

3次元シミュレータを用いた 対話ロボットによる参照表現生成の評価システム

A Evaluation System of Referring Expression Generation by Dialogue Robots Using 3D Simulation

小澤 優太^{*1}
Ozawa Yuta

船越 孝太郎^{*2}
Funakoshi Kotaro

中村 友昭^{*3}
Nakamura Tomoaki

中野 幹生^{*1}
Nakano Mikio

^{*1}東京電機大学大学院
Graduate of Tokyo Denki University

^{*2}(株) ホンダ・リサーチ・インスティテュート・ジャパン
Honda Research Institute Japan Co, Ltd.

^{*3}電気通信大学
The University of Electro-Communications

Robots that communicate with humans in a real environment are required to generate referring expressions. They must generate expressions easy to understand for humans according to the surrounding environment. In this study, we propose a virtual-reality system that can evaluate referring expression generation methods through human-robot interaction on a 3D simulator.

1. はじめに

人間と実環境でコミュニケーションし、部屋の片付けや救助活動といった共同でタスクを行うロボットの構築のためには、周囲の環境に対応した音声対話機能の実現が重要である。実環境でインタラクションを行う中で、話し手が聞き手に対して言語表現を用いて特定の事物を指し示す状況は頻繁に現れる。そのような言語表現は参照表現と呼ばれ様々な研究がなされてきた [Dale 1989]。人とインタラクションを行いながら協力してタスクを行うロボットは、参照表現の適切な理解に加えて、環境を考慮した人間にとってわかりやすい参照表現の生成を行うことが必要である。特に実環境の参照表現生成ではマルチモーダル情報を扱うことが要求されることから、位置関係や視線を考慮し指差し行動などと組み合わせることで、人間にとってより自然な参照表現を扱うことが可能な能力が望まれる。しかしながら、身体情報を含むマルチモーダルな環境での人間の命令を理解するロボットの研究は盛んであるのに対して、ロボットによる参照表現生成の研究はあまり行われていない。その理由の一つとして、生成された参照表現自体の評価が容易ではないことがあげられる。参照表現を人間が評価する必要があるが、生成された文章から直接評価を行うことは、生成されたものを1つずつ確認しなければならない点でのコストや、動的な環境を考慮した評価方法を定めにくいことが参照表現生成の研究を困難にしている。

そこで本研究では実環境の対話に向けた、環境が動的に変化する参照表現生成を評価するためのシステムの実現を目指す。そして、配置された物体を人間に取ってもらうタスクを用いた、3次元シミュレータに基づく参照表現生成の評価システムを提案する。

2. 関連研究

参照表現生成を含む自然言語生成の評価を行う研究として、複数の物体の写真うちの一つを指す参照表現を生成する TUNA Project [Deemter 2007] がある。TUNA Project では、実際に人間が作成した表現を集めたコーパスを用いることで、人間が生成した表現にどのくらい近いかで評価を行っている。対象と

している環境は静止画であり、3次元環境での参照表現は扱っていない。



図 1: TUNA Project (Gatt et al., ENLG'07)

また、特定の目的を達成するための動作を言語生成によって人間に指示する GIVE Challenge [Striegnitz 2011] がある。GIVE Challenge ではバーチャル空間で生成された参照表現によってユーザがタスクを行い、その達成度で参照表現の評価を行う。ユーザがバーチャル空間で行動するという点ではより



図 2: GIVE Challenge (Striegnitz et al., ENLG'11)

複雑な環境であるが、ロボットの实体は想定されていなく、ロボットと人の位置がダイナミックに変わるような実環境に近い複雑な状況は想定されていない。

3. 提案システム

本研究では、ダイナミックに変化する環境での参照表現生成の評価を行うことを目的としたシミュレーション環境を提案する。ベースのシミュレーション環境として SIGVerse 上で再現された RoboCup@Home の世界環境を用いる。

