

## 動作とオブジェクトの関係に基づく機能カテゴリーを表す単語のオンライン学習

中村 慎也<sup>1</sup> 長井 隆行<sup>1</sup> 岩橋 直人<sup>2</sup> 麻生 英樹<sup>3</sup> 佐藤 健<sup>4</sup><sup>1</sup>電気通信大学電子工学専攻 <sup>2</sup>国際電気通信基礎技術研究所 <sup>3</sup>産業技術総合研究所 <sup>4</sup>国立情報学研究所

## 1. はじめに

日常我々が用いている単語には、知覚特徴から直接的に形成される知覚特徴単語と、直接的には形成されない抽象的単語の2種類がある。例えば、オブジェクトそのものやその知覚特徴の一部を指示するもの、動作を指示する単語などが知覚特徴単語であり、抽象的単語は知覚特徴単語を基に形成されるものである。本稿では、「持ち上げる」などの動作を指示する単語を基に形成される「持ち上げられるもの」といった機能カテゴリーを指示する抽象的単語の学習について考える。文献[1]では、抽象的語意をバッチ学習するベイズ学習機構を提案している。しかし単語の学習は、本来一括にデータを与えることでなされるわけではなく、その単語が使用されるたびにインクリメンタルに学習されるものであると考えられる。そこで本稿では、文献[1]をベースに、単語の学習をオンラインで行うことを検討する。

具体的な単語の学習には、変分法を援用したベイズ学習である変分ベイズ法[2]を用いる。知覚特徴単語または機能カテゴリーを指示する単語を学習する場合、その時点でのレキシコンを表すグラフィカル・モデルをベイズ規準に基づいて拡張した複数の候補モデルを作成し、その中から最尤の解釈のモデルを選択する。本稿ではさらに、事前知識などの条件付けを様々に変更し、多数の条件下で実験を行い提案手法の有効性を検討する。

## 2. 学習機構の概要

## 2.1. グラフィカル・モデルとレキシコン

本稿で用いるグラフィカル・モデルは図1の通りである。動作は動かされるオブジェクトと基準となるオブジェクトの位置関係によって表現され、それぞれトラジェクタとランドマークと呼ぶ。知覚特徴単語と機能カテゴリーを表す単語の学習の際には学習サンプルとして、その単語を表すラベルとオブジェクトの色( $L^*a^*b^*$ , 3次元)、面積(1次元)、輪郭(フーリエ記述子, 8次元)の12次元のベクトルで表される画像特徴量  $X_T$ (或いは  $X_L$ )を与えるが、それらの概念を指示する単語をレキシコンに追加するとそれぞれ異なった構造のグラフィカル・モデルが得られる。その単語が知覚特徴単語を指示していると解釈した場合、知覚特徴概念とそれを指示する単語を表す  $W_T=[W_{T,1}, \dots, W_{T,mg}]$ (あるいは  $W_L=[W_{L,1}, \dots, W_{L,mg}]$ )に新たに要素が1つ追加される。トラジェクタ機能、あるいはランドマーク機能を指示する単語だと解釈した場合は、トラジェクタ機能を指示する単語を表す  $W_{FT}=[W_{FT,1}, \dots, W_{FT,ma}]$ か、またはランドマーク機能を指示する単語を表す  $W_{FL}=[W_{FL,1}, \dots, W_{FL,ma}]$ の要素の1つに単語が対応付けられる。なお、 $mg$  は学習した知覚特徴単語の数、 $ma$  は動作の数である。 $W_{FT}$  と  $W_{FL}$  の各要素は、動作を指示する単語を表す  $A=[a_1, \dots, a_{ma}]$ の各要素と対応しており、それぞれ条件付確率のパラメータの分布を共有する。

## 2.2. オンライン学習

新たな学習サンプルが入ってくる度に、逐次インクリメンタ

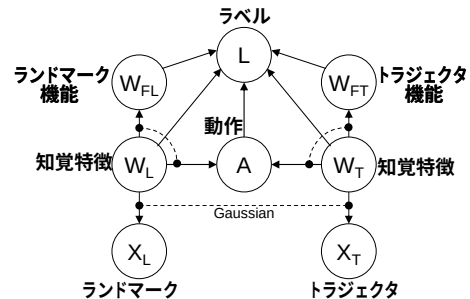


図1. レキシコンのグラフィカル・モデル

ル学習を行うことが本手法の特徴である。前項でも触れた通り、学習サンプルが与えられるとその単語の解釈を行い、ベイズ規準に基づいて新たな語意を獲得する。このとき、そのサンプル中のラベルが示す単語を既に学習済みの場合、その時点で学習している他の単語のパラメータは固定して、新たに与えられたサンプルを含む同一ラベルの画像特徴量セットを用いて、その単語を再学習する。従って、与えたサンプル数が少ない段階で誤った語意として学習していた場合にも、サンプル数の増加に伴い正しく語意を学習することが可能となる。なお、サンプル中のラベルが未学習の単語を示す場合には、その単語は新しい単語であるとして学習を行う。

