

反辞書確率モデルによる不整脈検出手法を利用した機械学習教育用ツールの検討

A Learning Support System for Machine Learning Using Arrhythmia Detection System based on Antidictionary Probabilistic Models

高松聖^{*1a}, 太田隆博^{*1b}, 森田啓義^{*2}, 野瀬裕昭^{*1c}

Hijiri TAKAMATSU^{*1}, Takahiro OTA^{*1}, Hiroyoshi MORITA^{*2}, Hiroaki NOSE^{*1}

^{*1}長野県工科短期大学校情報技術科

^{*1}Dept. of Computer and Systems Engineering, Nagano Prefectural Institute of Technology

^{*2}電気通信大学大学院情報理工学研究科

^{*2}Graduate School of Informatics and Engineering, The University of Electro-Communications

Email: n17a414^{*a} / ota^{*b} / nose^{*c} @cse.pit-nagano.ac.jp, morita@uec.ac.jp^{*2}

あらまし：IoT(Internet of Things)の技術発展に伴い、AI(Artificial Intelligence)が様々な分野で用いられている。AIの基盤技術として機械学習があり、機械学習の知識をもつ技術者の養成が求められる。本稿では、機械学習の実践的な教材の作成を目的として、心電図の異常波形（不整脈）検出を行う機械学習教育用ツールの開発を行う。開発に当たっては、省メモリ・リアルタイムに不整脈検出を行える反辞書確率モデルを用いた手法を利用する。機械学習による不整脈検出の予備実験として、教師用データとして必要なパラメータを調べるために、手でパラメータを設定し不整脈検出を行い、その性能確認を行う。実験としては、最初に、1種類の異常値(心室性期外収縮)とそれ以外の2種類に分類を行い、次に、2種類の異常値(心室性期外収縮, 心房性期外収縮)とそれ以外の3種類に分類を行った。実験結果から、3種類の分類では、心室性期外収縮は感度と特異度がそれぞれ100%と95%、心房性期外収縮では感度と特異度がそれぞれ91.3%と58.4%が得られた。予備実験から、教師あり機械学習に必要なパラメータに関する知見が得られた。

キーワード：AI, 機械学習, 不整脈, 分類, 反辞書, データ圧縮

1. はじめに

IoTによりモノがインターネットに繋がることでリアルタイムに情報をやり取りし、非常に多くのデータ(ビッグデータ)を収集・蓄積ができるようになった。このビッグデータをAIが分析し、規則性を見つけだし、これを用いて判断・予測などを行う。例えば、医療分野では多数の患者のカルテのデータを蓄積することで、AIが自動で病気診断⁽¹⁾や感染症の感染拡大予測⁽²⁾に関する研究が行われている。さらに、ロボット制御にも応用されている⁽³⁾。このように、AIは様々な分野で用いられている。AIの開発に当たっては、その基盤技術として機械学習があるため、機械学習の知識を持つ技術者の養成が求められている。

本稿では、機械学習の実践的な教材の作成を目的として、心電図の不整脈検出を通して機械学習を学ぶツールの開発を目的として、反辞書確率モデルを用いて不整脈の分類を行う手法⁽⁴⁾、⁽⁵⁾を教育用ツールに拡張するための予備実験の結果を報告する。

2. 機械学習

機械学習は反復的に学習し、学習結果を法則化(モデル化)する手法であり、その結果を利用して、未知のデータの予測や識別を行う。機械学習の分類としては、大きく分けると教師あり学習、教師なし学習、強化学習の3種類がある。本稿では、不整脈検出に

用いる教師あり学習と強化学習に着目する。

教師あり学習とは、ラベルを付けたデータ(教師データ)をAIに繰り返し学習させてモデルを生成する手法であり、生成したモデルを利用して、未知のデータを学習したラベルに分類を行う。強化学習とは、モデルの行動によって、報酬を与えることで、その結果をフィードバックしてモデルを改善し、より最適な振る舞いを学習する手法である。

3. 反辞書確率モデルによる不整脈検出手法

入力記号列に現れない極小禁止語の集まりを反辞書とよび、それを用いて構築できる確率モデルを反辞書確率モデルと呼ぶ。この確率モデルを用いたデータ圧縮手法⁽⁴⁾が提案されており、それを応用して、省メモリかつリアルタイム処理が可能な不整脈検出手法が提案されている⁽⁵⁾、⁽⁶⁾。従来手法では、正常洞調律の反辞書から反辞書確率モデルを構築する。このモデルは、正常洞調律は、効率よくデータ圧縮できるが、不整脈はうまくデータ圧縮できない特性を持つ。これは、不整脈では、正常洞調律に出現しないはずの極小禁止語の出現により、圧縮率が一時的に著しく劣化することによる。

例として、反辞書を{11, 000, 10101}とした場合の反辞書確率モデルを図1に示す。反辞書確率モデルは、反辞書の要素のパターン照合用のオートマトンの状態遷移に確率を与えた構造となっている。矢印

のラベル a/p は、記号 a の遷移確率 p を表す。四角の状態への遷移が極小禁止語の出現と同値になり、その場合には空記号 ε により遷移する。入力データにより初期状態 0 から遷移を行い、枝に振られた遷移確率を符号化する。図 1 のオレンジの線は、入力データが 01001000 の場合の遷移を表す。

