社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
PRMU2000-105 (2000-11)

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†

羽岡 哲郎[1] 岩橋 直人[2]

[1] 東京工業大学大学院情報理工学研究科 〒152-8552 東京都目黒区大岡山 2-12-1
[2] ソニーコンピュータサイエンス研究所 〒141-0022 東京都品川区東五反田 3-14-13
E-mail: {haoka, iwahashi}@csl.sony.co.jp

† 本研究は（株）ソニーコンピュータサイエンス研究所において行われた.

あらまし

日常的な環境において自然言語を用いて機械と自然で円滑なコミュニケーションを実現するためには，言語と実世界との関連性を重視した新しい言語情報処理手法を開発する必要がある．本報告では，参照点に依存した空間的移動の概念の学習法と，学習した概念に基づいて文法を獲得する手法について述べる．概念の学習では，参照点が非観測である動画像データから，「ちかづく」「はなれる」「のる」等の空間的移動の概念を表す確率モデルをEMアルゴリズムを用いて学習する．文法の学習では，動画像とそれを記述する文の関連付けにより，動画像中の複数の対象の間の関係に関する概念構造を抽出することで，文の構文構造を推定し原初的な文法を学習する．実験により良好な結果が得られることを示す.

キーワード 認知，言語，概念，文法，獲得，EMアルゴリズム

Learning of the reference-point-dependent concepts on movement for language acquisition††

Tetsuo Haoka[1] Naoto Iwahashi[2]

[1] Graduate School of Information Science and Engineering,
Tokyo Insitute of Technology 2-12-1 O-okayama, Meguro-ku, Tokyo, 152-8552 Japan
[2] Sony Computer Science Labs. Inc.
Takanawa Muse Bldg. 3-14-13 Higashi-Gotanda Shinagawa-ku, Tokyo, 141-0022 Japan
E-mail: {haoka, iwahashi}@csl.sony.co.jp

††This research was carried out at Sony Computer Science Labs. Inc.

Abstract

In order to realize the natural and smooth communication with machines using natural language in ordinary environment, the new approach of language processing which focuses on the relationship between language and real world is necessary. In this report, the learning method of the concepts dependent on reference points on movement, and the learning method of grammar based on the learned concepts, are described. In the concept learning, the probabilistic models of the concepts are learned by EM algorithm using the observation data of the dynamic scenes in which reference points are unobserved. In the grammar learning, by extracting the conceptual structures based on the association between the dynamic scenes and the sentences corresponding to them, syntactic structures of the sentences are inferred to acquire the primitive stochastic grammar. The validity of the methods is confirmed by experiments.

key words Cognition, Language, Concept, Grammar, Acquisition, EM-algorithm

1 はじめに

日常的な環境において自然言語を用いた機械との自然で円滑なコミュニケーションを実現するためには，シンボリックな処理に重点をおいた従来の自然言語処理のアプローチではなく，言語とアナログ的な実世界との関連性を重視した新しいアプローチが必要である．日常言語は，離散的で抽象的なシンボルの自律的な記号系として言語主体から独立して存在するものではなく，言語主体による外部世界の解釈を反映するアナログ的な経験のパターンから，カテゴリー化のプロセスを介して発現した抽象的な記号系であるとみなせる [1]．人間同士の日常的なコミュニケーションは，対話者間で共有されている認知的な経験を基盤として成立するものである．そのようなコミュニケーションを人間-機械間でも可能とするためには，人間-機械コミュニケーションにおける身体性を反映した認知的な経験に基づいて形成される言語知識の表現，獲得，およびその運用のための情報処理手法，いわば認知言語情報処理手法を確立する必要がある [2]．

このような研究として，音声と非言語知覚情報を結びつけることで，音韻，単語，および実世界に関する概念を確率モデルとして獲得するアルゴリズムが提案されている [3]．[3] ではオブジェクトとしての縫いぐるみと単語音声の組の集合が与えられて，視覚情報に基づいた個々のオブジェクトに付随する概念とそれに対応した単語が獲得された．一方，複数の単語を用いて情報を的確に伝達するためには，文法知識が人間-機械間で共有されていなければならない．文法は，複数の対象間の関係を文として表現する規則を含んでおり，そうした規則を実環境における経験から学習するには，複数の対象間の関係に関する概念が必要である．

本報告では，関係の概念として，参照点に依存したオブジェクトの空間的移動の概念を扱い，このような概念を表す確率モデルの学習手法，および，この確率モデルを用いた文法の学習手法について述べる．

本報告の構成は次のとおりである．まず，2 節で参照点に依存した空間的移動の概念の獲得についての問題点を明らかするとともに，その帰納的学習手法を提示する．3 節でこの学習手法の有効性を検証する実験結果を示す．最後に 4 節で文法の学習手法を述べ実験結果を示す．

