# 映像に基づく人物行動の文法学習

木谷 クリス 真実† 佐藤 洋一† 杉本 晃宏††

† 東京大学 生産技術研究所 〒 153-8505 東京都目黒区駒場 4-6-1
†† 国立情報学研究所 〒 101-8430 東京都千代田区一ツ橋 2-1-2
E-mail: †{kitani,ysato}@iis.u-tokyo.ac.jp, ††sugimoto@nii.ac.jp

**あらまし** 自然言語の構文解析に用いられている確率文脈自由文法は, 映像による人物の行動解析にも使われており, その有効性が報告されている. しかしながら, 文の単語列と異なり, 映像から得られる人物行動の記号列には多くのノイズが含まれているため, 行動文法の学習が困難になる. 従って, 高精度の文法学習を行うためには, ノイズ記号を除外した終端記号集合を特定する必要がある. そこで本研究では, 最小記述長原理にもとづき, ノイズを除外した終端記号集合とそれに伴う文法の獲得手法を提案する. 提案手法では, 終端記号の全組合せを評価し, 各々の部分集合の下で得られた文法の複雑さと観測データの記号列尤度とのトレードオフを定量化する. これにより, 評価値の高い終端記号集合と文法の候補を特定することができ, 記号列に含まれるノイズを除去しつつ行動文法の基本構造を獲得することが可能となる. 実験により, 提案手法の有効性を示す.

**キーワード** 文法学習，構文解析，最小記述長原理，行動認識

# Learning the Grammar of Human Activity from Video

Kris M. KITANI†, Yoichi SATO†, and Akihiro SUGIMOTO††

† The University of Tokyo, Institute of Industrial Science, 4-6-1- Komaba, Meguro, Tokyo 153-8505 JAPAN
†† National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo 101-8430 JAPAN
E-mail: †{kitani,ysato}@iis.u-tokyo.ac.jp, ††sugimoto@nii.ac.jp

**Abstract** Stochastic Context-Free Grammars (SCFG) have been shown to be useful for applications beyond natural language analysis, specifically vision-based human activity analysis. Vision-based symbol strings differ from natural language strings, in that a string of symbols produced by video often times contains noise symbols, making grammatical inference very difficult. In order to obtain reliable results from grammatical inference, it is necessary to identify these noise symbols. In our work, we propose a new technique for identifying the best subset of non-noise terminal symbols and acquiring the best activity grammar. Our approach uses the Minimum Description Length (MDL) principle, to evaluate the trade-offs between model complexity and data fit to quantify the difference between the results of each terminal subset. The evaluation results are then used to identify of a class of candidate terminal subsets and grammars that remove the noise and enable the discovery of the basic structure of an activity. In this paper, we present the validity of our proposed method based on experiments with real data.

**Key words** Grammatical Inference, Syntactic Analysis, Minimum Description Length Principle, Action Recognition

## 1. はじめに

文脈自由文法 (Context-Free Grammar, 以下 CFG) は自然言語のモデルとして広く知られているが, 人物の行動解析にも有用である. 近年, 映像による人物の行動解析に CFG が適用され, その有効性が立証されている ( [1], [2], [3], [4], [5]).

CFG は記号集合と生成規則で形成され, 与えられた文 (単語列) の文法的な構造を階層的に表現できることから, 自然言語のモデルとして相応しい. 一方, 知覚心理学の観点から, 人物行動は階層構造を持つ行動の記号列として解釈できるとされていることから [6], CFG は人物行動を表すモデルとしても相応しい. 人物の行動解析には隠れマルコフモデルや有限状態機械といった一次構造を持つモデルも提案されているが, 階層構造を持つ CFG の方が人物行動を表すモデルとして適しているといえる.

CFG の応用にあたって, 重要な課題の一つとして, 文法の定義があげられる. これまでの研究は主に文法を事前に与え, その文法の応用に着目しているが, 文法の学習には触れていない. 例えば, Ivanov ら [1] は映像から人物の単位行動を HMM で出力し, 確率文脈自由文法 (Stochastic Context-Free Grammar, 以下 SCFG) で指揮者の手ぶりから拍子を認識しているが, そこでは文法は手で与えられている. Moore ら [2] は SCFG を用いて Black Jack のゲームを認識し, さらに文法の事前知識を利用して, 単発的な誤り (記号挿入, 置換と削除) に対処したが, 文法は予め決められていた. 同様, Minnen ら [3] は事前に定義した SCFG から得られる事前知識を利用し, 人物行動の一時的な遮蔽に対応した行動認識システムを実装している.

一方, 映像から得られた人物行動の記号列から文法を学習する研究例は少なく, 自然言語解析で使用されている既存のアルゴリズムをそのまま適用することにとどまっている. その例として Wang ら [7] は HMM の出力を単位行動として表し, その行動を記号化し, COMPRESSIVE [8](CFG の学習アルゴリズム) を用いて文法を学習するという手法を提案した. 但し, 彼らの手法は低次の画像処理結果の精度に完全に頼っているため, 画像処理の部分で問題が発生した場合, 文法の学習精度が落ちるという問題を抱えている. つまり, COMPRESSIVE は, 与えられた記号列には誤りが存在しないことを前提にしているため, 画像処理結果から出力される確率的な (ノイズのある) 記号列には対応していない.

これに対して, 本研究ではノイズを考慮した行動文法の獲得手法を提案する. 以下, 最初に提案手法の基本的なアイデアについて述べたのちに, 提案手法の詳細を説明する. 次に提案手法を用いた実験結果を報告し, 結びを述べる.