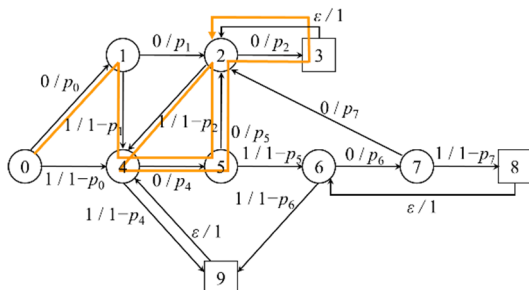


図 1 反辞書確率モデル

4. 提案ツールと予備実験

4.1 実装する機能

教師あり学習として、正常洞調律と不整脈(心室性期外収縮と心房性期外収縮)のラベルをつけた波形を与えて、学習モデルを作成する。強化学習としては、反辞書確率モデルにおいて、正常洞調律と不整脈検出に寄与が大きい状態への遷移に報酬を与えて、モデルの改善を行えるようにする。

ここで、心室性期外収縮と心房性期外収縮を選んだ理由としては、この 2 種類の不整脈の発生回数が他の不整脈に関して多いことと、心室性期外収縮が悪性なものとして分類されており、医的に検出が求められるためである⁽⁷⁾。

4.2 予備実験

従来手法^{(5), (6)}では、心室性期外収縮とそれ以外の 2 種類の分類についての波形数については知見が得られているが、心房性期外収縮を含めた 3 種類の分類についての波形数については知られていない。ここでは、3 種類の分類について精度で行われるための教師あり学習に必要な波形数について調べることを目的とする。実験には、MIT-BIH Arrhythmia データベースの心電図データを用いた。この心電図には、すべての波形に対して医師の診断がラベル付けされている。実験には、心電図データ(105, 200, 201, 202, 205, 215, 221, 222, 223, 228)の 10 ファイルを用いた。

検出の評価は、感度・特異度として評価する。感度・特異度を求めるために準備として表 1 に医師による事前判定と実験での検出結果の関係を示す。表 1 において、陽性と陰性とは、与えられた波形種別に対して、それぞれ、その波形種別であるかそうでないかを表す。表 1 から感度と特異度を次のように求められる。

$$\text{感度} = TP / (TP + FN) * 100 (\%)$$

$$\text{特異度} = TN / (FP + TN) * 100 (\%)$$

なお、感度が高い場合は、未検出が少なく、特異

度が高い場合には誤検出が少ない。

表 1 医師の事前判定と実験の検出結果の関係

		医師による事前判定	
		陽性	陰性
検出手法による 検出結果	陽性	真陽性(TP)	偽陽性(FP)
	陰性	偽陰性(FN)	真陰性(TN)

予備実験結果から、心電図データ 201 に関して最もよい結果が得られて、心室性期外収縮に関しては、感度 100%、特異度 95%であり、心房性期外収縮に関しては、感度 91.3%、特異度 58.4%となった。教師あり学習に必要な波形数は 209 拍(約 2 分 26 秒)であった。

5. まとめと今後の課題

実験結果から、感度に関しては、最もよい場合であるが、心室性期外収縮と心房性期外収縮に関して 90%以上となった。また、教師あり学習に必要な波形数に関しては、実験した 10 ファイルで最大 253 拍のラベル付けされた波形があればよいことがわかった。今後の課題としては、予備実験で与えられた波形種別の検出に寄与が大きい状態の集まりがわかったので、強化学習では、その状態への遷移に報酬を与えることで、強化学習の効果についての予備実験を行う必要がある。また、その結果を踏まえて、教育用ツールの作成を行う。

謝辞

本研究の一部は、JSPS 科研費 基盤研究(C) 17K00147 と 17K00400 の助成を受けた。

また、本研究にご協力いただいた長野県工科短期大学の卒業生である篠原知美氏に感謝する。

参考文献

- (1) E. Choi, M.T. Bahadori, A. Schuetz, W.F. Stewart, J. Sun: "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks", Proc. of the first Machine Learning for Healthcare Conference, pp. 301-318, Feb. (2016)
- (2) K. Rishav: "Epidemic Outbreak Prediction Using Artificial Intelligence", International Journal of Information Technology and Computer Science, vol. 10, no. 4. pp. 49-64, Aug. (2018)
- (3) 伊藤一之: "実装強化学習 C によるロボットプログラミング", オーム社 東京 (2018)
- (4) M. Crochemore, F. Mignosi, A. Restivo, and S. Salemi: "Data Compression Using Antidictionaries", In Proc. of IEEE, pp. 1756-1768, Nov. (2000)
- (5) G. Frias, H. Morita, and T. Ota: "Anomaly Detection on Quantized ECG Signals by the Use of Antidictionary Coding", Proc. of the 41st Symposium on Information Theory and its Applications, pp. 587-591, Oct. (2018)
- (6) T. Ota, H. Morita, and A.J. van Wijngaarden: "Real-Time and Memory-Efficient Arrhythmia Detection in ECG Monitors Using Antidictionary Coding", Vol. E96.A, no. 12, pp. 2343-2350, Dec. (2013)
- (7) 新谷富士雄: "はじめのはじめの心電図", メディカ出版, 東京 (1997)