

自動運転の対話的操作を実現するための 自然言語の実環境へのグラウンディング

Grounding natural language to the real world
to realize the interactive operation of autonomous cars

大田原 菜々^{*1}

Nana Otawara

塚原 裕史^{*2}

Hiroshi Tsukahara

小林 一郎^{*1}

Ichiro Kobayashi

^{*1}お茶の水女子大学

Ochanomizu University

^{*2}株式会社デンソーアイティラボラトリ

Denso IT Laboratory, Inc.

Recently, practical applications of automatic driving have been rapidly developing. In the future, it must be necessary to enable interactive operation by natural language in order to easily operate autonomous cars. We therefore attempt to realize the correspondence relationship, i.e., grounding, between the driving instructions expressed in natural language and the objects in the real world recognized by the sensors equipped with a car, and then convert the driving instructions into the particular spatial meaning description to operate autonomous cars. In this study, we particularly focus on the parking operation of a car.

1. はじめに

近年、車の自動運転の実用化に向けた活動が急速に進展している。今後、自動運転車の操作を容易に行うために、自然言語による対話的な操作を可能にすることが必要であると考えられる。そこで、本研究では、稲子ら [1] によって提案された自然言語表現を空間意味記述に変換する手法を用い、自然言語で表現される運転指示を空間意味記述に変換した表現と、車に備え付けられたセンサーによって認識される実世界との対応付け（グラウンディング）を行い、空間意味記述に変換された運転指示から、駐車の際の車の操作を推定する手法を提案する。

2. 関連研究

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 [2] は Spatial Description Clauses (SDC) という意味構造を提案した。Tellex ら [3] は 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である Generalized Grounding Graphs (G^3) を開発し、SDC の構造に意味解析された言葉と実世界を結びつけることで、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。また、稲子らは言葉を Combinatory Categorical Grammar (CCG) という空間意味のカテゴリを用いる文法により、空間意味の依存木を抽出し、SDC に変換する手法を提案している。この研究では、Kollar らが定義した 4 つの SDC のタイプ (EVENT, OBJECT, PLACE, PATH) に加え、STATE と VIEW の 2 つを追加している。

Howard ら [4] は、 G^3 を発展させたモデルである Distributed Correspondence Graph (DCG) を提案した。DCG は、環境内の空間に目標地点、通行可能な領域、通行不可能な領域という計画制約を推定することで、計算時間を削減し、 G^3 よりも高い正解率でのグラウンディングを可能にしている。他の DCG を拡張したモデルとして、Paul ら [5] は Adaptive Distributed Correspondence Graph (ADCG) を提案している。ADCG は抽象的、階層的な概念を理解するためのモデルであり、環境内に並んでいる複数の物体のグループや、複数の物体の内から一

つの物体をグラウンディングしたい場合に有効である。これらはグラフィカルモデルから尤度を計算し、確率の高い行動やオブジェクトを推定している。また、深層学習を用いてグラウンディングに取り組んでいる研究も存在する。Rohrbach ら [6] は、ある句から最も関連のある領域を推定する Grounding のモデルと、推定された領域から元の句を再現する Reconstruction の 2 つのモデルを組み合わせることで、アノテーション（画像への言葉のラベル付け）なしで物体と言語の結び付けを行なっている。

これらのグラフィカルモデルを用いた手法はいずれもグラウンディンググラフの尤度を、一つの言葉と実世界の組み合わせごとに独立して計算し、ロボットの操作や経路を推定している。グラフ全体の尤度を要素ごとの尤度に独立して求めることで近似しており、正確な推定が出来ているとは限らない。そこで、本研究では、稲子らの研究によって SDC に解析された言葉と環境情報からグラウンディンググラフを自動生成し、確率伝搬法により、各ノードのグラフ全体を考慮した確率を計算する手法を提案する。