## 2. 基本アイデア

いま, 人物行動を表す記号列 $S$ が与えられたとする. 我々の目標は $S$ の構造を説明する, 簡潔かつ曖昧性の少ない, 文法を求めることである (a や b は単位行動を表す記号).

$$\mathtt{S} \rightarrow \mathtt{a\ x\ b\ y\ c\ a\ b\ x\ c\ y\ a\ b\ c\ x}$$

文法を学習するためには, 記号列から, 何らかの規則を持つパターンを探す必要がある. 但し, 記号列にはノイズ記号も含まれているため, パターンを探すのは容易ではない. 実際, 上記の $S$ の場合も, ノイズの影響で明らかな規則性が見えない.

例えば, ノイズ記号がパターンを隠していると想定すると, ノイズ記号を削除すれば, パターンが見えてくると考えられる. そこで, この記号列に含まれているノイズを発見するためには, $S$ の中から任意に記号を削除し, 規則性が発見されるかどうかを試してみればよい. この例の場合, 任意に $S$ の中から y をノイズ記号と仮定する. そして, その仮定のうえで記号 y を省略すると, 以下のような記号列が得られる.

$$\mathtt{S} \rightarrow \mathtt{a\ x\ b\quad c\ a\ b\quad x\quad c\ a\ b\quad c\ x}$$

$S$ には部分記号列 c a b が 2 回繰返されているが, これではまだ明らかな規則性が見えていない.

さらに, もう一つの記号 x をノイズ記号と仮定すると, 今度は記号列が明らかなパターン (基本構造) で構成されていることが見えてくる.

$$\mathtt{S} \rightarrow \mathtt{a\ b\ c\quad a\ b\ c\quad a\ b\ c}$$

すなわち, 記号列は部分記号列 a b c を 3 回繰返したものである. 新たな規則 $A$ で $S$ を置き換えると, より少ない記号数で元の記号列の基本的な構造を表すことができる.

$$\mathtt{S} \rightarrow \mathtt{A\ A\ A}$$
$$\mathtt{A} \rightarrow \mathtt{a\ b\ c}$$

この例では, x と y をノイズ記号と仮定することによって, $S$ の規則性が明らかになり, 簡潔な規則 ($A \rightarrow\ a\ b\ c$) が発見された. そして, その発見された規則は, 記号列 $S$ を的確に表現している (曖昧性がないという意味で). $S$ を的確に表す簡潔な規則が得られた結果から, x と y は高い可能性で, ノイズ記号であることがいえる.

## 3. 提案手法

このような考えに基づき、提案手法では各々の記号をノイズ記号と仮定し, その仮定を最小記述長原理で検証する. そして, その検証結果から最も単純かつ曖昧性の少ない文法を導く, 最適な終端記号集合を特定することにより, 学習データから基本的な行動文法を得る. このようにして, 画像処理及び人物行動に起因するノイズを含む観測からでも行動文法を獲得することを実現する.

本節では, 前述で紹介した基本アイデアの適用方法を具体的に説明し, ノイズの特徴について論じたのち, 文脈自由文法の応用を説明し, ノイズを含んだ行動記号列からの文法の獲得手法を紹介する. 但し, 本手法は映像から得られた記号列を用いた, 文法獲得に着目しているため, 単位行動記号の抽出方法に対する議論は, 本論文の範囲外であるとして行わない.

### 3.1 ノイズの特徴

映像から得られた単位行動の記号列を用いて, 行動文法を学習する際, 二つの要因によるノイズが記号列に含まれていると考えられる.

第一の要因は, 画像処理によるシステムノイズである. 例えば, 照明変動や物体の遮蔽による, 記号の見逃し (削除) や誤検出 (挿入) があげられる. このような, 頻繁に見逃される記号や誤検出される記号は学習に望ましくない.

第二の要因は, 人物行動に含まれている余計な動きによる挿入ノイズである. 人間の行動を認識するためには重要な単位行動を観察することが重要である. しかし, 人物の行動には, ばらつきがあり, 認識に重要な単位行動 (一次的な単位行動記号) と不要な単位行動 (二次的な単位行動記号) が混在している. この二次的な記号は, 重要な一次的な記号の間を埋める特徴を持ち, また不定期に発生する傾向があるため, 文法学習の障害となる.

すべてのノイズの種類 (削除と挿入) を同時に扱うことは困難であることから, 本手法では, いくつかの仮定を設ける.

**Context-Free Grammar G**

$\mathbf{T} = \{walk, appear, disappear\}$
$\mathbf{N} = \{S, ENTER, EXIT\}$
$\mathbf{R} = \{S \rightarrow ENTER\ \ EXIT$
$ENTER \rightarrow appear\ \ walk$
$EXIT \rightarrow disappear\}$

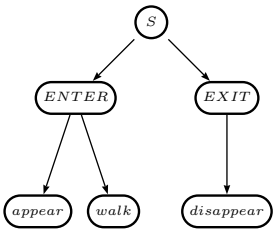


図 1　文脈自由文法が表す人物行動の階層構造.

(1)　挿入によるノイズ記号のみを考慮する.
(2)　記号列にはノイズ記号と非ノイズ記号が混在している (ノイズ記号のみ, 非ノイズ記号のみの記号列はない) ことを仮定する.
(3)　記号はノイズ又は非ノイズのいずれかである (一つの記号が同時にノイズと非ノイズにはなれない) と仮定する.
(4)　非ノイズ記号の出現には規則性があると仮定する.

