

カテゴリ数を学習するファジィ分類アルゴリズム

A fuzzy clustering algorithm for learning the number of category

西田 豊

Yutaka NISHIDA

大阪大学 大学院基礎工学研究科

Graduate School of Engineering Science, Osaka University

nishida [at] sigmath.es.osaka-u.ac.jp

Abstract

Prototype and exemplar model in category learning are described as a mixture model with different number of mixing distribution. For constructing a computational model with high descriptive validity, we need appropriate number of mixing element in mixture model. Generally appropriate number of mixing element is not given so that we need to estimate the number of mixing element as a parameter. Although DP-means is a good algorithm to achieve this purpose, estimate value of membership is obtained as a binary value. The membership parameter should be represented as fuzzy continuous value in a cognitive model. In this paper, we developed DP-means with fuzzy membership parameter.

Keywords — category learning, number of category, nonparametric Bayes, DP-means algorithm, fuzzy membership

1. 概念の計算的認知モデル

概念は認知心理学において記憶や注意と並び古くからの研究トピックの一つでありさまざまな研究アプローチがとられてきたが、計算論的なアプローチが活発に行われているのが特徴でもある (Pothos and Wills, 2011). 概念とは無数の情報から将来有効に活用できるように形成された知識といえる (松香・本田・吉川, 2010).

人間は概念を形成する学習をどのように行っているかについては類似性をベースとした2つのモデルが有名である。1つはプロトタイプモデルと呼ばれるもので、以下のような特徴がある。

- カテゴリに属する事例からそのカテゴリの情報を要約しプロトタイプを形成する。
- 新規事例はプロトタイプとの類似性によって、カテゴリに属するか否かが判断される。

- カテゴリは単一のプロトタイプによって表現される。

もう1つは事例モデルと呼ばれるもので、以下のような特徴がある。

- これまでに観察・経験した事例を記憶として保持している。
- 観察した個々の事例と新規事例との比較によってカテゴリ化を行う。
- カテゴリは個々の事例として記憶されている。

どちらのモデルも事例 \mathbf{x}_i と参照点 \mathbf{r}_j との心理的距離 d_{ij} を以下のように定義する。

$$d_{ij} = \|\mathbf{x}_i - \mathbf{r}_j\|^2 \quad (1)$$

心理的距離 d は次式で類似性 s に変換される。 λ は勾配を決定するパラメータ。

$$s_{ij} = \exp(-\lambda d_{ij}) \quad (2)$$

事例 \mathbf{x}_i をカテゴリ k に分類する確率は

$$P(R = k|x) = \frac{\sum_{j \in k} s_{ij}}{\sum_c \sum_{j \in c} s_{ij}} \quad (3)$$

で得られる。

プロトタイプモデルと事例モデルの違いは分類対象である新規事例と比較する参照点の個数の違いとまとめることができる (Rosseel, 2002).

カテゴリ学習を密度推定の問題と捉えるとプロトタイプモデルは線形判別、事例モデルはカーネル密度推定と考えることができる (Jäkel, Schölkopf and Wichmann, 2008). ある概念が平均 $\boldsymbol{\mu}$, 分散共分散行列 $\boldsymbol{\Sigma}$ を持つ正規分布で表現できるとすると、プロトタイプモデルは

$$p(\mathbf{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (4)$$

を考え、事例モデルは

$$p(\mathbf{x}) = \frac{1}{n_c} \sum_{k=1}^{n_c} N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (5)$$

を考えている。 n_c はカテゴリ c に含まれる事例数である。すなわち、プロトタイプモデルと事例モデルは各カテゴリごとの混合要素数が1の場合と全事例数 n_c の場合という混合分布モデルにおける両極に位置するモデルであり統一的に扱うことが可能である。

どちらの説が正しいのかという検証は実験的に行われてきたが、それぞれの説で支持する結果が得られている。おそらく、人間は状況にあわせて、カテゴリ化の方略を変えていてどちらの説も正しいと考えられる。しかし、プロトタイプモデルは単純すぎて、人間が学習できるカテゴリを再現できないことも多い。また、事例モデルは観察されたすべての対象を記憶しているという仮定が人間の心理モデルとしては妥当ではない。ちょうどいくらかの混合要素数 k を持つモデルが必要となる。

一般的にカテゴリを学習する状況において、あるカテゴリがさらに k 個のサブカテゴリに分類されるというような教師ラベルは与えられない。このようなネスト構造を学習するためには、教師なし学習によって達成させなければならない。すなわち、カテゴリを学習する際の混合要素数 k は教師なし分類モデルによって推定されるパラメータである。