3. 確率伝搬法を用いたグラウンディング

3.1 提案手法概要

本研究では、運転指示に対応する SDC と、実世界において観測された環境データからグラウンディンググラフを生成する。SDC は、EVENT, PATH, PLACE, OBJECT, STATE, VIEW の 6 つのタイプの内 1 つを持ち、relation, landmark, figure, condition, view の 5 つの要素で構成されている。EVENT は行動の連続、PATH は実世界の経路やその一部、PLACE は実世界の場所、OBJECT は実世界の物体、STATE は物体や場所の状態、VIEW は視点を表現している。

グラウンディンググラフの構造と因子関数 $\Psi_\alpha(x_\alpha)$ から、確率伝搬法により、因子ノード α の周辺確率分布を計算する。次に、周辺確率分布と学習データから計算される経験分布を用いて、素性の重みをパラメータとしてパラメータの更新を行う。この更新を、全てのパラメータについて更新の幅が一定以下になる、または更新の回数が一定以上に達するまで繰り返す。推定の際には、学習したパラメータを用いて確率伝搬を行い、 $\gamma_i \in \Gamma$ に対応する確率変数ノードの周辺確率分布を計算することで、自然言語に対応する操作や実世界のオブジェクトを推

連絡先: 大田原菜々, お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース, 〒112-8610 東京都文京区大塚 2-1-1 小林研究室, g1420511@is.ocha.ac.jp

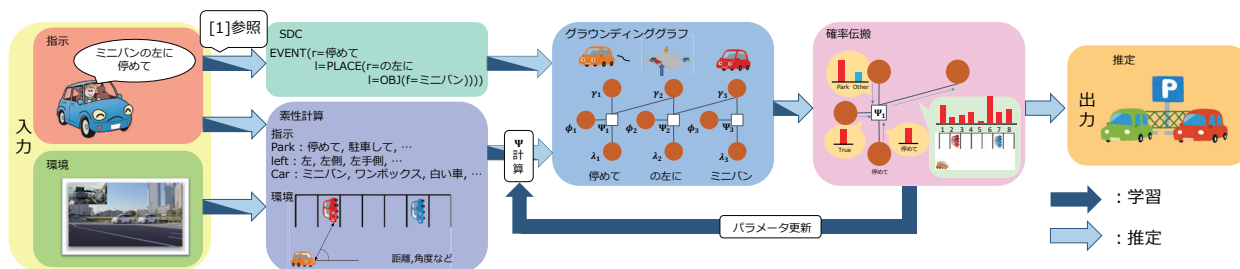


図 1: 提案手法概要

定する. 図 1 に提案手法の概要を示す.

3.2 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, 確率変数ノード γ , λ , ϕ と因子ノード (確率変数の関数のノード) Ψ から構成される因子グラフ型モデルである. ここで, 因子グラフとは, 確率変数ノードと因子ノードによって構成される二部グラフであり, 多変数関数の因数分解を表している. γ は車の操作や実世界のオブジェクトを表すノードであり, 複数の操作やオブジェクトの候補の中から確率伝搬により周辺確率分布を計算し, 最も確率が高いものを選択する. λ は SDC の単位に分割された言葉を表すノードである. ϕ は, ある因子ノードに接続している実環境 γ と言語 λ の対応関係が正しければ *True*, 間違っていれば *False* を表す二値関数である. グラフの因子関数 Ψ は次のように設定する.

$$\Psi_{\alpha}(x_{\alpha}|\theta_{\alpha}) = \exp(\langle \theta_{\alpha}, f_{\alpha}(x_{\alpha}) \rangle) \quad (1)$$

ここで, α は因子グラフの因子ノード, x_{α} は α に接続する確率変数ノードの集合, $f_{\alpha}(x_{\alpha})$ は素性, θ_{α} は素性に対する重みである. $\langle \cdot, \cdot \rangle$ はベクトルの内積を表している. SDC と環境データからグラウンディンググラフを生成する際, SDC の入れ子の構造によって, グラウンディンググラフの構造を決める. 以下, グラウンディンググラフの生成アルゴリズム (Algorithm 1) を示す. アルゴリズム内の括弧は図 2 左の SDC の具体例である.