### 3.2　文脈自由文法を用いた人物行動モデル

冒頭で述べたように, 文脈自由文法は人物の階層的な行動を表すモデルとして適切である. CFG は 4 項組 $\mathbf{G} = \{\mathbf{T}, \mathbf{N}, S, \mathbf{R}\}$ によって定義される. $\mathbf{T}$, $\mathbf{N}$, $\mathbf{R}$, はそれぞれが終端記号 $\{t_1, \ldots, t_m\}$, 非終端記号 $\{N_1, \ldots, N_n\}$, 生成規則 $\{R_1, \ldots, R_p\}$ の集合であり, 開始記号 $S$ は $S \in \mathbf{N}$ である. 生成規則は $A \rightarrow \alpha$ という形を持ち, $A$ は非終端記号, $\alpha$ は終端記号と非終端記号の記号列を表す. 各々の生成規則に確率 $P(A \rightarrow \alpha)$ が付加される場合, $\sum_k P(A \rightarrow \alpha_k) = 1$ を満たし, 確率的文脈自由文法と呼ばれている.

本研究では, 終端記号 $t$ は単一の行動 (単位行動) を表し, 非終端記号 $N$ は (一つもしくは複数の) 単位行動を表している. 開始記号 $S$ は一つの行動シーケンス (行動の記号列) を表す記号である. 従って, 文脈自由文法 $G$ を用いることにより, 人物の行動から得られる記号列の階層構造を表すことができる (図 1 参照).

### 3.3　ノイズ仮説の準備

映像から得られた学習データ $\mathbf{W}$ から文法を学習するためには, ノイズ記号を特定したい. 但し, 学習データ $\mathbf{W}$ は $m$ 個の行動記号列 $W_1 \cdots W_m$ であり, 行動記号列 (行動シーケンス)$W_i$ は $k$ 個の単位行動記号 $w_1, \ldots, w_k$ から形成されている. ノイズ記号を直接的に探すことは困難であるため, 本手法では, 各々の記号に対して順番に, ノイズであるという仮説を立て, すべての仮説の妥当性を最小記述原理の枠組みで検証する (図 2 参照). つまり, 全数検索を行い, $2^n - 2$ の仮説を検証する ($n$ は記号の種類の数). 例えば, 3 つの記号種類, a, b,c が与えられた場合, 6 つの仮説が得られる: {a}, {b}, {c}, {a,b}, {a,c}, {b,c} ({ · } はノイズ記号を表すとする).

まず最初に, 一つの仮説 (例えば, $c$ がノイズであるという仮説) に対して, 仮説を反映した文法 (以下, 仮説文法) を学習するため, ノイズ記号に関する仮説を表現する以下のような初期文法を事前に準備する.

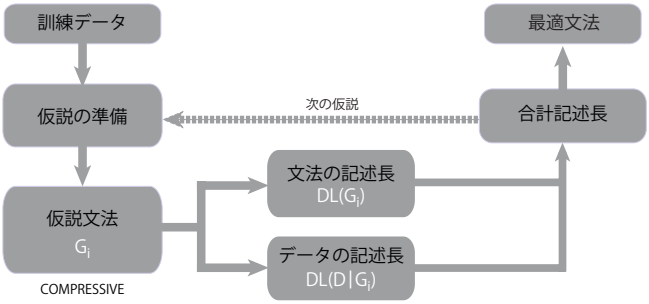


図 2　最小記述長原理を用いた文法学習の枠組み.

$$\mathbf{G}_0 = \left\{ \begin{array}{ll} N_1 \rightarrow w_1^+, & * \rightarrow w_1^-, \\ \ldots & \ldots \\ N_u \rightarrow w_u^+, & * \rightarrow w_v^-, \\ S \rightarrow \mathbf{W}', & * \rightarrow ** \end{array} \right\}.$$

初期文法の準備は, 学習データ $\mathbf{W}$ の前処理とも考えられ, $\mathbf{W}$ を, 仮説を反映した記号列 $\mathbf{W}'$ に変換する処理である.

初期文法 $\mathbf{G}_0$ の生成規則 $N_i \rightarrow w_i^+$ は非ノイズ記号 $w_i^+$ が非終端記号 $N_i$ によって生成されることを表し, この生成規則は $w_i^+$ の存在を $\mathbf{W}'$ に残す効果がある (図 3 では, $A \rightarrow a$ によって $a$ が $A$ で置き換えられている). $u$ 種類の非ノイズ記号に対して $u$ 個の生成規則を準備する.

次に, 生成規則 $* \rightarrow w_i^-$ はノイズ記号 $w_i^-$ がノイズを意味する非終端記号 $*$ によって生成されることを表し, この生成規則によりどのノイズ記号も同じ非終端記号で表現することができる. $v$ 種類のノイズ記号に対して, $v$ 個の生成規則を準備する.

生成規則 $* \rightarrow **$ は隣接するノイズ記号が非終端記号 $*$ によって生成されることを表し, 記号列内で隣接するノイズ記号を一つの非終端記号 $*$ で省略する効果がある (図 3 ではすべての $c$ が $*$ で置き換えられ, 隣接している $c$ は一つの $*$ にまとめられている).

最後の生成規則 $S \rightarrow \mathbf{W}'$ は, 仮説の反映した学習データ $\mathbf{W}'$ が開始記号 $S$ によって生成されることを表し, 仮説の影響を学習データに反映する. $\mathbf{W}'$ は, $\mathbf{W}$ の各々の単位行動 $w_i$ を, 対応する生成規則で置き換えたものである (図 3 の $\mathbf{G}_0$ の規則 $S$ を参照). なお, 各記号列 (行動シーケンス)$W_i$ の開始点を記す, 固有な記号を, 先頭に付加する (図 3 では, 1, 2, 3). この開始記号は学習後の処理で使用される.

