

一般化期待情報量基準最小化法と応用

Generalized Expectation Minimization of Information Criterion Algorithm and Applications

藤巻 遼平*1 森永 聡*1
Ryohei Fujimaki Satoshi Morinaga

*1 NEC 情報・メディアプロセッシング研究所
NEC Information Media Processing Laboratory

We propose generalized expectation minimization of information criterion (GEMIC) to resolve the model selection for heterogeneous mixture models (MSHM). For MSHM, we must learn the number of components and the types of components as well as their parameters. GEMIC makes a significant improvement on the combinatorial scalability issue of MSHM, especially when the number of component candidates itself grows explosively. Further, GEMIC makes us possible to use a broad range of information criteria including Akaike's information criterion, minimum description length, etc. In addition, we explain its two applications; sparse covariance selection and feature subset selection. Based on artificial, benchmark and real world data, we show that GEMIC performs well in the above two applications.

1. はじめに

本稿は、異種多様なコンポーネントから構成される混合モデル（異種混合モデルと呼ぶ）に対するモデル選択問題を解決するために、一般化期待情報量基準最小化法（generalized expectation minimization of information criterion; GEMIC）を提案する。異種混合モデルのクラスには、金融や医療分野でしばしば観察される、対数正規分布と正規分布の混合分布モデルのような、比較的単純なモデルから、異なる次数をもつ混合自己回帰モデル [Wong 00]、階層隠れ変数モデル [Bishop 98] といった複雑なモデルが含まれ、金融、ライフログ分析、システム同定など、様々な応用分野でその重要性が高まっている。

異種混合モデルの学習問題（異種混合モデル選択と呼ぶ）では、混合比や各コンポーネントのパラメータだけでなく、適切なコンポーネントの種類およびその数をデータから推定する必要がある。この問題では、モデルの探索空間（モデル候補の数）が、コンポーネント候補*1の組合せによって決まり、コンポーネント数とコンポーネント候補数の増加と共に、モデルの探索空間（モデル候補の数）が爆発的に広がるため、効率的な探索アルゴリズムが必要となる。

この問題に対する一つの有力なアプローチとして、期待情報量基準最小化法（Expectation Minimization of Information Criteria; EMIC）が挙げられる [藤巻 09]。EMIC は EM アルゴリズムをモデル選択へ拡張したアルゴリズムで、E ステップで完全変数（観測変数とデータのコンポーネント割当を表す隠れ変数）の情報量基準（赤池情報量基準 [Akaike 73]、最小記述長 [Rissanen 78] など）の隠れ変数の事後分布に対する期待値を計算し、M ステップでコンポーネントの種類とパラメータを最適化する事で、コンポーネント候補の組合せ問題を回避し、コンポーネント候補数とコンポーネント数に対して線形時間で異種混合モデル選択を解く事が可能である。EMIC の弱点としては、M ステップの最適化の計算コストが高い場合*2に、

連絡先: 藤巻遼平, NEC 情報・メディアプロセッシング研究所 データマイニングリサーチグループ, 神奈川県川崎市中原区下沼部 1753, r-fujimaki@bx.jp.nec.com

*1 上述の対数正規分布と正規分布の混合の例では、コンポーネント候補は対数正規分布と正規分布。

*2 コンポーネント候補の数が大きい、コンポーネント間に相互作用が存在する等。

計算コストが高くなる点が挙げられる。

本稿では、上述の弱点を改善するために、EMIC を一般化 EMIC (Generalized EMIC; GEMIC) へ拡張する。この拡張は、EM アルゴリズムが一般化 EM アルゴリズム [McLachlan 07] に拡張されるのと同様の考えに基づき、M ステップにおける最適化問題を緩和する比較的単純なものであるが、これによってより複雑な異種混合モデル選択問題を現実的な計算時間で解く事が可能となる。本稿では、GEMIC の適用例として、混合正規分布の疎共分散選択問題と、混合分類モデルの疎属性選択問題を説明し、人工データと UCI データ [Asuncion 07] によって、GEMIC の有効性を示す。

