

自動運転車の対話的操作を実現するための 自然言語の空間意味表現に基づくグラウンディング

大田原 菜々[†] 塚原 裕史[‡] 櫻 惇志[‡] 小林 一郎[†]

[†]お茶の水女子大学 [‡]株式会社デンソーアイティラボラトリ

[†]{g1420511, koba}@is.ocha.ac.jp, [‡]{htsukahara, akeyaki}@d-itlab.co.jp

1 はじめに

近年、自動運転の実用化に向けた活動が急速に進展しており、既に一部の機能に特化はしているものの、市場に投入された車種も存在する。しかし、現状の自動運転車には、例外的な交通状況に対する操作性にはまだ解決しなくてはならない多くの課題が存在する。その解決方法の一つとして、誰にとっても身近なインタラクション手段である発話によって対話的に運転指示を行えるようにすることで、システム側で認識が難しい例外的な交通状況が発生した場合に対処する技術が望まれる。この技術を実現するためには、対話の中で現れる場所や物体などの言葉と、実世界上の場所や物体を結び付ける技術が必要であると考えられる。このことを踏まえ、本研究では、自動運転車に対話的に操作するための、自然言語と実世界空間のグラウンディング(対応付け)を行い、運転指示に応じた行動を出力するモデルの構築を目指す。

2 グラウンディングにおける推論

図1に提案手法の概要を示す。本研究では、Kollarら[1]により提案されたSDC(Spatial Description Clause)と呼ばれる空間関係を表すことが出来る意味構造に変換された運転指示と、車に取り付けられたセンサーにより認識される環境表現を入力として与える。SDCの構造に基づいた確率的グラフィカルモデルを生成し、確率を計算することで、自然言語と実世界空間のグラウンディングを行う。グラフィカルモデルは、Tellexら[2]により提案された G^3 (Generalized Grounding Graph)を使用する。図2にSDCとそのSDCから生成された G^3 の例を示す。

なお、運転指示からSDCへの変換は、稲子ら[3]による既存手法によるものとする。 G^3 の確率の計算は、

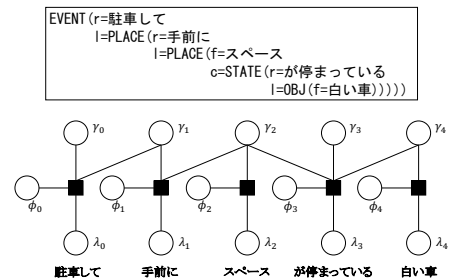


図2: SDC と生成された G^3

確率伝搬法と呼ばれる手法を使用する。Tellex らの手法では、それぞれの環境と言葉の組み合わせを独立に扱い、それぞれの確率を最大にするような近似的な計算を行っていたため、グラフ全体の確率について必ずしも最大化されているとは限らなかった。本研究では、確率伝搬法による近似を用いてグラフ全体の繋がりを考慮した周辺確率を計算し最大化を行う。

2.1 グラウンディンググラフの生成

グラウンディンググラフは、SDCに変換された駐車指示から自動的に生成する。SDCの要素を前から順に見て行き、ルールに従って G^3 を生成することで、SDCの階層的な構造を反映している。

2.2 グラフにおける素性の設定

素性は、空間に関する素性と、言語に関する素性をそれぞれ用意する。空間に関する素性は、因子ノードに接続する環境変数ノードの数と種類ごとに考える。また、空間については、車と駐車場の間の具体的な距離のような定量的な素性と、位置の順序関係のような定性的な素性の2種類を用意した。言語に関する素性

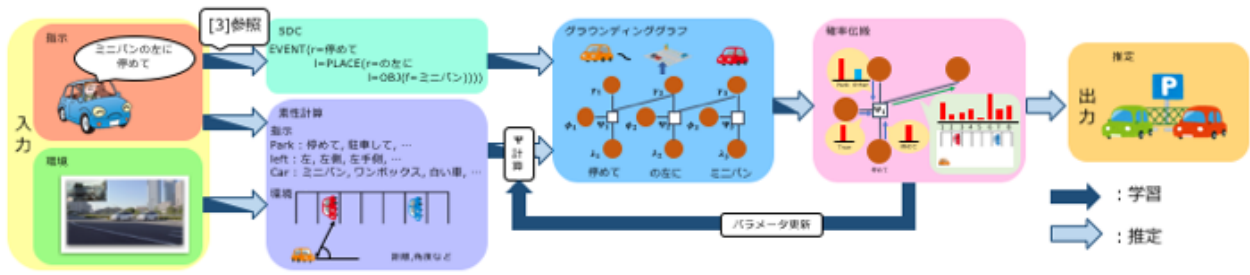


図 1: 提案手法概要

では、辞書を人手で作リ、異なる表現ではあるが同じ意味を表す言葉を吸収する。

2.3 確率伝搬法による周辺確率分布の計算

確率伝搬法は、ノードからノードへメッセージを伝搬させることで、観測されているノードの情報を元に、観測されていないノードの周辺確率を計算する手法である [4]。Algorithm 1 が確率伝搬法のアルゴリズムである。

