

対数線形化された無限混合正規分布に基づく 未知クラス推定確率ニューラルネットの構造最適化

南部 穰汰[†] 迎田 隆幸[‡] 柏木 僚太[§] 島 圭介[¶]
 横浜国立大学[†] 神奈川県立産業技術総合研究所[‡] 横浜国立大学[§] 横浜国立大学[¶]

1 はじめに

教師の存在するパターン分類問題において、学習時点で発生しうる全ての事象を学習することは困難であり、学習時に考慮しない未知事象の発生に対する誤分類が問題となる。この問題に対し、未知事象を含めた多クラス分類を実現できるオープンセット認識 (OSR: Open Set Recognition) 手法が数多く提案されており、我々の研究グループも、未知クラス推定確率ニューラルネット (NACGMN: Normal and Complementary Gaussian Mixture Network) [1] と呼ばれる確率的 OSR 手法を提案した。NACGMN は、混合正規分布 (GMM: Gaussian Mixture Model) と混合余事象分布 (CGMM: Complementary Gaussian Mixture Model) を内包した確率 NN (図 1) であり、少量の学習データからも高精度な識別が可能という特徴を有する。しかしながら、NACGMN が内包する正規分布の数などのモデルパラメータを経験的に設定する必要があり、NACGMN を含む通常の OSR 手法では単純に異常を検知するだけで異常値に関する情報の獲得までは想定されていない。そこで、本稿では適応的に異常値に関する情報を獲得し、モデル構造を修正しながら運用可能な新たなオープンセット認識手法を提案する。提案法では NACGMN の学習段階に無限混合正規モデル (IGMM: Infinite Gaussian Mixture Model) [2] を導入し、学習データに対して適当なハイパーパラメータの学習的獲得を実現する。さらに、IGMM を学習データに適用することで OSR 手法の教師なし学習、未知事象サンプルに適用することで新規クラスの獲得まで可能になる。IGMM の実行時に懸念される学習時間の長さについても、提案モデルでは冗長な確率演算を対数線形化 [3] により線形演算に変換し、学習の高速化を試みる。

2 FAST-IGMM を利用した NACGMN の構造最適化

2.1 GMM と CGMM に基づくオープンセット認識

まず、NACGMN の構造について述べる。NACGMN は既知クラスに対応する正規分布 $g(\mathbf{x}; k, m)$ と未知クラスに対応する余事象分布 $h(\mathbf{x}; k, m)$ を内包した識別モデルであり、これらの確率分布を用いて未知クラスを含む各クラスの事後確率を算出する。既知クラス ω_k と未知クラス ω_{unknown} の事後確率はそれぞれ次式によって表される。

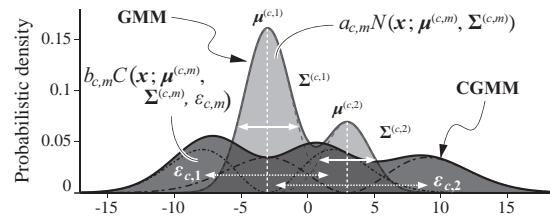


図 1 混合正規分布と混合余事象分布の例 [1]

$$P(s_i|\mathbf{x}) = \begin{cases} \frac{1-p_0}{F(\mathbf{x})} \sum_{m=1}^{M_k} \alpha_{k,m} g(\mathbf{x}; k, m) & (s_i = \omega_k) \\ \frac{p_0}{F(\mathbf{x})} \sum_{k'=1}^K \sum_{m'=1}^{M'_k} \beta_{k',m'} h(\mathbf{x}; k', m') & (s_i = \omega_{\text{unknown}}) \end{cases} \quad (1)$$

$$F(\mathbf{x}) = (1-p_0) \sum_{k'=1}^K \sum_{m'=1}^{M'_k} \alpha_{k',m'} g(\mathbf{x}; k', m') + p_0 \sum_{k'=1}^K \sum_{m'=1}^{M'_k} \beta_{k',m'} h(\mathbf{x}; k', m') \quad (2)$$

ただし、 s_i は i 番目のデータ $\mathbf{x}_i \in \mathbb{R}^D$ の所属クラス、 p_0 は未知クラスの事前確率、 K は既知のクラス数、 M_k はクラス ω_k の混合正規分布が含む正規分布の数に相当するコンポーネント数、 $\alpha_{k,m}, \beta_{k,m}$ は各分布の混合度である。