2. 一般化期待情報量基準最小化法

X を D 次元の確率変数とし、 $P(\bullet; \star)$ を \star でパラメトライズされる \bullet の確率密度（質量）関数とする。本節では紙面の都合からデータの確率モデル $P(X; \theta)$ の場合についてのみ説明をするが、同様の議論はターゲット Y に対する条件付分布 $P(Y|X; \theta)$ の場合についても成立する。また、本節では GEMIC の一般形についてのみ説明をする。赤池情報量基準や最小記述長を利用した場合の具体形は、[Fujimaki 09, 藤巻 09] を参照。

2.1 異種混合モデル選択

コンポーネント候補集合を $S = \{V_j | j = 1, \dots, |S|\}$ と表記する。ただし、 $|\circ|$ は集合 \circ の要素数を表す。 $V_j = \{P(X; \phi^{V_j}) | \phi^{V_j} \in \Phi_j\}$ はパラメトリックな確率分布族とし、 $\phi^{V_j} = (\phi_1^{V_j}, \dots, \phi_{J_{V_j}}^{V_j})$ はパラメータ集合 Φ_j の元とする。異種混合モデルでは、 V_j に応じて、 $P(X; \phi^{V_j})$ の関数形や ϕ^{V_j} の次元 J_{V_j} が異なる点で注意が必要である。

S に対して、異種混合モデル H は、

$$\left\{ P(X; \theta) = \sum_{c=1}^C \pi_c P(X; \phi_c^{S_c}) \right\}, \quad (1)$$

と定義される*3。ただし、 C はコンポーネント数、 $S_c \in S$ ($c = 1, \dots, C$) は各コンポーネントの種類、 π_c は混合比を表わす。

*3 これ以降、表記の簡略化のため、 θ^H の上付文字 H は省略する。

また、モデル候補の集合を $\mathcal{H} = \{H_i | i = 1, \dots\}$ と表記する事とする。異種混合モデル選択問題は、適切なモデル $\bar{H} \in \mathcal{H}$ およびパラメータ $\bar{\theta}$ を推定する問題と定義する事ができる。

2.2 完全データの期待情報量基準

本節では、GEMIC で中心的な役割を果たす、完全データの期待情報量基準を導入する。まず、観測データおよびその隠れ変数を、 $\mathbf{x}^N = \mathbf{x}_1, \dots, \mathbf{x}_N$ および $\mathbf{z}^N = z_1, \dots, z_N$ とする。ただし、 $z_n = (z_{n1}, \dots, z_{nC}) \in \{1, 0\}^C$ は、コンポーネント割当を表すとする。なお、 $(\mathbf{x}^N, \mathbf{z}^N)$ の組は完全データと呼ばれ、対する \mathbf{x}^N は不完全変数データと呼ばれる。

GEMIC では、 $P(\mathbf{x}^N | \mathbf{z}^N; \theta)$ 及び $P(\mathbf{z}^N; \theta)$ に関する情報量基準の和 $IC_{x|z}(\mathbf{x}^N | \mathbf{z}^N; H) + IC_z(\mathbf{z}^N; H)$ を考える。ここで、各 $P(\mathbf{x}^N; \phi_c^{S_c})$ に隠れ変数を条件とした独立を仮定すると、 $IC_{x|z}(\mathbf{x}^N | \mathbf{z}^N; H)$ は、

$$IC_{x|z}(\mathbf{x}^N | \mathbf{z}^N; H) = \sum_{c=1}^C IC_c(\mathbf{x}^{N_c}; \phi_c^{S_c}), \quad (2)$$

と書く事が出来る。ただし、 $\mathbf{x}^{N_c} = \mathbf{x}_{n_1}, \mathbf{x}_{n_2}, \dots (z_{n_i c} = 1)$ は c 番目のコンポーネントから生成されたデータの列で、 N_c はその数を表す。

完全変数の情報量基準を学習の目的関数とした場合には、 S_c と $\phi_c^{S_c}$ を個別に最適化可能であり、コンポーネント候補の組合せによる計算量の爆発を回避できるが、実際には隠れ変数を観測できないため、 $IC_{x|z}(\mathbf{x}^N | \mathbf{z}^N; H)$ と $IC_z(\mathbf{z}^N; H)$ を計算する事が出来ない。そこで、GEMIC では EM アルゴリズムと同様に、隠れ変数の事後分布^{*4}、

$$P(\mathbf{z}^N | \mathbf{x}^N, \theta^*, H^*) \propto \prod_{n=1}^N \prod_{c=1}^C (\pi_c P(\mathbf{x}_n; \phi_c^{S_c^*}))^{z_{nc}}. \quad (3)$$

