3B1-OS22c-7

# 階層ディリクレ過程を用いたマルチモーダル物体概念の形成

Forming Multimodal Object Concept by Hierarchical Dirichlet Process

中村 友昭*1 Tomoaki NAKAMURA
長井 隆行*2 Takayuki NAGAI
岩橋 直人*3 Naoto IWAHASHI

*1電気通信大学電子工学専攻
Dept. of Electronic Engineering, The University of Electro-Communications

*2電気通信大学知能機械工学専攻
Dept. of Mechanical Engineering and Intelligent Systems, The University of Electro-Communications

*3情報通信研究機構
National Institute of Information and Communications Technology

In this paper, we propose a nonparametric Bayesian framework for categorizing multimodal sensory signals such as audio, visual, and haptic information by robots. The robot uses its physical embodiment to grasp and observe an object from various viewpoints as well as listen to the sound during the observation. The multimodal information enables the robot to form human-like object categories that are bases of intelligence. The proposed method is an extension of Hierarchical Dirichlet Process (HDP), which is a kind of nonparametric Bayesian models, to multimodal HDP (MHDP). MHDP can estimate the number of categories, while the parametric model, e.g. LDA-based categorization, requires to specify the number in advance. As this is an unsupervised learning method, a human user does not need to give any correct labels to the robot and it can classify objects autonomously. At the same time the proposed method provides a probabilistic framework for inferring object properties from limited observations. Validity of the proposed method is shown through some experimental results.

## 1. はじめに

事物のカテゴリ分類は，人間の知的活動において重要な役割を果たしている．人間はカテゴリを形成することにより，カテゴリを通した未観測情報の予測を可能とし，これが事物の理解の基礎となっている．すなわちロボットにおいても，このようなカテゴリ分類する能力が非常に重要であると考えられる．

そこで著者らは，これまで pLSA (probabilistic Latent Semantic Analysis)[Hofmann 01] 及び LDA (Latent Dirichlet Allocation)[Blei 03] を拡張したマルチモーダルカテゴリゼーションを提案し，複数のモダリティを用いることにより，より人間の感覚に近いカテゴリを形成することが可能となることを示した [Nakamura 07, Nakamura 09]．提案手法は確率モデルに基づいており，学習したグラフィカルモデルを用いることで，未学習物体のカテゴリ認識が可能である．さらに，学習したモデルを用いた未観測のモーダル情報の推定を可能とした．例えば，物体を見ることで得られる視覚情報から，その物体の硬さやどのような音がするかといった情報を確率的に推定することが可能である．これはまさに人間が日々行っている物体のカテゴリを通した機能の予測であり，提案手法によりその機能をロボットに実装することが可能となった．

しかし，pLSA や LDA はカテゴリ数をあらかじめ人手で与えなければならないという欠点がある．このような問題に対し，本稿では，カテゴリ数の推定が可能なノンパラメトリックベイズモデルである Hierarchical Dirichlet Process(HDP)[Teh 06] をマルチモーダルカテゴリゼーションへと拡張する．HDP では，事前分布にディリクレ過程を導入することで，自由度の高い事前分布が設定でき，さらにデータの複雑さに応じてカテゴリ数を自動で決定することが可能である．

関連研究として，視覚情報のみを用いた物体カテゴリの教師なし学習に関する研究がある [Sivic 05, Fergus 03, Fei-fei 05, Wang 09]．しかしカテゴリは，常に視覚的な情報のみで決定できるわけではなく，人間が物体をカテゴリに分類する際には，他にも様々な情報を用いていると思われる．

