2C4-2

SIAC:SLAMアルゴリズムを応用した実世界対話システムの提案

SIAC:Proposed of new algorithm appliying SLAM algorithm to real-world dialogue system

松元　崇裕[*1]　大澤　博隆[*2]　大村　廉[*3]　今井　倫太[*4]
Takahiro Matsumoto　Osawa Hirotaka　Ren Ohmura　Imai Michita

[*1]慶應義塾大学　理工学研究科
Graduate School of Science and Technology, Keio University

[*2]国立情報学研究所　コンテンツ科学研究系
National Institute of Informatics Digital Content and Media Sciences Research Division

[*3]豊橋技術大学　情報・知能工学系
Toyohashi University of Technology Knowledge-based Information Engineering

[*4]慶應義塾大学　理工学部
Faculty of Science and Technology, Keio University

This paper proposes a new algorithm named SIAC for natural language dialogues between a human and an agent system, using contextual information. One of the most difficult problems in constructing conversation robots is the mutually related constraints between the meaning of the word and its context; the meaning of the word is determined depending on the context and the context consists of the set of the interpreted words. SIAC determines the both information simultaneously. We employ the method similar to SLAM algorithm to develop SIAC. We also show a conversation robot based on SIAC to demonstrate its ability to solve problems with mutually related constraints between the meaning of the word and its contexts.

1. はじめに

本研究は実世界において人とコミュニケーションをとりながら人間の活動を支援するロボットの実現を目指し、ロボットと人とのジェスチャを用いた音声対話について扱う。ヒューマンロボットインタラクション（HRI）についての研究は現在まで数多く行われてきた [B.Jensen 02][T.Kanda 04][T.Watababe 04]。今までの HRI 研究の知見から、ヒューマノイドロボットが人間に対しジェスチャや発話を行いながらコミュニケーションをとることの重要性が明らかになりつつある。重要性の理由として、ジェスチャ・発話といった人間同士で行う自然なコミュニケーション方法をロボットが取ることで、ユーザはロボットに対する特別な知識を持たなくても扱うことができることが挙げられる。

しかしながら、ジェスチャ・発話を用いた対話システムを構築する際の問題の１つとして、対話における単語と実世界の対象の「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。例えば、ユーザがロボットに対し「本を運ぶ」というタスクを与えたとする。タスクを解決するためには、ロボットはユーザから指し示されている「本」が実空間上にあるどの物体のことを指しているのかを理解しなければならない。以上の例から単語は実世界における物体・概念・知識との間で参照関係を取ることで初めて意味が特定されることがわかる。

ユーザ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、対話時における環境情報・ジェスチャ情報を考慮する手法が提案されている [P.Lison 08]。提案手法ではユーザ発話・ジェスチャの解釈を境情報を含めて行うことで、発話のみを扱う手法で解決できない単語の参照関係の特定を可能にしている。

連絡先: 連絡先：慶應義塾大学理工学部情報工学科安西・今井研究室〒 223-0061 神奈川県横浜市港北区日吉 3-14-1 E-mail: matumoto@ayu.ics.keio.ac.jp

しかしながら、上記の手法では環境情報・ジェスチャ情報を用いても単語の対象が特定できない場合、単語の参照関係を解決できない。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は、単語と先行詞の関係を注目した照応処理を行う研究 [C.Taylor 06] が多く行われてきた。しかしながら、単語と実世界の対象における参照関係の特定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。したがって, 単語の参照関係を求めるためには以下の相互依存制約を解決する必要がある。

- 単語の参照関係を特定するためには文脈が必要である
- 文脈を知るためには意味解釈されたユーザ発話・行動が必要である

そこで本研究では、上記の相互依存関係を解決するモデル SIAC (Simultaneous Interpretation And Contextualizing) を提案する。相互依存関係の解決法として、ロボットの自律移動の分野において提案されている SLAM(Simultaneous Localization and Mapping) アルゴリズムを制約充足アルゴリズムとして捉え、自然言語システムへの応用することで行う。また、環境情報・ジェスチャ情報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を含む対話を扱う能力を示す。

2. SLAM アルゴリズム

SLAM 問題は自律移動ロボットが地図作成を行う際における自己位置の特定をする際にできた課題であり、現在高い完成度で自己位置を推定できるアルゴリズムが提案されている。SLAM アルゴリズムの本質は以下の相互依存関係における制約を解決することである。

