

周期的生体信号からのリアルタイム異常検知

上山 拓也[†] Goutam Chakraborty[‡] 馬淵 浩司[‡] 松原 雅文[‡]

岩手県立大学大学院ソフトウェア情報学研究科[†] 岩手県立大学ソフトウェア情報学部[‡]

1 はじめに

近年、個人向けの医療アプリケーションが日常的に使用されている。これらのアプリケーションは携帯端末上での動作を想定しており、携帯端末は内蔵メモリが少ない。そのため、リアルタイムでの異常検知や複雑な解析は困難である。これら問題を解決するために、高速な異常検知手法が提案されている。

既存手法 [1] では、Mother Signal と呼ばれる部分時系列集合の平均を用いて部分時系列の距離を測定することにより、高速な異常検知を実現している。しかし、周期的時系列でも、すべての部分が基本周期と同じ長さではない。時系列が長くなるにつれて、この問題は顕著となる。この問題を解決するために、複数の部分時系列集合から生成される Mother Signal を用いる必要がある。

本研究では複数の Mother Signal を用いた周期的生体信号からの異常検知の有効性について示す。

2 時系列と異常部分の定義

時系列と異常部分の定義時系列全体を T 、部分時系列の長さを m 、部分時系列の数を n と定義する。 m を T の自己相関から得られる基本周期とする。 p 番目の部分時系列（部分時系列 p ）の異常値を部分時系列 p と残りの部分時系列の中で最大限マッチングした部分時系列 q とのユークリッド距離として示す。ただし、 $|p - q| \geq m$ の条件を満たす時のみ、距離を計算する。

すべての部分時系列を平均 0、標準偏差 1 になるように正規化する。すべての部分時系列の距離は 0 から 1 に正規化し、閾値以上である距離が得られた部分時系列を異常部分と定義する。異常部分を全探索で算出する場合、 $O(n^2)$ の計算量を要する。

3 Mother Signal を用いた異常検知

3.1 概要

既存手法 [1] では、Mother Signal を用いてすべての部分時系列の距離計算が行われた。部分時系列の距離は、 m 回の循環処理が施された Mother Signal の中で、部分時系列と最大限マッチングした時の距離となる。循環処理とは、部分時系列の末尾の要素を部分時系列の先頭へ移動させる処理である。計算量は $O(mn)$ であるが、Mother Signal と 2 種類のヒューリスティック探索

を用いることで、より高速な異常検知を実現している。

3.2 Mother Signal の生成

Mother Signal を生成するために、部分時系列への循環処理と階層型クラスタリングにより部分時系列を抽出する。その後、閾値を決定し、要素数が最大となるクラスタの平均を Mother Signal としている。

このクラスタに含まれる循環処理が施された部分時系列は、もとの時系列から得られる部分時系列と類似した部分時系列が生成される。循環処理によって、もとの時系列から得られる部分時系列より多くの部分時系列を取得することが可能になる。

3.3 ヒューリスティック探索

既存手法で用いられている 2 種類のヒューリスティック探索の 1 つ目は、1 から m 回まで循環処理が施された Mother Signal と部分時系列の距離の推移が滑らかであり、最大限マッチングした時の距離が顕著に小さいことを利用した探索である。1 から適切な間隔で m 回まで循環処理が施された Mother Signal と部分時系列との距離を計算する。その後、得られた距離の中で、2 番目に小さい距離が得られた循環数から、適切な間隔で 3 番目に得られた循環数の間まで、循環処理が施された Mother Signal と部分時系列との距離を計算する。

2 つ目は、部分時系列 p の距離が、 i 回目の循環処理が施された Mother Signal から得られた時、部分時系列 $p+1$ と最大限マッチングする Mother Signal は、多くの場合 $i+1$ 回目の循環処理が行われた時の距離となることを利用した探索である。

これら 2 つのヒューリスティック探索を用いることで、既存手法は高速な異常検知を実現した。

4 提案手法

4.1 既存手法との違い

本研究では、Mother Signal を複数用いた異常検知手法について提案する。既存手法では、Mother Signal 生成のために階層型クラスタリングを用いたが、提案手法では、SAX [2] で得られた結果をもとに Mother Signal を生成する。また、部分時系列の距離を測定する際、Mother Signal と部分時系列のフーリエ変換後の距離を用いる。これらより、Mother Signal 生成までの過程と距離の測定が高速化される。

4.2 SAX

SAX とは時系列を符号化する手法である。部分時系列は時間軸方向に任意の数だけ等しく分割され、各区間の平均値を計算する。時間軸方向の分割数をワードサイズと呼ぶ。その後、正規分布の面積が等しくなるように、任意の数だけ分割する。この分割数をアルファベットサイズと呼ぶ。計算された平均値により、対応する文字が割り当てられる [2]。符号化された部分時系列を図 1 に示す。

Real-time Anomaly detection Continuously Monitored Periodic Bio-signals

Takuya KAMIYAMA[†], Goutam CHAKRABORTY[‡], Hiroshi MABUCHI[‡] and Masafumi MATSUHARA[‡]