## 2.3. 学習アルゴリズム

新たに学習サンプルが与えられた場合、作成される候補モデルは次の3種類である。即ち、与えられたデータの示す語意を知覚特徴単語としたモデル  $M_G$ 、語意をトラジェクタ機能としたモデル  $M_{T,1}, \dots, M_{T,ma}$ 、語意をランドマーク機能としたモデル  $M_{L,1}, \dots, M_{L,ma}$  であり、候補モデルの数は  $(1+2ma)$  個となる。これらのモデルから最尤のものを選択し、語意の解釈を行う。また、その単語が既に学習済みであった場合、新たに与えられた学習サンプルを用いて単語の語意の再推定を行う。なお、学習には変分ベイズ法を用いており、各モデルの自由エネルギーを計算し、最大となるものを選択する。

## &lt; 学習アルゴリズム &gt;

1. ある単語を指示する画像特徴量セット  $X_i$  とラベル  $L_i$  の組を1サンプル取得する
2. ラベルの有無を確認し、保持していた場合その時点での全同一ラベルの画像特徴量をデータセットとし、保持していない場合は新たに与えられた画像特徴量をデータセットとする
3. 学習する単語に対して、 $(1+2ma)$ 個の候補モデルを作成する
4. データセットを用いて、変分ベイズ法により各モデルのパラメータを更新、及び自由エネルギーを計算する
5. 自由エネルギーが最大となるモデルを選択し、語意の解釈とする

ここで、トラジェクタ機能を指示する単語の場合の自由エネルギーは次の式で表される。 $Z$  は非観測変数、 $\Theta$  はモデルのパラメータである。 $p(\cdot)$  は事前確率、 $q(\cdot)$  は変分事後分布を示す。ランドマーク機能を指示する単語の場合は  $W_T$  と  $F_{WT}$  をそれぞれ  $W_L$  と  $F_{WL}$  に置き換えればよい。

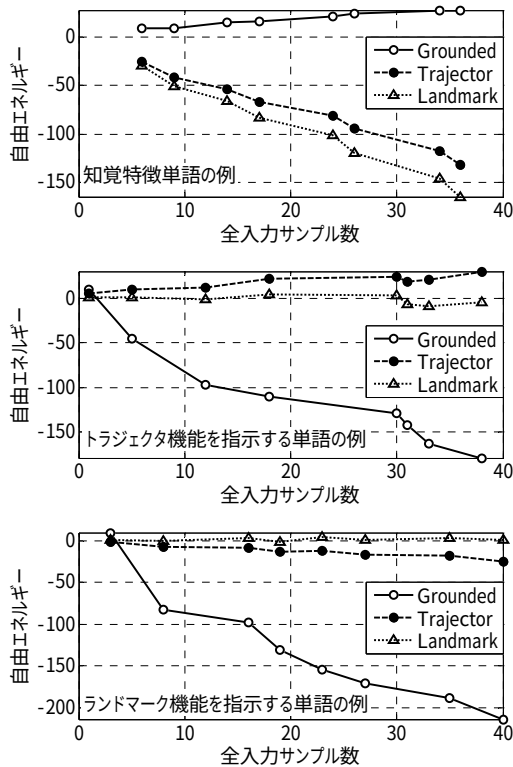


図2. オンライン学習

$$F[q(Z, \Theta)] = \left\langle \log p(X, F_{WT}, W_T | \Theta) \right\rangle_{q(F_{WT}, W_T)} - \left\langle \log q(F_{WT}, W_T) \right\rangle_{q(F_{WT}, W_T)} - KL(q(\Theta), p(\Theta))$$

### 3. 実験

学習サンプルに用いた画像特徴量は、1つのオブジェクトに対して照明や位置など複数の条件下で取得した大きなデータセットの中から任意に選択したものであり、ラベルは各単語に対して1つ割り当てた。学習サンプルはランダムに逐次与えるが、最終的に、1つの単語に対して学習サンプルは8サンプルとなるようにした。