- ロボットの自己位置を知るためには地図が必要
- 地図を作成するためにはロボットの位置情報が必要

ロボットが自己位置を推定するためには、これら双方の制約を同時に満たす形で処理を行う必要がある。SLAM 問題を解決するアルゴリズムでは、ロボットの行動モデル (例えば、ロボットの車輪がどのくらいの確率でスリップするか) と計測モデル (距離センサの計測はどのくらいの誤差をもつのか) を駆使することで、地図作成と自己位置推定の相互依存関係を解決する。この詳しい手法は後に詳細する。

本稿では SLAM アルゴリズムが 2 つの制約条件を同時に扱うことで相互に満たす解を考える点について着目し、特にロボット自身が移動し動的に制約を解消する点を注目する。

つづいて、簡単に SLAM における前提条件と確率的 SLAM アルゴリズムについて述べる。

2.1 SLAM アルゴリズム概要

SLAM アルゴリズムにおける地図とは周囲環境における目印の位置の関係であると考える。具体的には、ロボットがセンサにより目印となる物体との距離を測定するこで、自身の位置と計測による方向・距離の関係から地図を作成していく。SLAM による地図作成の様子を示したのが図 1 である。

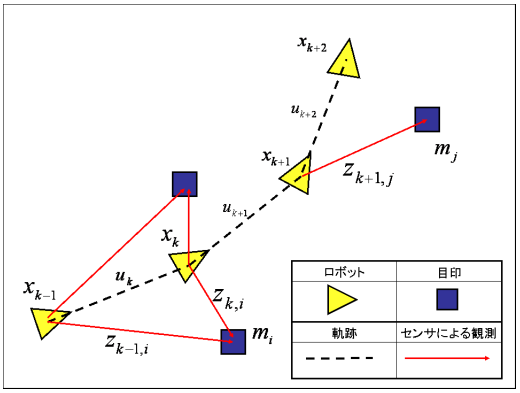

図 1: SLAM 問題における要点

図 1 におけるそれぞれの変数は時間 t において、ロボットの絶対座標と方向のベクトル x_t、時間 $t-1$ から t におけるロボットの移動コントロールベクトル u_t、地図情報となる目印物体の絶対座標ベクトル m、ロボットの目印に対する距離の観測値 z_t である。

コントロールベクトル u_t は車輪の空転やセンサによる誤差で正確に値を求めることは出来ない。また、レーザレンジファインダ等により求める目標物との距離 z_t も同様に誤差を含んでいる。したがって、ロボットの正確な絶対位置 x_t というのは明らかではない。そこで、SLAM では行動モデル (移動による誤差を含んだ確率モデル) と計測モデル (計測による誤差を含んだ確率モデル) の 2 つを用いて自己位置 x_t と地図 m_i を同時に求めることで、移動距離の測定と対象物との距離測定において生じる誤差を修正している。

行動モデル、計測モデルは次の式 (1)(2) のように定義することが出来る。

$$P(z_t|x_t,m) \tag{1}$$

$$P(x_t|x_{t-1},u_t) \tag{2}$$

上記の (1)、(2) より以下の (3) が導出できる。これにより、前回の位置と地図の推定結果 $P(x_{t-1},m|z_{1:t-1},u_{1:t-1})$ に対し具体的な行動モデルと計測モデルを適用することで同時推定を可能にしている。

$$P(x_t,m|z_{1:t},u_{1:t}) = \alpha P(z_t|x_t,m)\int P(x_t|x_{t-1},u_t) P(x_{t-1},m|z_{1:t-1},u_{1:t-1})dx_{t-1} \tag{3}$$

3. 自然言語対話システムにおける各要素変数

自然言語対話システムの構築時において考慮するべき要素には、以下の物が存在する。

- 文脈:c_t
- 対話相手のレスポンス (発話、行動):u_t
- 人間の発話, 行動の解析結果:z_t
- ロボットの発話や行動:a_t

c_t は、対話の進行によって情報が付加されていく。具体的には対話により語られた実世界の状況、やシステムの発話に対する人間側の解釈（信念情報として扱う）により構成される。また u_t はロボットのセンサにより獲得された人間の行動や発話である。ここで u_t は、ユーザの発話・行動から得られたセンサ値であり、対話で扱うためには現在の文脈に照らし合わせながら意味解釈をする必要がある。そのため、解釈結果である z_t とは分けて取り扱う。一方で a_t の内容は、システム内部において決定される。したがって、システムは行った発話・行動の意味を一意に取り扱うことが出来ため、A_T は文脈に照らし合わせることなしに意味解釈を行う。