に関する $IC_{x|z}(\mathbf{x}^N | \mathbf{z}^N; H) + IC_z(\mathbf{z}^N; H)$ の期待値を、モデル選択の最適化基準とする。

2.3 最適化アルゴリズム

GEMIC の最適化アルゴリズムを概説すると、1) $C = 1, \dots, C_{\max}$ に対して、 $\bar{\pi}$ 、 \bar{S}_c 及び $\bar{\phi}_c^{S_c}$ を、以降で説明する EM ステップによって計算する、2) 計算された $\bar{\pi}$ 、 \bar{S}_c 及び $\bar{\phi}_c^{S_c}$ に対して、不完全データの期待情報量基準 $IC_x(\mathbf{x}^N; H)$ を最小化する事で、最適なコンポーネント数 \bar{C} を決定する。ここで、2) のステップでは、 $IC_x(\mathbf{x}^N; H)$ を最適化の基準としている点で注意が必要である。EM アルゴリズムが完全データの期待対数尤度の最大化によって、不完全データの対数尤度を最大化可能であると同様に、赤池情報量基準、最小記述長などを含む複数の情報量基準について、1) の EM ステップによって $IC_x(\mathbf{x}^N; H)$ が最小化可能である事が示されている [藤巻 09]。

以降では、上付文字の (t) は t 回目の EM ステップを、 $E_Z^{(t)}[\bullet]$ は \bullet の $P(\mathbf{z}^N | \mathbf{x}^N, \theta^{(t-1)}, H^{(t-1)})$ に関する期待値を表すとする。

E ステップ: E ステップでは、 $E_Z^{(t)}[IC_{x|z}(\mathbf{x}^N | \mathbf{z}^N; H) + IC_z(\mathbf{z}^N; H)]$ の計算を行う。このために、多くの情報量基準で隠れ変数の期待値 $E_Z^{(t)}[z_{nc}]$ が、また最小記述長などでは $E_Z^{(t)}[\log(\sum_{n=1}^N z_{nc})]$ の計算が必要となる。まず、前者は EM アルゴリズムと同様に容易に計算が可能である。次に、後者を直接計算すると $\mathcal{O}(2^N)$ の計算コストが必要となるが、計算

コストが $\mathcal{O}(N)$ である効率的な近似アルゴリズムが提案されている [Fujimaki 09]。

M ステップ: M ステップでは、 $\pi^{(t)}$ 、 $S_c^{(t)}$ 及び $\phi_c^{S_c^{(t)}}$ の最適化を行う。まず、混合比 π は、以下のように推定される。

$$\bar{\pi}^{(t)} = \arg \min_{\pi} \{E_Z^{(t)}[IC_z(\mathbf{z}^N; H)]\}. \quad (4)$$

次に、EMIC では $S_c^{(t)}$ 及び $\phi_c^{S_c^{(t)}}$ を

$$S_c^{(t)}, \phi_c^{S_c^{(t)}} = \arg \min_{S_c, \phi_c^{S_c}} E_Z^{(t)}[IC_c(\mathbf{x}^{N_c}; \phi_c^{S_c})]. \quad (5)$$

なる最小化問題を解く事で得る。

GEMIC では、(5) の最小化問題を解く事が計算的に困難な場合に、

$$E_Z^{(t)}[IC_c(\mathbf{x}^{N_c}; \phi_c^{S_c^{(t)}})] < E_Z^{(t)}[IC_c(\mathbf{x}^{N_c}; \phi_c^{S_c^{(t-1)}})] \quad (6)$$

の不等式が成立するように、 $E_Z^{(t)}[IC_c(\mathbf{x}^{N_c}; \phi_c^{S_c})]$ を減少させる $S_c^{(t)}$ 及び $\phi_c^{S_c^{(t)}}$ を探索し、M ステップを終了する。GEMIC では、(5) が計算的に扱いづらい場合でも、適切な探索アルゴリズムを利用する事で (6) を解き、最適化を効率的に解く事が可能である。なお、(6) の単調減少性と、[藤巻 09] の定理 1, 2 を利用する事で、GEMIC の EM ステップが $IC_x(\mathbf{x}^N; H)$ の局所最適解を得る事が証明可能であるが、本稿では紙面の都合上割愛する。

