

## 新生児心電図におけるクラスタリング手法の比較

### Comparison of Clustering Methods for Neonatal Electrocardiograms

○江原大起<sup>1</sup>, 川村駿太<sup>2</sup>, 関弘翔<sup>3</sup>, 佐藤洋明<sup>4</sup>, 細野裕行<sup>3</sup>

\*Daiki Ehara<sup>1</sup>, Shunta Kawamura<sup>2</sup>, Hiroto Seki<sup>3</sup>, Hiroaki Sato<sup>4</sup>, Hiroyuki Hosono<sup>3</sup>

**Abstract:** The aim of this study is to analyze heart rate variability in neonates. In this paper we apply clustering using X-Means and TimeSeriesKMeans to heart rate (HR) data from approximately 500 neonatal electrocardiograms and interpret the results.

#### 1. まえがき

新生児の自律神経系の成熟過程は、身体や精神の発達のように目に見える変化ではなく、成長を捉えることが難しいが、成人と比較して未成熟であり、身体成長とともに発達する。この成熟過程を把握することは、自律神経系の未成熟さが一因とされる乳幼児突然死症候群 (Sudden Infant Death Syndrome, SIDS) などの疾患の早期発見に寄与する可能性がある。自律神経遮断後に心拍変動がほぼ消失することから、自律神経の影響により心拍間隔が微細に変動することが知られている<sup>[1]</sup>。先行研究<sup>[2][3]</sup>では、心拍変動を通じて自律神経活動を読み取る手法が検討されてきた。心拍変動は新生児に対して非侵襲的に計測可能な心電図から解析できる利点がある。心臓が収縮と拡張を繰り返す際に発生する微弱な電流を体に装着した電極で捉えた波形が心電図である。心電図は医療やヘルスケアで広く用いられ、不整脈や心筋梗塞などの診断に利用される。

林らの研究<sup>[3]</sup>では、生物が24時間周期のサーカディアンリズムに基づいて体温やホルモン分泌を調節することから、12時間区切りの波形に基づく少人数(2人)の心電図データを画像化し、教師なし機械学習手法の主成分分析(PCA)<sup>[4]</sup>による次元削減とk-means++法を用いたクラスタリングが検討された。この研究では、65~70%の精度で個人識別可能であったとされている。一方でサンプル数の少なさや、クラスタリング手法の改善、クラスタリング結果の解釈の点で課題が残った。

本研究は、500人規模の新生児の心拍数(HR)データに対して、クラスタリングの比較とその結果の解釈を行い、新生児の心拍変動を解析すること目的とする。

#### 2. 方法

本研究は自治医科大学附属さいたま医療センター臨床研究に関する倫理委員会の承認(臨S23-071)を受けて実施した。

##### 2.1. データクリーニング

Fluclet®にて抽出した新生児のHRデータに対して、

データクリーニングを実行する。データクリーニングは12時間間隔データ域抽出、誤検出値の修正、外れ値の除去の3段階で行う<sup>[5]</sup>。その後、先行研究<sup>[3]</sup>と同様にHRデータを画像化したデータセット(HR Images)および、画像化せずに時系列データとして扱うデータセット(HR Series)を作成した。データセットはそれぞれ、446件のデータで構成されている。

##### 2.2. クラスタリング手法

本稿では、HR Imagesに対してPCAにより抽出した特徴を用いたX-means法によるクラスタリング、HR Seriesに対してPCAにより抽出した特徴を用いたX-means法によるクラスタリング、HR Seriesに対するTimeSeriesKMeans法<sup>[6]</sup>によるクラスタリングの3パターンで比較を行った。

X-meansとは、k-meansを繰り返し適用することでクラスタ数を自動決定するクラスタリング手法である。TimeSeriesKMeansとは、時系列データに対し、k-meansを適用する手法である。通常のk-meansは、ユークリッド距離に基づいてデータをクラスタに分割するが、TimeSeriesKMeansでは、動的時間伸縮(Dynamic Time Warping)距離等の時系列データ特有の類似性を考慮した距離を用いてクラスタリングを行うことができる。

#### 3. クラスタリング結果

##### 3.1. 画像データに対するX-meansクラスタリング

HR Imagesデータセットに対し、X-meansクラスタリングを実行した。Fig. 1にクラスタリング結果をPCAの第1主成分を横軸に、第2主成分を縦軸に取ることで可視化した図を示す。X-meansを実行した結果、クラスタ数は4と決定された。

##### 3.2. 時系列データに対するX-meansクラスタリング

HR Seriesデータセットに対し、X-meansクラスタリングを実行した。Fig. 2にクラスタリング結果をPCAの第1主成分を横軸に、第2主成分を縦軸に取ることで可視化した図を示す。X-meansを実行した結果、クラスタ数は4と決定された。