[†]Graduate School of Software and Information Science, Iwate Prefectural University
020-0693, Iwate, Japan

[‡]Faculty of Software and Information Science, Iwate Prefectural University
020-0693, Iwate, Japan

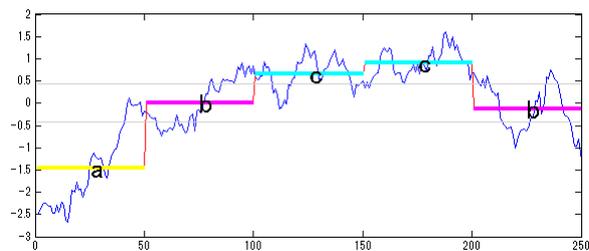


図 1: SAX による符号変換

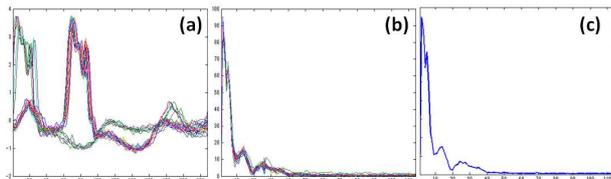


図 2: Mother Signal の生成過程

適切なワードサイズとアルファベットサイズを与えることで、出現頻度の少ない SAX コードを持つ部分時系列は、多くの場合、異常な部分時系列である。しかしながら、出現頻度の多い符号化された部分時系列集合の中には、異常な部分時系列が含まれることもある。従って、Mother Signal を用いて、このような異常な部分時系列を発見する必要がある。

4.3 Mother Signal の生成

提案手法では、すべての部分時系列を SAX により符号化する。出現頻度の多い符号化が施された部分時系列集合の上位 k 個をもとに、Mother Signal を k 個生成する。その後、同じ符号化された部分時系列のフーリエ変換後を平均したものを Mother Signal として生成する。“cbb” に符号化された部分時系列集合をもとに生成される Mother Signal を図 2 に示す。“cbb” に符号化された部分時系列集合 (図 2a) をそれぞれフーリエ変換 (図 2b) し、その平均 (図 2c) を Mother Signal として使用する。

部分時系列の異常値を計算するためには、部分時系列のフーリエ変換の結果と最大限マッチングする Mother Signal との距離を計算する必要がある。この計算量は $O(kn)$ となる。

5 実験

5.1 実験条件と評価方法

既存手法と提案手法について計算時間と F 尺度から評価を行う。実験では、長さの異なる MIT-BIH 不整脈データベース [3] の ECG データ 3 件を用いる。使用したデータを表 1 に示す。

5.2 結果と考察

実験結果を表 2 に示す。提案手法は、既存手法と同程度の精度で、少ない計算時間で異常検知をすることが可能になった。既存手法では、長い時系列ほど F 尺度が大きく減少している。提案手法では、時系列の長さにほとんど影響を及ぼされることなく、どのデータサイズでも同程度の精度で異常を検知することができ

表 1: 実験に使用したデータ

データ	n	m	異常部分時系列数
1	3524	251	328
2	5050	351	402
3	16129	372	1307

表 2: 計算時間と F 尺度の比較

データ	提案手法		既存手法	
	時間	F 尺度	時間	F 尺度
1	0.51	0.83	2.41	0.99
2	1.21	0.81	5.31	0.87
3	2.50	0.80	14.7	0.81

た。1つの MotherSignal では異常とみなされる部分時系列が、他の Mother Signal を用いることで、より正確な距離を測定し通常の部分時系列とみなしているからである。

今回、出現頻度の上位 k 個の Mother Signal を使用していた。しかし、出現頻度の少ない符号化された部分時系列から生成される Mother Signal を選択することで、より精度を高められると考えられる。

6 まとめ

本研究では、複数の Mother Signal を用いた異常検知手法を提案した。部分時系列のフーリエ変換の距離を比較することで高速な異常検知を実現した。また、複数の通常の部分時系列から複数の Mother Signal を生成することで、時系列が長い場合でも精度を維持することが可能となった。

今後の予定として、メモリを考慮した実験を行う。また、適切な Mother Signal を選択することで、精度が向上するかどうかを確認する。提案手法ではワードサイズやアルファベットサイズ、Mother Signal の 3 つのパラメータを与える必要がある。適切なパラメータを自動算出する必要がある。

謝辞

本研究の一部は JSPS 科研費 15K00155 の助成を受けたものである。

参考文献

- [1] Takuya Kamiyama, and Goutam Chakraborty: Real-time Anomaly Detection of Continuously Monitored Bio-signals Like ECG, etc., TSDAA 2015.
- [2] Jessica Lin, Eamonn Keogh, Stefano Lonardi and Bill Chiu. : A Symbolic Representation of Time Series, with Implications for Streaming Algorithms,” In SIGMOD workshop, 2003.
- [3] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101 (23):e215-e220