

機械学習を用いた 混合正規分布の成分数削減手法に関する研究

A Study on Gaussian Mixture Reductions Using Machine Learning

風間 春輝・ネットワーク分科会・中央大学

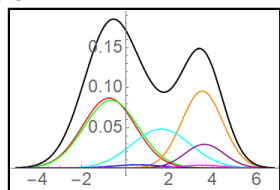
Gaussian mixture model(GMM) is a useful distribution for statistical methods, but the number of components of GMM increases by statistical operations. Hence, the number of components must be reduced to around 2 in order to repeat operations effectively and efficiently. Although several methods for reducing the number of components have been proposed, each of them has strength and weakness in accuracy and time complexity. Therefore, selecting an appropriate reduction method for an input distribution is a practical way for reducing the number of components. This paper proposes a selection method using support vector machine and evaluates its performance.

混合正規分布(GMM)と統計的演算

- GMMは正規分布の確率密度関数(PDF)の線形和で表される。

$$f(D) = \sum_{i=1}^m P_i \cdot \frac{1}{\sigma_i} \cdot \phi\left(\frac{D - \mu_i}{\sigma_i}\right) \cdot \phi(x): \text{標準正規分布のPDF} \cdot \sum_{i=1}^m P_i = 1$$

- ✓ 非正規分布を的確に表現できる
- 統計的な解析で良く用いられている
 - ・LSI設計における統計的静的遅延解析
 - ・組電池の寿命解析
 - ・防犯カメラ映像から異常行動を検知
- ・統計的最大値演算により、**成分数増大**そのまま演算を繰り返すと

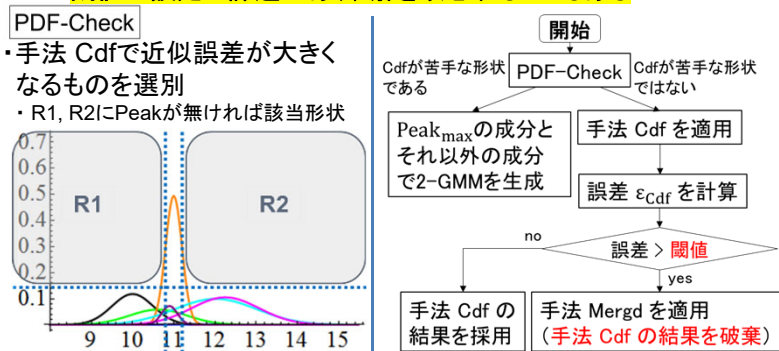


統計的最大値演算
 $\text{Max} [D_A, D_B] = D_M$
 m-GMM m-GMM $2m^2$ -GMM
 成分数増大

- × 非効率的
- × 実際の最大値分布との誤差増大[1]
- ◆ **成分数削減を行う必要がある**

既存手法 Selct [4]

- 手法 Mergdと手法 Cdfを使い分けて近似誤差 ϵ を減らす手法
- ⇒ **閾値の設定に課題があり、解を改悪することもある**



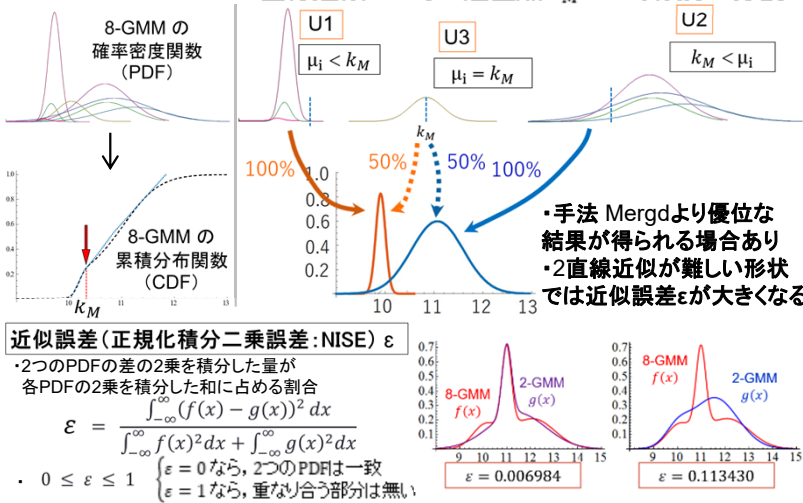
既存手法 Mergd [2]

- 成分間距離が最小の2成分を1つの成分に併合
- ⇒ 成分間距離を求めながら1つずつ成分数を削減し2-GMMを得る



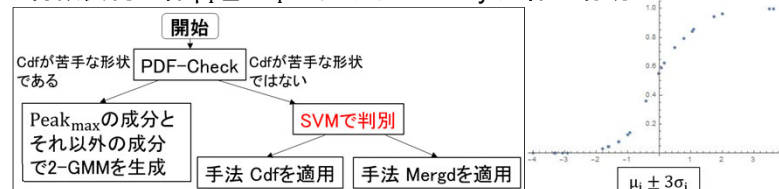
既存手法 Cdf [3]

- 8-GMMのCDFを2直線近似した時の屈曲点 k_M により成分を分割



提案手法

- SVMで手法を判別し、使い分けることで、解の改悪の減少を図る
- PDF-Checkは良好な形状判別ができていたため、利用する
- 特徴表現は各 $\mu_i \pm 3\sigma_i$ におけるCDFの y 座標が有効だった



提案手法の評価

- 2-GMM D_A を30種, 2-GMM D_B を5,400種用意し, 162,000個の8-GMM $D_M = \text{Max} [D_A, D_B]$ を求める。 D_M に各手法を適用し, 近似誤差を小さくする手法を記録しSVMで利用するクラスラベルとする。
- PDF-Checkで該当形状でなく, 学習で使わない20,000個で評価
- ✓ **総識別率の向上(66.9% → 94.8%)**
- ✓ **改悪率の減少(14.9% → 2.3%)**
- ✓ **良い近似を行った個数の増加**
- 総当たりで最適な分割を発見する手法の近似誤差 ϵ の1割増し以下の ϵ になっていれば, 良い近似

まとめ

SVMを用いて手法を識別することで, 手法 Selctよりも精度よく手法の選別が可能になり, 改悪率も減少した。

識別結果を評価する中で, 手法 Cdfも手法 Mergdも良い近似を行わないものが存在することが分かった。これらに対する効果的な成分数削減手法の構築が今後の課題である。

* 本ポスターでは, 中央大学大学院理工学研究科 築山研究室において風間が修士論文として纏めている研究について紹介した。

参考文献

- [1] D.Azuma and S.Tsukiyama, "A new algorithm to determine covariance in statistical maximum for Gaussian mixture model," IEICE Trans. Fund., vol.E100-A, no.12, pp.2834-2841, 2017.
- [2] A.R.Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," IEEE Trans. Aerosp. Electron. Syst., vol.43, no.3, pp.989-999, 2007.
- [3] S.Tsukiyama and M.Fukui, "A statistical maximum algorithm for Gaussian mixture models considering the cumulative distribution function curve," IEICE Trans. Fund., vol.E94-A, no.12, pp.2528-2536, 2011.
- [4] N.Yokoyama, D.Azuma, S.Tsukiyama, M.Fukui, "A new algorithm for reducing components of a Gaussian mixture model," IEICE Trans. Fund., vol.E99-A, no.12, pp.2425-2434, 2016.