連絡先: 中村 友昭，電気通信大学電子工学専攻，〒 182-8585 東京都調布市調布ヶ丘 1-5-1，naka_t@apple.ee.uec.ac.jp

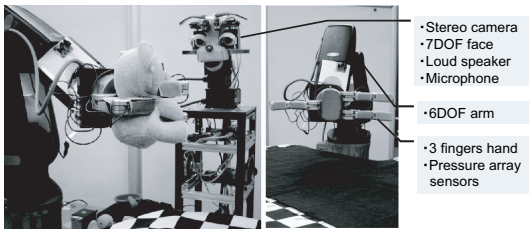


図 1: ロボット

## 2. マルチモーダル情報

ここでは，ロボット (図 1) が取得するマルチモーダル情報（視覚・聴覚・触覚情報）について述べる．

### 視覚情報

ロボット (図 1 左) は頭部にカメラを搭載しており，物体を観察することで得られる画像を視覚情報として利用する．画像は，物体毎に複数枚取得する (後に示す実験では，各物体に対して 50 枚の画像を用いた)．各画像から抽出する特徴量として，36 次元の PCA SIFT[Ke 04] を用いる．これは局所的な特徴である SIFT[Lowe 04] における記述子の代わりに，特徴点周辺の画素値に対して PCA を行い，上位 36 個の主成分を 36 次元の記述子として用いる．PCA SIFT は，SIFT に比べ高い表現力があることが知られている．

PCA SIFT により，1 つの画像から多数の特徴ベクトルを得ることができる．特徴ベクトルの数は，画像により異なるため，このままでは物体の特徴量としては扱いにくい．そこで，これらの特徴ベクトルは，500 の代表ベクトルによりベクトル量子化することで，500 次元のヒストグラムとする．

### 聴覚情報

物体を振ることで発生した音をマイクにより取得し，聴覚情報として利用する．ひとつの物体を観測している間に得られる音声信号をフレームに分割し，フレーム毎に 13 次元の MFCC(Mel-Frequency Cepstrum Coefficient) を計算する．これにより，各フレームは 13 次元の特徴ベクトルとなる．最

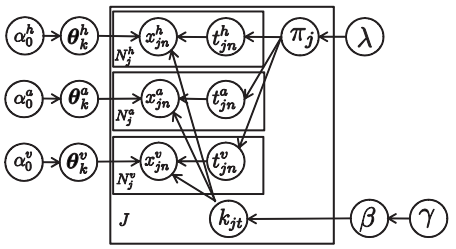

図 2: マルチモーダル HDP のグラフィカルモデル

終的にこの特徴ベクトルも，ベクトル量子化を行い，50 次元のヒストグラムとする．

**触覚情報**

触覚情報の取得には，3 本指のロボットハンドに取り付けられた 162 個のセンサーから構成されている触角アレイセンサーを用いた (図 1 右)．ロボットが実際に物体を把持することで得られるセンサーの時系列データの近似を行い，その近似パラメータを各センサーの特徴ベクトルとして扱う [中村 10]．さらに，この特徴ベクトルをベクトル量子化することで，15 次元のヒストグラムを触覚情報として用いる．

## 3. Multimodal Hierarchical Dirichlet Process

図 2 が Hierarchical Dirichlet Process(HDP)[Teh 06] をマルチモーダルカテゴリゼーションへ適用したグラフィカルモデルである．図において，$x^v_{jn}, x^a_{jn}, x^h_{jn}$ がそれぞれ $j$ 番目の物体の視覚･聴覚･触覚情報の $n$ 番目の特徴を表している．モダリティ $m(\in \{v, a, h\})$ の情報は $\boldsymbol{\theta}^m_k$ をパラメータとする多項分布から生成され，その多項分布は $\alpha^m_0$ をパラメータとするディリクレ事前分布によって決まる．HDP と同様に，物体を中華料理店，各モダリティの情報を客と考えることで，Chinese Restaurant Franchise により生成過程を表現することができる．

### 3.1 Chinese Restaurant Franchise

Multimodal Hierarchical Dirichlet Process では，モダリティ $m$ を客層（子供・男性・女性等）と考え，料理 $k$ は，各客層向けの食べ物が置かれた 1 つのオードブルとして各テーブルに提供される．各客 $x^m_{jn}$ は，テーブル $t^m_{jn}$ に座り，その客層 $m$ に応じた料理を食べる．この生成過程は，客がテーブルを選択する過程と，テーブルに置く料理を選択する過程の二つの Chinese Restaurant Process(CRP)[Aldous 85] で表現される．