Algorithm 1: グラウンディンググラフ生成アルゴリズム

1. 最初の γ ノード, ψ ノード, ϕ ノードを生成 (γ_0, ψ_0, ϕ_0)
2. SDC の要素 (figure, relation, landmark, view, condition) を順に調べる
 - 2-1 要素が 1 つ目の言葉の場合 (r =駐車して)
 - 新しく λ を生成し (λ_0), ψ と接続 (ψ_0)
 - 2-2 要素が 2 つ目の言葉の場合
 - 新しく ψ, ϕ, λ を生成し, γ に接続
 - 2-3 要素が EVENT または STATE SDC の場合 (c =STATE(...))
 - 新しく γ, ψ, ϕ を生成 (γ_3, ψ_3, ϕ_3) し, γ に接続 (γ_2), 2 に戻り次の要素を見る
 - 2-4 要素が上記以外の SDC の場合 (l =PLACE(...))
 - 新しく γ, ψ, ϕ を生成 (γ_1, ψ_1, ϕ_1) し, ψ に接続 (ψ_0), 2 に戻り次の要素を見る

このアルゴリズムにより生成したグラウンディンググラフの例を図 2 に示す. 図の左側が「白い車が止まっているスペース

(の) 手前に駐車して」, 右側が「白いバンの車から見て左側に駐車して」という指示の SDC と, SDC を元に生成したグラウンディンググラフである.

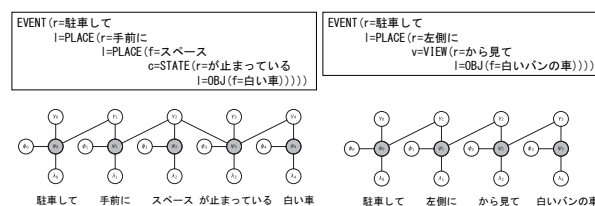


図 2: SDC と生成されたグラウンディンググラフ

3.3 グラフにおける素性の設定

素性は, 言語に関する素性と環境内の空間関係に関する素性の 2 つを組み合わせたものを用いる. 以下, それぞれの素性の設定について説明する.

3.3.1 言語に関する素性

言語に関する素性は, 事前に辞書を用意し, 同じ行動やオブジェクトを表すが異なる表現である言語を吸収する. 例えば, 「の奥に」, 「の向こう側に」, 「の一つ奥側に」は表現は異なるが, 全て「自分から見たとき, ある物体の奥」という同じ場所を表しているため, 「奥」という素性に吸収される.

3.3.2 空間に関する素性

空間に関する素性は, 因子ノードのタイプ, つまり, 因子ノードに接続している確率変数ノード γ の数と SDC のタイプごとに考える. 今回, 因子ノードが 1 つの γ に接続し, SDC のタイプが OBJECT であるファクター (タイプ ID1), 因子ノードが 2 つの γ に接続し, SDC のタイプが PLACE であるファクター (タイプ ID2), 因子ノードが 2 つの γ に接続し, SDC のタイプが EVENT であるファクター (タイプ ID3) の 3 種類について考える. 空間的な素性には, 自転車から見た距離や角度, 自転車の向いている方向とオブジェクトを見たとき右側にあるか左側にあるか, などを用意し, それらを距離の順序 (定性的な素性) と実際の距離の数値 (定量的な素性) の 2 つに分けて考えた. 表 1 に空間的な素性の一覧を示す.

3.4 確率伝搬法と周辺確率の計算

確率伝搬法は, グラフィカルモデルの周辺確率分布を計算する手法である [7]. 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を元に, 観測されていないノードの周辺確率分布の計算を可能にする. グラウンディンググラフの周辺確率分布の計算では, まず, 全ての因子ノード α から確率変数ノード i へのメッセージについて, 1 に初期化する. 次に, 因子ノードから確率変数ノードへのメッセージの計算と,