連絡先: 船越 孝太郎, funakoshi@jp.honda-ri.com

3.1 RoboCup@Home での環境

RoboCup@Home は日常的な実世界空間でロボットと人間が自然にコミュニケーションを行いながら特定の目的を達成する行動を評価する競技である [RoboCup@Home] . RoboCup@Home では、指定した場所への物体の移動をはじめ、指示された日用品の未知環境での探索、ユーザの特定及び追従などのいくつかのタスクで構成されており、各チームのロボットが制限時間内での達成度を競う。SIGVerse は RoboCup@Home のような環境での、人間と知能ロボットのインタラクションを検証するプラットフォームとしてのシミュレータである [SIGVerse] .



図 4: シミュレーション環境の外観 (SIGVerse)

3.2 SIGVerse によるシミュレーション環境

SIGVerse は 3 次元空間上で物理的な力学計算 (運動)、知覚 (センサー)、対話 (コミュニケーション) のシミュレーションを同時に行うことが可能な統合シミュレータであり、ソフトウェアによる知能ロボットの動作が可能なことや多人数の参加型の実験が可能であることから様々な研究に用いられている [谷口 2013] . SIGVerse は展開の 1 つとして RoboCup@Home でのシミュレーションリーグを立ち上げており、仮想空間でのインタラクションによるタスクの検証によって、ハードウェアの開発にかかる負担の大幅な軽減や、参加者の拡大、認識部分や学習部分などのより局所的な研究へのリソースの集中を目指している [稲邑 2014] . 提案シミュレーション環境で用いた SIGVerse 上の世界環境では、全体の空間が、ダイニング、寝室、2 つのリビングからなる 4 つの部屋で構成されており、人間とロボットがそれぞれ配置される .

3.3 システムの構成

提案システムでは、SIGVerse による世界環境と、Module として実装したそれぞれの機能によってシステムが構成されている . システムの構成図を図 3 に示す . 提案シミュレーション

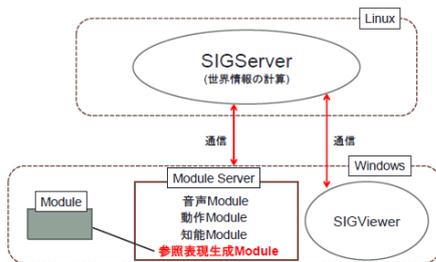


図 3: システム全体の構成

環境の特徴として、TCP/IP を用い、ロボットの知能や動作、音声処理、言語生成処理などの独立したモジュールが通信しながら動作している点あげられる . これにより独立モジュールとして参照表現生成を実装することが可能になる . したがって、参照表現生成を様々なモデルに置き換えて検証することが可能になる . システム上では、知能モジュールが表現生成モジュールや動作モジュールなどを扱うことで、物探しタスクの進行を実現している . 音声モジュールは音声合成及び語彙認識、動作 Module は SIGverse のエージェントの移動等を制御する .

3.4 物探しタスク

提案システムでは生成された参照表現の評価タスクとして、物探しタスクを扱う . 探す対象となる物体として、飲み物やぬいぐるみなどのいくつかの種類のものがランダムな位置に配置されている . 図 4 に例を示す . 物探しタスクはロボットによ

て指示された物体を人間に探してもらって非常にシンプルなタスクとなっている . 物探しタスク全体の流れを図 5 に示す . は

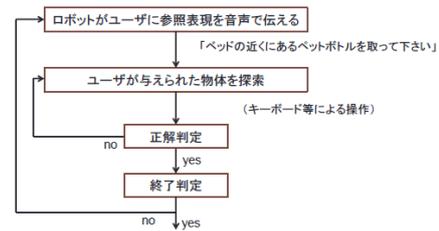


図 5: 物探しタスクの流れ

じめにロボットが人間 (ユーザ) に対して探索する対象となる物体について参照表現を生成し、音声合成によってユーザに伝える . そしてユーザは、キーボードによる操作で探索を行い、ロボットによって参照された物体だと思いのものをクリックで判定する . これを規定回数繰り返すことによって、操作回数、判定回数、経過時間等の情報を取得する .