**テーブル選択過程：** $j$ 番目の店に新たな客が来店した際には，客 $x^m_{jn}$ がテーブル $t$ に座る事前確率は，CRP より次のようになる．

$$P(t_{jn} = t|\lambda) = \begin{cases} \frac{N_{jt}}{\gamma+N_j-1} & (t = 1, \cdots, T_j) \\ \frac{\lambda}{\gamma+N_j-1} & (t = T_j + 1) \end{cases} \quad (1)$$

ただし，$T_j$ は中華料理店 $j$ 内のテーブルの数である．$N_{jt}$ は $j$ 番目のチェーン店においてテーブル $t$ に座っている客の人数であり，そのテーブルの人気に相当する．すなわち，客は人気のあるテーブルに座る確率が高くなる．また，その客が料理 $k$ を好む確率は次式となる．

$$P(x^m|\boldsymbol{X}^m_k) = \frac{N^m_{kx^m} + \alpha^m_0}{N^m_k + d^m\alpha^m_0} \quad (2)$$

ただし，$N^m_k$ は料理 $k$ を食べている人数，$N^m_{kx^m}$ は客 $x^m$ のうち料理 $k$ を食べている人数であり，$d^m$ はモダリティ $m$ の次元数である．$\boldsymbol{X}^m_k$ は，$x^m_{jn}$ と同じ客層の客が，全店で料理 $k$ を食べている客の集合である．したがって，客 $x^m_{jn}$ がテーブル $t$ に座る事後確率は，ベイズの定理より

$$\begin{aligned} P(t^m_{jn} = t|\boldsymbol{X}, \lambda) &\propto P(x^m_{jn}|\boldsymbol{X}^m_{k=k_{jt}})P(t^m_{jn}|\lambda) \\ &= \begin{cases} P(x^m_{jn}|\boldsymbol{X}^m_{k=k_{jt}})\frac{N_{jt}}{\gamma+N_j-1} & (t = 1, \cdots, T_j) \\ P(x^m_{jn}|\boldsymbol{X}^m_{k=k_{jt}})\frac{\lambda}{\gamma+N_j-1} & (t = T_j + 1) \end{cases} \end{aligned} \quad (3)$$

となり，この確率にしたがって客はテーブルを選択する．$t = T_j + 1$ のテーブルを選択することは，新たなテーブルを生成することを意味する．新たなテーブルが選択された際には，次に説明する料理選択過程により料理が決定される．

**料理選択過程：** テーブル $t$ に料理 $k$ を置かれる事前確率は，CRP より次のようになる．

$$P(k_{jt} = k|\lambda) = \begin{cases} \frac{M_k}{\gamma+M-1} & (k = 1, \cdots, K) \\ \frac{\gamma}{\gamma+M-1} & (k = K + 1) \end{cases} \quad (4)$$

ただし，$K$ は現在提供されている料理数，$M_k$ は全店で料理 $k$ が置かれているテーブルの数であり，料理の人気に相当する．すなわち，人気のある料理は選択される確率が高くなる．また，既にテーブル $t$ に座っている客の集合 $\boldsymbol{X}_{jt}$ が料理 $k$ を好む確率は $P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)$ で表される．したがって，テーブル $t$ に料理 $k$ を置く事後確率は，ベイズの定理より

$$\begin{aligned} P(k_{jt} = k|\boldsymbol{X}, \gamma) &= P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)P(k_{jt} = k|\gamma) \\ &= \begin{cases} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{M_k}{\gamma+M-1} & (k = 1, \cdots, K) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{\gamma}{\gamma+M-1} & (k = K + 1) \end{cases} \end{aligned} \quad (5)$$

