依存関係の解決を行う．

SLAMを自然言語対話に応用する方法として，地図作成と自然言語対話の各変数要素の対応関係をとる．表1にSLAMとSIACにおける各変数の対応を示した．

表1　SLAMとSIACの変数の対応表

SLAM	SIAC
ロボットの位置 x_t	文脈 c_t
ロボットの行動 u_t	ロボットの発話・行動 a_t
センサによる計測 z_t	人間の発話・行動の解釈 z_t
地図 m	ユーザ発話, 行動の解釈ドメイン r

SIACではSLAMにおける位置を文脈に対応させる．これは，ロボットが発話行動を行うことで文脈が進展することと，移動により位置情報が進展していくことが対応していると考えるモデルである．c_t は文脈を表し環境情報・信念情報からなる．信念情報は，対話開始から時間 t までのユーザの対話行動の意図解釈の結果である．

また, ユーザの発話行動の解釈ドメインは解釈対象（物体・概念・知識,etc）のセットである.

このように変数の対応を取ることで SLAM における行動モデルと計測モデルは、それぞれ SIAC における発話・行動モデルと解釈モデルと考えることが出来る.

- 発話・行動モデル: $p(c_t|c_{t-1}, a_t)$
- 解釈モデル： $p(z_t|c_t, r)$

これら 2 つのモデルにより SLAM の式を置き換えることで (1) 式となる. このようにすることで SIAC は単語ー発話解釈ドメイン間の参照関係と文脈の同時推定を実現する.

$$P(c_t, r|z_{1:t}, a_{1:t}) = \alpha P(z_t|c_t, r) \int P(c_t|c_{t-1}, a_t) P(c_{t-1}, r|z_{1:t-1}, a_{1:t-1}) dc_{t-1} \quad (1)$$

4 対話システム構築

SIAC を元に対話システムの構築を行う.

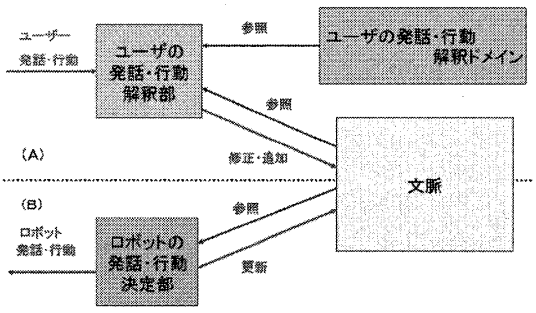

図 1 対話システムアーキテクチャ

図 1 に本システムのアーキテクチャを示す. 図 1 においてユーザの発話・行動解釈部は解釈モデル, ロボットの発話・行動決定部は発話・行動モデルにあたる. 文脈は環境情報・信念情報からなり, 環境情報には対話において対象となる物体や属性を持つ. 一方で信念情報は, ユーザ発話の解釈結果や解釈結果に対する尤度, ロボットとユーザの間で「何に注目して対話を行っているか」ということに対する尤度などを持つ.

本対話システムの全体の流れを A・B の 2 つに分けて述べる. (A) ユーザからの発話・ジェスチャによる入力を得ると, ユーザの発話・行動解釈部はドメイン・文脈を参照して入力に対し意味解釈を行う. その後, 解釈の結果に従い文脈の追加・修正を行っていく. その際, 単語の参照関係が上手く取れなかった場合は文脈の解釈結果の尤度を低く修正する. (B) ロボットの発話・行動部では文脈を元にロボット発話・ジェスチャを生成する. 文脈における解釈結果の尤度が低い場合は, 尤度を高めるため参照関係の正否を確認する発話・ジェスチャを生成する. 最後に, 生成の内容に基づき文脈を更新し次のユーザ発話・ジェスチャを待つ.

5 検証

検証では SIAC において単語の参照関係を解決することで, 図 2 の状況におけるロボット (R) と人 (H) の間の以下のような対話を扱えることを確認する.

H-1 あの本取ってください. (指差しをしながら)
R-1 この、赤いのですか？
H-2 いいえ、白いのです.
R-2 わかりました.

通常この対話例では,[H-1] の人の発話・ジェスチャの意味解釈時において「あの本」の参照関係を特定することが出来ない. 本システムでは [H-1] の後も文脈を進めていくことで,「あの本」=「白い本」の参照関係を特定を行いユーザ発話の意図解釈を可能にしている.

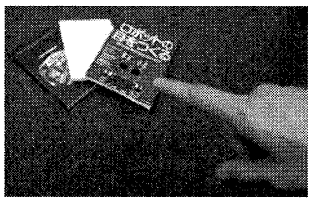

図 2 対話例

図 3 対話検証の様子

検証は図 3 に示したロボットと人との対面環境で行う. 人は 8 つの物体のうち 1 つを選び, その物体が欲しいことを発話・ジェスチャによりロボットへ伝える. 単語の参照関係の解決能力を検証するため, 人は,1 つの発話だけでは参照関係の特定できない単語を含むように対話を行う.

なお, 実装環境として音声認識には julius3.2 [2] を使用し, あらかじめ解釈時に必要な文法・語彙においては設計者による入力を行った. 人のジェスチャ認識はロボットのカメラによる簡単な画像認識により指差しジェスチャの方向を取った.

検証の結果, 人が 8 つの物体のうちどの物体を選び対話を行った場合においても, 単語の参照関係を文脈より特定することでロボットは選ばれた物体を認識することが出来た.

6 まとめ

本稿では文脈と単語の参照関係における相互依存関係を解決するモデル SIAC を提案し, 実際のロボット対話システム構築することでその能力を検証した. 今後はより複雑な環境や対話において SIAC を適応することを目指す.

参考文献

[1] Pierre Lison and Greet-Jan Kruijff: "Salience-driven Contextual Priming of Speech Recognition for Human-Robot Interaction", European Conference on Artificial Intelligence(ECAI) 2008, pp.636-640

[2] 李晃伸, 鹿野清宏: "複数文法の同時認識および動的切り替えを行う認識エンジン julius/julian-3.3", 日本音響学会研究発表講演論文集,Vol.2002, 3-9-12

[3] 東中 竜一郎, 中野 幹生, 相川 清明: "複数文脈を用いる音声対話システムにおける統計モデルに基づく談話理解法", 情報処理学会研究報告,SLP-45-17,pp101-106,2003