4. SIAC

本章では SLAM の手法を人間とロボットシステムの対話に応用する手法 SIAC の提案を行う。SIAC は、SLAM おける各変数要素と自然言語対話における変数要素の対応を取ることで構築する。SIAC では SLAM における位置を文脈に対応させる。これは、ロボットが発話や行動を行うことで対話状況が進展することがロボットの移動により位置情報が進展していくことに対応していると考えるモデルである。また、SLAM のセンサによる計測における地図情報の更新は、人間の発話や行動に対する解釈の結果によるドメインの更新に対応付ける。SLAM がロボットの移動距離の計測と周囲環境のセンシングによって自己位置を進展・修正することに対し、SIAC ではロボットの発話とユーザの発話・行動の解釈の結果、文脈の進展修正を行っていく。この関係を表 1 に SLAM と SIAC の変数の対応表に挙げる。

表 1 に挙げたように SLAM と自然言語対話の対応を取ると,SLAM における行動モデルと計測モデルは、自然言語対話における発話・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: $p(c_t|c_{t-1},a_t)$ (4)

解釈モデル： $p(z_t|c_t,r)$ (5)

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	人間の発話, 行動の解釈ドメイン r

ここで、ユーザ発話・行動の解釈ドメインは $r = r_1, r_2, ..., r_n$ からなる解釈対象（物体・概念・知識、etc）のセットである。これらの式は式 (1)、(2) に対応する。式の (3) を (4)、(5) で置き換えると (6) のように表せる.

$$P(c_t, r|z_{1:t}, a_{1:t}) = \alpha P(z_t|c_t, r) \int P(c_t|c_{t-1}, a_t) P(c_{t-1}, r|z_{1:t-1}, a_{1:t-1}) dc_{t-1} \quad (6)$$

式 (6) は次のような意味を持つ。現時点から 1 つ前での文脈 c_{t-1} と解釈ドメイン r の推測結果 $p(c_{t-1}, r|z_{1:t-1}, a_{1:t-1})$ に対して、発話・行動モデルと解釈モデルを適応することで修正更新を行い現在の推測結果としている。また、発話・行動モデルでは文脈がロボットの発話 a_t によって進展することを示す。解釈モデルでは人間からの発話・行動 u_t に対し、文脈 c_t と解釈ドメイン r によって最も確信度の高い解釈結果 z_t が得られ、解釈結果を元に推定されていた文脈及び解釈ドメインが修正される。

実際の計算においては、ロボットの発話・行動に対する対話相手の解釈結果についてのコーパスと対話相手の発話に対する対話相手の意図についてのコーパスを用いることで、それぞれ発話・行動モデルと解釈モデルにおける文脈と意味解釈の確率算出モデルを構築することができる。

5. 対話システム構築

本章では SIAC アルゴリズムを使用したロボット対話システムにおけるシステムの設計について述べる。

本対話システムはロボットと対話者の対面環境において両者が指差しと発話を通して対話を進める。両者の間には RFID タグを着けた 8 つの物体（本・コップ）を配置し、対話者にはマイクと RFIT タグを、ロボットには RFID タグを装着する。対話者は配置された 8 つの物体のうち 1 つを任意で選び、ロボットに欲しいことを対話により伝える。（以降、物体参照タスクとする）対話者の発話は対話者に着けたマイクにより取得を行い、指差しジェスチャはロボットの持つステレオカメラによる画像として取得する。

図 2 に本システムのアーキテクチャを示す。

図 2 において、ユーザ発話・行動の解釈ドメインは対話の対象となる概念・対象・対象の属性の 3 つからなり、今回の物体参照タスクに応じたドメインの構築を設計者の手により行った。

文脈情報は環境情報と信念情報の 2 つからなり、環境情報はドメイン情報を元に構築されたセマンティック・センサネットワーク [B.Guo 07][M.Imai 06] から取得する。セマンティック・ネットワーク (以下ＳＳN) はセンサデータとメタデータ（環境記述）をセットで管理する。具体的には RFID タグとタグを取り付けた対象の関係をデータベースに管理することで環境の状態を保持することができる。メタデータはドメインを元に 2 つのクラス（対話者クラス・物体クラス）にしたがって記述さ