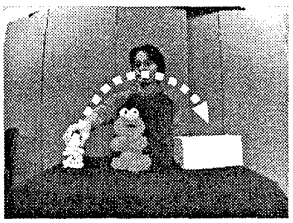

図 1: カメラ入力動画像．

2 参照点に依存した空間的移動の概念の獲得

2.1 参照点に依存した空間的移動の概念

図 1 に示す動画像中で，人間によって動かされている縫いぐるみの移動は，画面中央で静止している縫いぐるみを参照点にとれば「とびこえる」という概念の例であり，画面右側の箱を参照点にとれば「のる」という概念の例になる．このように空間的移動の概念には参照点に依存しているものがある．たとえば，認知言語学の基本的枠組みにおいては，認知される領域において焦点化される存在のうち，相対的により際立って認知されるオブジェクトはトラジェクター，これを背景的に位置付けるオブジェクトはランドマークとして区別することで，複数の存在の関係に基づく概念を記述している．この動画像の場合，移動している縫いぐるみがトラジェクターで，参照点として機能する静止している縫いぐるみや箱が「とびこえる」と「のる」の概念のそれぞれにおけるランドマークとみなせる．

ここで，このような参照点に依存する移動の概念を，移動の軌道に関する時空間上での確率モデルとして学習することを考える．このとき，適切な確率モデルを学習するためには，移動の軌道上の各点の位置を表す座標系を概念やその参照点に依存して設定することが必要であると考えられる．たとえば，「とびこえる」の概念は，参照点の位置を原点として座標軸が垂直および水平方向である直交座標系を用いれば適切に表現できると考えられる (図 2(1))．また，「ちかづく」の概念では，近づく対象となっているランドマークを参照点にして，これを原点としてランドマークからトラジェクターへ向かうベクトルの方向をひとつの座標軸の向きとした直交座標系を選ぶのが妥当であろう (図 2(2))．このように，参照点は概念および動画像に依存して変わるのに対して，座標軸の向きの設定方法は動画像に関わらず各概念に固有なものであると考えらる．また，そのような座標軸の向きの設定方法としては，ここであげたような基本的な方法が数種類存在し，そのなかから適切なものが選択されると考えられる．

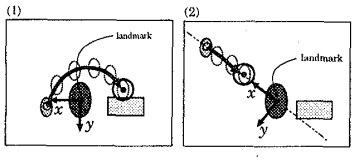

図 2: 座標系の設定方法: (1)「とびこえる」の概念の場合，(2)「ちかづく」の概念の場合．

しかしながら，学習データとして動画像のほかにこのような座標系に関する情報を与えるのは自然でない．したがって，概念の確率モデルは，その概念の確率モデルを学習す

るのに適切な座標系を探索しながら，学習されなければならない．

2.2 学習手法

移動するオブジェクトが一つで複数の静止オブジェクトが存在する動画像から抽出された特徴量データ集合 $V = \{V_1, V_2, \ldots, V_N\}$ が学習データとして与えられるものとする．ただし，各動画像 V_i は，画像中に仮に固定した2次元座標系における，移動オブジェクトの軌道点列 $\boldsymbol{X}_i = \boldsymbol{x}_1^i \boldsymbol{x}_2^i \cdots \boldsymbol{x}_{T_i}^i$，および，各静止オブジェクトの位置ベクトルの集合 $\boldsymbol{O}_i = \{\boldsymbol{o}_1^i, \boldsymbol{o}_2^i, \ldots, \boldsymbol{o}_{S_i}^i\}$ の組として，

$$V_i = (\boldsymbol{X}_i, \boldsymbol{O}_i) \tag{1}$$

と表記される．ここで，各静止オブジェクトと軌道開始点，画像中心点を参照点の候補とし，その集合を $\boldsymbol{L}_i = \{\boldsymbol{l}_1^i, \boldsymbol{l}_2^i, \ldots, \boldsymbol{l}_{M_i}^i\}$ とおく．

各 V_i に設定する座標系は2次元とし，参照点および軌道開始点の情報を用いて2つの座標軸の向きが設定され，参照点を原点とすることによって決定するものとする．ただし，可能な座標軸の向きの設定方法は K 種類であるとし，それぞれ $1, 2, \ldots, K$ とインデックスされているものとする．座標軸の向きの設定方法が概念に固有なものであるのに対して，参照点は各学習データ V_i に対して選択される．

ここで，参照点 $\boldsymbol{l}$ およびトラジェクタ軌道 $\boldsymbol{X}$ と座標軸の向き設定 k によって決定される座標系でのトラジェクタ軌道を，$F(\boldsymbol{X}, k, \boldsymbol{l})$ と表記する．

この設定のもと，最適な座標軸の設定方法 k および参照点を探索しながら，軌道に関する確率モデルのパラメータ λ を尤度最大化基準により学習する．つまり，

$$(\tilde{\lambda}, \tilde{k}, \tilde{\boldsymbol{m}}) = \arg\max_{\lambda, k, \boldsymbol{m}} \sum_{i=1}^{N} \log P(F(\boldsymbol{X}_i, k, \boldsymbol{l}_{m_i}^i); \lambda). \tag{2}$$

ここで，m_i は参照点 $\boldsymbol{l}_{m_i}^i$ を選択することを示す．また $\boldsymbol{m} = (m_1, m_2, \ldots, m_N)$ とおいた．

一般的に，座標軸の向きの設定 k の選択肢の個数 K はそれほど多くないと考えられるのに対して，$\boldsymbol{m}$ の組合せは非常に多く，$M_1 \times M_2 \times \cdots \times M_N$ 通りである．このように k に関する最適化と $\boldsymbol{m}$ に関する最適化では，その難しさが大きく異なる．そこで，まず k を固定して，

$$(\tilde{\lambda}_k, \tilde{\boldsymbol{m}}_k) = \arg\max_{\lambda, \boldsymbol{m}} \sum_{i=1}^{N} \log P(F(X_i, k, l_{m_i}^i); \lambda) \tag{3}$$

