

自動運転の言葉による指示を対象にした空間的意味表現の構造化への取り組み

Spatial Semantic Representation of Natural Language Commands In Automatic Driving

稲子 明里^{*1} 塚原 裕史^{*2} 小林 一郎^{*1}
Akari Inago Hiroshi Tsukahara Ichiro Kobayashi

^{*1}お茶の水女子大学 ^{*2}株式会社デンソーアイティラボラトリ
Ochanomizu University Denso IT Laboratory, Inc.

This paper describes a method to transform verbal instructions to a self-driving car into a spatial semantic description used to be grounded onto the objects or the phenomena in the real world. As for spatial semantic representation, we use spatial description clause (SDC) to represent spatial semantics and develop a parser that can transform natural language instructions to a parse tree represented in the categories of SDC. To build the parser, we firstly define a grammar based on Combinatory Categorical Grammar (CCG) which enables us to represent the dependency relation of spatial semantic categories as a syntactic tree, and then develop a Shift-Reduce parser to transform the instructions into SDC trees using the defined grammar.

1. はじめに

近年、車の自動運転の実用化に向けた動きが活発化している。車を運転できない人でも自動運転車を操作するために、より操作を容易にする手法の開発が期待されている。なかでも、口頭指示による対話的な操作は車への意思疎通を行う手段として、今後必要になることが予想される。そこで本研究では、自然言語で表現された操作指示内容と実世界との対応づけ(グラウンディング)を目的とし、駐車指示内容を車の操作へと結びつける空間意味記述へ変換する手法を提案する。

2. 関連研究

物理環境と言語の結びつけに対する取り組みは SHRDLU にはじまり、限定された環境下での自然言語による命令により行動を生成する研究が行われてきた。空間意味記述を用いたグラウンディングの手法として、Kollar ら [Kollar 10] は Spatial Description Clause(SDC) という意味構造を提案した。SDC の構造は Jackendoff (1983) [Jackendoff 83], Landau and Jackendoff (1993)[Landau 93], Talmy (2005)[Talmy 05] らの研究で提案された意味構造にもとづいている。Tellex ら [Tellex 11] はこの SDC の構造に従った確率的グラフィカルモデル(グラウンディンググラフ)を生成するアルゴリズムを開発し、空間情報を含む自然言語命令を理解するためのシステムを構築した。Kollar らは SDC を系列ラベリング問題として近似し条件付き確率場 (CRF) で扱う手法 [Kollar 10] を提案し、また Tellex らは the Stanford Parser[Marneffe 06] により構文解析した結果を SDC に変換する手法 [Tellex 11] を提案している。一方、我々は、自然言語文の構文情報により、与えられた言語指示に対する SDC のラベルを正解に付与することができると考え、構文解析に対して SDC のタイプにより定義された CCG を文法とする Shift-Reduce パーザを構築する。また、そこで得られた結果を SDC への変換規則を用いることによりグラウンディンググラフを生成することを目的とする。

3. 提案手法

本研究では、言語指示に対する SDC を生成する前処理として、文における空間的意味の依存関係が構文木として表示できるような CCG による文法を定義し、この構文木から SDC への変換を行う。指示内容と観測した物体とのグラウンディングには、Generalized Grounding Graphs[2] を拡張して用いる。図 1 に提案手法の概要を示す。

4. 空間意味構造を用いたグラウンディング

4.1 Spatial Description Clause

Spatial Description Clause(SDC) とは、figure, relation, landmark の 3 要素から構成される空間意味記述のことである。空間表現を含む言語指示は SDC の木として表現できる。また各 SDC は Event, Object, Place, Path の 4 つのタイプに分けられる。Event は実世界で実行される行動のシーケンス、Object は自動車や人間などの物体、Place は場所、Path は経路や方向を表す。先行研究では、空間におけるナビゲーションが目的だったため 4 つのタイプで十分であったが、自動車を操作する多様な言語指示の中に現れる表現に対応するため、View と State という新たな 2 つの要素を追加する。ここで、View は視点、State は物体や場所の状態や条件を表す。図 2 と 3 に View と State を含んだ駐車指示を SDC 木に変換した例をそれぞれ示す。