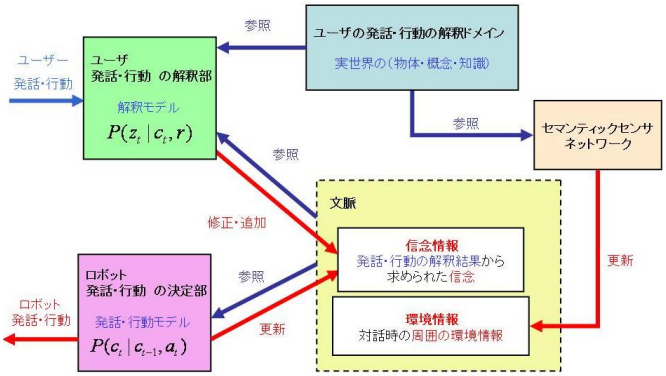

図 2: 対話システムアーキテクチャ

れ、タグを取り付けた対象のクラスごとに属性要素値を対応付けたインスタンスを生成する。各インスタンスは物体を示す変数と結び付けられる。

信念情報はユーザの発話・行動の解釈結果による追加・修正とロボット自身の発話・行動結果による更新により進展する。今回のシステムでは信念情報の要素にロボットと対話者の間での共同注意対象の推定結果と対話者が選んだ対象の推定結果の 2 つの尤度を持つ。本対話の目的はユーザが選んだ対象とロボットがユーザが選んだと推定を行った対象が一致し、かつ推定結果に高い尤度を与えることである。

ユーザの発話・行動解釈部では、まず取得したユーザ発話の音声情報を音声認識エンジン Julius3.2[A.Lee 01] を用いることで文字列に変換する。また、ユーザの指差しジェスチャは取得された画像に対し簡単な画像認識処理を行うことでユーザが正面・右方向・左方向のいずれかを指しているかを識別することができる。獲得した音声・画像の認識結果はユーザのレスポンス u_t にあたる。次に、システムは u_t に対して解釈ドメイン・文脈を元に u_t の意味解釈結果 z_t を求める。解釈時に発話内の単語や指差しジェスチャの指差し対象と解釈ドメインとの参照関係が複数考えられる場合、解釈モデルは文脈に応じて z_t に対する信頼度 $P(z_t|c_t, r)$ を決定する。最後にユーザの発話・行動部は解釈結果 z_t と信頼度 $P(z_t|x_t, r)$ を元に文脈 c_t の修正・追加を行う。

一方、ロボットの発話・決定モデルは返答用の単語・文法から文脈を元に文字列による返答文を生成し、返答文を音声合成することで発話を行う。また、返答文において指し示したい物体が存在する場合、ロボットは物体とロボット自身の位置関係から指差しと視線の方向を決定しジェスチャを行う。発話とジェスチャの結果、システムはロボットが発話・ジェスチャにより指し示した対象に関してユーザと共同注意を持ったという信念を持つことができる。共同注意に対する信念は行動・解釈モデル $P(c_{t+1}|a_t, c_t)$ を元に更新される.

今回のシステムでは、SIAC の動作確認を主として発話・行動モデルと解釈モデルはシステム作成者の手入力データにより構築を行った。

6. 対話例

本節ではシステムが実際の動作した際における対話例を示す。対話は図 3 に示すような対話時の対象となる物体の配置を行い、ロボットと対話者が向かい合った状況で行った。SSN には環境情報としてそれぞれの物体に対しては種類・色・所有者の属性情報、対話者・ロボットには名前の属性情報を与えた。

各物体・ロボット・対話者の位置属性 (x,y,z) は、それぞれに着けられた RFID タグの位置を検出することで行った。図 3 内の Left,Right,Center の囲いはそれぞれロボットの指差し認識の限界であり、対話者の指差しに対し、どの Zone を指差しているかの認識はできるが、どの物体を指差しているかまでは認識することができない。

以下に SIAC を用いた対話システムにおいて、上記の環境の中で人間 (H) とロボット (R) の間で実際に行われた対話例と対話の様子図 4 を示す。

R-1 何を取りましょうか？
H-1 本を取ってください。
R-2 どの本ですか？
H-2 黒いのです。
R-3 これですね？(Book5 を指差しながら)
H-3 はい。
R-4 わかりました。

今回の対話条件において、ロボットは対話者の発話内における「本」「黒いの」という単語に対しては、同属性の物体がいくつか存在するため単語の参照関係を一意に決めることはできない。対話例において、SIAC はその文のみでは参照関係を特定できない単語に対しても、文脈を進めることで参照関係を求めユーザ意図を特定している。また、次の対話例は [H-3] において否定された例である。

H-3' いいえ, こっちです.(左方向への指差し)
R-4' それでは, こちらですね.(Book2 を指差しながら)
H-4' はい.
R-5' わかりました.