表 1: 空間的な素性の一覧

タイプ ID	定量 or 定性	素性の名前	説明	計算
1	定性	(n)th_near_from_fp	自車と物体の間の距離の順序 ($0 \leq n < 4$)	-
1	定性	right(left)_side_from_fp	自車から見て物体が右(左)側にあるか	-
1	定量	size_of_obj_(s)	物体の体積 ($0 \leq s < 20$)	-
1	定量	dist_fp_to_obj_(i)	自車と物体の間の距離 ($0 \leq i < 25$)	$i \leq \text{get_dist} < i + 1$
1	定量	angle_of_fp_and_obj_(θ)	自車を基準としたときの 自車の向きと物体の間の角度 ($0 \leq \theta < 360$)	$\theta \leq \text{get_angle} \leq \theta + 5$
2	定性	is_overlap	駐車場に車が停まっているか	-
2	定性	(n)th_near_from_fp	自車と物体 1 の間の距離の順序 ($0 \leq n < 7$)	-
2	定性	right(left)_side_from_fp	自車から見て物体 1 が右(左)側にあるか	-
2	定性	right(left)_side_(n)th	自車と物体 1 の間の距離の順序, 自車から見て物体 1 が右(左)側にあるか の組み合わせ ($0 \leq n < 7$)	-
2	定量	dist_obj1_to_obj2_(i)	物体 1 と物体 2 の間の距離 ($0 \leq i < 25$)	$i \leq \text{get_dist} < i + 1$
2	定量	angle_of_obj1_and_obj2_(θ)	自車を基準としたときの 物体 1 と物体 2 の間の角度 ($0 \leq \theta < 360$)	$\theta \leq \text{get_angle} \leq \theta + 5$
2	定量	angle_of_obj1_and_hd_(θ)	自車を基準としたときの 物体 1 と自車の方向の間の角度 ($0 \leq \theta < 360$)	$\theta \leq \text{get_angle} \leq \theta + 5$
3	-	is_park	行動が「駐車」であるか	event=="Park"
3	定性	is_overlap	駐車場に車が停まっているか	-
3	定量	dist_fp_to_obj_(i)	自車と物体の間の距離 ($0 \leq i < 25$)	$i \leq \text{get_dist} < i + 1$
3	定量	angle_of_hd_and_obj_(θ)	自車を基準としたときの 自車の向きと物体の間の角度 ($0 \leq \theta < 360$)	$\theta \leq \text{get_angle} \leq \theta + 5$

確率変数ノードから因子ノードへの計算を、収束するか step 数が 5000 に達するまで交互に繰り返す。全てのメッセージについて、 t 回目のメッセージと $t+1$ 回目のメッセージの差が ϵ 以下になったとき、更新を終了する。収束したメッセージを用いることで、因子ノードの周辺確率と確率変数ノードの周辺確率が計算できる。因子ノードの周辺確率分布は学習の際に、確率変数ノードの周辺確率分布は車の操作やオブジェクトの推定の際に用いる。確率伝搬のアルゴリズム (Algorithm 2) と式を以下に示す [8]。

Algorithm 2: 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。
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計算 (2) と、確率変数ノードから因子ノードへのメッセージの計算 (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。

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

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

3.5 パラメータの更新

次に、離散状態変数のモデルの場合において、各 θ_{α} ごとに対数尤度関数を最大化していく、IPF アルゴリズム (Iterative

Proportional Fitting algorithm)[9] に正則化項を含めた更新式 (6) を適用し、全てのパラメータ (素性の重み) について更新を行い、収束するかどうか調べる。

$$\theta_{\alpha}^{(t+1)} \leftarrow \theta_{\alpha}^{(t)} + \eta \log \frac{\hat{p}_{\alpha}(z) - \lambda \theta_{\alpha:z_{\alpha}}^{(t)}}{p_{\alpha}(z|\theta^{(t)}) + \lambda} \quad (6)$$