という問題を解くことを考える．しかし，$\boldsymbol{m}$ の組合せの多さを考慮すると，これを解くことは現実的ではない．そこで，この $\boldsymbol{m}$ に関する制約条件を緩めて問題を解きやすくするために，各 V_i に対して，l_m^i が参照点として選択される重みを w_m^i とおいて，(3) 式の離散最適化問題を次の連続最適化問題

$$(\tilde{\lambda}_k, \tilde{\boldsymbol{w}}_k) = \arg\max_{\lambda, \boldsymbol{w}} \sum_{i=1}^{N} \log \left[\sum_{m=1}^{M_i} w_m^i P(F(X_i, k, l_m^i); \lambda) \right] \tag{4}$$

で近似する．ただし，

$$\sum_{m=1}^{M_i} w_m^i = 1 \qquad (i = 1, 2, \ldots, N) \tag{5}$$

であり，$\boldsymbol{w}_i = (w_1^i, w_2^i, \ldots, w_{M_i}^i)$，$\boldsymbol{w} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_N\}$ とした．こうすることによって，参照点の選択 $\boldsymbol{m}$ は隠れ変数とすることができ，(4) 式は EM アルゴリズム [4] を適用して効率良く解くことができる．特に，確率モデルとして隠れマルコフモデル (HMM) を用いた場合の解法を付録 A で解説する．

このように，k を固定した上で $\tilde{\lambda}_k, \tilde{\boldsymbol{w}}_k$ を求め，最終的に次式によって求まる k と $\tilde{\lambda}_k$ が求める確率モデルのパラメータである

$$\sum_{i=1}^{N} \log \left[\sum_{m=1}^{M_i} \tilde{w}_{km}^i P(F(X_i, k, l_m^i); \tilde{\lambda}_k) \right] \to \max. \tag{6}$$

3 空間的概念の学習実験

3.1 実験条件

動画像はカメラ入力によるカラー画像を使用する．この動画像に，オブジェクトの色を基にした追跡アルゴリズムを適用することによって，オブジェクトの位置を得る．

学習する概念は，

{ あがる, ちかづく, はなれる, まわる,
ならぶ, のる, おりる, さがる, とびこえる }

の計9個とした．それぞれの概念につき 20 個の動画像を撮影し，そのうち 15 個学習データ，5 個をモデルの検証用オープンデータとした．

確率モデルは隠れマルコフモデル (HMM) を用いる．HMM の状態数は cross validation によって決定する．

座標軸の向きの設定方法の種類は，学習する概念に対して必要であると想定される次の 4 つとした．

1. ランドマークから軌道開始点方向の水平な軸と，垂直方向の軸をとる方法．
2. ランドマークから軌道開始点方向に向ける軸と，それと直交する軸をとる方法．
3. 軌道開始点を参照点として水平軸，垂直軸をとる方法．

4. 画面中心を参照点とするからの水平軸，垂直軸をとる方法．

それぞれ 1:ランドマーク・垂直型，2:ランドマーク・トラジェクタ方向型，3:トラジェクタ・垂直型，4:正面・垂直型と呼ことにする．トラジェクタ・垂直型と正面・垂直型は参照点の選択を必要としない．

3.2 空間的概念の学習実験結果

図 3(2) のように，全体の尤度が上昇するなかで，一つの学習データにおける，ランドマーク (参照点) 選択の例を図 3(1) に示す (横軸:パラメータ再推定回数，縦軸:ランドマーク選択，重みを黒い正方形の大きさで示す)．重みパラメータは，1 または 0 に収束した．この例は収束に比較的時間がかかった数少ない例である．他の多くのデータでは 2 回の再推定で収束した．この結果から，参照点を選択しながら尤度を高めることに成功したことがわかる．

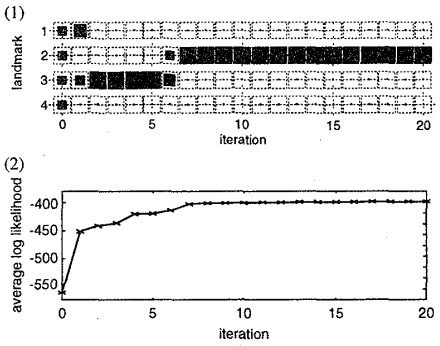

図 3: 「のる」の学習過程における，(1) 尤度の変化，(2) 一画像データにおける参照点選択重みの変化の例．

次に，各概念における座標軸の設定方法の選択結果を図 4 に示す．それぞれの概念に対する 4 本のグラフは，座標軸設定方法の選択に関する対数尤度を示していおり，そのうち最も尤度の高いものが選択されたことを示している．「あがる」，「まわる」，「さがる」では “トラジェクタ・垂直型” が，「ちかづく」，「はなれる」，「ならぶ」では “ランドマーク・トラジェクタ方向型” が，「のる」，「おりる」，「とびこえる」では “ランドマーク・垂直型” が，それぞれ選択された．このように，全体として適切であると考える座標系の設定方法が選択された．ただし，「おりる」では座標軸設定方法の選択に関する尤度の差が他のものと比較して小さい．このことは用意した 4 つの座標軸設定方法のなかに「おりる」という概念に適したものがなかったことを示していると考えられる．