2 番目の対話例においても、ロボットは左方向への指差しと「こっち」という単語のみでは、「こっち」に対し左に位置する３物体のどれが参照関係をとるかを特定することはできないが、SIAC により文脈から確率的に推定を行うことでユーザの指定する物体を特定することができている。

図 3: 対話対象物体と配置

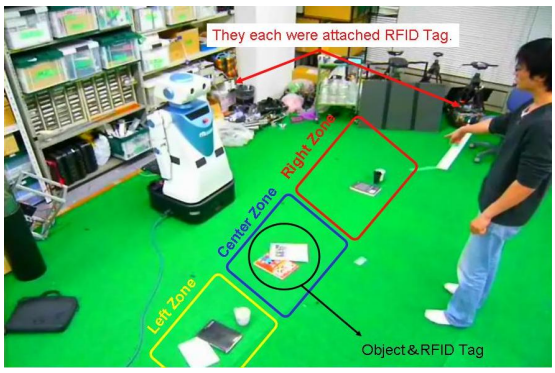

図 4: 対話例

7. 結論と今後の展望

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動作を確認した。今後は発話・行動モデル、解釈モデルを実コーパスから起こすことを行っていく。

参考文献

[S.Thrun 05] S.Thrun, W.Burgard and D.Fox: "Probabilistic Robotics", The MITPress, September 1, 2005

[B.Jensen 02] B.Jensen, G.Froidevaux, A.Lorotte, M.Meisser, G.Ramel, R.Siegwart; "The interactive autonomous mobile system RoboX", IROS 2002, IEEE Press, pp.1221-1227.

[B.Guo 07] B.Guo, M.Imai; "Home-Explorer: Search, Localize and Manage the Physical Artifacts Indoors", AINA pp.378-385, 2007.5.

[M.Imai 06] M.Imai, Y.Hirota, S.Satake, H.Kawashima; "Semantic Sensor network for Physically Grounded Applications" Proceedings in ICARCV, pp.1637-1642, 2006

[C.Taylor 06] C.Taylor, A.Rahimi, J.Bachrach, H.Shrobe,and A.Grue: "Simultaneous Localization, Calibration, and Tracking in an ad Sensor Network", In IPSN '06 pp.27-33.

[H.Whyte 06] H.Whyte and T.Baliey: "Simultaneous Localization and Mapping(SLAM) Part 1 The Essential Algorithms" IEEE Robotics & Automation Magazine 2006 Citeseer.

[A.Haasch 05] A.Haasch, N.Hofemann, J.Eritsch, and G.Sagerer "A Multi-Modal Object Attention System for a Mobile Robot" In Int. Conf. on Intelligent Robots and Systems, pages 1499-1504, August 2005

[P.Lison 08] P.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

[T.Iio 09] T.Iio, M.Shiomi, K.Shinozawa, T.Miyashita, T.Akimoto and N.Hagita: "Lexical Entrainment in Human-Robot Interaction: Can Robots Entrain Human Vocabulary?", IEEE IROS October 11-15, 2009 St.Lousis, USA. pp.3727-3734

[T.Kanda 04] T.Kanda, H.Ishiguro, M.Imai and T.Ono: "Development and Evaluation of Interactive Humanoid Robots", Proceedings of the IEEE vol92 No11 pp.1839-1850, 2004

[T.Watababe 04] T.Watanabe; "Embodied Communication System for Mind Connection", Proc, IEEE RO-MAN'04, Kurashiki, Japan, 2004

[A.Lee 01] A.Lee, et. al; "Julius - an Open Source Real-Time Large vocabulary continuous speech recognition Engine", Proc. EUROSPEECH ,pp.1691-1694 (2001)

[T.Isobe 03] T.Isobe, S.Hayakawa, H.Murao, T.Mizutani, K.Takeda, F.Itakura, "A Study on Domain Recognition of Spoken Dialogue Systems", Proc. EUROSPEECH, pp.1889-1892(2003)