この際、式 (5) により求めた周辺確率を用いる。パラメータの更新では、全ての因子について、経験分布 $\hat{p}_{\alpha}(z)$ とモデルの周辺確率分布 $p_{\alpha}(z|\theta^{(t)})$ が近付く方向にパラメータの更新を行い、 θ を決める。パラメータ θ は素性に対する重みである。経験分布 $\hat{p}_{\alpha}(z)$ は素性とし、0 または 1 の二値をとる。周辺確率分布 $p_{\alpha}(z|\theta^{(t)})$ は式 (5) で求められる因子ノードの周辺確率である。また、 η は学習率、 z は一つの素性、 λ はパラメータの絶対値が大きくなり過ぎないように調節するための正則化項である。学習率はシグモイド関数をベースにした、以下の式で定義する。

$$\eta = 2 \left(\frac{1}{\exp(-\text{maxstep}/(2c(t+1)))} - 0.5 \right) \quad (7)$$

ただし、 $\text{maxstep} = 5000$ 、 c は減衰の速度を決める値で $c = 0.25$ とする。学習率は 1.0 から始まり、学習を進めるにつれて減衰していくように設定を行った。

収束する場合は更新を終了し、収束しなければ求めたパラメータを用いて、全ての因子ノードについて式 (1) を計算する。式 (1) で求めた値を使い、式 (2) と式 (3) をメッセージが安定するまで繰り返し、式 (5) によって周辺確率分布を求める。求めた周辺確率分布を用いて式 (6) で更新を行い、収束するかどうか判定する。これを収束するまで繰り返すことによって、各素性に最適な重みが得られる。

3.6 推定

パラメータの更新によって求めた重みを使い、式 (4) によって $\gamma_i \in \Gamma$ の確率変数ノードの周辺確率分布を求めることによって、各 γ の候補の中からどれが確率が高いかを推定する。

4. 実験

駐車指示データと外部情報のデータを入力とし、実装したモデルの評価実験を行なった。

4.1 実験設定

対象データは、クラウドソーシングによって集めた、駐車場に 7 つの駐車スペースがあり、そのうち 2 つに自動車が停まっている画像と駐車したい位置が与えられたときの駐車指示 206 文と、それに対応する環境 4 つのうち 1 つのペアで与えられるため、データサイズは 824 となる。今回、駐車指示は、自車から見た場合の指示を行なっているものに限定した。つまり、「(自分から見て) 白いワゴンの右隣に停めて」という文は使用するが、「(電気自動車から見て) 電気自動車の左に駐車して」と言っている文は除外した。このデータに人手で正解ラベルを作成した。評価は 5 分割交差検定により求めた、ラベルに対する推定の正解率のマクロ平均である。結果を図 3 に示す。図 3(a) が定量的な素性と定性的な素性を組み合わせた際の結果、(b) が定量的な素性を用いた場合の結果、(c) が定性的な素性を用いた場合の結果である。それぞれの結果の左上の all は、1 つの指示に対する全てのグラウンディングが正しかった場合、右上の event は指示のうち動作に関するグラウンディングが正しかった場合、place は駐車場所が正しかった場合、object は指

示に含まれる物体のグラウンディングが正しかった場合正解とした。黒い線で表される正解率が最も確率の高いものを正解とみなした場合、赤い線の正解率は、最も確率の高いものと2番目に確率の高いものを正解とみなした場合である。グラフの縦軸が正解率、横軸が学習のステップ数である。また、最終ステップでのそれぞれの正解率を表2に示した。

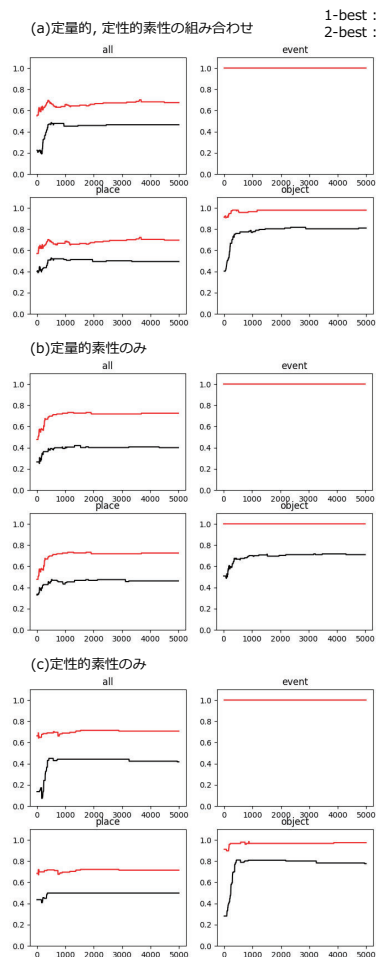


図 3: 実験結果 (横軸 : step 数, 縦軸 : 正解率)