4. 参照表現生成の実装

参照表現生成モデルとして参照表現ページネットワーク (REBN) に基づくモデルを実装した . REBN は記述表現、照応表現、直示表現を統一的に扱う参照表現の理解のモデルであり、対話文脈や状況情報を統合し、多様な参照表現を扱うことが可能である [Funakoshi 2012] . REBN は、与えられた対象世界の情報と参照表現に対して参照解決を行い、参照表現が対象世界の中の事物を参照していると思われる確からしさを確率 (参照確率) の形で出力する .

このモデルを用いて generate-and-test 方式で参照表現生成を行うことができる . すなわち、候補となる参照表現を複数生成し、それを REBN を用いて評価することで、最適な参照表現を選択する . ただし後述のように、REBN が出力する参照確率が最大となる表現を選択するだけでは、参照表現生成には不十分であり、付加的な評価モデルが必要となる .

本稿では実際に人間が使用する参照表現を参考に生成可能な表現のパターンを定め、それによって候補表現の生成を行う . 人間による参照表現の収集のため、実装に先立って同様のシミュレーション環境を用いた Web アンケート調査によって参照表現を収集した . 東京電機大学の学生 15 名に回答してもらい、一人につき 20 問として合計 300 個の参照表現のデータを収集した . アンケート調査によって得られた参照表現の例を以下に示す .

表 1: 実装した参照表現のパターン

(L) (R) (T)
(L) (R) (A) (T)

- ギター持ってきて
- 後ろのテーブルの上のぬいぐるみを取って
- スピーカーの近くにあるうさぎのぬいぐるみを取って来て
てください
- そのテーブルの上にある黄色いぬいぐるみを取ってほ
しい
- チョコ頂戴
- そのテーブルの上の手前側のペットボトルを取って
ください
- あなたの目の前に見えてる右側のペットボトルとって
- 目の前のテーブルにあるカンを取って

収集された参照表現の中には、係り受けが複雑なもの、聞き手に対して動作を求めるもの（後ろを向いたら見えるテーブルの上）など、実装上複雑となる表現も数個見られたが、概ね表 1 に示すパターンに収まることが確認された。ここで、参照対象物を (T)、対象物が持っている情報としての属性表現を (A)、対象物を示す際の目印となる情報としてのランドマークを (L)、対象物 (T) とランドマーク (L) との関係を表す関係表現を (R) で表している。例えば、「ソファの上にある右の本」という表現では、「ソファ」が (L)、「の上にある」が (R)、「右の」が (A)、「本」が (T) となる。

環境によっては (L) に該当する物体や表現可能な関係のパリエーションが多数あり、考えられる候補全てを生成しようとした場合に膨大な組み合わせが発生する。そのため、何かしらの方法で候補表現生成の計算量を抑える工夫が必要となるが、本稿ではランドマーク (L) の数を決定することで上限を定め、対象物 (T) からユークリッド距離で近い順に選択している。

対象物 (T) が与えられた時の参照表現生成の流れを図 6 に示す。参照表現生成では各パターンを用いて生成された候補表現全てを REBN によって参照解決し、後述する参照スコアが高い表現を採用することで最終的な参照表現を決定する。

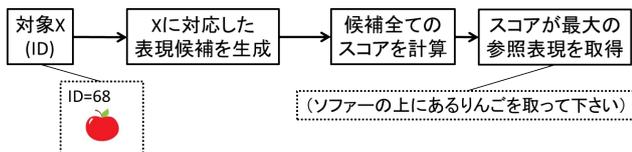


図 6: 参照表現生成の流れ

REBN によって参照解決を行うにあたっては、参照対象部の候補が参照表現に現れる概念に適合する度合いを表現する適合度関数を実装する必要がある。また、参照ドメインとよばれるグループの概念を用いて相対的な表現を自然に生成することができる。以下で、実装した適合度関数、参照ドメインについて説明する。

