2D3-OS-12b-3

ロボットによる実世界情報を用いた付属語の獲得

Acquisition of Attached Words using Real World Information by Robots

植田紗也佳 *1　岩橋直人 *2　國島丈生 *2
Sayaka Ueda　Naoto Iwahashi　Takeo Kunishima

*1岡山県立大学大学院　情報系工学研究科
Okayama Prefectural University Graduate School of Computer Science and Systems Engineering

*2岡山県立大学　情報工学部
Okayama Prefectural University Faculty of Computer Science and System Engineering

In this paper, using the moving image data and text for teaching the operation to learn the meaning of the syntax that including attached words to the robots. Mapping word and deep case by mutual information, and created syntax templates. Interconvertible between the deep cases and text.

1. はじめに

ロボットが人と対話を行うためには、ロボットが言語と実世界の情報との対応関係を理解している必要がある。言語と実世界情報の対応はあらかじめ人手で決めておくのでなく、人が実生活の中で自ら言語を獲得していくように、ロボットが周囲の環境に応じて自ら学習をすることが望ましい。岩橋 [Iwahashi 03] の研究では、人が物体名・動作を付属語を含まない文で発話を行い、発話の内容に沿う動作をロボットに見せることにより言語獲得を可能とした (図 1)。本研究では岩橋の研究をもとに、付属語を含むより自由な文法で発話を行い、ロボットに構文の意味を獲得させることを目的とする。また獲得した情報によって、発話と行動との相互変換が行えることを示す。

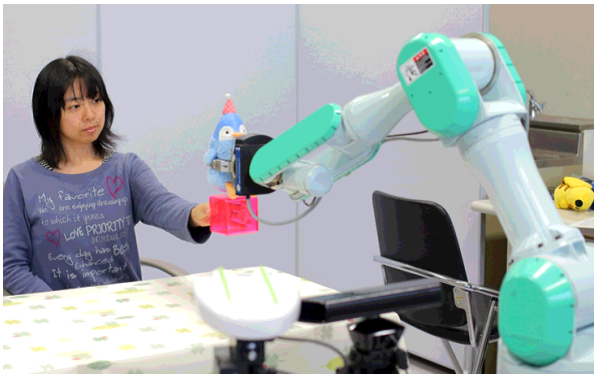

図 1: ロボットとのインタラクションの様子

2. 学習データの取得

学習データには次のデータセットを使用する。

$$D = \{V_i, S_i | i = 1, ..., M\} \quad (1)$$

連絡先: 植田紗也佳，岡山県立大学大学院 人工知能学研究室，〒 714-1197 岡山県総社市窪木 111，TEL 0866-94-2111，E-mail:cd26006b@c.oka-pu.ac.jp

V_iは2.1 節で説明する動画像内の深層格情報のデータ、S_i は 2.2 節で説明する発話のテキストである。

2.1 動画像内の深層格情報の取得

本研究において動画像内の深層格情報とは「NOSERU Obj01_Trj Obj02_Lnd」のように表記される、動画像内の物体の動きを深層格 [Fillmore 75] を明らかにしてテキスト列で書き表したものとする。深層格情報は「NOSERU」「CHIKAZUKU」等の動作のクラスを表す深層格要素、「Obj01」等の物体のクラスを表す深層格要素、フィルモアの格における対象格であるトラジェクタを表す「Trj」もしくは目標格、源泉格、場所格であるランドマークを表す「Lnd」の、物体の格を表す深層格要素で構成される。二つの深層格要素は"_"(アンダーバー) で繋ぐことで「組」となり要素の意味を修飾できる。組「Obj01_Trj」は物体「Obj01」が動作の中でトラジェクタの役割を担っていることを表す。よって「NOSERU Obj01_Trj Obj02_Lnd」は「物体 01 を物体 02 に対して NOSERU という動作をする」ことを表す。組の順序は不同であり、例えば「NOSERU Obj01_Trj Obj02_Lnd」と「Obj02_Lnd Obj01_Trj NOSERU」は同じ動作を表す。深層格情報を用いると、参照点に依存した HMM[Haoka 00][Sugiura 11] によってアーム型ロボットが物体を動かすための軌道を生成することができる。

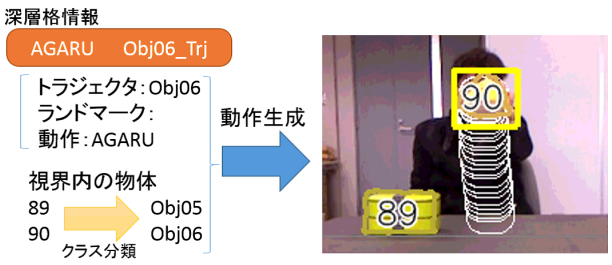

図 2: 深層格情報を用いた軌道生成の例

次に動画像からの深層格情報の取得について述べる。撮影には Kinect を用いる。カメラから一定距離にある物体を認識し、それらの色・形・大きさといった特徴を 12 次元の特徴ベ