例えば「はさみ」は、「何かを切ることが出来るもの」であると同時に「持ち上げることが出来るもの」でもあるように、あるオブジェクトが単一の機能しか持たないことは稀である。つまり、ある機能を指示する単語を学習する場合、そこで提示されるオブジェクトは複数の機能を持っているため、誤った語意を学習する可能性がある。学習する単語が増加するにつれこのような状況は起こるため、この実験では、個々のオブジェクトを複数の機能に対応させた状態で学習を行い、その結果について検討した。

まず、初期モデルとして、知覚特徴単語30個と動作を指示する単語4個(「のせる」「近づける」「飛び越えさせる」「持ち上げる」(この単語のみランドマークを持たない))を学習した。ここで、例えばオブジェクト1を「のせる」と「近づける」のトラジェクタに登録するというように、オブジェクトを複数の機能で共有した。新たに学習させる単語として、知覚特徴単語8個、トラジェクタ機能を指示する単語4個、ランドマーク機能を指

示する単語3個の、計15単語を学習させた。ここでトラジェクタ機能を指示する単語とは、4種類の動作とそれぞれ関連付けられたものであり、ランドマークも同様である。それぞれの単語を指示するサンプルを8サンプルずつ、計120サンプルを与えて学習を行った。この時、動作数を4としているので、それぞれの単語の解釈として作られる候補モデルの数は9個となった。

実験の結果、ほとんどの単語で正しく語意の学習が行われた。結果の一例を図2に示す。グラフ中の各点は、与えられたサンプルが全サンプル中で幾つ目なのかを表している。オンライン学習ではサンプルが与えられるとともに逐次学習を行うため、1サンプルが1つ与えられた時点では知覚特徴単語と学習し、その後サンプル数が増えるにつれて正しい解釈を行った。これは、与えられたオブジェクトが1つであった場合、その単語はそのオブジェクトを指す名前であると解釈することが最も自然であるためだと考えられる。また、誤って学習した例としては、オブジェクト1, 2に共に「のせる」と「近づける」というトラジェクタ機能を持たせて、この2つのオブジェクトの画像特徴量のみで「のせられるもの」を学習させようとした場合があった。この場合には、提示された単語が「のせられるもの」を指示する単語なのか「近づけられるもの」を指示する単語なのかの区別が出来なかった。しかし、これは人でも判断することが出来ない問題であると考えられるため、当然の結果であると考えられる。このような場合には、その機能として特徴的なオブジェクト、あるいはその機能に固有のオブジェクトというような、語意を判断できるオブジェクトの画像特徴量を与えることによって、誤って学習を行っていた場合にも正しく語意を再学習することが可能であった。また、オンライン学習はバッチ学習に比べ、正しく語意を学習するためには機能が重複していないオブジェクトを多少多く必要とする傾向があった。また、動作を指示する単語を4単語から変化させて、同様の実験を行った。特に動作数を増やした場合、必然的にオブジェクトの各機能間での共有は増加したが、上記の点に留意することで、正しく語意を学習させることができた。

### 5. まとめ

本稿では、知覚特徴単語および知覚特徴に基づく抽象的単語のオンライン学習に関して検討した。また、学習条件を様々な変化させることで学習アルゴリズムを評価した。本実験では特にオブジェクトに複数の機能を持たせた状態での学習について検討を行ったが、学習サンプルの全ての機能が同一であるような特殊な場合を除いて、正しく学習が行われることがわかった。しかし本実験では実世界に比べて動作数や学習する単語が少ないため、これは機能カテゴリーを指示する単語の学習が行われ易い状態であると考えられる。従って今後は動作数や登録するオブジェクトをさらに増やすなどして実験を行う必要がある。

**謝辞** 本研究は、国立情報学研究所共同研究「高次元環境知覚データにおける情報構造の発見的認識に関する研究」による研究助成を受け実施したものである。

### 参考文献

- [1] 岩橋, 佐藤, 麻生: バイジアンモデル選択に基づく知覚特徴量を用いた抽象的語意の学習, 信学技報, vol. 105, no. 673, PRMU2005-234, pp. 7-12, 2006年3月
- [2] Attias, H.: Inferring Parameters and Structure of Latent Variable Models by Variational Bayes, Proc. 5th conference on Uncertainty in Artificial intelligence, 1999