前述のように、REBN の出力する参照確率だけでは適切な参照表現を生成することができないため、最終的なスコア計算では適合度関数を再度適用する。この点について本節の最後で説明する。

表 2: 実装した概念と言語表現

分類	言語表現	概念記号
属性概念	右の	RIGHT
属性概念	左の	LEFT
属性概念	奥の	FRONT
属性概念	手前の	BACK
関係概念	の右にある	TO_RIGHT_OF
関係概念	の左にある	TO_LEFT_OF
関係概念	の前にある	IN_FRONT_OF
関係概念	の後ろにある	BEHIND
関係概念	の近くにある	NEAR
関係概念	の上にある	ON
関係概念	の下にある	UNDER

4.1 適合度関数

参照表現に現れた概念を適切に解釈するためには適合度を実装する必要がある。例えば“TABLE”という概念についての適合度関数は、「テーブル」と表現された物体 x がその概念にどれくらい当てはまるのかということを表す。適合度は $[0, 1]$ の実数値で表現され、0 は「まったく適合しない」、0.5 は「どちらともいえない」、1 は「強く適合する」を意味する。

本稿では、位置関係に関する属性概念と関係概念に絞って実装した（“TABLE”などの対象概念についての適合度は該当するかどうかの 0/1 で表現した。）実装した属性概念と関係概念を表 2 に示す。参照枠は、すべて話し手からの視点で設定し、対象物から話し手方向が「前」、その反対が「奥」、話し手からみて対象物の右側が「右」、その反対が「左」となっている。距離のみに依存する関係概念として「の近くの」という表現では、(L) と (T) によって得られる距離を用いることで適合度を実装した。距離の適合度 Fit_{dist} に関しては次のような減衰関数

$$\text{Fit}_{\text{dist}}(x) = e^{-\lambda x} \quad (1)$$

として実装し、本稿のシミュレーション空間では $\lambda = 0.01$ とした。

向き持つ関係概念 TO_RIGHT_OF, TO_LEFT_OF, IN_FRONT_OF, BEHIND では、(L) と (T) によって得られる距離に加えて方向を用いることで適合度を実装した。方向に関しては円周上に定義された連続確率分布であるフォン・ミーゼス分布によって適合度関数を定義しており、平均角度 μ に対応した確率密度関数は次の式で表され、値を正規化することで適合度として扱う。

$$f(\theta) = \frac{\exp\{\beta \cos(\theta - \mu)\}}{2\pi I_0(\beta)} \quad (2)$$

式 (1) と式 (2) によって得られた距離と方向に対応する値の積によって、向きに関する関係概念の適合度を算出する。

上下関係に関する関係概念としての ON, UNDER では、(L) の情報と (T) の情報から、大きさを考慮した高さを調べることで 0 か 1 を返す 2 値の適合度関数として実装した。

4.2 参照ドメイン

REBN では参照ドメインを考慮した解決が可能のため、参照ドメインを考慮した参照解決を実装した。ここでの参照ドメインは、人が視覚的に認識する物体のグループに相当する。大きさや色、物体の名称といった情報を考慮した教師なし学習などによってクラスタリングすることで参照ドメインを自律的



図 7: 実装した参照ドメイン

に獲得することが望ましいが、本稿では予めシステム側で参照ドメインの定義を行っている。定義した参照ドメインは、ベッドの上の2つのクッション(図7左)、キッチンテーブルの上にある2つの赤いペットボトル(同中央)、テーブルの上にある3色のペットボトル(同右)とした。それぞれに対応する物体の配置を図に示す。