となり，この確率にしたがって料理が選択される．$k = K+1$ の料理を選択することは，新たな料理を生成することを意味する．

### 3.2 マルチモーダルカテゴリゼーション

物体の分類は，Gibbs Sampling により，テーブルの割り当て $t_{jn}$，料理の割り当て $k_{jt}$ を推定することによって実現される．Gibbs Sampling では，客 $x^m_{jn}$ を除いた客の集合 $\boldsymbol{X}^{-mjn}$ を条件とした条件付確率からテーブル $t^m_{jn}$ をサンプリングする．

$$t^m_{jn} \sim P(t^m_{jn}|\boldsymbol{X}^{-mjn}, \lambda) \quad (6)$$

テーブル $t$ に置かれる料理 $k_{jt}$ は，テーブル $t$ に座っている客 $\boldsymbol{X}_{jt}$ を除いた客の集合 $\boldsymbol{X}^{-jt}$ を条件とした条件付き確率からサンプリングする．

$$k_{jt} \sim P(k_{jt}|\boldsymbol{X}^{-jt}, \gamma) \quad (7)$$

式 (6) によるサンプリングを全店の全客に対して繰り返し，式 (7) によるサンプリングを全店の全テーブルに対して繰り返すことで，パラメータが推定される．最終的に，収束した値を用い，物体 $j$ がカテゴリ $k$ に属する確率は

$$P(k|\boldsymbol{X}_j) = \frac{\sum^{T_j}_t \delta_k(\hat{k}_{jt})\hat{N}_{jt}}{N_j} \quad (8)$$

となる．ただし，$\boldsymbol{X}_j$ は $j$ 番目の物体の全特徴量であり，$\hat{N}_{jt}$, $\hat{k}_{jt}$ はそれぞれパラメータ推定アルゴリズムによって収束した $N_{jt}$, $k_{jt}$ の値である．$\delta_a(b)$ は $a = b$ の時 1，さもなくば 0 となるデルタ関数である．

以上の学習アルゴリズムを以下にまとめる．

**[マルチモーダル HDP のパラメータ推定アルゴリズム]**
while 収束するまで
　for all $j, m, n$

1. テーブル $t = t^m_{jn}$ から，客 $x^m_{jn}$ を除外し，$k = k_{jt}$ のパラメータを更新

$$N^m_{kx^m_{jn}} --, N_{jt} --, N^m_k --$$

2. 新たなテーブル $t$ を事後分布からサンプリング

$$t \sim \begin{cases} P(x^m_{jn}|\boldsymbol{X}^m_{k=k_{jt}})\frac{N_{jt}}{\gamma+N_j-1} & (t=1,\cdots,T_j) \\ P(x^m_{jn}|\boldsymbol{X}^m_{k=k_{jt}})\frac{\lambda}{\gamma+N_j-1} & (t=T_j+1) \end{cases}$$

3. $t=T_j+1$ ならば新たなテーブルを生成し，新たなテーブルに置く料理をサンプリング

$$k \sim \begin{cases} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{M_k}{\gamma+M-1} & (k=1,\cdots,K) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{\gamma}{\gamma+M-1} & (k=K) \end{cases}$$
$$T_j = T_j+1$$

4. テーブル $t^m_{jn}=t$，$k_{jt}=k$ としパラメータを更新

$$N^m_{kx^m_{jn}}++, N_{jt}++, N^m_k++$$

5. 空のテーブルを削除

end for
for all $j,t$

1. 店舗 $j$ のテーブル $t$ の客 $x^m \in \boldsymbol{X}_{jt}$ を料理 $k=k_{jt}$ を食べている客から除外し，パラメータを更新

$$N^m_k --, N^m_{kx^m} -- \quad \text{for all} \quad x^m \in \boldsymbol{X}_{jt}$$
$$M_k --$$

2. 料理をサンプリング

$$k \sim \begin{cases} P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{M_k}{\gamma+M-1} & (k=1,\cdots,K) \\ P(\boldsymbol{X}_{jt}|\boldsymbol{X}_k)\frac{\gamma}{\gamma+M-1} & (k=K+1) \end{cases}$$

