

## 自由連想ネットワーク上の幾何学的な性質

日高 昇平 ([shhidaka@indiana.edu](mailto:shhidaka@indiana.edu))

Department of Psychology and Brain Sciences, Indiana University

1101 E. Tenth Street, Bloomington, IN 47405-7007 USA

カテゴリ化の最も基礎的な機能は、特徴を選択し情報を圧縮する事である。カテゴリの構造を調べる事は、自然な概念形成を理解するための基礎となると考えられる。幼児は新奇な物体に新奇語が付与された場合に、既存の知識を活用して新奇語の意味する事例を推定すると考えられている[4, 6, 7, 8]。従って、新奇語の汎用は、幼児が活用している自然カテゴリに関する知識を調べる一つの方法である。このような幼児の新奇語汎用に関する先行研究に基づき、著者は、自然カテゴリでは類似したカテゴリが類似した特徴空間における事例の分布パターンを持つのではないかと仮説を立て、その構造の発生する計算論的モデルを提案した[2]。この自然カテゴリの大域構造を「滑らかなカテゴリ」と呼ぶ。滑らかなカテゴリの一つの利点は、新奇カテゴリに関する予測性の高さである。滑らかなカテゴリにおいて、類似したカテゴリが類似した事例の分布を持つので、十分な数の類似カテゴリに基づけば、新奇カテゴリの事例分布を予測することができる[3]。

本研究では、自然カテゴリにおいて滑らかな構造が存在するか、具体的にデータを分析し検討した。本研究で検討したデータは成人の自由連想課題[4]である。自由連想課題では、協力者がある手がかり語(e. g., Monkey)が与えられ、その語から連想する語(e. g., Banana)を答える。この課題において協力者の答えに制約は無く、論理的にはあらゆる連想パターンが可能である。そのため、制約の小さい自由連想課題から、それでもカテゴリの滑らかな構造が見つければ、提案仮説の強い証拠となると考えられる。データの分析には、多次元正規分布を用いて連想パターンを近似する方法を用いた。このモデルでは、個々の単語の分布を多次元正規分布とみなし、二つの単語分布の確率的な重なりを連想確率とする。つまり、多くのカテゴリと連想関係をもつ単語の分布はそれらの単語分布と重複をもつと仮定し、連想パターンからカテゴリの特徴空間における配置を決定する。モデルに自体には滑らかな構造を生成する要因はなく、連想パターンから推定されるパターンから滑らかな構造の有無を分析した。

### データ

本研究では University of Southern Florida Word Association Norm[4]を用いた。これは英語の自由連想課題の結果を収録したもので、5000 以上の手がかり語、10000 以上の連想語を 6000 人以上の協力者から得ている。本研究では全ての語ではなく、最も連想される頻度の高い 100 語(全ての連想回答のうち 22%に相当)と同一の手がかり語を対象とした。

## 分析方法

与えられるデータにおいて、手がかりカテゴリ  $i (i=1,2,\dots,N)$  から連想カテゴリ  $j (j=1,2,\dots,N)$  が連想される確率を  $Q(j|i)$ 、それに対応するモデルでの連想確率を  $P(j|i)$  とする。モデルの連想確率  $P(j|i)$  はカテゴリ  $i$  の事前確率  $P(i)$ 、カテゴリ  $i, j$  間の確率的な重複を表す  $F_{ij}$  を用いて以下のように表現される。

$$P(j|i) = \frac{P(j)\exp(F_{ij})}{\sum_j^M P(j)\exp(F_{ij})}$$

ただし、 $F_{ij}$  はButtercherrya Bound[1]で、以下のように、多次元空間上でのカテゴリの平均ベクトル  $\mu_i$  および共分散行列  $\sigma_i$  を用いて表される。

$$F_{ij} = -\frac{1}{4}(\mu_i - \mu_j)^T (\sigma_i + \sigma_j)^{-1} (\mu_i - \mu_j) - \log\left(\frac{1}{2}|\sigma_i + \sigma_j|\right) + \frac{1}{2}\log(|\sigma_i||\sigma_j|)$$

パラメタ  $\{\mu_i, \sigma_i, P(i)\} (i=1,2,\dots,N)$  の推定は以下の対数尤度  $L$  を最大化することで行った。

$$L = \sum_j^M \sum_i^N Q(j|i) \log(P(j|i))$$

本研究では10次元特徴空間を分析に用いた。

## 滑らかさ指標

もし、カテゴリが滑らか(2つのカテゴリが類似であるほど、類似の分布パターンを持つ)であれば、カテゴリの中心(平均)間のユークリッド距離とカテゴリ尤度(共分散行列)のユークリッド距離には正の相関があるはずである。従って、滑らかさの指標として、2カテゴリの平均ベクトル間距離と共分散行列間距離の相関を用いた。

## 結果・考察

最も連想される頻度の高い100語の連想確率行列を、個々のカテゴリを多次元正規分布と仮定するモデルに基づき分析した。10次元特徴空間における平均・共分散によって単語を表現するモデルは、与えられた連想確率パタンの全分散のうち28%を説明可能であった