図 5 に学習データの一部 (各概念につき 5 個) と，学習の結果選択された座標系を表す矢印で表示する．「のる」では台となるオブジェクトが参照点 (ランドマーク) として

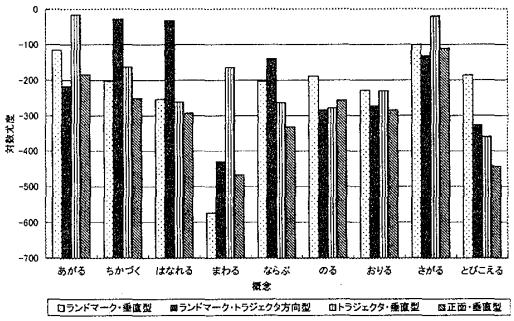

図 4: 座標軸設定方法の選択に対する尤度．

選択され，「とびこえる」では飛び越えられるオブジェクトが参照点として選択された．その他を見ても，ほぼ適切なランドマークが選択されたことがわかる．ただし，「おりる」に関しては，適切な座標系が選択されたとは言えず，実際にどのような座標系が適切なのか推定することも困難であった．

3.3 概念確率モデルの運用実験

適切なモデルとなっているかどうかを確認するために，ランドマーク認識，空間的概念の認識，軌道の生成の各実験を行なった．

ランドマーク認識実験 ランドマークを有する空間的概念について，概念と動画像データが与えられ，尤度を最大にする参照点となるランドマークを求める実験を行なった．

実験には，各概念ごとに 5 個 のオープンデータをテストデータ (図 6) として使用した．図 6 のテストデータ上に，適切であると想定したランドマークを白抜き矢印，認識の結果選択されたランドマークを黒点で示す．

ランドマークの認識の失敗 (認識結果が想定したものと異なった場合) は 25 (5 概念 ×5 データ) 回のテストで 1 つのみ (「はなれる」の図左端のデータ) であった．

空間的概念の認識実験 動画像データが与えられたときに，尤度が高い概念モデルと参照点を上位 5 位まで求める実験を行なった．

図 7 に示す 3 種類のオープンデータ (data(1-3)) に関する認識結果を表 1 に示す．ただし，項目 “LM” は，ランドマークとして選択されたオブジェクトの番号 (“—” はランドマークが必要ないことを示す) である．

data(1) では，トラジェクタが下がっているので，「さがる」が 1 位で選ばれた結果は適切である．同時にトラジェクタはオブジェクト 2, 3 に近づいてもいるので，2，3 位

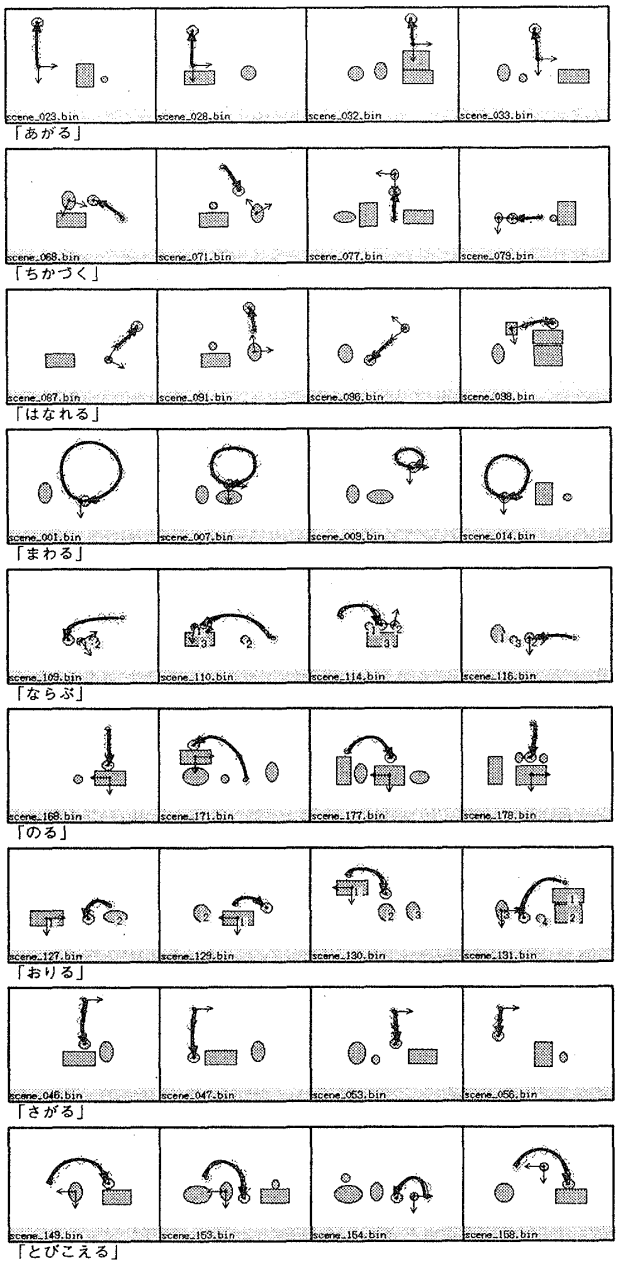

図 5: 学習データの一部と学習の結果選択された座標系.