このような認知モデルとして DP-means (Kulis and Jordan, 2012) を応用することを考える。DP-means は2つの特徴を持った、教師なし学習アルゴリズムである。まず、事前にクラスタ数を定める必要がないため、データからちょうどいいサブカテゴリの数 k を推定することが可能である。さらに、データを学習する順番が分類結果に影響するという点が、認知モデルとしての記述妥当性を高めると考えられる。

2. ディリクレプロセス混合モデルと DP-means

DP-means の目的関数は以下のようにかける。

$$\min \|\mathbf{x} - \boldsymbol{\mu}_c\|^2 + \lambda k \quad (6)$$

この目的関数を最小化するアルゴリズムは以下で与えられる。

- クラスタペナルティ λ をセットする。
 - クラスタ数 k を1, クラスタ中心 $\boldsymbol{\mu}_1$ を全平均とする。
- 1 すべてのデータポイント \mathbf{x}_i に対し
 - $c = 1, \dots, k$ について $d_{ic} = \|\mathbf{x}_i - \boldsymbol{\mu}_c\|^2$ を計算する。
 - もし $\min_c d_{ic} > \lambda$ であるならば, $k = k + 1$, データ i の所属をクラスタ k , クラスタ k の平均 $\boldsymbol{\mu}_k$ を \mathbf{x}_i とする。
 - そうでなければ, データ i の所属を $\arg \min_c d_{ic}$ とする。
 - 2 各クラスタでクラスタ平均 $\boldsymbol{\mu}_j = \frac{1}{|\ell_j|} \sum_{\mathbf{x}_i \in \ell_j} \mathbf{x}$ を求める。
- 1. と 2. を収束するまで繰り返す。

DP-means はディリクレプロセス (DP) 混合モデルのある極限として導くことができる。

正規混合モデルにおける EM アルゴリズムにおいて、混合要素の分布の分散をゼロに近づけるといふ極限を考えたとき、分類を考えているデータとクラスタ中心との距離が最も近いクラスタへの所属確率は1に収束し、それ以外のクラスタへの所属確率はゼロに収束する (Bishop, 2006)。これは K-means と同等である。

同様の議論が DP 混合モデルでも可能である。DP 混合モデルの混合要素の分布の分散をゼロに近づけるといふ極限を考えたとき、最も近いクラスタ平均を持つクラスタに分類する。もし最も近いクラスタとのユークリッド距離の2乗が λ よりも大きければ、新しいクラスタを生成する。すなわち λ は新しいクラスタの生成しやすさを調節するパラメータとして解釈することができる。

DP-means の特徴として、まず K-means アルゴリズムとよく似ている。ただし、データポイントと既存のクラスタ中心との距離が λ よりも遠い場合は新しいクラスタを生成する。DP-means と K-means を比較すると、DP-means では全平均をクラスタ中心とするシングルクラスタを初期値としてスタートするのに対し、K-means はランダムな初期値をとる。また、DP-means は分類結果が処理されるデータの順序に依存するのに対し、K-means は分類結果が初期値に依存する。

3. DP-means のファジィ拡張

DP-means ではディリクレプロセス (DP) 混合モデルのクラスタ内分散を 0 に近づけることによって、ハードクラスタリング (所属のパラメータがバイナリ) を達成する。ベイズ的な事後分布の計算が必要なくなり、K-means によく似たアルゴリズムとなるため計算速度は向上するが、事後確率として得られていたクラスタへの所属度が所属するかしないかの 2 値になってしまった。

認知的なモデルを考えると、カテゴリの所属度としてファジィなパラメータを推定できたほうがよい。しかし、DP-means の K-means ライクな高速のアルゴリズムはそのままのほうがよい。DP 混合モデルと DP-means は連続な所属度の解を得るか計算速度をとるかどちらか一方しか得られない。そこで、DP-means をエントロピー正則化の枠組み (Miyamoto, Ichihashi and Honda, 2008) を用いて、ファジィな所属度パラメータを得られるよう拡張する。

DP-means の目的関数にファジィ正則化項を加え、拡張を行う。ファジィ DP-means の目的関数は以下のように書ける。

$$\min \sum_{j=1}^k \sum_{i=1}^n u_{ij} \|\mathbf{x}_i - \boldsymbol{\mu}_j\|^2 + \lambda k + \eta^{-1} \sum_{j=1}^k \sum_{i=1}^n u_{ij} \log u_{ij} \quad (7)$$