### 3.4　仮説文法の取得

本節では, 前節で構築した初期文法を基に, 文脈自由文法の発見的学習アルゴリズム, COMPRESSIVE [8] を用いた仮説文法の学習方法を簡単に紹介する. COMPRESSIVE は, 学習データ内に頻繁に出現する部分記号列から, 新たな生成規則を作り, 逐次的に部分記号列を, その生成規則で置き換えて, 文脈自由文法を学習する方法である. 新しい生成規則を作るための最適な部分記号列 $\nu$ は, 文法全体の長さの減少 $\Delta DL$ を最大にする部分記号列である.

映像から得られた記号列 $\mathbf{W}$:
$\mathbf{W} = W_1 W_2 W_3$
但し,
$W_1 = c\ a\ b\ a\ a\ b\ c$
$W_2 = a\ b\ a\ c\ c\ a\ b$
$W_3 = c\ a\ b\ a\ a\ b\ c\ c$

仮説:
終端記号集合 $\{c\}$ がノイズ.

仮説を反映した初期文法 $\mathbf{G}_0$:
$S \to 1 * ABAAB *\quad 2ABA * AB\quad 3 * ABAAB *$
$A \to a$
$B \to b$
$* \to c$
$* \to * *$

学習された仮説文法 $\mathbf{G}$:

| 生成規則 | 確率 |
|---|---|
| $S \to\ C$ | (2/3) |
| $S \to\ D * E$ | (1/3) |
| $A \to a$ | (9/9) |
| $B \to b$ | (6/6) |
| $C \to * D\ E\ *$ | (2/2) |
| $D \to E\ A$ | (2/2) |
| $E \to A\ B$ | (3/3) |
| $* \to c$ | (7/8) |
| $* \to * *$ | (1/8) |

図 3　仮説文法の学習の例.

$$\arg\max_{\nu} \Delta DL = \arg\max_{\nu} \left\{ M_\nu \cdot N_\nu - (M_\nu + 1) - N_\nu \right\}. \quad (1)$$

但し, $M_\nu$ は部分記号列の長さ, $N_\nu$ は部分記号列の出現回数である (図 4 参照). なお, 本手法では, 生成規則の適用回数を保存することにより, 確率的な文脈自由文法の学習に拡張している (図 3 の文法 $\mathbf{G}$ の生成規則には確率が付加されている).

COMPRESSIVE が収束したのち, 生成規則の確率の計算と規則 $S$ の分解を行う. 前節で述べたように規則 $S$ の右辺 $\mathbf{W}'$ は $m$ 個の行動記号列の結合であるため, 最後に $\mathbf{W}'$ を分ける必要がある. 具体的には, 開始記号を用いて, 規則 $S \to \mathbf{W}'$ を個々の規則 $S \to W'_1, \ldots, S \to W'_n$ に分ける処理を行うということである ($n \leqq m$). 但し, $S \to W'_i$ は 1 種類の行動を表す規則であるため, 重複する規則 $S$ は一回のみ文法に加えられる. 各生成規則の確率は出現頻度の比を用いて式 (2) で計算する.

$$P(N \to \nu^*_i) = \frac{c(N \to \nu^*_i)}{\sum_j c(N \to \nu^*_j)}, \quad (2)$$

但し, $c(\cdot)$ は規則の使用回数, $N$ は任意の非終端記号, 規則の右辺 $\nu^*$ は一つまたは複数の記号を意味する.

### 3.5　最小記述長原理による仮説文法の評価

本研究の目標は映像から得られた記号列を, 簡潔かつ的確に表す文法を探すことである. この目標を達成するために, 最小記述長原理の枠組みを用いて, 仮説文法の簡潔さを表現する文法記の記述長 $DL(\mathbf{G})$ と文法の的確さを表現するデータの記述長 $DL(\mathbf{W}|\mathbf{G})$ を最小にする, 最適な文法 $\hat{\mathbf{G}}$ を求める (式 (3)). シャノンの符号法により, 確率変数 $X$ の記述長は $X$ の確率の対数としても表せるため, 最小記述長原理を対数の和としても

入力記号列:

$S \to a\ b\ c\ d\ a\ b\ c\ d\ b\ c\ d\ a\ b\ a\ b$

| 繰り返しパターン $\nu$ | 出現回数 $N_\nu$ | 長さ $M_\nu$ | 圧縮 $\Delta DL$ |
|---|---|---|---|
| $b\ c\ d$ | 3 | 3 | **2** |
| $a\ b\ c\ d$ | 2 | 4 | 1 |
| $a\ b$ | 4 | 2 | 1 |
| $c\ d$ | 3 | 2 | 0 |

(1) 圧縮を最大にする $\nu$ で新しい生成規則を作成:

$A \to b\ c\ d$

(2) 記号列を新しい生成規則で符号化:

$S \to a\ A\ a\ A\ A\ a\ b\ a\ b$

(3) ステップ (1) と (2) を繰り返す

図 4　COMPRESSIVE アルゴリズムの例.

表現できる (式 (4)).

$$\hat{\mathbf{G}} = \arg\min_G \{ \underbrace{DL(\mathbf{G})}_{\text{文法の記述長}} + \underbrace{DL(\mathbf{W}|\mathbf{G})}_{\text{データの記述長}} \} \quad (3)$$

$$= \arg\min_G \{ -\log P(\mathbf{G}) - \log P(\mathbf{W}|\mathbf{G}) \}. \quad (4)$$

本手法では, [9] で提案された手法を用いて, 仮説文法の記述長を計算し, データの記述長はデータ尤度 $P(\mathbf{W}|\mathbf{G})$ から求める [10]. 本節では, この二つの記述長の計算方法と意味を説明する.