4.2 Generalized Grounding Graph

Generalized Grounding Graph (G^3) は、ファクターグラフである。ファクターグラフとは、分布の因数分解の構造を表した二部グラフを指す。 G^3 では、言葉とグラウンディングの対象可能な組み合わせから式 1 で表されるように全体の尤度が最大になるものが探索される。

$$\Gamma^* = \arg \max_{\Gamma} p(\Phi = True | commands, \Gamma) \quad (1)$$

Γ はグラウンディング γ_i の集合、 Φ は言語指示の一部がグラウンディングと一致していたら True を返す二値関数 ϕ_i の集合、commands は自然言語による指示文である。式 (1) の確率分布はファクタ Ψ_i と正規化項 Z を用いて以下のように変形

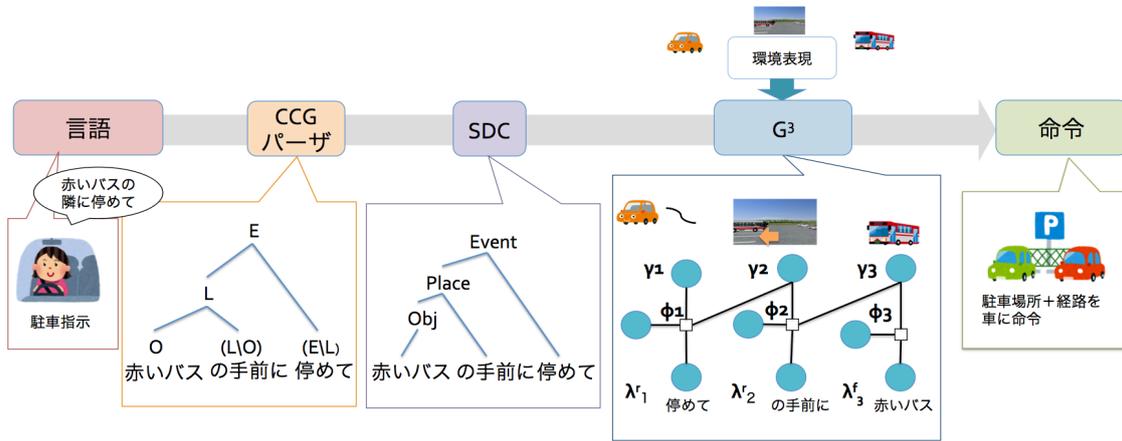


図 1: 提案手法の概要

EVENT₁(r = 停めて
 l = PLACE₂(f = スペース
 r = 右側の
 v = VIEW₃(r = に向かって
 l = OBJ₄(f = コンビニ)))

図 2: View を含んだ SDC 木

EVENT₁(r = 停めて
 l = PLACE₂(l = PLACE₃(r = 側に
 c = STATE₄(r = がいない
 l = OBJ₅(f = 歩行者)))
 r = の
 l = OBJ₆(f = 白いクルマ)))

図 3: State を含んだ SDC 木

される。

$$p(\Phi | commands, \Gamma) = p(\Phi | SDCs, \Gamma) \\ = \frac{1}{Z} \prod_i \Psi_i(\phi_i, SDC_i, \Gamma) \quad (2)$$

SDC_s は、言語指示から変換された SDC である。グラフは、確率変数 (ϕ, γ, λ) とファクタという 2 つのタイプのノードから構成される。各確率変数は 1 つかそれ以上のファクタノード Ψ_i に接続する。図 4 は「赤い車の隣に停めて」という駐車指示の G^3 の例である。

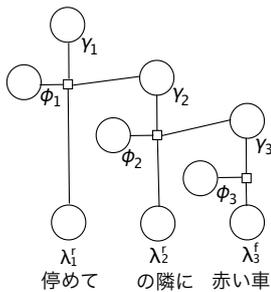


図 4: G^3 の例

5. CCG を用いた SDC 木への変換

5.1 Combinatory Categorical Grammar

Combinatory Categorical Grammar (CCG) [3] とは、カテゴリに対する関数適用により構文解析を行う語彙化文法の 1 つである。CCG のカテゴリは、基底カテゴリと、それらを演算子 “\”, “/” で組み合わせて再帰的に定義される関数型カテゴリから成る。また、CCG は関数適用規則のような組み合わせ規則をもつ。本研究では、SDC のタイプに従い E(Event), O(Object), L(Place), P(Path), V(View), S(State) を基底カテゴリとし、 $X/Y \ Y \Rightarrow X$ と $Y \ X \backslash Y \Rightarrow X$ の 2 つの関数適用規則のみを用いる。図 5, 6, 7 に「赤い車の隣に停めて」という駐車指示を SDC 木に変換した例と CCG による構文木に変換した例、また図 3 の SDC を CCG による構文木に変換した例をそれぞれ示す。