3. $k_{jt}=k$ とし，客 $x^m \in \boldsymbol{X}_t$ を料理 $k$ を食べている客に追加

$$N^m_k++, N^m_{kx^m}++ \quad \text{for all} \quad x^m \in \boldsymbol{X}_t$$
$$M_k++$$

新たな料理が生成された場合は $K=K+1$ とする．

4. テーブルに置かれていない料理を削除

end for
end while

### 3.3 未知物体の認識

学習したモデルを用いることで，未学習物体のカテゴリを認識することができる．認識では，学習時に収束したパラメータ $\hat{N}^m_k$ と $\hat{N}^m_{kx^m}$ を固定し，新たな中華料理店（未知物体）内の客（特徴）$\bar{x}^m$ が料理 $k$ を好む確率を次のように計算する．

$$P(\bar{x}^m|\hat{\boldsymbol{X}}_k) = \frac{\hat{N}^m_{kx^m}+\alpha^m_0}{\hat{N}^m_k + d^m\alpha^m_0} \tag{9}$$

ただし，$\hat{\boldsymbol{X}}_k$ は，学習時に料理 $k$ を食べた客の集合である．また，新たな中華料理店のテーブル $t$ についた客 $\bar{\boldsymbol{X}}_t$ が食べる料理 $\bar{k}_t$ は次式からサンプリングする．

$$\bar{k}_t \sim P(\bar{\boldsymbol{X}}_t|\hat{\boldsymbol{X}}_k) \tag{10}$$

すなわちカテゴリ数を固定し，テーブルの選択のみを行うことになる．以上の 2 式を用いて，前述のアルゴリズムによりパラメータの推定を行う．最終的に未知物体 $\bar{\boldsymbol{X}}$ のカテゴリは，学習時と同様に式 (8) により計算可能である．

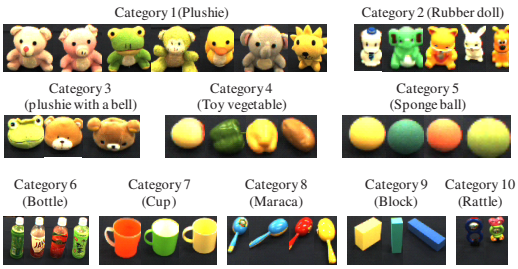


図 3: 実験に用いた物体

### 3.4 未観測モダリティの予測

提案するマルチモーダルカテゴリゼーションの有効性は，物体のカテゴリ分類だけでなく，あるセンサ情報を得ることによって他のセンサ情報を推測することができる点にもある．つまり，物体を見ることによって，その硬さや，それが音を出すかどうか，またどのような音を出すかなどの情報を推測することができる．例として，視覚情報から聴覚情報を推定する場合を考える．視覚情報 $\bar{\boldsymbol{x}}^v$ が与えられた時，聴覚情報 $x^a$ が発生する確率は次のようになる．

$$P(x^a|\bar{\boldsymbol{x}}^v) = \sum_k P(k|\bar{\boldsymbol{x}}^v)P(x^a|\hat{\boldsymbol{X}}_k) \tag{11}$$

ただし，$P(k|\bar{\boldsymbol{x}}^v)$ は，前述の認識アルゴリズムにより認識した結果である．

## 4. 実験

物体の分類性能を検証するため実験を行った．実験では，図 3 に示したコップやヌイグルミなどの 10 カテゴリの物体，計 39 個を使用した．マルチモーダル情報は，これらの物体をロボットが実際に掴み・振り・観察することで得られるマルチモーダル情報を用いた．