表 2: 最終ステップでの正解率

	all	event	place	object
組み合わせ	46.5	100.0	49.3	80.9
定量	40.0	100.0	46.1	70.9
定性	41.6	100.0	49.8	77.5

4.2 実験結果と考察

定量的な素性、定性的な素性、両方を組み合わせた素性の3つの場合について実験を行なった。学習の最後の5000ステップ目では、定量的な素性を用いた場合が一番正解率が低く40.0%、両方を組み合わせた素性が、定性的な素性よりも正解率が高く、全体で46.5%であった。行動については、今回の実験設定では駐車かそれ以外かの二択からのグラウンディングであり、使った正解データに駐車の場合だけが含まれているため、正解率は100%である。物体では、両方を組み合わせた素性で80.9%、定

量的な素性で70.9%、定性的な素性で77.5%であった。場所の推定はどの素性を用いても正解率が低く、一番正解率の高い(a)でも49.3%だった。全体的に定量的な素性を用いた場合よりも定性的な素性を用いた場合の方が正解率が高くなっている。定量的な素性では、自転車から見たオブジェクトの座標が環境によって微妙にずれているため、観測している素性は異なっているが駐車したい場所と駐車指示は同じであるという場合が多く存在する。対照的に、定性的な素性では、距離や角度などの具体的な数字は用いず、位置関係だけを捉えているため、環境が異なる場合にも対応出来ていると考えられる。

IPF アルゴリズムによる更新式は、学習データがスパースで経験分布が0になる場合があると、数値的に不安定になり、学習が収束しなかったが、正則化項を取り入れることによって、安定した学習が可能になった。

5. まとめと今後の課題

本研究では、駐車指示を空間意味記述に変換した表現からグラウンディンググラフを生成し、グラウンディンググラフと環境表現から、確率伝搬法を用いることでグラフ全体の確率を考慮して言葉と行動やオブジェクトをグラウンディングする手法を提案し、モデルの実装を行った。行動と物体については、高い確率で正しいものを推定することができたが、駐車場所の推定では40%から50%の正解率であり、実車に導入するには改良の必要がある結果であった。素性については定量的な素性と定性的な素性を組み合わせた場合が最も正解率が高かったが、正解率向上のために、どのような素性を用いるべきか検討していきたい。現在は特定のSDCの型に変換できる指示文のみで学習を行なったが、今後は全ての駐車指示で学習を行えるモデルの実装を進め、多クラス分類ロジスティック回帰による手法との比較を行なっていきたい。

参考文献

[1] 稲子 明里, 塚原 裕史, 小林 一郎: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, 人工知能学会全国大会, 2017.

[2] 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.

[3] 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.

[4] Thomas M. Howard, Stefanie Tellex, Nicholas Roy: A Natural Language Planner Interface for Mobile Manipulators, ICRA, 2014

[5] Rohan Paul, Jacob Arkin, Nicholas Roy, Thomas Howard: Grounding Abstract Spatial Concepts for Language Interaction with Robots, IJCAI, 2017

[6] Anna Rohrbach1 Marcus Rohrbach2,3 Ronghang Hu2 Trevor Darrell2 Bernt Schiele1: Grounding of Textual Phrases in Images by Reconstruction, ECCV, 2016

[7] Jonathan S. Yedidia, William T. Freeman, Yair Weiss:Generalized Belief Propagation, 2001.

[8] 渡辺 有祐:グラフィカルモデル, 講談社, 2016.

[9] Stephen E. Fienberg, Michael M. Meyer:Iterative Proportional Fitting., 1981