クトルとして取得する。また、認識した物体それぞれの座標の変位を記録する。ロボットはあらかじめ物体・動作のクラスを学習により持っているものとし、多次元正規分布によって特徴ベクトルから物体のクラス分類を、参照点に依存したHMM[Haoka 00][Sugiura 11] によって物体の座標の変位から動作の識別を行う。これらの操作により動画像から図 3 のように動画像内の深層格情報を表すデータが取得できる。

図 3: 動画像からの深層格情報の抽出

2.2 発話テキストの取得

本研究では発話の音節列を書き起こしたテキストを、教師なし単語分割 [Ueda 14] により単語ごとに分割したものを使用する。例えば「こっぷをはこにのせて」という発話について音節を正しく書き起こして単語分割を行うと「こっぷ/お/はこ/に/のせて」のようになる。

3. 提案手法

提案手法の流れを図 4 に示す。テキスト内の単語と、深層格情報内の深層格要素のマッチングをとり、それを元にテキストと深層格情報の構造を一般化した構文テンプレートを生成する。単語と深層格要素のマッチングは「共起の相互情報量の計算」を行い、「重みの更新」と「マッチング候補の取得」を繰り返す。

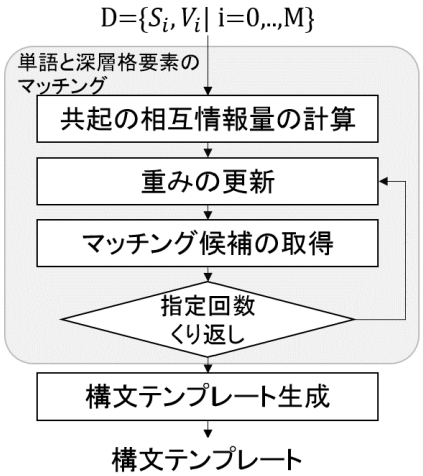

図 4: 構文テンプレート生成の流れ

3.1 共起の相互情報量計算

単語 W と深層格要素 L が同データ内に共起する確率の相互情報量を計算する。

$$X_W = \begin{cases} 1(W \text{ が } S_i\text{に含まれる}) \\ 0(W \text{ が } S_i\text{に含まれない}) \end{cases} \quad (2)$$

$$X_L = \begin{cases} 1(L \text{ が } V_i\text{に含まれる}) \\ 0(L \text{ が } V_i\text{に含まれない}) \end{cases} \quad (3)$$

$$I(X_W, X_L) = \sum_i \sum_j p(X_W = i, X_L = j) \times \log \left[\frac{p(X_W = i, X_L = j)}{p(X_W = i)p(X_L = j)} \right] \quad (4)$$

3.2 重みの更新

相互情報量を計算するだけでは正しく単語と深層格要素を結びつけることはできない。例えば、助詞の「から」と動詞の「はなして」は必ず共起する。よって動作の深層格要素「HANASHITE」との相互情報量はどちらも同じとなってしまう。そこで単語の並びと深層格要素の組の関係によって重みを計算する。各単語 W と深層格要素 L の組み合わせについて、(5) 式で重み $\alpha(W, L)$ を計算する。

$$\alpha(W, L) = \sum_{i=1}^{N_W} \prod_{l \in \psi_{Wi}} Bigram(W, W_i) pair(Class(L), Class(l)) \quad (5)$$

ここで、$Class(L)$ は深層格要素 L が物体のクラス・動作のクラス・物体の格のクラスのいずれであるかを表す。$pair(Class(L), Class(l))$ は二種類の深層格要素が組を形成している確率を示す。N_W は単語の種類数である。

3.3 マッチング候補の取得

相互情報量 $I(X_W, X_L)$ に、単語の並びと深層格の組の関連度による重み $\alpha(W, L)$ を掛けた値 $\alpha(W, L)I(X_W, X_L)$ によってマッチング候補を得る。

単語 W について、$\alpha(W, L)I(X_W, X_L)$ が高くなる最大 5 つの要素 L からなる集合 ψ_W を得る。ただし、ノイズを防ぎ ψ_W を収束させるため、$\alpha(W, L)I(X_W, X_L)$ が $\max\{\alpha(W, L)I(X_W, X_{L_i}|i = 1, ..., N_L)\}$ の 1/5 に満たない要素 L は ψ_W に含めない。N_L は深層格要素の種類数である。