#### 3.5.1　文法の記述長

最小記述長原理の最初の項は文法の記述長である. $DL(\mathbf{G})$ は文法の簡潔さまたは複雑さを表す値であり, 間接的にデータの規則性も表している. 最小記述長原理の枠組みでは, 簡潔な文法程, 文法の記述長が短い.

文法の記述長は, 文法の構造 $G_S$(例えば, $A \to a$) と文法のパラメータ $\Theta_G$(例えば, $P(A \to a) = 0.5$) に分けることができる.

$$\begin{aligned} DL(\mathbf{G}) &= -\log P(G_S, \Theta_G) \\ &= -\log P(\Theta_G|G_S) - \log P(G_S) \\ &= \underbrace{DL(\Theta_G|G_S)}_{\text{パラメータの記述長}} + \underbrace{DL(G_S)}_{\text{構造の記述長}}, \end{aligned} \quad (5)$$

さらに, 文法構造の記述長 $DL(G_S)$ は, 各々の生成規則 $R$ の, 長さの記述長 $DL(l_R)$ と記号の記述長 $DL(s_R)$ の和として計算する.

$$DL(G_S) = \sum_{R \in \mathbf{G}} \big( DL(l_R) + DL(s_R) \big). \quad (6)$$

生成規則の長さ $l_R$ の最小限は 2 であり (例えば, $S \to a$ の長さは 2), ほとんどの生成規則の長さは 2 から 5 の間である性質を持つ. この性質から, $l_R$ は, ポアソン分布から生起すると近似することができる [9]. 従って, $DL(l_R)$ を次の式で表す.

$$DL(l_R) = -\log p(l_R - 1; \mu) = -\log \frac{e^\mu \mu^{(l_R - 1)}}{(l_R - 1)!}. \quad (7)$$

但し, $\mu$ は生成規則の長さの事前平均である (実験では $\mu = 3$).

文法で使用されている各々の記号は, 文法中同じ確率で使用されると仮定すると, 一つの記号の記述長は $\log_2 |\Sigma|$ ビットとなり ($\Sigma$ は記号の集合), 長さ $l_R$ を持つ生成規則 $R$ は $l_R \log |\Sigma|$ ビットの記述長を要する.

$$DL(s_R) \ = \ l_R \log |\Sigma|. \tag{8}$$

次に, パラメータの記述長 $DL(\Theta_G|G_S)$ を, パラメータの確率 $P(\Theta_G|G_S)$ から求める. 任意の非終端記号 $N$ は, 一つまたは複数の生成規則の左辺であり, これらの生成規則集合 $\mathbf{R}_N$ のパラメータ (確率) は多項分布 $\theta_N$ である (図 3 では, 文法 $\mathbf{G}$ の非終端記号 $S$ のパラメータ $\theta_S = (\theta_{S\to C}, \theta_{S\to D*E})$ は $(\frac{2}{3}, \frac{1}{3})$ である). 左辺 $N$ を持つ生成規則集合 $\mathbf{R}_N$ のパラメータ $\theta_N$ は, どのような値をとるのか事前に分からないため, 多項分布の事前確率を, 一様に分布したディリクレ分布で表す (ディリクレ分布の妥当性に関する議論は [9] を参照). 従って, パラメータ $\theta_N = (\theta_1, \ldots, \theta_n)$ の確率 $P_N(\theta_N|\mathbf{R}_N)$ は, 式 (9) で表される.

$$P_N(\theta_N|\mathbf{R}_N) = \frac{1}{B(\alpha_1, \ldots, \alpha_n)} \prod_{i=1}^{n} \theta_i^{\alpha_i - 1}. \tag{9}$$

但し, $\alpha_1, \ldots, \alpha_n$ は各生成規則の事前の重み, 正規化項の $B$ はベータ分布である. なお, [9] と同様, 各々の生成規則の確率 $\theta_i$ は一様に分布しており, 事前の重み $\alpha_i$ も一様に分布しているものとして扱う. 最後に, 各々の非終端記号 $N$ のパラメータ記述長の和をとることによって, 文法全体のパラメータ記述長が次の式から得られる.

$$DL(\Theta_G|G_S) = \sum_{N\in\Sigma} -\log P_N(\theta_N|\mathbf{R}_N). \tag{10}$$

**3.5.2** データの記述長

小さい記述長を持つ仮説文法は簡潔である反面, 表現力に欠けている場合も多く, その表現力を評価することが必要である. 最小記述長原理の第 2 項はデータの記述長 $DL(\mathbf{W}|\mathbf{G})$ であり, 文法の記述長への影響を抑える役割を果たし, データを的確に表す文法に高い評価を与えるものである. 逆に, データの記述長を, 文法の曖昧性に対するペナルティ値としてみなすこともできる. 曖昧性がない文法の記述長は 0 であり, 曖昧性が高い程, 文法の記述長が大きくなる.

文法 $\mathbf{G}$ が与えられたときの行動記号列 $W_i$ の尤度 $P(W_i|\mathbf{G})$ は, 一般的に内側アルゴリズムで求めるが, 本手法で学習された文法はチョムスキー標準形ではないため, Pynadath らが提案した, ベータ確率を用いて尤度を求める (詳細は [10] 参照). すべての記号列 (学習データ) の尤度 $P(\mathbf{W}|\mathbf{G})$ は式 (11) で求まる. 最後に式 (12) で尤度を記述長に変換する.