に「ちかづく (LM: 2)」「ちかづく (LM:3)」が選択された.

data(2) は,「とびこえる」または「のる」という概念が選択された. オブジェクト 1, 5 をランドマークとするならば「とびこえる」, オブジェクト 3 をランドマークとするならば「のる」という概念となるので, これらの結果は適切であるといえる.

data(3) では「まわる」が 1 位で選ばれた. 軌道が円を描いているので適切な結果であるといえる. また, 1 位との尤度差が大きいが, 2 位に「のる」が選ばれた. これは, 軌道の終点の下にオブジェクトがあるためで, 正しい選択であるといえる.

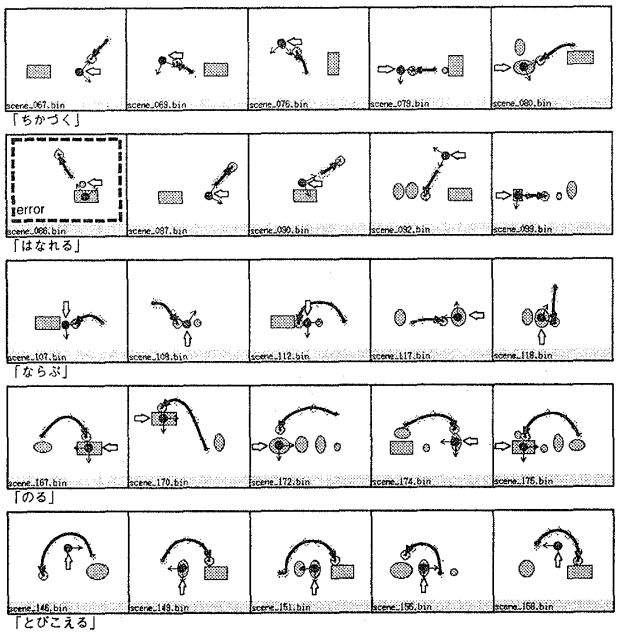

図 6: ランドマーク認識結果

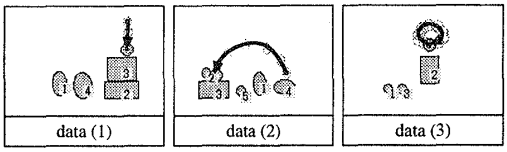

図 7: 空間的概念の認識テストデータ

軌道の生成 トラジェクタ初期位置と参照点, および, 概念モデルが与えられるときに, 軌道を生成する実験を行なった. 軌道の生成は, 軌道開始点 $\boldsymbol{x}_0$ の制約条件つき尤度 $P(F(\boldsymbol{X}, k, \boldsymbol{l})|\boldsymbol{x}_0, \lambda)$ を最大にする軌道 $\boldsymbol{X}$ を [5] による最適化法を用いて求めることによって行なった.

図 8 に,「とびこえる」の軌道を求めた結果を示す. 図上段が軌道を求める前の静止オブジェクトの位置を示している. 図下段の 3 つの情景が, 実験の結果求まった軌道を加

data(1)

順位	概念	LM	尤度
1	さがる	—	-48.76
2	ちかづく	2	-75.12
3	ちかづく	3	-84.39
4	のる	3	-290.53
5	のる	2	-408.07

data(2)

順位	概念	LM	尤度
1	とびこえる	1	-467.19
2	のる	3	-492.86
3	とびこえる	5	-688.78
4	のる	2	-1036.24
5	のる	5	-1394.15

data(3)

順位	概念	LM	尤度
1	まわる	—	-196.49
2	のる	2	-1412.09
3	はなれる	2	-1502.15
4	はなれる	3	-1566.32
5	はなれる	1	-1697.07

表 1: 動画像データの認識結果

えたものである．3つの異なるトラジェクタとランドマークの選び方それぞれに対して適切な軌道が生成された．

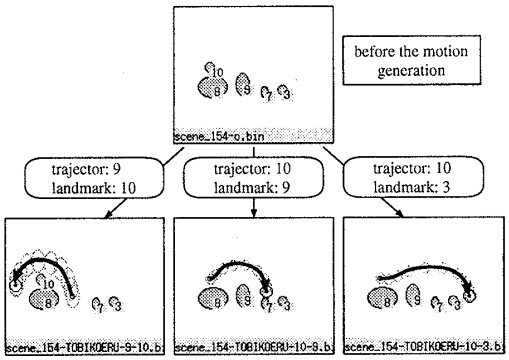

図 8: 「とびこえる」軌道生成の結果．

これらの結果から，適切なモデルとなっていると言える．

4 文法の獲得

4.1 概念構造に基づく文法の学習

前節において，参照点に依存した概念の確率モデルを用いることで，動画像からオブジェクトの動きに関するトラジェクタとランドマークの関係であらわされる概念構造を抽出できることが示された．複数の対象の関係を文として記述するときの規則を表す文法は，このような概念構造を反映している．ここでは，動画像とそれを記述する文の組の集合を学習データとし，文の構文構造を，参照点に依存した概念の確率モデルを用いて動画像から認識した概念構造に基づいて推定しながら，文法を学習する手法について述べる．

ここで扱う文は，空間的移動の概念を表す句 A，トラジェクタとなるオブジェクトを記述する句 W，ランドマークとなるオブジェクトを記述する句 L から構成されるものとする．たとえば，図1の動画像に対して，「のる おおきい カーミット 茶色 箱」という文が与えられる．ここで，A は「のる」，T は「おおきい カーミット」，L は「茶色 箱」である．なお，空間的移動の概念を表す句は3節で学習した概念を表す一単語（「のる」など）からなり，「おおきい」や「カーミット」等の単語で表されるオブジェクトの静的概念の確率モデルは別途与えられるものとする．そして，学習すべき文法はこれらの句の順序を規定するものとする．