2.2 パラメータ推定部

パラメータ推定部では、NACGMN で識別を行う際に必要な GMM のコンポーネント数、教師ラベルを決定するために IGMM を用いたクラスタリングを行う。IGMM では、中華料理店過程 (CRP: Chinese Restaurant Process) によってクラス $k = (1, \dots, K)$ を決定する。

$$s_i | \alpha \sim CRP(\alpha) \quad (3)$$

$$(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) | \phi \sim \mathcal{N}(\boldsymbol{\mu}_k; \boldsymbol{\mu}_k, (\beta \boldsymbol{\Lambda}_k)^{-1}) \cdot \mathcal{W}(\boldsymbol{\Lambda}_k; \nu, \mathbf{S}) \quad (4)$$

$$\mathbf{x}_i | s_i = c \sim \mathcal{N}(\boldsymbol{\mu}_k; \boldsymbol{\Lambda}_k^{-1}) \quad (5)$$

ここで、CRP における α は集中度パラメータ、 $\mathcal{N}(\cdot), \mathcal{W}(\cdot)$ は正規分布とウィシャート分布を表し、 $\phi = \{\boldsymbol{\mu}_0, \beta, \nu, \mathbf{S}\}$ はハイパーパラメータであり、全てのクラスは平均ベクトル $\boldsymbol{\mu}_k$ と精度行列 $\boldsymbol{\Lambda}_k$ を持つ。式 (3) においてデータ \mathbf{x}_i のクラス番号 s_i を決定し、同時に識別クラス数の増減を行う。その後、正規ウィシャート分布から $\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k$ をサンプリングし、これらのパラメータを用いてモデル全体の事後確率を以下の式から算出する。

$$\nu = \log P(s) + \sum_{i=1}^c \left(\log G_0(\theta_i) + \sum_{\mathbf{x}_k \in \omega_i} \mathbf{w}_i \mathbf{X}_i \right) \quad (6)$$

$$\mathbf{X} = \left[\mathbf{1}, \mathbf{x}_i^T, x_1^2, x_1 x_2, \dots, x_1 x_D, x_2^2, x_2 x_3, \dots, x_D^2 \right]^T \quad (7)$$

Structural Optimization of Probabilistic Neural Network Based on Log Linearized Infinite Gaussian Mixture Model

[†] Jota Nambu, Yokohama National University
[‡] Mukaeda Takayuki, Institute of Industrial Science and Technology

[§] Ryota Kashiwagi, Yokohama National University

[¶] Keisuke Shima, Yokohama National University

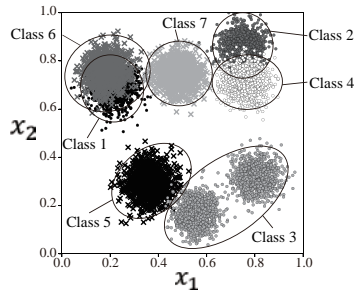


図2 学習対象とした人工データ

表1 学習初期段階における構造修正結果

		Predicted label							Predicted label		
		Class 1	Class 2	Class 3	Class 4	Unknown			Class 1	Class 3	Unknown
True label	Class 1	897	0	0	0	3	True label	Class 1	897	0	3
	Class 2	0	886	0	0	14		Class 3	0	886	14
	Class 3	0	0	436	450	14		Unknown	922	13	1765
	Class 4	0	883	0	0	17					
Unknown	922	2	0	13	1763						

図3 新規クラス獲得後の構造修正結果

		Predicted label									Predicted label			
		Class 1	Class 3	Class 5	Class 6	Class 7	Unknown			Class 3	Class 5	Class 7	Unknown	
True label	Class 1	890	0	5	0	0	5	True label	Class 3	899	0	0	1	
	Class 3	0	0	0	445	454	1		Class 5	0	850	45	5	
	Class 5	0	850	0	0	45	5		Class 7	0	0	900	0	
	Class 6	895	0	0	0	0	5		Unknown	0	3	0	1797	
	Class 7	0	0	900	0	0	0							
Unknown	0	0	3	0	0	1797								