$$P(\mathbf{W}|\mathbf{G}) \ = \ \prod_{i=1}^{n} P(W_i|\mathbf{G}), \tag{11}$$

$$DL(\mathbf{W}|\mathbf{G}) \ = \ -\log P(\mathbf{W}|\mathbf{G}). \tag{12}$$

**3.5.3** 合計記述長の重み付け

文法の記述長とデータの記述長は本質的に異なるものを表しているため, それぞれ変化の傾向が異なる. 人工データ及び実データの解析結果によると, データの記述長より, 文法の記述長が激しく変動し, 最小記述長原理の観点から, 小さい文法が優先されるという傾向がある [11]. この現象に対応するために, 記述長に重みを付加することが可能である [9] [12]. 本手法では, 文法とデータ尤度の記述長範囲の比

$$\lambda_x = \frac{DL_{max}(\mathbf{G}) - DL_{min}(\mathbf{G})}{DL_{max}(\mathbf{W}|\mathbf{G}) - DL_{min}(\mathbf{W}|\mathbf{G})} \tag{13}$$

を用いて, 文法の記述長を正規化する. 但し, $\lambda_x$ は $x$ 個の非ノイズ記号を用いた文法 $\mathbf{G}$ の正規化係数である. 重み $\lambda_x$ は複雑な文法に対するペナルティを減少する効果があるため, 本手法では $\lambda_x$ は文法の事前重みと呼ぶことにする. 合計記述長は次のようになる.

$$DL(\mathbf{G}|\mathbf{W}) \ = \ \lambda_x DL(\mathbf{G}) + DL(\mathbf{W}|\mathbf{G}). \tag{14}$$

式 (3) により合計記述長 (14) を最小にする仮説文法が, 最適な文法として得られる.

**3.5.4** 最適な文法 $\hat{\mathbf{G}}$ の決定

ここまでは, 各々の仮説文法 $\mathbf{G}$ の, 簡潔さ (文法の記述長) と的確さ (データの記述長) を, 定量化する手法を説明し, 最適な文法 $\hat{\mathbf{G}}$ は, 重み付合計記述長 $DL(\mathbf{G}|\mathbf{W})$ を最小にする文法として求まることを導いた. しかしながら, 実利用を考慮した際, アルゴリズムの出力は一つの最適な文法ではなく, 複数の文法を要する場合も考えられる. 例えば, ユーザが用途に応じて, 複数の文法から適当な文法を選択したい場合やユーザのニーズに合わせて使用する非ノイズ記号の数を予め決めたい場合などである. このような要求に対して本手法は, 任意の仮説文法に対する合計記述長の計算が可能であるため, 仮説文法の合計記述長の順に並べてユーザに表示することができる. 又, 非ノイズ記号の数を固定し, 上位の文法を提示することも可能である. 何らかな評価基準に基づいて並べられている結果からユーザが選択するという手法は, 自動文書要約の研究でも見られるものである [13]. 但し, 評価の観点からは, 一つの文法, つまり記述長を最小にする文法を特定する必要があるため, 本論文では最適な文法に対して議論を進める.

## 4. 実験結果

実データを用いて本提案手法の有効性を検証するために, 某店舗内のレジ上に監視カメラを設置し, 人物行動の文法学習を行った. 図 5 に表示されているように, CCD カメラをレジの上部に設置し, 3 日に渡って 6 時間の店員と客の行動パターンを映像として録画した.

本実験では色情報を用いて複数の物体 (手やトレイやお金等) を検出し, 10 種類の単位行動 (終端記号) を出力するプログラムを準備した. 手とトレイの検出例を図 6 で示す. プログラムから得られた単位行動記号の説明を表 1 で示す. 映像の処理はオフラインで行い, 映像データ中の 60 人の取引から, 60 個の記号列が自動的に得られた. 但し, 各取引 (各々の記号列 $W_i$) の分割は手で行った. 記号列の最大長は 11 で, 最小長は 3 であり, すべての記号列を合わせて長さ 429 個の記号が得られた. この


図 5　頭上の CCD カメラで 6 時間の行動を録画.

記号列を入力 (学習データ) として, 本手法では, この記号列を解析した.

各々のノイズ仮説に対して, 学習データから仮説文法を学習し, 最小記述長原理に基づき, 仮説文法を評価した. すなわち, 10 種類の記号で生成される, 1022 のノイズ仮説を評価し, 全数の合計記述長を計算した.

評価から得られた, 最適文法を含む, 上位の候補文法を表 2 に示す. 数少ない非ノイズ記号で形成されている文法は簡潔である一方, 文法にノイズ記号が多く含まれているため, 文法が曖昧であり, データ記述長の大きさにより, 合計記述長が最小ではない. その反面, 数多い非ノイズ記号を利用している文法は行動記号列を細かく表現する力があるものの, 文法は複雑であり, この文法の合計記述長も最小ではない.

一方, 最適解として得られた文法 (表 2 の $x = 3$) は簡潔かつ曖昧性の低い文法であることが分かる. 最適文法を表 3 に示す. 学習された文法は, 10 種類の記号の中から, 非ノイズ記号 `EMP_ReturnedScanner` と `EMP_TookReceipt` と `EMP_TookScanner` を用いて, 数少ない記号で, 取引行動を簡潔に記述している. さらに, ここで使用されている記号は, いずれも店員の単位行動であるということにも注目されたい. 取引行動を的確に表すために, 最も規則性のある店員の行動が, 学習文法で使用されている. つまり, 本手法は文法の簡潔さのみならず, 文法の表現力も考慮している. すなわち, 本手法は行動の構造 (規則性) を出来るだけ詳細に説明する文法を学習する. 学習データに, 頻繁に観察された取引行動 $H$ の構文木を図 7 に示している. 図 7 の構文木から「店員は取引の前半 (非終端記号 $E$) では何らかの行動を行い ($*$), スキャナを取る ($C$). 後半 (非終端記号 $D$) ではスキャナを返し ($A$), 領収書を発行する ($B$).」という行動が, 学習された文法で, 説明されているのが分かる. つまり, 学習された文法は, 一次的な記号の構造を的確に表現し, 二次的な記号はノイズ記号 $*$ として構造内で省略している.