### 4.1 マルチモーダルカテゴリゼーション

まず，視覚・聴覚・触覚情報それぞれの分類を行った．分類は各 100 回行い，合計 100 個のモデルを学習した．

視覚情報のみによる分類では，カテゴリ数 12 が最も多く発生した．この時の分類の精度は 79.5%であった．スポンジや一部のぬいぐるみ，野菜のおもちゃ等見た目が似ているものを誤分類する傾向にあった．聴覚情報のみの分類では，カテゴリ数 3 が最も多く発生し，マラカスは区別することができた．しかし，他の音がなる物体を 1 つのカテゴリとして分類してしまっている．分類の精度は，38.4%であった．触覚情報のみによる分類では，カテゴリ数 12 が最も多く発生し，正解のカテゴリより多く推定された．触覚情報はノイズが多く，1 つの物体のみで形成されるカテゴリが多く存在した．しかし，ぬいぐるみやスポンジのボール等は，材質毎に正しく分類された．分類の精度は，53.8%であった．

次に，3 つのモダリティを用いて 39 個の物体の分類を行った．分類は 100 回行い，100 個のモデルを学習した．図 4 が 100 個のモデルのカテゴリ数のヒストグラムであり，図 5 が各カテゴリ数において最尤なモデルによる分類の精度である．図 4 より，正解であるカテゴリ数 10 が最も多く発生しており，正しくカテゴリ数を予測することができた．さらに，図 5 よりカテゴリ数 10 の時の分類の精度が最も高く，正解の分類と完全に一致した．このように 3 つのモダリティを用いることで，正しいカテゴリを推定できる．

また，比較として Gibbs Sampling によるパラメータ推定を用いた LDA[Griffiths 04] を，マルチモーダルカテゴリゼーションへと拡張したマルチモーダル LDA により分類を行った．ただし，マルチモーダル LDA ではカテゴリ数をあらかじめ与える必要があるため，カテゴリ数を 10 に固定した．その結果，マルチモーダル HDP と同様に全物体は正しく分類された．そ

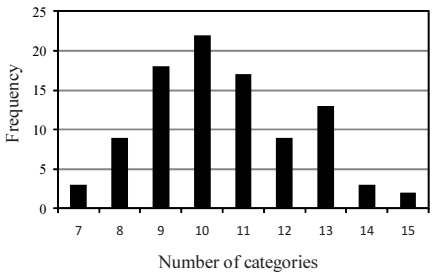


図 4: カテゴリ数と発生頻度

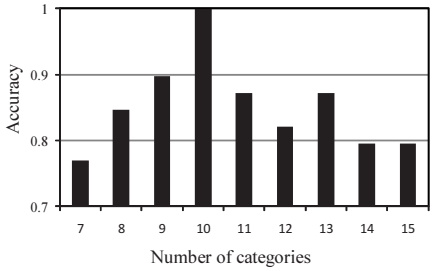


図 5: カテゴリ数と分類精度

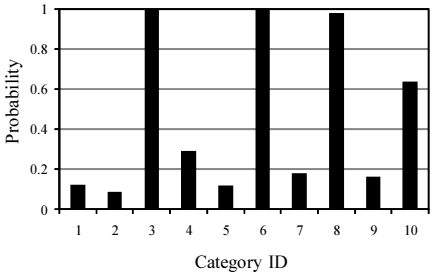


図 6: 未学習物体の音情報の予測

の際の分類の対数尤度は $\mathcal{L}_{lda} = -115270$ であった．一方，マルチモーダル HDP による分類の対数尤度は $\mathcal{L}_{hdp} = -113629$ と，マルチモーダル LDA より高い尤度となり，十分な分類性能があるといえる．さらに，HDP ではカテゴリ数の推定も行うことができ，LDA に比べ有効であるといえる．

### 4.2 未知物体の認識

次に，39 個の物体のうち 29 個を学習し，残りの 10 個の物体のカテゴリを，画像情報のみから認識した．認識用の物体は各カテゴリから 1 物体をランダムに選択した．その結果，認識率は 90%となり，未知の物体であっても正しく分類できることが分かる．誤りはカテゴリ 10 の学習物体数が 1 つとなってしまい，正しく汎化することが出来なかったことが原因として挙げられる．