4.2 学習方法

文法を G は，句 A,T,L のならび B_j を確率 $P(B_j)$ で生成するものとする．ここで，可能な句のならび全ての集合を $\{B_j\}_{j=1}^{N_B}$ とおく．

動画像特徴量データ V_i に対応して，文法 G に従う文を構成する単語列 W_i が学習データとして与えられるものとする．ここで，V_i は，トラジェクタ軌道 $\boldsymbol{X}_i$，トラジェクタのオブジェクト特徴量 $\boldsymbol{x}_i$，静止オブジェクトの特徴量の集合 $\boldsymbol{O}_i = \{\boldsymbol{o}_1^i, \boldsymbol{o}_1^i, \ldots, \boldsymbol{o}_{M_i}^i\}$ によって，

$$V_i = \{\boldsymbol{X}_i, \boldsymbol{x}_i, \boldsymbol{O}_i\} \tag{7}$$

と表記される．

次式で示す尤度最大化基準に基づき文法 G を求める

$$\sum_{i=1}^{N} \log P(V_i|W_i, G) \rightarrow \max. \tag{8}$$

ここで，

$$P(V_i|W_i, G) \equiv \max_{m,C_i} P(\boldsymbol{X}_i|\boldsymbol{o}_m^i, H_A(C_i))P(\boldsymbol{x}_i|H_T(C_i)) \times P(\boldsymbol{o}_m^i|H_L(C_i))P(C_i|W_i, G) \tag{9}$$

と表せる．ここで，

$$C_i = [S, (U_1, W_i^{(1)}), (U_2, W_i^{(2)}), (U_3, W_i^{(3)})] \tag{10}$$

は，文 W_i の句への分割を表しており，U_1, U_2, U_3 は句の属性 $(A, T, L$ のいずれか$)$ を表す．また，$H_A(C_i)$ は，C_i 中の句 A の単語列を表し，同様に，$H_T(C_i), H_L(C_i)$ は，それぞれ C_i 中の句 T, L の単語列を表す．$P(\boldsymbol{X}_i|\boldsymbol{o}_m^i, H_A(C_i))$ の値は，単語列 $H_A(C_i)$ が表す空間的概念の確率モデルの尤度として計算することができ，$P(\boldsymbol{x}_i|H_T(C_i))$ と $P(\boldsymbol{o}_m^i|H_L(C_i))$ の値は，静的概念の確率モデルを用いて計算することができる．よって，

$$(\tilde{m}_i, \tilde{C}_i) = \arg\max_{m,C_i} P(\boldsymbol{X}_i|\boldsymbol{o}_m^i, H_A(C_i))P(\boldsymbol{x}_i|H_T(C_i)) \times P(\boldsymbol{o}_m^i|H_L(C_i))P(C_i|W_i, G) \tag{11}$$

を求めることにより，(8) 式は，

$$\sum_{i=1}^{N} \left[\log P(V_i|\tilde{C}_i) + \log P(\tilde{C}_i|W_i, G)\right] \tag{12}$$

となる．第一項は定数となるので，第二項をに関して句のならびの生成確率値の最尤推定値は，

$$P(B_j) = \frac{(\tilde{C}_i \text{ の句のならびが } B_j \text{ である場合の数})}{(\text{学習データ数: } N)} \tag{13}$$

となる．

ここで，単語 w が表す静的概念の確率モデルのパラメータを λ_w とするとき，単語列 $W = w_1 w_2 \cdots w_N$ がオブジェクト特徴量 $\boldsymbol{o}$ を記述する尤度 $P(\boldsymbol{o}|W)$ は，

$$P(\boldsymbol{o}|W) = \prod_{n=1}^{N} P(\boldsymbol{o}|\lambda_{w_n}) \tag{14}$$

のように各単語に対する尤度の積として計算する．

4.3 句順序生成確率の学習実験

4.2 節で述べた学習手法により，句順序の生成確率を学習する実験を行なった．

単語列生成の際に用いた文法は，「$(A\,T\,L)$ を基本形とし，T，または L，あるいは両方の省略を許す」というものである．72 個の動画像データに対して，この文法に基づいた文を構成する単語列を生成し，学習データとした．句 T, L 内の単語数は 0〜2 個とした (0 個は句の省略を示す)．文の平均単語数は 3.5 であった．

オブジェクト特徴量としては，色と大きさおよび形状を表す計 6 次元のものを使用し，そのメンバーシップ関数としてガウス分布を用いた．

また，この実験では，4.2 節(11) 式による句構造 C_i の推定を簡単にするために，次式を用いた

$$\arg\max_{m,C_i} P(\boldsymbol{X}_i|\boldsymbol{o}^i_m, H_A(C_i))P(\boldsymbol{x}_i|H_T(C_i))P(\boldsymbol{o}^i_m|H_L(C_i)). \tag{15}$$