EVENT₁(r = 停めて
 l = PLACE₂(r = の隣に
 l = OBJ₃(f = 赤いバス)))

図 5: SDC 木

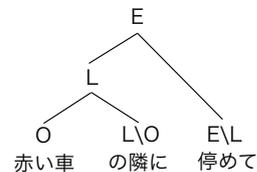


図 6: SDC の意味カテゴリによる構文木

5.2 Shift-Reduce 法に基づくパーザの構築

Shift-Reduce 法とは、スタックを利用して深さ優先探索でボトムアップの構文解析を行う手法である。その手法では、スタックに対して 2 つの操作 (アクション) を行う。1 つは、入力文から単語を一つ取り出してスタックにプッシュする Shift と呼ばれる操作、もう 1 つは、スタックの先頭に文法規則の左辺と一致する単語があるときにそれらを文法規則の右辺のカテゴリに変換する Reduce とよばれる操作である。本研究では、文

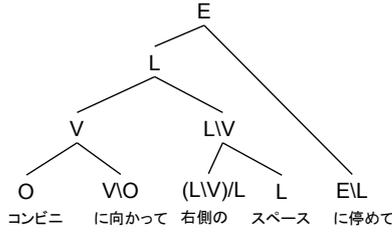


図 7: View を含んだ SDC の意味カテゴリによる構文木

法規則に CCG を用いた Shift-Reduce 法に基づくパーザを構築する。

5.3 未知語処理

辞書にない単語がある場合, CRF を用いてカテゴリの推定を行う。使用する素性は, 前後 2 語の単語, 品詞細分類, カテゴリである。予測されたカテゴリに従って構文解析できる場合は, そのカテゴリを単語とともに文法規則に追加することで未知語の処理を行う。

5.4 パージング処理

Algorithm1 にパーザの手順を示す。まず, 未知語の処理を行い, 元の文法規則と照らし合わせて未知語がある場合は, 新たな規則を文法規則に追加する。その後, 入力文の形態素解析を行いキューに入れ, Shift-Reduce 法により構文解析を行う。デコーディングにはビームサーチを用いる。Algorithm1において, item はスタックとキューの状態構成され, 初期状態には空のスタックと入力文の形態素が入ったキューが入る。各ステップで文法規則に従ってアクションを適用し, Reduce 時のスコア (5.5 に詳述) のみを対象に累加する。スタックにカテゴリのみが残り, キューが空になった場合に構文解析終了とし, 累加したスコアを Reduce の回数で割りスコアとして採用する。最も高いスコアを持つ item を解とする。

Algorithm 1 パーザの処理手順

```
function: DECODE(input, agenda, list, N, grammar,
               candidate_output):
    agenda.clear()
    agenda.insert(GETSTARTITEM(input))
    candidate_output = NONE
    while !agenda.empty() do
        list.clear()
        for action to grammar.getActions(item) do
            new_item = item.apply(action)
            if new_item.F == true then
                new_item.score = new_item.score / reduce_num
                if candidate_output == NONE or new_item.score
                > candidate_output.score then
                    candidate_output = new_item
                else
                    list.append(new_item)
                end if
            end if
        end for
        agenda.clear()
        agenda.insert(list.best(N))
    end while
```

5.5 スコアモデル

スコアはビームサーチによって与えられる Shift を除いたアクションの系列の対数尤度の和をとることで計算される。one versus rest 法を用いた Shift と 11 のカテゴリに対する Reduce のアクションを識別するモデルにはロジスティック回帰を適用した。素性はスタックの先頭 4 個とキューの先頭 3 個の単語,

表 1: 素性の型

型	素性内容
1	$S_0pw, S_0w, S_1pw, S_1c, S_1cw, S_2cw, S_3cw$
2	Q_0pw, Q_1pw, Q_2pw
3	$S_0wS_1w, S_0wS_1c, S_0wQ_0pw, S_0wQ_0p, S_1cwQ_0pw, S_1cQ_0pw, S_1cwQ_0p, S_1cQ_0p$
4	$S_0wS_1cQ_0p, S_0wS_1cwQ_0p, S_0wS_1cQ_0pw, S_0pS_1pQ_0p, S_0wQ_0pQ_1p, S_0pQ_0pQ_1p, S_0wQ_0pwQ_1p, S_0pQ_0pQ_1pw, S_0wS_1cS_2c, S_0wS_1cwS_2c, S_0wS_1cS_2cw, S_0pS_1pS_2p$