3. 適用例

3.1 混合正規分布における疎共分散選択

正規分布の疎共分散選択は、データから共分散行列を推定する際に、非対角成分の多くがゼロとなるように推定し、比較的少数のデータから安定的に共分散行列を推定し、属性間の依存構造の分析や特徴的な部分空間の発見を行う技術である [Marlin 09]。混合正規分布における疎共分散選択では、各コンポーネントが異なる共分散構造 (つまり非ゼロ要素の位置や数) を考慮する必要があり、異種混合モデル選択として捉える事が可能である。

異種疎共分散をもつ混合正規分布 (異種混合正規分布と呼ぶ) は、

$$P_{mg}(X; \theta_{mg}, \Gamma) = \sum_{c=1}^C \pi_c \mathcal{N}(X; \mu_c, \Gamma_c \bullet \Sigma_c), \quad (7)$$

と表現される。ただし、 $\mathcal{N}(X; *, *)$ は平均 $*$ 、共分散行列 $*$ をもつ正規分布、演算 \bullet は行列やベクトルの要素同士の積、 Γ_c は $D \times D$ のバイナリ行列とする。また、 $\theta_{mg} = (\pi, \mu_1, \dots, \mu_C, \Sigma_1, \dots, \Sigma_C)$ とする。

(7) では、コンポーネント候補 S は、可能な Γ_c 全体 (非対角成分の 1, 0 の組合せ) である。したがって、コンポーネント候補数 $|S|$ のオーダーは $\mathcal{O}(2^{D^2})$ となり、次元と共に急速に増加するため、(5) を (6) へ緩和する事で、計算爆発を回避して解を得る事が必要となる。本稿では、(6) を満たす $S_c^{(t)}$ 及び $\phi_c^{S_c^{(t)}}$ を得る方法として、1 回の M ステップで Γ_c の非対角要素を個別に 1 回ずつ更新する単純なアルゴリズムを実験で評価する。

4 一時的なモデル H^ およびパラメータ θ^* が与えられている場合。

3.2 混合分類モデルにおける属性選択

属性選択問題は機械学習の最重要問題の一つで、特に、近年は疎属性の選択が注目を集めている [Liu 09, Kawahara 09]. GEMIC の重要な応用として、混合ロジスティック回帰、混合エキスパートモデル [Jacobs 91] などの混合分類モデルに対する疎属性選択問題が挙げられる。この問題は、混合する各分類モデルごとに、適切な疎属性を選択するという観点で、異種混合モデル選択と捉える事ができる。本節では、基本的なモデルとして、コンポーネントごとに異なる疎属性を利用する混合ロジスティック回帰モデル (異種混合ロジスティック回帰と呼ぶ) を説明する。

$Y \in \{1, \dots, M\}$ を X のクラスラベルを表す確率変数とし、 M はクラス数とする。異種混合ロジスティック回帰モデルは、

$$P_{ml}(Y = m|X; \theta_{ml}, \alpha) = \sum_{c=1}^C \pi_c \frac{\exp(\alpha_{cm} \cdot \beta_{cm} \bar{X})}{\sum_{k=1}^M \exp(\alpha_{ck} \cdot \beta_{ck} \bar{X})}, \quad (8)$$

と表現される。ただし、 $\bar{X} = (1, X)$ である。また、 $\alpha_{cm} \in \{0, 1\}^{D+1}$ は属性のインジケータベクトル (d 番目の属性を利用する場合には $\alpha_{cmd} = 1$)、 β_{cm} は c 番目のコンポーネントの m 番目のクラスに関する重みベクトルである。また、 $\theta_{ml} = (\pi, \beta_1, \dots, \beta_C)$ 、 $\alpha = (\alpha_1, \dots, \alpha_C)$ 及び $\alpha_c = (\alpha_{c1}, \dots, \alpha_{cM})$ とする。

(8) では、コンポーネント候補 S は、可能な α_c の全体 ($M(D+1)$ 要素の 1,0 の組合せ) である。したがって、コンポーネント候補数 $|S|$ のオーダーは $\mathcal{O}(2^{M(D+1)})$ となり、異種混合正規分布の場合と同様に、 $|S|$ が次元およびクラス数と共に次元と共に急速に増加するため、(5) を (6) へ緩和する事で、計算爆発を回避して解を得る事が必要となる。本稿では、異種混合正規分布の場合と同様に、1 回の M ステップで α_c の各要素を個別に 1 回ずつ更新する単純なアルゴリズムを実験で評価する。