Algorithm 1 確率伝搬アルゴリズム

1 メッセージの初期化

全てのメッセージを 1.0 に初期化する。

$$m_{\alpha \rightarrow i}^{(0)} = 1, \Gamma m_{i \rightarrow \alpha}^{(0)} = 1 \quad (1)$$

2 メッセージの計算

以下の式で、因子ノードから確率変数ノードへのメッセージの伝搬 (2) と、確率変数ノードから因子ノードへのメッセージの伝搬 (3) を行う。

$$m_{\alpha \rightarrow i}^{(t+1)}(x_i) \propto \sum_{x_{\alpha \setminus i}} \Psi_{\alpha}(x_{\alpha}) \prod_{j \in x_{\alpha} \setminus i} m_{j \rightarrow \alpha}^{(t)}(x_j) \quad (2)$$

$$m_{i \rightarrow \alpha}^{(t+1)}(x_i) \propto \prod_{\beta \in \partial i \setminus \alpha} m_{\beta \rightarrow i}^{(t)}(x_i) \quad (3)$$

3 周辺確率の計算

収束したメッセージを $m_{\alpha \rightarrow j}^*$ として、確率変数ノードの周辺確率 (4) と因子ノードの周辺確率 (5) を以下の式で計算する。

$$p(x_i) \propto \prod_{\alpha \ni i} m_{\alpha \rightarrow i}^*(x_i) \quad (4)$$

$$p(x_{\alpha}) \propto \Psi_{\alpha}(x_{\alpha}) \prod_{j \in x_{\alpha}} m_{j \rightarrow \alpha}^*(x_j) \quad (5)$$

2.4 重みパラメータの学習

式 (6) で与えられる重み因子により G^3 の周辺確率を計算し、得られた周辺確率を用いてパラメータ $\theta_{\alpha, s}$ の更新を行う。

$$\Psi_{\alpha}(x_{\alpha} | \theta_{\alpha}) = \exp \left(\sum_s \theta_{\alpha, s} t_s(x_{\alpha}) \right) \quad (6)$$

α は因子ノードを区別するインデックスである。パラメータの更新には、IPF アルゴリズム (Iterative Proportional Fitting Algorithm) [5] を用いる。更新式は、式 (7) で与える。

$$\theta_{\alpha, s}^{(q+1)} \leftarrow \theta_{\alpha, s}^{(q)} + \eta \log \frac{\hat{p}_{\alpha}(t_s(x_{\alpha})) - \mu(\theta_{\alpha, s}^{(q)} - 1)}{p_{\alpha}(t_s(x_{\alpha}) | \theta_{\alpha}^{(q)}) + \mu} \quad (7)$$

ここで、 $\hat{p}_{\alpha}(t_s(x_{\alpha}))$ は学習データから与えられる経験分布、 $p_{\alpha}(t_s(x_{\alpha}) | \theta_{\alpha}^{(q)})$ は、確率伝搬法により与えられる、因子の周辺分布である。 μ は、パラメータがオーバーフローすることを防ぐ正則化項であり、この値を導入することで安定した学習を行うことができる。また、 η は学習率であり、1 から始まり、学習が進むにつれて減衰するように設定をした。

3 実験

実験では、駐車指示内に含まれる物体や位置と、実環境に存在する物体の対応付けを行った。素性は定性的な素性のみを用いた場合、定量的な素性のみを用いた場合、両方を用いた場合の 3 種類について結果を出した。また、駐車指示から SDC への変換には稲子ら [3] の手法を用いた。

3.1 実験設定

対象データは、環境データと SDC に変換した駐車指示の組み合わせを用いる。環境データは、LiDAR で

表 1: SDC の構造とグラウンディンググラフの例

ID	SDC の例	グラウンディンググラフの例	数
1	EVENT(r=停めてください l=PLACE(r=の右隣りに l=OBJ(f=白いワゴン)))		$\frac{261}{1000}$
2	EVENT(r=とめて l=PLACE(r=の手前に l=OBJ(f=白い大きな車 r=左側の)))		$\frac{116}{1000}$
3	EVENT(r=停めてください l=PLACE(r=一番手前に l=PLACE(r=の右側 l=OBJ(f=白い乗用車))))		$\frac{48}{1000}$
⋮	⋮	⋮	⋮

センシングされたデータを yaml ファイルで記述したデータである。環境内には駐車場が 7 ヶ所、車が 4 台あり、2 台は駐車場に、2 台は駐車場の正面の少し離れた場所に停まっている。環境は合計 8 種類あり、うち 4 種類は自車が駐車場の手前、残りの 4 種類は自車が駐車場の正面にあり、自車の座標を基準に駐車場や他の車の座標や向きが表されている。駐車指示データは、クラウドソーシングを用い、100 人のクラウドワーカーに自車から見た環境データの写真を見せ、ある位置に駐車したいときの駐車指示 1000 文を収集した。なお、今回の実験では、“左の白い軽自動車と、乗用車の真ん中に停めて。両方の車から、一台分けて”のような駐車指示が 2 文以上にまたがる文は除外している。また、駐車指示の視点について、“白いワゴンの(自分が乗っている車から見て)右隣に停めて”のように、視点が自分の場合と、“白いワゴンの(白いワゴンから見て)右隣に停めて”のように、視点が物体の場合の両方が存在していた。本実験では、視点が自分の場合以外は除外して計算を行っている。更新式の正則化パラメータについては、値を変えて複数回実験を行い、最も精