( $R^2 = 0.284$ )。推定された10次元のうち、2次元におけるカテゴリの分布パターンを図1に示す。図1では、中心にあるカテゴリほど分散が小さく、また周辺にあるカテゴリは中心方向に大きな分散を持つ傾向があった。この傾向は、10次元全ての特徴空間で一貫しており、ある特定の連想関係を持つ語が方向付けられた(ある特徴軸に特化した)分布を持ち、より一般的に連想される単語が中心に密集することを示している。次に、カテゴリの滑らかさ指標として、カテゴリ平均間の距離と共分散行列間距離の相関を分析した結果、有意な正の相関が得られた( $R=0.67$ )。これは、連想語のネットワークにおいて、カテゴリの滑ら

かな構造、すなわち類似のカテゴリが類似の分布パターンを持つ構造を示唆している。カテゴリの滑らかな構造は、理論的には、カテゴリの識別性、一般性を最適化した結果発生する大域的構造であると考えられる[3]。これを踏まえると、自由連想は、課題としては何ら制約を持たないが、そのパターンは意味知識の幾何学的な最適化を反映しているのではないかと考えられる。

(本文2073字)

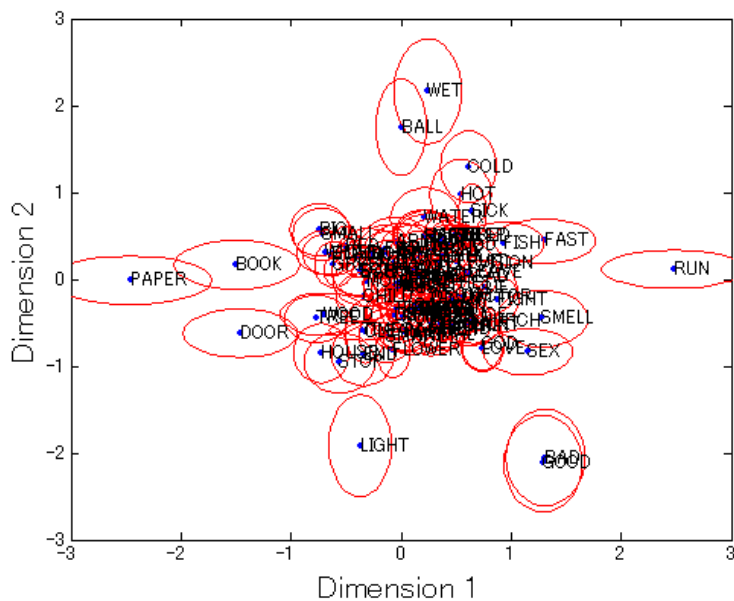


図 1: 推定された特徴空間(次元 1, 2)。カテゴリ平均を点(および単語)で、共分散行列(0.3 標準偏差領域)を楕円で示す。

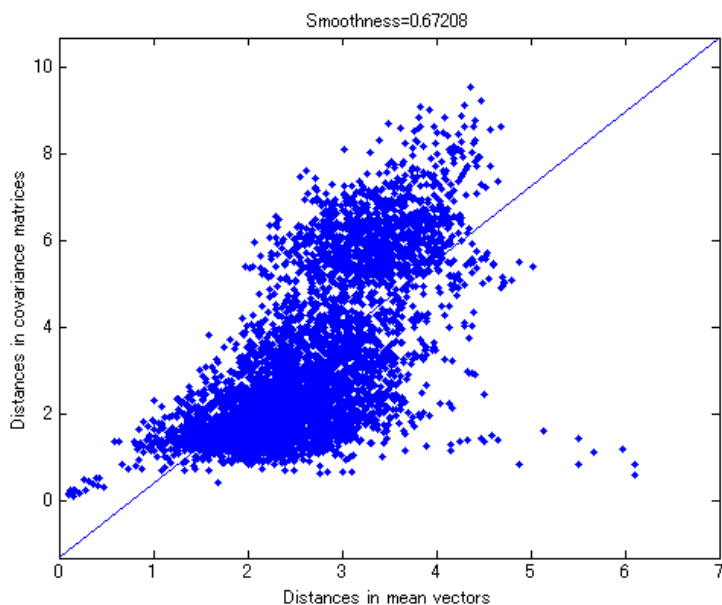


図 2: カテゴリ平均間のユークリッド距離(x 軸)と共分散行列間のユークリッド距離(y 軸)。

### 参考文献

- [1] Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification (2nd ed.)*. New York: John Wiley and Sons.
- [2] Hidaka, S. Saiki, & Smith, L. B. (2006). Semantic Packing as a Core Mechanism of Category Coherence, Fast Mapping and Basic Level Category. In *Proceedings of Twenty Eights Annual Cognitive Science Society*, 1500-1505.
- [3] Hidaka, S. & Smith, L. B.(2008) How Features Create Knowledge of Kinds. In *Proceedings of The Thirtyth Annual Meeting of Cognitive Science Society*, 1029–1035.
- [4] Markman, E. (1989). *Categorization and Naming in Children: Problems of Induction*. Cambridge, MA: MIT Press.
- [5] Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- [6] Samuelson, L., & Smith, L. B. (1999). Early Noun Vocabularies: Do Ontology, Category Structure and Syntax Correspond? *Cognition*, 73, 1-33.
- [7] Smith, L. Jones, S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object Naming Learning Provides On-the-job training for attention. *Psychological Science*, 13, 13-19.
- [8] Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological Categories Guide Young Children's Inductions of Word Meanings: Object term and substance terms. *Cognition*, 38, 179-211.