表 2 に，各句順序が推定された回数と生成確率の推定結果を示す．ただし，括弧内は学習データを生成したときに実際に使用した句の順序を数えあげた数字である．

このように，学習データの実際の値と比較しても，良い推定結果が得られていることがわかる．

句順序	回数	確率
$S \to A\,T\,L$	27 (33)	0.38 (0.46)
$S \to A\,L$	18 (18)	0.25 (0.25)
$S \to A$	12 (12)	0.17 (0.17)
$S \to A\,T$	12 (9)	0.17 (0.13)
$S \to A\,L\,T$	3 (0)	0.04 (0.00)
その他	0 (0)	0.00 (0.00)
計	72	–

表 2: 句順序の生成確率の推定結果 (括弧内は正解値)．

5 考察

概念や文法等の人間の認知にかかわる知識を獲得する方法においては，概念の座標系の設定や文法の制約等の認知に関する知識に基づいくモデルの構成法と，そのパラメータをデータを用いて学習する情報理論的手法を，効果的に組み合わすことで必要である．参照点に依存した概念の学習手法では，座標系の選択という認知のモデルと，座標系の選択が非観測であるときの学習を可能とした情報理論的学習手法が，効果的に組み合わされて良好な結果が得られたといえる．実験結果はおおむね良好であったものの，一部，たとえば，「おろす」の概念の獲得等については問題が残った．実験によって明らかになったそのような問題については，モデルの構成法および数理的学習法の両面からの改良に取り組む必要がある．

参照点が非観測である条件での概念学習における最適化アルゴリズムは，非観測である情報選択を隠れ変数としてみなすという学習原理に基づいている．この原理は，与えられている多くの情報から適切に情報を選択しながらモデルを学習することが求められる，より一般的な学習問題にも適用できるであろう．

文法獲得では，実世界に関する概念を用いることにより，文だけからは得ることが難しい構文情報が推定され文法の獲得に利用された．こうした学習のメカニズムは，人間が言語を獲得する際にも，少量の文の提示だけで文法を獲得できることに寄与していると考えられている [6]．また，このようなメカニズムで機械によって獲得された文法は，実世界での運用において優れたものになるであろうと考えている．

6 まとめ

関係の概念のひとつとして参照点に依存した空間的移動の概念を学習し，関係の概念によるカテゴリ化を介して認知的経験を反映した原初的な文法を獲得する方法を示した．今後は，さまざまな関係に関する概念の学習，および，それらの概念と文法獲得との関わりについて研究していく予定である．

A 参照点推定を伴う概念モデルの学習

はじめに，EM アルゴリズムを適用するときに自然になるように，2.2 節で用いた表記を変更する．2.2 節の(3) 式右辺に含まれている確率モデルの尤度 $P(F(\boldsymbol{X}_i, k, \boldsymbol{l}^i_{m_i}); \lambda)$ の値は，k, m_i, λ が決まった上での軌道 $\boldsymbol{X}_i$ の確率尺度であると言える．つまり，k, m_i, λ の条件付きの確率として，

$$P(F(\boldsymbol{X}_i, k, \boldsymbol{l}^i_{m_i}); \lambda) = P(\boldsymbol{X}_i|m_i, \boldsymbol{L}_i, k, \lambda) \tag{16}$$

と表記することができる．$\boldsymbol{L}_i$ も条件の側に入っている理由は，参照点の選択 m_i は，選択肢 $\boldsymbol{L}_i$ そのものがなければ意味を持たないことを考慮したからである．同様にこのことを考慮すると，参照点選択 m_i の確率尺度は，$P(m_i|\boldsymbol{L}_i, \boldsymbol{w})$ と表記でき，その値はパラメータ $w^i_{m_i}$ そのものである．このようにすると，(16) に $w^i_{m_i}$ が掛かったものを

$$\begin{aligned} & w^i_{m_i} P(F(\boldsymbol{X}_i, k, \boldsymbol{l}^i_{m_i}); \lambda) \\ &= P(m_i|\boldsymbol{L}_i, \boldsymbol{w}) P(\boldsymbol{X}_i|m_i, \boldsymbol{L}_i, k, \lambda) \\ &= P(\boldsymbol{X}_i, m_i|\boldsymbol{L}_i, k, \lambda, \boldsymbol{w}) \end{aligned} \tag{17}$$

と表記しなおすことができる．ここで，表記を簡単にするために，パラメータを一つにまとめて $\Lambda = (\lambda, \boldsymbol{w})$ とおく．

以上の表記を用いると，2.2 節(4) 式は，

$$\begin{aligned}\tilde{\Lambda} &= \arg\max_{\Lambda} \sum_{i=1}^{N} \log \left\{ \sum_{m=1}^{M_i} P(\boldsymbol{X}_i, m | \boldsymbol{L}_i, k, \Lambda) \right\} \\ &= \arg\max_{\Lambda} \sum_{i=1}^{N} \log P(\boldsymbol{X}_i | \boldsymbol{L}_i, k, \Lambda) \end{aligned} \quad (18)$$

となり，これが解くべきものである．

ここで，確率モデルとして HMM を用いるとする．参照点の選択 $\boldsymbol{m} = (m_1, m_2, \ldots, m_N)$ と HMM の状態遷移 q_i を隠れ変数として，EM アルゴリズムを適用すると，(18) 式の最適化問題は，

$$\begin{aligned} & Q(\Lambda, \Lambda') \\ &= \sum_{i=1}^{N} \sum_{m_i=1}^{M_i} \sum_{\boldsymbol{q}_i} P(m_i, \boldsymbol{q}_i | \boldsymbol{X}_i, \boldsymbol{L}_i, k, \Lambda') \\ &\qquad \times \log P(\boldsymbol{X}_i, m_i, \boldsymbol{q}_i | \boldsymbol{L}_i, k, \Lambda) \end{aligned} \quad (19)$$