4. 実験と考察

4.1 疎共分散選択

本節では、人工データを用いて、混合正規分布に対する疎共分散選択に対する GEMIC と EMIC の比較を行う。情報量基準は、赤池情報量基準 (AIC) と最小記述長 (MDL) を利用し、各アルゴリズムを $GEMIC_{AIC}$ 、 $GEMIC_{MDL}$ 、 $EMIC_{AIC}$ 、 $EMIC_{MDL}$ と表記する。

4.1.1 実験設定

上付 * は真のモデルを表すとする。 π 、 μ_c^* 及び Σ_c^* は $[0.1, 1]$ 、 $[-10, 10]$ 及び $[0.2, 0.8]$ からランダムに生成した。 Γ_c^* は、非対角成分が 0 である確率を $[0, 1]$ からランダムに決定し、その確率にしたがって対角要素を同じくランダムに決定した。次元 D の増加に対する比較をするために、コンポーネント数と探索する最大コンポーネント数は $C^* = 5$ 及び $C_{max} = 10$ と固定した。訓練データおよびテストデータの数は、 $N = 1000$ 及び $N_{te} = 10000$ とし、真のモデルにしたがってランダムに生成した。実験結果は 10 回の試行に対する平均値であり、各試行で局所解を避けるために 10 回のリスタートを行っている。

4.1.2 評価基準

評価基準は、1) 学習に要した CPU 時間 (計算コストの評価)、2) Γ_c^* と $\bar{\Gamma}_c$ の非対角成分の一致する割合 (モデル選択性能の評価)、3) 真のモデルと推定モデルで計算されるテストデータに対する対数尤度の比 (汎化性能の評価)、の 3 つの観

表 1: UCI データのまとめ。

data	N	D	M	data	N	D	M
iris	150	4	3	phoneme	5405	5	2
heart	267	44	2	spambase	4601	57	2
onehr	1848	72	2	wine_qual	6497	11	3
pima	768	8	2	coil	9822	85	2

点で実施した。より正確には、2 つ目の評価基準は、

$$T(\bar{H}, H^*) = \frac{2}{C^* D (D-1)} \sum_{c=1}^{C^*} \sum_{i=1}^D \sum_{j=i+1}^D |(\bar{\Gamma}_c)_{ij} - (\Gamma_c^*)_{ij}|, \quad (9)$$

によって算出し、3 つ目の評価基準は、

$$L(\bar{H}, H^*) = 1 - \frac{\log P(\mathbf{x}_{te}^{N_{te}}; H^*) - \log P(\mathbf{x}_{te}^{N_{te}}; \bar{H})}{|\log P(\mathbf{x}_{te}^{N_{te}}; H^*)|}. \quad (10)$$

によって算出した。

4.1.3 結果

まず、次元 D の増加と共に EMIC (AIC と MDL) は CPU 時間が急激に増大し、6 次元以上では計算時間の関係で評価ができなかった*5 (図 1 左)。一方で、GEMIC (AIC と MDL) では、次元とともに緩やかに計算時間が増加し、計算量において大きなメリットがある事が確認できる。

次に、モデル選択性能 (図 1 中央) については、 $EMIC_{AIC}$ と $EMIC_{MDL}$ の方が $GEMIC_{AIC}$ と $GEMIC_{MDL}$ よりも、5 から 10% 程度 $T(\bar{H}, H^*)$ が高い事が確認できる。これは、EMIC が M ステップでモデルの全探索をするのに対して、GEMIC では局所的な探索をしているため、妥当な結果と言える。高次元では計算コストの問題で EMIC を利用する事は難しいため、低次元では EMIC、高次元では GEMIC という、利用可能な計算コストに応じた使い分けが重要と言える。AIC と MDL の比較については、GEMIC と EMIC の両方で、MDL を利用した方がモデル選択性能が高いという結果であった。

汎化性能 (図 1 右) については、すべての手法がほぼ同等で、真のモデルと比較して 1 から 2% 程度対数尤度が低いという結果であった。なお、図 1 では訓練データ数を $N = 1000$ と固定しているが、訓練データ数を増加させると、 $T(\bar{H}, H^*)$ と $L(\bar{H}, H^*)$ は両者とも 1 に近づく傾向であった。