## 5. 結　　び

本論文では, 人物行動を映像から得られたノイズを含む記号列から, 最適な文法を学習する枠組みを提案した. 各記号の組合せに対して, その記号集合がノイズであるという仮説を立て, その仮説を最小記述長原理で評価した. 最小記述長の枠組みでは, 仮説文法の簡潔さと, データに対する文法の表現力の両方を評価し, 仮説文法の妥当性を計った. 最小記述長原理の文法記

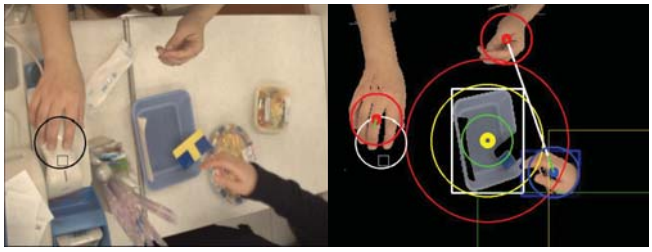

図 6　画像処理によって 10 種類の単位行動記号を抽出.

表 1　終端記号の定義.

| NO. | 終端記号 | 記号の意味 |
|---|---|---|
| 1 | CUS_AddedMoney | お客がトレイにお金を置く |
| 2 | CUS_MovedTray | お客がトレーを動かす |
| 3 | CUS_RemovedMoney | お客がトレイからお金を取る |
| 4 | EMP_HandReturns | 店員の手がしばらくしてから戻る |
| 5 | EMP_Interaction | 店員がお客とやり取りをする |
| 6 | EMP_MovedTray | 店員がトレーを動かす |
| 7 | EMP_RemovedMoney | 店員がトレイからお金を取る |
| 8 | EMP_ReturnedScanner | 店員がスキャナを返す |
| 9 | EMP_TookReceipt | 店員がレシートを取る |
| 10 | EMP_TookScanner | 店員がスキャナを取る |

表 2　上位に現れた候補文法. 最適文法は太字で表示.

| $x$ | Non-noise Symbols | $\lambda_x$ | $\lambda_x$DL(G)+DL(W\|G) |
|---|---|---|---|
| 1 | EMP_TookScanner | 0.383 | 1279.0016 |
| 2 | EMP_ReturnedScanner<br>EMP_TookScanner | 0.2897 | 1160.7076 |
| **3** | **EMP_ReturnedScanner**<br>**EMP_TookReceipt**<br>**EMP_TookScanner** | **0.3096** | **1140.0563** |
| 4 | CUS_MovedTray<br>EMP_ReturnedScanner<br>EMP_TookReceipt<br>EMP_TookScanner | 0.3847 | 1211.0414 |
| 5 | CUS_MovedTray<br>CUS_RemovedMoney<br>EMP_ReturnedScanner<br>EMP_TookReceipt<br>EMP_TookScanner | 0.4246 | 1260.2536 |
| 6 | CUS_MovedTray<br>CUS_RemovedMoney<br>EMP_MovedTray<br>EMP_ReturnedScanner<br>EMP_TookReceipt<br>EMP_TookScanner | 0.4859 | 1353.1436 |
| 7 | CUS_AddedMoney<br>CUS_MovedTray<br>CUS_RemovedMoney<br>EMP_MovedTray<br>EMP_RemovedMoney<br>EMP_ReturnedScanner<br>EMP_TookScanner | 0.5335 | 1523.8244 |
| 8 | CUS_MovedTray<br>EMP_HandReturns<br>EMP_Interaction<br>EMP_MovedTray<br>EMP_RemovedMoney<br>EMP_ReturnedScanner<br>EMP_TookReceipt<br>EMP_TookScanner | 0.6228 | 1784.4875 |

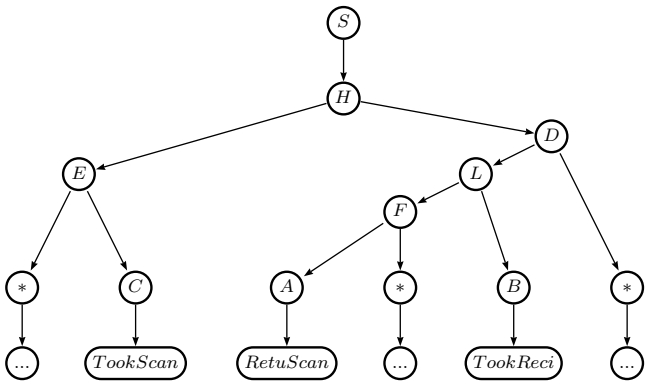

図 7　学習データに頻繁に現れた行動の構文木.