3.4 構文テンプレート生成

得られた単語と深層格要素のペアを用いて構文テンプレートを作成する。構文テンプレートは表層表現と深層構造と変数データから構成される。入力データの発話テキスト・動画像の深層格情報の中で、$\alpha(W, L)I(X_W, X_L)$ の高い単語-深層格要素のペアから順番に変数記号に置換し、そのつどテンプレート候補に追加する。変数記号に置換された箇所に元々あった単語・深層格要素を記録する。図 5、6 では変数記号を”#”(シャープ) から始まる文字列で表している。

入力データとマッチング情報によりテンプレートを作成する例を図 5 に示す。①で、入力データ内で最も $\alpha(W, L)I(X_W, X_L)$ が高い「はこ」と「Obj01」を変数記号「#o02」に置換してテンプレート候補とする。②では①の置換に加え、次に

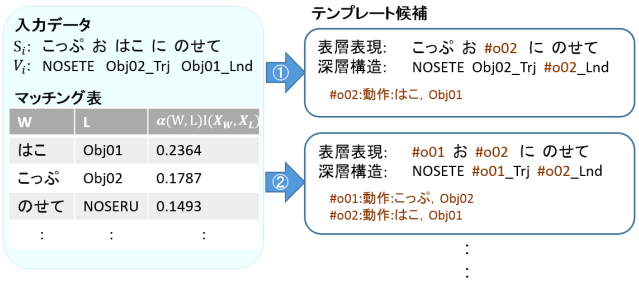

図 5: 構文テンプレートの生成例

$\alpha(W,L)I(X_W,X_L)$ の高い「こっぷ」と「Obj02」のペアを変数記号「#o01」に置換し、テンプレート候補に追加する。この操作を、データ内の置換可能な部分がすべて置換されるまで繰り返す。

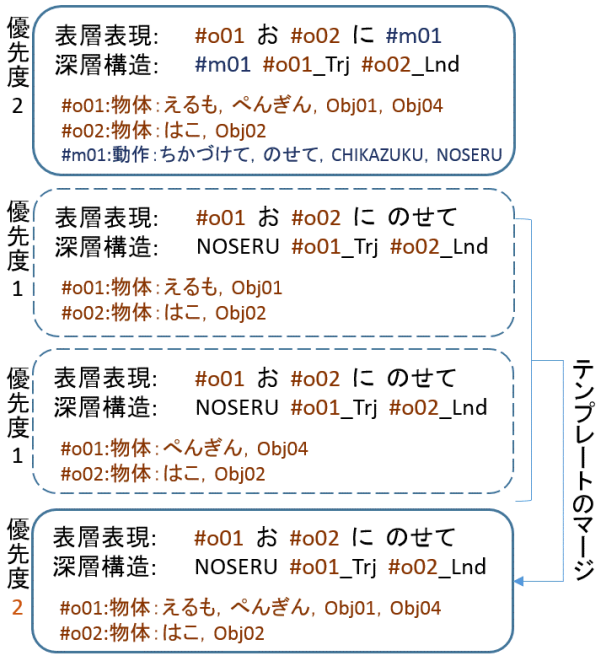

図 6: 構文テンプレートのマージ例

図 6 は構文テンプレートのマージについて示している。表層表現と深層構造が一致しているテンプレート候補をマージする。ただし深層構造は組の順序が入れ替わっていても一致しているとみなす。多くマージされたテンプレート候補ほど優先順位が高くなる。

4. 実験

提案手法の有効性を検証するため、助詞を含む文と動作情報の相互変換実験を行った。構文テンプレートの生成には図 6 に示す物体と、表 1 に示す言い回しからなる命令文と動画像情報のペアを 200 セット用いた。単語と深層格要素をマッチングするときの反復回数は 5 とした。

図 6: 使用した物体一覧

表 1: 動作・言い回し一覧

(物体) お (物体) に のせて
(物体) に (物体) お のせて
(物体) お (物体) に ちかづけて
(物体) に (物体) お ちかづけて
(物体) お (物体) から はなして
(物体) から (物体) お はなして
(物体) お もちあげて
(物体) もちあげて
(物体) お まわして
(物体) まわして
(物体) お (物体) のうえお とびこえさせて
(物体) のうえお (物体) お とびこえさせて

4.1 命令文テキストから動作情報を得る実験

上記の物体・動作・言い回しから命令文を 50 作成して入力し、各命令文に対して適した深層格情報が出力されるかを検証する。入力された命令文テキストと表層表現が一致している構文テンプレートに、各単語とペアである深層格要素を当てはめることで深層格情報を作成する。