以上の結果から、EMIC と比較して GEMIC は大きな次元の異種混合正規分布の疎共分散選択を解くのに適していると言える。なお、 $D = 100$ に対しては、CPU 時間が $4.55e+4$ 、 $T(\bar{H}, H^*)$ が 0.176、 $L(\bar{H}, H^*)$ が 0.864 であった。 $T(\bar{H}, H^*)$ が低い点については、1) GEMIC の探索アルゴリズムが単純なため高次元で局所解につかまりやすい、2) $N = 1000$ に対して共分散行列の非対角成分が約 5000 であり本質的に難しい、という 2 つの問題があり、今後は GEMIC の M ステップでより洗練された方法を利用する事が課題として挙げられる。

4.2 属性選択

本節では、混合ロジスティック回帰 (MixLogit) に対する属性選択について、GEMIC と EMIC の比較を行う。

4.2.1 UCI データ

評価データとして、表 1 に示される 10 種類の UCI データ [Asuncion 07] を利用した。なお、wine quality データは、元々は 10 クラスであるが、高品質、低品質のクラスに属するデータが少数のため、本実験では品質が低 (品質が 4 以下)、普通 (品質が 5,6)、高 (品質が 7 以上) と 3 クラスに再分類した。

*5 6 次元でのコンポーネント候補数は $2^{15} = 32768$ 、10 次元では $2^{45} = 3.51843721 \times 1e13$ 。

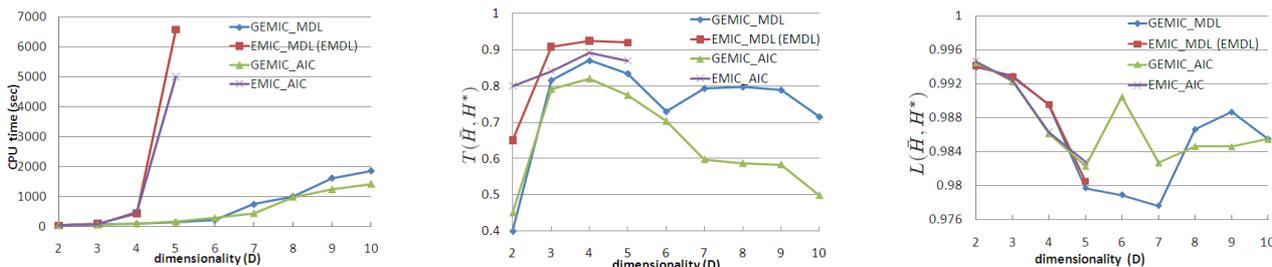


図 1: GEMIC と EMIC の CPU 時間 (左), モデル選択性能 (中央), 汎化性能 (右) の比較. 横軸はデータの次元.

表 2: 各手法に対する分類性能 (acc) と属性の疎性 (spar) の比較. 太字は acc が最大のもので, 信頼水準 95% の片側 t 検定で, 最大なものよりも有意に悪くないものを表す.

モデル選択	MixLogit		MixLogit		MixLogit		SparLogit		SparSVM	
	無		GEMIC _{MDL}		GEMIC _{AIC}		L ₁		L ₁	
	acc	spar	acc	spar	acc	spar	acc	spar	acc	spar
iris	0.980	×	0.960	0.585	0.953	0.627	0.953	0.927	0.953	0.980
heart	0.802	×	0.794	0.247	0.802	0.266	0.772	0.902	0.790	0.993
onehr	0.950	×	0.967	0.075	0.968	0.107	0.967	0.360	0.968	0.600
pima	0.772	×	0.768	0.633	0.762	0.656	0.775	1.00	0.772	0.989
phoneme	0.787	×	0.782	0.721	0.788	0.831	0.757	1.00	0.759	1.00
spambase	0.901	×	0.913	0.528	0.917	0.769	0.920	0.957	0.919	0.983
wine_qual	0.774	×	0.775	0.346	0.773	0.432	0.767	0.844	0.773	0.956
coil	0.939	×	0.940	0.195	0.939	0.361	0.940	0.605	0.940	0.749

4.2.2 評価指標

疎属性選択の目的は, 1) 疎属性を抽出し少数の属性で分類を行う事で分類の解釈性を向上する, 2) ノイズとなる属性を除去し分類性能を向上する, という 2 点が挙げられる. ここで, 第 1 の目的のために α の疎性 (非ゼロ要素の割合), 第 2 の目的のために分類に成功したテストデータの割合を評価した. なお, 評価は 10 分割の交差検定によって実施した.