ここで、 \mathbf{x}_i は事例、 $\boldsymbol{\mu}_j$ はサブカテゴリの平均、 u_{ij} は事例 i がサブカテゴリ j に所属する程度を意味する。 n と k はそれぞれ事例数とサブカテゴリ数である。 λ と η はそれぞれクラスタ数調整パラメータとファジィ調整パラメータである。アルゴリズムは DP-means とほぼ同じであり、所属度とカテゴリ平均のパラメータ推定にエントロピー正則化ファジィ K-means の方法を用いて以下のように求まる。

$$u_{ij} = \frac{\exp(-\lambda \|\mathbf{x}_i - \boldsymbol{\mu}_j\|^2)}{\sum_{l=1}^k \exp(-\lambda \|\mathbf{x}_i - \boldsymbol{\mu}_l\|^2)} \quad (8)$$

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^n u_{ij} \mathbf{x}_i}{\sum_{i=1}^n u_{ij}} \quad (9)$$

ファジィ DP-means であれば、K-means ライクな計算速度と連続的なファジィ所属度パラメータの両方を実現する。

4. データ解析

アヤメデータをファジィ DP-means を用いて解析し、提案手法の挙動を確認する。ファジィ DP-means を用

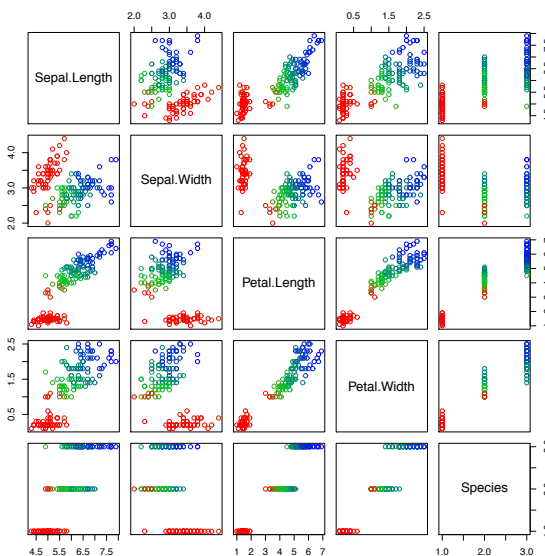


図1 ファジィ罰則: $\eta = 0.2$, クラスタ罰則: $\lambda = 10$ のときの結果. 推定されたクラスタ数は $k=3$ となった.

いて解析した結果を示す (図1)。ファジィ罰則 $\eta = 0.2$, クラスタ罰則 $\lambda = 10$ のときクラスタ数 $k = 3$ と推定された。クラスタ数を直接指定することなく、クラスタリングが実行可能というノンパラメトリックベイズの特徴を持ちつつ、DP-means では推定できなかった連続的な所属度を推定することが可能となった。

5. まとめ

概念学習におけるプロトタイプモデルと事例モデルは、混合モデルにおいて、混合要素数のパラメータがとりうる最小値か最大値かの両極に位置すると考えられる。記述妥当性の高い認知的なモデルを考えるにあたっては、適当な混合要素数のパラメータを持つ混合モデルを構成すればよいことになる。しかし、一般的に混合要素数は与えられないため推定が必要になる。DP-means は適当な混合要素数を推定するのに適したアルゴリズムであるが、事例がカテゴリに所属するかしないかという 2 値的な判断になってしまう。しかし、人間が判断する場合、カテゴリへの所属はファジィさを持って表現されるべきである。そこで、ファジィ所属度を推定できるように拡張を行った。ファジィ DP-means はあらかじめ設定しなくてはならないパラメータが 2 つ存在する。これらのパラメータをどのように設定するかはさらなる研究が必要となる。

文献

- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Jäkel, F., Schölkopf, B., and Wichmann, F.A. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review*, **15**, 256–271.
- Kulis, B. and Jordan, M.I. (2012). Revisiting k-means: New Algorithms via Bayesian Nonparametrics. *Proceedings of the 29th International Conference on Machine Learning*.
- 松香敏彦・本田秀仁・吉川詩乃 (2010). プロトタイプ理論再考. 『認知科学』, **17**, 95–109.
- Miyamoto, S., Ichihashi, H. and Honda, K. (2008). *Algorithms for Fuzzy Clustering*. Springer-Verlag.
- Pothos, E. M. and Wills, A. J. (2011). *Formal Approaches in Categorization*. Cambridge University Press.
- Rossee, Y. (2002). Mixture Models of Categorization. *Journal of Mathematical Psychology*, **46**, 178–210.