表 3　最小記述長で選ばれた学習文法 (3 つの非ノイズ記号を使用).

| | | | |
|---|---|---|---|
| **S** → D | (0.02) | D → L * | (1.000) |
| **S** → H | (0.16) | E → * C | (1.000) |
| **S** → G | (0.18) | F → A * | (1.000) |
| **S** → N * | (0.04) | G → C D | (1.000) |
| **S** → J | (0.13) | H → E D | (1.000) |
| **S** → Q | (0.05) | I → * B * | (1.000) |
| **S** → * | (0.02) | J → C F | (1.000) |
| **S** → N | (0.02) | K → * D | (1.000) |
| **S** → R | (0.05) | L → F B | (1.000) |
| **S** → J B | (0.02) | M → C * | (1.000) |
| **S** → M L | (0.04) | N → E A B | (1.000) |
| **S** → M A H | (0.02) | O → E * | (1.000) |
| **S** → C K | (0.04) | P → E I | (1.000) |
| **S** → C A M F | (0.02) | Q → E K | (1.000) |
| **S** → O F | (0.02) | R → E L | (1.000) |
| **S** → M | (0.02) | | |
| **S** → O L | (0.02) | * → * * | (0.309) |
| **S** → O | (0.02) | * → CUS_AddMoney | (0.153) |
| **S** → P | (0.05) | * → CUS_MovedTray | (0.006) |
| **S** → I | (0.04) | * → CUS_RemMoney | (0.003) |
| **S** → K | (0.04) | * → EMP_HandReturn | (0.080) |
| A → EMP_ReturnedScanner | (1.00) | * → EMP_Interaction | (0.275) |
| B → EMP_TookReceipt | (1.00) | * → EMP_MovedTray | (0.028) |
| C → EMP_TookScanner | (1.00) | * → EMP_RemMoney | (0.147) |

述長 $DL(\mathbf{G})$ に対して, 事前重み $\lambda$ を付加し, 文法記述長のバイアスを正規化した. さらに, 提案手法を実データを用いて適用し, 最も小さい, かつ表現力のある文法を発見した.

今後の研究の展開として, 三つの課題がある.

一つ目は計算量の課題である. 本手法では, 仮説の全数検索を行ったが, 記号の種類が増えるにつれ, 仮説の数が指数的に増加するため, 他の検索方法を検討する必要がある. 例えば, $n-1$ の非ノイズ記号を用いる仮説から高く評価された仮説を選び, 次に来る $n-2$ 個の非ノイズ記号の仮説評価は, その高く評価された仮説の部分仮説のみを評価するといった方策が考えられる.

二つ目の課題は記号の削除への対応である. 本手法では, ノイズ記号の挿入のみを扱ったが, 実際画像処理の観点からは行動の見逃しも重要な課題である. しかし, 見えているノイズ行動の処理に比べて, 見えていない行動の推定は一層難しいため, 他の事前情報を手掛かりに, この問題に取り組まなければならない. 一つ手掛かりとなる情報は単位行動の検出器の出力であると考えられる. 認識アルゴリズムが統計的なモデルだとすれば, 単位行動の有無情報のみならず各行動の確率を利用することによって起こりうる行動を検知することができる. もう一つの手掛かりとして使用できる方法は, 本手法で得られた文法である. 文法の事前知識を基に記号列の尤度を計算することができ, どの記号を追加することによって尤度があがるかも分かる. 記号列マッチング研究で使われている編集距離などを利用して, 最適な記号列を復元することも考えられる.

三つ目の課題は, 行動文法のモデルの選択である. 文脈自由文法は, 一次的なモデルと比べて, 表現力が高いが, 文脈自由文法には表現できない行動もある. 例えば, 重なり合う行動や, 同時に起こる行動は現在の枠組みでは, 表現することができない. 次の段階として, 他のモデルの導入も考えなければならない.

## 文　献


[1] Y. A. Ivanov and A. F. Bobick: "Recognition of Visual Activities and Interactions by Stochastic Parsing", IEEE Transactions on Pattern Analysis and Machine Intelligence, **22**, 8, pp. 852–872 (2000).

[2] D. J. Moore and I. A. Essa: "Recognizing Multitasked Activities from Video Using Stochastic Context-Free Grammar", Proceedings of the Eighteenth National Conference on Artificial Intelligence, American Association for Artificial Intelligence, pp. 770–776 (2002).

[3] D. Minnen, I. A. Essa and T. Starner: "Expectation Grammars: Leveraging High-Level Expectations for Activity Recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. II: 626–632 (2003).

[4] 三富, 藤原, 山本, 佐藤："習慣的な行動の確率文脈自由文法に基づくベイズ識別", 電子情報通信学会論文誌, **J88-D2**, 4, pp. 716–726 (2005).

[5] K. M. Kitani, Y. Sato and A. Sugimoto: "Deleted Interpolation using a Hierarchical Bayesian Grammar Network for Recognizing Human Activity", Proceedings of the Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 239–246 (2005).

[6] J. M. Zacks and B. Tversky: "Event Structure in Perception and Conception", Psychological Bulletin, **127**, pp. 3–21 (2001).

[7] T. Wang, H. Shum, Y. Xu and N. Zheng: "Unsupervised Analysis of Human Gestures", Proceedings of the IEEE Pacific Rim Conference on Multimedia, Vol. 2195, pp. 174–181 (2001).

[8] C. G. Nevil-Manning and I. H. Witten: "Online and Offline Heuristics for Inferring Hierarchies of Repetitions in Sequences", Proceedings of IEEE, No. 11 in 88, pp. 1745–1755 (2000).

[9] A. Stolcke: "Bayesian Learning of Probabilistic Language Models", PhD thesis, University of California at Berkeley (1994).

[10] D. V. Pynadath and M. P. Wellman: "Generalized Queries on Probabilistic Context-Free Grammars", IEEE Transactions on Pattern Analysis and Machine Intelligence, **20**, 1, pp. 65–77 (1998).

[11] K. M. Kitani, Y. Sato and A. Sugimoto: "An MDL Approach to Learning Activity Grammars", 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, **106**, 376, pp. 19–24 (2006).

[12] J. R. Quinlan and R. L. Rivest: "Inferring Decision Trees Using the Minimum Description Length Principle", Information and Computation, **80**, 3, pp. 227–248 (1989).

[13] K. Knight and D. Marcu: "Statistics-Based Summarization - Step One: Sentence Compression", Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, pp. 703–710 (2000).