4.3 適合度を考慮した参照表現生成

最終的な参照表現を決定する際には、図6にあるようにREBNで参照解決して得られた参照確率とともにスコアを算出する。REBNの参照解決から得られる参照確率はREBNが物体を同定できる確からしさであるので、対象物(T)に該当するものが1つしかない場合に不自然な表現を含め全ての表現候補に対する参照確率が1になってしまい、候補表現の比較ができなくなってしまうためである。そのため、ある参照対象物に対する、ある参照表現のスコアとして、次式のように参照確率 P に対して各概念の適合度の積をとったスコア S を用いる。

$$S \leftarrow P \cdot \text{Fit}(L) \cdot \text{Fit}(T) \cdot \text{Fit}(R) \cdot \text{Fit}(A) \quad (3)$$

ここで、 $\text{Fit}(L)$ 、 $\text{Fit}(T)$ 、 $\text{Fit}(R)$ 、 $\text{Fit}(A)$ は、 L, T, R, A に該当する概念それぞれについての参照対象物の適合度を表す。これによって不自然な表現候補のスコアは小さくなり、自然な参照表現生成が可能になる。

ただしこのままでは、長い参照表現ほどスコアが小さくなってしまいうために、短いものが選択される傾向が高くなってしまふ。対象物が目の前にあればよいが、そうでなければ聞き手が探索しやすいように、冗長でも長い表現のほうが好ましい場合もある。このような要因の検討・実装は今後の課題である。

5. 提案システムによる評価

提案システムによって物探しタスクを行い、その際に生成された参照表現と対応するシミュレーション環境での状況を表3と図8に示す。リモコンについては、ロボットの正面にある扇風機に対して、ロボット側にリモコンがある状況で、扇風機の前にあるリモコンを取って下さいという表現が生成された。関係概念では扇風機に対して自分側を「前」とするのを自然な表現として実装してあり、正常な生成がされることを確認した。クッション及びペットボトルでは、参照ドメイン内で特定しやすい属性概念が用いられた表現が生成されることを確認した。

6. おわりに

本稿では人間とロボットの位置などの環境情報が動的に変わる複雑な環境で行う参照表現生成の評価のためのシミュレーション環境を提案し、幾つかの状況において、実際に適切な参照表現を生成できることを確認した。今後は様々な状況での、参照表現生成モデルの有効性の検証や比較をこのシミュレーション環境を用いて行っていく。

表 3: 対象とそれに対して生成された表現結果

対象	生成された表現
リモコン (図8右上)	扇風機の前にある リモコンを取って下さい
ペットボトル (図8下)	キッチンテーブルの上にある 左のペットボトルを取って下さい
クッション (図8左上)	ベッドの上にある 奥のクッションを取って下さい



図 8: 参照表現生成時の状況

参考文献

- [Dale 1989] Dale, R. Cooking up referring expressions. Proc. ACL, pp. 68–75, 1989.
- [RoboCup@Home Japan] “RoboCup@Home Japan”
<https://sites.google.com/site/robocuphomejapan>
- [SIGVerse] “SIGVerse”
<http://sigverse.org/sigverse/main>
- [谷口 2013] 谷口彰, 他, “SIGVerse を用いた自己位置と位置概念の同時推定に関する研究”, 2013 年度人工知能学会全国大会予稿集, 2013.
- [稲邑 2014] 稲邑, 他, “大規模長時間のインタラクションを可能にするロボカップ@ホームシミュレーションの構想とその基盤技術”, 知能と情報 Vol.26, No.3, pp.698–709, 2014.
- [Deemter 2006] K. van Deemter. TUNA: Towards a Unified Algorithm for the Generation of Referring Expressions, Technical report, Aberdeen University. <http://www.csd.abdn.ac.uk/research/tuna/pubs/TUNA-final-report.pdf>, 2007.
- [Striegnitz 2011] K. Striegnitz, et al. “Report on the Second Second Challenge on Generating Instructions in Virtual Environments (GIVE-2.5)”, Proc. the Generation Challenges Session at the 13th ENLG, 2011.
- [Funakoshi 2012] K. Funakoshi, et al. “A Unified Probabilistic Approach to Referring Expressions”, Proc. SIGDIAL, pp.237–246, 2012.