1: 日大理工・学部・情報 2: 日大理工・院(前)・情報 3: 日大理工・教員・情報 4: 自治医科大学附属さいたま医療センターNICU

### 3.3. 時系列データに対する TimeSeriesKMeans クラスタリング

HR Series データセットに対し、TimeSeriesKMeans を実行した。X-means の結果を踏まえ、本稿ではクラスタ数  $k=4$  として実行した。Fig. 3 にクラスタリング結果を示す。なお、分類されたデータの先頭から 5 例のみを表示している。

### 3.4. クラスタリング結果の比較

Fig. 1, 2, 3 より、それぞれの手法が異なるデータ特性を強調することがわかる。TimeSeriesKMeans は時系列データの動的な側面を把握するのに適しており、時間的変動が重要な場合に効果的である。一方、X-Means はデータを広範囲に分類し、異なる特徴を持つクラスタ間の構造を視覚化するのに適している。Table 1 に Fig. 1, 2 の各クラスタ分類データ一致数を示す。Table 2 に Fig. 2, 3 の各クラスタ分類データ一致数を示す。Table 1 より、画像および、時系列データセットに対して X-Means を実行した場合、各クラスタに分類されたデータはほぼ同一である。このことから、画像および時系列データセットのどちらでも同様の解析結果が得られる可能性が明らかとなった。Table 2 より、時系列データセットに対して X-Means 及び、TimeSeriesKMeans を実行した場合、分類データ一致数は低い。このことから、使用するデータが一緒ではあるが 2 つのクラスタリング手法では異なるクラスタ構造を強調しているとわかる。

### 4. まとめ

本研究では、新生児の心電図データから自律神経系の成熟過程を把握するために、教師なし学習によるクラスタリングを行った。X-Means および TimeSeriesKMeans を用いてクラスタリングを実施し、得られた結果を比較した。現在は定性評価しか行っていないため、今後は XAI を実装して各クラスタリング手法での分類根拠可視化や評価を明確に行いたい。

### 参考文献

[1] N. Craft, et al., Am J Physiol., 268(4), pp.1441-1452, 1995.  
 [2] 福崎, 他, BME, 12 巻, 7 号, pp.79-88, 1998.  
 [3] 林, 他, 第 67 回理工学部学術講演会, G-8, pp.303-304, 2003.  
 [4] H. Abdi, et al., Wiley interdisciplinary reviews: computational statistics, 2(4), pp.433-459, 2010.  
 [5] X. Huang, et al., Information Sciences, Vol. 367, pp1-13, 2016.

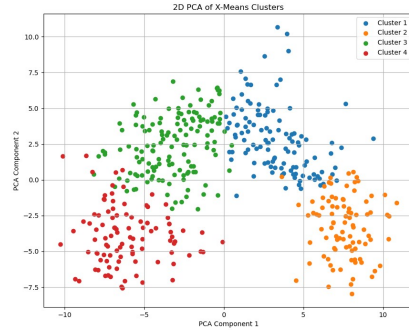


Figure 1. Result of X-Means Clusters(HR images)

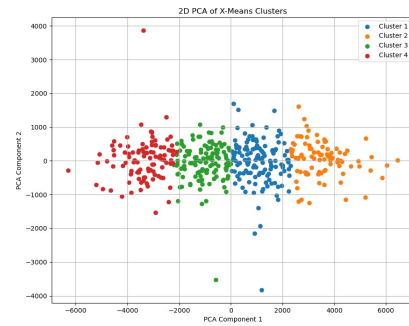


Figure 2. Result of X-Means Clusters(HR Series)

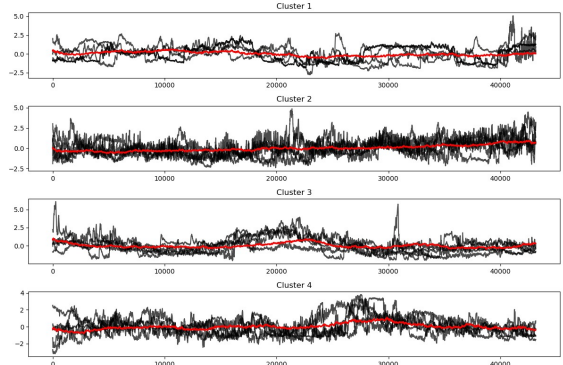


Figure 3. Result of TimeSeriesKMeans Clusters(HR Series)

Table 1. number of cluster matches(Fig.1, 2)

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	0	0	0	90
Cluster2	1	0	100	14
Cluster3	125	3	15	1
Cluster4	9	88	0	0

Table 2. number of cluster matches(Fig.2, 3)

	Cluster1	Cluster2	Cluster3	Cluster4
Cluster1	31	43	28	33
Cluster2	32	18	21	20
Cluster3	25	37	26	27
Cluster4	26	43	22	14