度が高かった $\mu = 0.1$ の結果を示す。

3.2 SDC の解析

図 3 に SDC の構造の分布を示す。横軸が SDC の構造の ID、縦軸がその ID の SDC が対象コーパスに現れた頻度である。駐車指示を SDC に変換した際の構文のパターンは、全部で 144 種類あった。最も単純な ID 1 の構造と、この OBJECT SDC に relation を追加した ID 2 の構造の数を合わせると 377 であり、1/3 以上をこの 2 種類が占めていた。本実験では、この 2 種類のグラウンディングの精度を計算している。また、表 1 は、それぞれの SDC の構造と、その SDC から生成されるグラウンディンググラフを表している。

3.3 結果と考察

表 2, 3, 4 に、学習を 5000 回行った際の結果を示す。表 2 は、表 1 における ID 1 の SDC のみ、表 3 は、ID 2 の SDC のみ、表 4 は ID 1, 2 の SDC を使った場合の

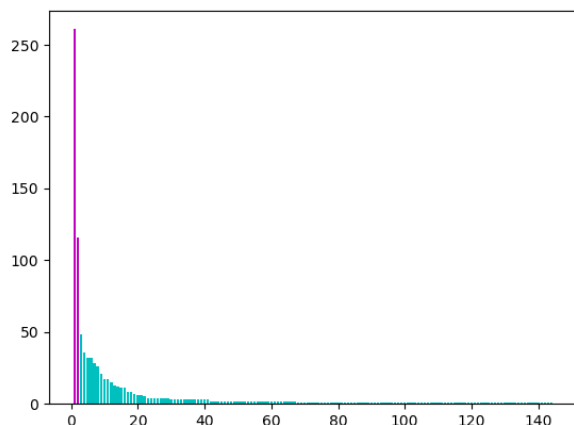


図 3: SDC の構造の分布

精度である。それぞれの表の all は、駐車指示に現れる全ての SDC に対するグラウンディングが正しい確率、event は駐車指示内の出来事に関するグラウンディング、place は場所に関するグラウンディング、object は物体に関するグラウンディングが正しい確率を表している。

ID 1 のみを用いた場合の精度については、2 つの素性を使った際は all で 79% と、高い精度での推論が行っていた。しかし、ID 2 については、2 つの素性を使った際の all で 49% であり、ID 1 と比較し大幅に精度が下がってしまった。この理由として、ID 2 の SDC は、“EVENT(r =とめて l =PLACE(r =の手前に l =OBJ(f =白い大きな車 r =左側の)))” のように、位置関係を表す言葉が 2 つ現れている。つまり、視点を持ち込まなければならないような言葉が複数出てしまい、位置関係が複雑になったため、ID 1 と比べ精度が下がったということがある。また、ID 2 で計算を行っている素性が環境表現の特徴を捉えておらず、推論が正しく行えていない可能性も考えられる。

表 2: ID 1 のみ

	all	event	place	object
組み合わせ	0.79	1.0	0.82	0.83
定性	0.61	1.0	0.66	0.73
定量	0.77	1.0	0.79	0.82

表 3: ID 2 のみ

	all	event	place	object
組み合わせ	0.49	0.91	0.42	0.39
定性	0.49	0.91	0.30	0.24
定量	0.48	0.91	0.46	0.39

表 4: ID 1 と ID 2

	all	event	place	object
組み合わせ	0.71	0.97	0.71	0.66
定性	0.56	0.97	0.54	0.42
定量	0.70	0.97	0.72	0.66

4 おわりに

本研究では、自動運転車を対話により操作することを目的とし、駐車シーンにおいて駐車指示を正しく理解し駐車場所を推定する手法を提案した。今後は、グラウンディングについて、学習可能な SDC の構造を増やし、様々な駐車指示に対応出来るようモデルを拡張したい。また、SDC を時間に依存する運転指示にも対応できるように、時空間意味記述へ拡張し、例えば「後ろの車が来る前に、隣のレーンに入って」のような動的な運転指示に対しても対応できるようにして行きたい。

参考文献

- [1] T.Kollar, S.Tellex, D.Roy, N.Roy: Toward Understanding Natural Language Directions, ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI), pp.259-266, 2010.
- [2] 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.
- [3] A.Inago, H.Tsukahara, I.Kobayashi: Parsing Parking Instructions for Self-driving Cars into Spatial Semantic Descriptions, ICICA, 2019
- [4] 渡辺 有祐:グラフィカルモデル, 講談社, 2016.
- [5] J. N. Darroch and D. Ratcliff, Generalized Iterative Scaling for Log-Linear Models, Ann. Math. Statist. 43 (1972) 1470-1480.