4.2 動作画像から命令文テキストを得る実験

上記の物体・動作・言い回しからなる命令を表す動作画像を 50 本撮影し、動画像から深層格情報を取得する。取得した深層格情報を入力し、入力に対して適した命令文が出力されるかを検証する。入力された深層格情報と深層構造が一致している構文テンプレートに、各深層格要素とペアである単語を当てはめることで命令文テキストを作成する。

4.3 重みの有用性

重み $\alpha(W, L)$ を更新せずに命令文と動作情報の相互変換を行い、重みを付加した場合の結果と比較する。

5. 結果

5.1 命令文テキストから動作情報を得る実験

すべての命令文に対して正しい深層格情報を得ることができた。

出力の例を以下に示す。

- 「ととろにえるもおちかづけて」
  →「Obj09_Lnd　Obj08_Trj　CHIKAZUKU」
- 「ぺんぎんのうえおぴかちゅうおとびこえさせて」
  →「Obj04_Lnd　Obj07_Trj　TOBIKOERU」
- 「ちょきんばこおはこにのせて」
  →「Obj10_Trj　Obj01_Lnd　NOSERU」

5.2 動作画像から命令文テキストを得る実験

多くの深層格情報に対して正しい命令文を得ることができたが、今回用いた物体「ぺんぎん」と「ととろ」の特徴量が似ていたために間違ったクラス分類が行われることがあった。動作認識についても、「はなして」「ちかづけて」の動作が誤って認識されることがあった。また、入力に対して適した構文テンプレートを見つけられず出力が生成できない事があった。今回入力した 50 データのうち、正しい出力が得られたものが 45 データ、クラス分類の間違いが 1 データ、動作認識の間違いが 2 データ、出力が生成できなかったものが 2 データであった。

正しい出力の例を以下に示す。

- 「AGARU　Obj03_Trj」
  →「こものいれおもちあげて」
- 「OBIKOERU　Obj01_Trj　Obj06_Lnd」
  →「はこおきんぎょのうえおとびこえさせて」
- 「MAWARU　Obj05_Trj」
  →「べんとおおまわして」

5.3 重みの有用性

5.3.1 命令文テキストから動作情報を得る実験

重みありの場合と同じように深層格情報を得ることができた。これは一意に単語と深層格要素が定まったペアによって「(物体)(付属語)(物体)(付属語)(動作)」という並びの有用な構文テンプレートが生成されたため、正しく深層格を当てはめることができたと思われる。

5.3.2 動作画像から命令文テキストを得る実験

重みなしの場合、「はこおえるもにとびこえさせて」「ぴかちゅうにべんとおおはなして」のように動作「とびこえさせて」と動作「はなして」において助詞と動詞がかみ合わない命令文が生成されてしまった。これは単語と深層格要素が一意に決まらなかったために必要な構文テンプレートが作成できなかったことと、深層格要素「Lnd」に対して複数の単語 (に、から、のうえお) が対応しているためであると考えられる。

6. まとめ

動画像から得られた深層格情報と動画像内の動作を教示するテキストを用いて、構文の意味を保存するテンプレートの作成を行うことでロボットに付属語を含む構文の意味を取得させることに成功した。今後は発話の音声データを用いての言語獲得や、より複雑な構文の意味取得を試みる。さらに、日本語以外でも適用可能であることを検証する。

参考文献

[Fillmore 75] Charles J. Fillmore , TOWARD A MODERN THEORY OF CASE & OTHER ARTICLES , (邦訳 : 田中春美 , 船城道雄 , 格文法の原理　言語の意味と構造 , 三省堂 (1975)) .

[Haoka 00] 羽岡哲郎，岩橋直人，” 言語獲得のための参照点に依存した空間的移動の概念の学習”，信学技報 TECHNICAL REPORT OF IEICE，PRMU2000-105，pp.39-46(2000)

[Iwahashi 03] 岩橋直人，” ロボットによる言語獲得　-言語処理の新しいパラダイムを目指して-”，人工知能学会誌，Vol.18, No.1, pp.49-58(2003)

[Sugiura 11] Komei Sugiura , Naoto Iwahashi , Hideki Kashioka and Satoshi Nakamura , Learning, Generation and Recognition of Motions by Reference-Point-Dependent Probabilistic Models, Advanced Robotics 25 , pp.825-848(2011)

[Ueda 14] 植田紗也佳，” ロボットによる言語獲得のための教師なし単語分割”，SSI2014，SS3-3(2014)