を補助関数とするパラメータ再推定 $(\Lambda' \to \Lambda)$ 問題となる．ここで，簡単のために，$\Xi_{im_i} = F(X_i, k, l^i_{m_i})$ とおく．ただし，$\Xi_{im_i} = \boldsymbol{\xi}_1^{im_i}, \boldsymbol{\xi}_2^{im_i}, \ldots, \boldsymbol{\xi}_{T_i}^{im_i}$ である．すると，

$$\begin{aligned} & \log P(\boldsymbol{X}_i, m_i, \boldsymbol{q}_i | \boldsymbol{L}_i, k, \Lambda) \\ &= \log w^i_{m_i} + \log \pi_{q^i_0} + \sum_{t=1}^{T_i - 1} \log a_{q^i_t q^i_{t+1}} \\ &\quad + \sum_{t=1}^{T_i - 1} \log b_{q^i_{t+1}}(\boldsymbol{\xi}_t^{im_i}) \end{aligned} \quad (20)$$

と，

$$P(m_i, \boldsymbol{q}_i | \boldsymbol{X}_i, \boldsymbol{L}_i, k, \Lambda') = \frac{(w^i_{m_i})'}{P_i} P(\Xi_{im_i}, \boldsymbol{q}_i | \lambda') \quad (21)$$

となる．ここで，

$$P_i = P(\boldsymbol{X}_i | \boldsymbol{L}_i, k, \Lambda'),$$

とおいた．また，(a_{ij}) は HMM の状態遷移確率行列，$b_j(\boldsymbol{\xi})$ は状態 j での $\boldsymbol{\xi}$ の出力確率密度関数である．

よって，(19) 式から次のように最適なパラメータ再推定式が導出される．

$$\bar{w}^n_m = \frac{w^n_m}{P_n} \sum_{n=1}^{N} \alpha^{nm}_{T_n}(), \quad (22)$$

$$\bar{\pi}_i = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{P_n} \sum_{m=1}^{M_n} w^n_m \alpha^{nm}_0(i) \beta^{nm}_0(i), \quad (23)$$

$$\bar{a}_{ij} = \frac{\sum_{n=1}^{N} \frac{1}{P_n} \sum_{m=1}^{M_n} w^n_m \sum_{t=1}^{T_n - 1} \alpha^{nm}_t(i) a_{ij} b_j(\boldsymbol{\xi}^{nm}_{t+1}) \beta^{nm}_{t+1}(j)}{\sum_{n=1}^{N} \frac{1}{P_n} \sum_{m=1}^{M_n} w^n_m \sum_{t=1}^{T_n - 1} \alpha^{nm}_t(i) \beta^{nm}_t(j)}. \quad (24)$$

ただし，α^{nm}, β^{nm} は，それぞれ特徴量時系列 $\boldsymbol{\xi}_1^{nm} \cdots \boldsymbol{\xi}_{T_n}^{nm}$ に関する HMM(λ) の前向き変数，後ろ向き変数である．

出力関数 b_j が単一ガウス分布の場合には，そのパラメータ (平均 $\boldsymbol{\mu}_j$, 共分散行列 $\boldsymbol{\Sigma}_j$) 再推定式は次のようになる．

$$\bar{\boldsymbol{\mu}}_j = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M_n} \sum_{t=1}^{T_n} \gamma^{nm}_t(j) \boldsymbol{\xi}^{nm}_t}{\sum_{n=1}^{N} \sum_{m=1}^{M_n} \sum_{t=1}^{T_n} \gamma^{nm}_t(j)}, \quad (25)$$

$$\bar{\boldsymbol{\Sigma}}_j = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M_n} \sum_{t=1}^{T_n} \gamma^{nm}_t(j) (\boldsymbol{\xi}^{nm}_t - \boldsymbol{\mu}_j)(\boldsymbol{\xi}^{nm}_t - \boldsymbol{\mu}_j)'}{\sum_{n=1}^{N} \sum_{m=1}^{M_n} \sum_{t=1}^{T_n} \gamma^{nm}_t(j)} \quad (26)$$

ただし，

$$\gamma^{nm}_t(j) = \frac{w^n_m}{P_n} \alpha^{nm}_t(j) \beta^{nm}_t(j). \quad (27)$$

である．

参考文献

[1] 山梨正明，“認知言語学原理”，くろしお出版, 2000.

[2] Iwahashi. N., “Language Acquisition through a Human-Robot Communication”, Proc. Int. Conf. Spoken Language Processing, 2000.

[3] 金景柱，岩橋直人，“知覚情報の統合に基づく言語音声単位の獲得アルゴリズム”，信学技報 TL200-21(2000-10), pp9-16.

[4] Dempster, A. P., Laird, N. M., and Rubin, D. B., “Maximum likelihood from incomplete data vie the EM algorithm”, Jornal of Royal Statistocal Society B **39** (1977) 1–38.

[5] Tokuda, K., Kobayashi, T., & Imai, S., “Speech paramter generation from HMM using dynamic features”, Proc. of International Conference on Acoustics, Speech and Signal Processing (1995) 660–663.

[6] Pinker, S., “Learnability and cognition”, Harvard University Press, 1989.