4.2.3 比較手法

GEMIC と 1) 属性選択無の MixLogit (疎性無), 2) L₁ 正則化ロジスティック回帰 (SparLogit), 3) L₁ 正則化サポートベクトルマシン (SparSVM) の 3 つの手法を比較した. なお, SparLogit と SparSVM の実装には, LIBLINER [Fan 08] を利用し, L₁ 正則化の超パラメータは交差検定によって決定した.

4.2.4 結果

表 2 を観察すると, 分類精度 (acc) については各手法で統計的に有意な差はほとんど見られなかった. 一方で, GEMIC_{MDL} と GEMIC_{AIC} の疎性 (spar) は, 他の手法と比較して大きく低い事が確認できる. この結果から, GEMIC が分類精度が同程度で, 解釈性の高いモデルを学習するのに適していると言える. なお, SparLogit と SparSVM についても, 超パラメータを大きくする事で, 疎性の高いモデルを得る事ができる. その場合には交差検定による超パラメータの決定の際に, 分類精度と属性の疎性のトレードオフを考慮する必要がある, GEMIC ではその点が自動化されている点を注記しておく. なお, 疎属性の選択問題は, より大規模で疎なデータに対して特に有効であり, そのようなデータへの GEMIC の適用が今後の課題として挙げられる.

5. まとめ

本稿では, 異種混合モデル選択のために, 一般化期待情報量基準最小化法を提案し, 混合正規分布に対する疎共分散選択, 混合分類モデルに対する疎属性選択への適用例を示した. 今後の課題としては, 複雑なモデルへの適用, M ステップの最適化の高度化, 大規模なデータへの適用等が挙げられる.

参考文献

[Akaike 73] Akaike, H.: Information theory and an extension of the maximum likelihood principle, in Petrov, B. N. and Caski, F. eds., *Proceedings of the 2nd International Symposium on Information Theory*, pp. 267–281 (1973)

[Asuncion 07] Asuncion, A. and Newman, D.: UCI Machine Learning Repository (2007)

[Bishop 98] Bishop, C. and Tipping, M.: A hierarchical latent variable model for data visualization, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp. 281–293 (1998)

[Fan 08] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J.: LIBLINEAR: A Library for Large Linear Classification, *Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874 (2008)

[Fujimaki 09] Fujimaki, R., Morinaga, S., Momma, M., Aoki, K., and Nakata, T.: Linear Time Model Selection for Mixture of Heterogeneous Components, in *Proceedings of 1st Asian Conference on Machine Learning*, pp. 82–97 (2009)

[Jacobs 91] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E.: Adaptive mixtures of local experts, *Neural Computation*, Vol. 3, No. 1, pp. 79–87 (1991)

[Kawahara 09] Kawahara, Y., Nagano, K., Tsuda, K., and Bilmes, J.: Submodularity cuts and applications, in *Proceedings of the 15th Ann. Conf. on Neural Information Processing Systems* (2009)

[Liu 09] Liu, J., Chen, J., and Ye, J.: Large-scale sparse logistic regression, in *the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 547–556 (2009)

[Marlin 09] Marlin, B. M. and Murphy, K. P.: Sparse Gaussian Graphical Models with Unknown Block Structure, in *Proceedings of the 26th International Conference on Machine Learning*, pp. 705–712 (2009)

[McLachlan 07] McLachlan, G. J. ed.: *The EM Algorithm and Extensions*, WILEY (2007)

[Rissanen 78] Rissanen, J.: Modeling By Shortest Data Description, *Automatica*, Vol. 14, pp. 465–471 (1978)

[Wong 00] Wong, C. S. and Li, W. K.: On a Mixture Autoregressive Model, *On a Mixture Autoregressive Model B*, Vol. 62, pp. 95–115 (2000)

[藤巻 09] 藤巻遼平, 森永聡, 門馬道也, 青木健児, 中田貴之: 線形時間異種混合モデル選択のための期待情報量基準最小化法, 第 12 回情報論的学習理論ワークショップ (2009)