ここで、 $P(s)$ はイーウェンスの抽出公式であり、 w_i は i 番目のクラス対数の対数線形化された重みである。

この高速化 IGMM を学習データ全体に適用すれば教師なし学習の実現が可能であり、未知クラスに識別されたデータ群に適用すれば新規クラスの獲得が可能となる。

2.3 NACGMN の適応的構造修正

識別部ではパラメータ推定部によって獲得された各正規分布のパラメータを NACGMN の重み係数に変換し、未知クラスを含めて識別を行う。識別するための既知クラスを新たに追加する際には、学習済み NACGMN にパラメータ推定部で得た新規クラスの GMM、および CGMM を対数線形化して並列に接続する。これによって、容易に識別器の拡張が可能である。クラスの削除も同様であり、提案モデルはその構造を適用的に変更することが可能である。また、この操作を繰り返すことで、データ構造や利用目的に対応した望ましい分類モデルを進化的に獲得することが可能である。

3 人工データを利用した基礎検証

3.1 実験条件

実験には図2に示す2次元正規分布に基づいた7クラス人工データセットを利用した。学習時には、各クラス1000点の中から無作為に100点ずつを抽出して学習データを、残りの900点ずつでテストデータを作成した。IGMMの初期化は教師なしデータに対する密度準拠クラスタリング手法であるDBSCANを用いて行い、終了条件は学習データ点数の150倍のループ数の実施、または最大尤度の更新が学習データ点数の30倍のループ数の間行われないこととした。また、コンポーネント数は1とした。初めに、図2中に円で示された1、

2、3、4クラスを学習クラスとし、その他の×印で示されたクラスを学習済みでない未知クラスとしてOSRの精度を評価した。

3.2 結果と考察

識別によって得られた混同行列を図1の左側に示す。混同行列から、Class 2とClass 4が同一のクラスとして識別され、Class 3が2つに分裂していることが確認された。よって、Class 2とClass 4は与えられたデータからは識別が困難である類似クラスであると判断し学習データから排除する。一方、Class 3における2つのクラスへの分裂は、このクラスが2コンポーネントの混合正規分布に由来するものであるというIGMMの推定を反映している。したがって、分裂している2クラス分のデータをClass 3に所属するデータであるとして統合して扱う。これらのデータ加工によって、混同行列は図1の右側に示される結果となる。加えて、データ加工前後において未知データの中にClass 1に誤識別しているデータが存在していることが確認された。

ここで、未知であったデータを学習に追加することで、データ全体のクラス構造を取得することが可能である。前回の学習に含めなかった5、6、7クラスを学習データに追加し、前回の学習結果に従ってClass 2とClass 4を学習データから取り除き再び学習を行った結果を図3の左側に示す。この混同行列から、Class 1とClass 6が同一であると識別され、Class 3は前回の学習同様にコンポーネント数が2であることを示している。これに従い、識別困難であるClass 1とClass 6を排除しClass 3のコンポーネントを統合することによって図3右側に示す混同行列となる。本実験で扱ったデータセットにおいて正確に識別できるクラスがClass 3とClass 5、Class 7であることが確認できる。

また、学習の初期段階において、提案モデルとIGMMの確率演算に対数線形化を用いない従来モデルで学習時間の比較を行った。それぞれ3回実行した平均値を比較した結果、提案モデルの実行時間は従来の約0.62倍であり、高速化を実現していることが確認された。

4 まとめ

本稿では、OSRの学習として対数線形化を取り入れたIGMMをモデルパラメータ算出に利用することでデータの構造を獲得する手法を提案した。識別結果に従い臨機応変にクラス数を調整することで、識別可能クラスを柔軟に最適化することが可能であることが確認できた。

参考文献

- [1] 迎田隆幸, 島圭介.: 混合余事象分布に基づく未学習クラス推定確率ニューラルネットワーク. 計測自動制御学会論文集, Vol. 56, No. 12, pp. 532 - 540, 2020.
- [2] Matza, A. and Bistriz, Y.: Infinite Gaussian Mixture Modeling With an Improved Estimation of the Number of Clusters. AAAI, 35, 8921-8929, 2021.
- [3] T. Tsuji, O. Fukuda, H. Ichinobe, and M. Kaneko.: A log-linearized Gaussian mixture network and its application to EEG pattern classification. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), Vol. 29, No. 1, pp. 60 - 72, 1999.