品詞とスタックの先頭より後の 3 個のカテゴリを組み合わせからなる。表 1 に識別器構築に使用される素性の型を示す。 S_0, S_1, S_2, S_3 はスタックの先頭 4 個のノードを表し, Q_0, Q_1, Q_2 はキューの先頭 3 個のノードを表す。p, w, c はそれぞれ品詞, 単語, カテゴリを表す。

6. 実験

駐車指示データを用いて実装したパーザの評価実験を行った。

6.1 実験設定

対象データには, クラウドソーシングによって集められたコンビニエンスストアの駐車場で駐車シーンにおける 6055 文 (内 S, V を含まないデータは 4078 文) の日本語の駐車指示を用いた。これらの文はランダムマークを含まない指示, ランダムマークを 1 つ含む指示, ランダムマークを 2 つ含む指示に分けられる。駐車可能なスペースの数は 3 つである。S や V を含まない文の例としては, 「黒い車から一番遠いところに停めて」「白い車の隣でコンビニ前のスペースに停めて」などがあり, S や V を含む文の例としては, 「一番右の空いているスペースに停めて」「コンビニから見て目の前に停めて」などが挙げられる。SDC で表現した際に Path のタイプを含む文は事例数が少ないため, 今回は除いた。未知語の処理には条件付き確率場 (CRF) を用いた。10 分割交差検定を行い, S や V を含むデータと含まないデータにおける未知語の処理を行ったパーザと行っていないパーザの精度を比較する。

6.1.1 実験結果と考察

精度の結果を表 2 に示す。未知語の処理を行わない場合に

表 2: パーザの精度

S や V	未知語処理	全体の正解率	構文解析できた確率	構文解析できた文の正解率
あり	あり	0.3523	0.4950	0.7159
あり	なし	0.3853	0.4808	0.8017
なし	あり	0.5343	0.6606	0.8097
なし	なし	0.5346	0.6415	0.8338

比べ, 行った方が構文解析できた割合が上がった。構文解析できた割合が上がったのは, 未知語の文法規則を追加したことにより, 元の文法規則だけでは構文解析不可能だった文が解析できるようになったことによる。構文木生成に成功した事例がまだ少ないことについては, CCG の規則として 2 つの関数規則しか用いていないため, 規則適用の制約が強すぎることを考えられる。また S や V を含まないデータの方が構文解析できる確率が高い原因としては, S や V を含む場合のカテゴリの数が 43 なのに対して, 含まない場合は 11 であることから多様性が少ないことが挙げられる。

7. まとめ

本研究では、操作指示文に対する SDC を得る前処理として、文における空間的意味の依存関係が構文木として表示可能な CCG による文法を定義し、この構文木から SDC への変換を行う手法を提案した。また、SDC のタイプに Vision や State を加えることで従来研究にはない操作指示も扱えるようにする拡張についても提案し、Shift-Reduce 法を用いたパーザの実装を行った。今後は、パーザの精度の向上や CCGBank の拡充、グラウンディンググラフの構築を目指す。また未知語の処理で予測するカテゴリを順位付けし、上位 n 位までのカテゴリを文法規則に追加していきたい。

参考文献

- [Kollar 10] T.Kollar, S.Tellex, D.Roy, N.Roy: Toward Understanding Natural Language Directions, ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI), pp.259-266, 2010.
- [Steedman 00] T.Kollar, M.Steedman: The Syntactic Process, MIT Press, 2000.
- [Marneffe 06] M.Marneffe, B.MacCartney, C.D.Manning: Generating Typed Dependency Parses from Phrase Structure Parses, Proceedings of the International conference on Language Resources and Evaluation (LREC), 2006.
- [Jackendoff 83] R.Jackendoff: Semantics and Cognition, MIT Press, 1983.
- [Talmy 05] L.Talmy: The fundamental system of spatial schemas in language, Hamp, B., ed., From Perception to Meaning: Image Schemas in Cognitive Linguistics, 2005.
- [Landau 93] B.Landau and R.Jackendoff: "What" and "where" in spatial language and spatial cognition, Behavioral and Brain Sciences 16, pp.217-265, 1993.
- [Tellex 11] S.Tellex, T.Kollar, S.Dickerson, M.R.Walter, A.G.Banerjee, S.Teller, N.Roy: Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation, AAAI, 2011.
- [Zhang 11] Yue Zhang, Stephen Clark: Shift-Reduce CCG Parsing, In Proc. ACL 2011, pp.683-692, 2011.