### 4.3 未観測モダリティの予測

次に，前節同様未学習物体 10 個を用い，未観測情報の予測を行った．未学習物体の視覚情報 $\boldsymbol{x}^v$ から，音がなる確率 $P(\boldsymbol{x}^a \in \{\text{sounder}\}|\boldsymbol{x}^{\text{v}})$ を求めた．10 カテゴリのうち，音がなる物体はカテゴリ 3，カテゴリ 6，カテゴリ 8，カテゴリ 10 である．その結果が図 6 である．図より，実際に音がなる物体の音がなる確率 $P(\boldsymbol{x}^a \in \{\text{sounder}\}|\boldsymbol{x}^{\text{v}})$ は 0.6 以上であり，音がならない物体に比べ高い値となった．カテゴリの認識では，認識を誤ったカテゴリ 10 の物体であるが，正しく音がなることを予測することができている．すなわち，カテゴリの認識が出来ない物体であっても，未観測の情報の予測が可能であることが分かる．

## 5. まとめ

本稿では，ノンパラメトリックベイズモデルの 1 つである Hierarchical Dirichlet Process を拡張し，マルチモーダル情報の分類へ応用した．ノンパラメトリックベイズモデルを用いることで，Latent Dirichlet Allocation ではあらかじめ与える必要のあったカテゴリ数を，データから適切に推定することが可能となった．実験によって，マルチモーダル情報を用いることで，物体カテゴリが適切に形成できることを示し，分類性能は LDA と同等以上の性能であることがわかった．さらに，マルチモーダルカテゴリゼーションの有効性は，未観測の情報の予測が可能な点にある．実験では視覚情報から聴覚情報の予測を行ったが，他の情報の予測も可能であり，このような能力こそが物体概念の理解に繋がると考えられる．

## 謝辞

本研究は，科研費（特別研究員奨励費 23-10330）の助成を受け実施したものである．

## 参考文献

[Aldous 85] Aldous, D.: Exchangeability and related topics, *École d'Été de Probabilités de Saint-Flour XIII-1983*, pp. 1–198 (1985)

[Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent dirichlet allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022 (2003)

[Fei-fei 05] Fei-fei, L.: A bayesian hierarchical model for learning natural scene categories, in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 524–531 (2005)

[Fergus 03] Fergus, R., Perona, P., and Zisserman, A.: Object Class Recognition by Unsupervised Scale-Invariant Learning, in *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 264–271 (2003)

[Griffiths 04] Griffiths, T. L. and Steyvers, M.: Finding scientific topics, *Proceedings of the National Academy of Sciences*, Vol. 101, No. Suppl. 1, pp. 5228–5235 (2004)

[Hofmann 01] Hofmann, T.: Unsupervised Learning by Probabilistic Latent Semantic Analysis, *Machine Learning*, Vol. 42, pp. 177–196 (2001)

[Ke 04] Ke, Y. and Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors, in *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 506–513 (2004)

[Lowe 04] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, p. 91 (2004)

[Nakamura 07] Nakamura, T., Nagai, T., and Iwahashi, N.: Multimodal Object Categorization by a Robot, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2415–2420 (2007)

[Nakamura 09] Nakamura, T., Nagai, T., and Iwahashi, N.: Grounding of word meanings in multimodal concepts using LDA, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3943–3948 (2009)

[Sivic 05] Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A., and Freeman, W. T.: Discovering Object Categories in Image Collections, in *IEEE International Conference on Computer Vision* (2005)

[Teh 06] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M.: Hierarchical Dirichlet Processes, *Journal of the American Statistical Association*, Vol. 101, No. 476, pp. 1566–1581 (2006)

[Wang 09] Wang, C., Blei, D., and Li, F.-F.: Simultaneous image classification and annotation, *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, Vol. 0, pp. 1903–1910 (2009)

[中村 10] 中村 友昭, 西田 匡志, 長井 隆行：把持動作による物体カテゴリの形成と認識, 情報処理学会全国大会, 5V–3 (2010)