

# 強化学習による心疾患臨床データの分析

## Analysis for Clinical Data of Cardiac Disease using Reinforcement Learning

城 真範\*<sup>1</sup>      神嶌 敏弘\*<sup>1</sup>      橋田 浩一\*<sup>1</sup>      赤穂 昭太郎\*<sup>1</sup>      麻生 英樹\*<sup>1</sup>      荒牧 英治\*<sup>2</sup>  
 Masanori Shiro      Toshihiro Kamishima      Kōiti Hasida      Shotaro Akaho      Hideki Asoh      Eiji Aramaki

興梠 貴英\*<sup>3</sup>  
 Takahide Kohro

\*<sup>1</sup>産業技術総合研究所

The National Institute of Advanced Industrial Science and Technology (AIST)

\*<sup>2</sup>京都大学デザイン学ユニット

Kyoto University, Design Unit

\*<sup>3</sup>東京大学医学部附属病院

The University of Tokyo Hospital

It is important to retrieve helpful patterns from clinical data which are daily stored, in order to upgrade medical treatments. In this report, we tried to unveil the best action of doctor's instructions by using a reinforcement learning. Records of laboratory values and drugs for patients who are affected the cardiac disease are used. We found acceptable results and some of future problems.

### 1. はじめに

近年、情報化の進展により、臨床医療の現場にも多くの情報機器が導入され、電子カルテ等の治療記録や各種検査値など、多様な情報が電子的に蓄積されるようになってきている。しかしながら、医療過程が高度化、複雑化、情報化してゆく中で、それらが十分に活用されているとは言えない。日常的な診療過程で蓄積される観察データを体系的に活用し、治療過程の質の向上につなげることは重要な課題である。

病気の治療過程は、医師と患者の間の継続的なインタラクションであり、対話の一種とみなすことができる。Hauskrechtらは、こうした観点から、虚血性心臓疾患の患者の状態を階層的な変数群で表現し、治療過程を POMDP (部分観測マルコフ決定過程) によってモデル化し、最適な治療方策を求めて、医師の行動と比較した [Hauskrecht 00]。

しかし、この研究において POMDP に使われたパラメータは既存の文献や医師からの聞き取り調査によるものであり、実データに基づく推定は行われていない。こうした治療過程に関して、実際のデータに基づいた治療過程のモデル化と分析を行うことは興味深い課題であろう。我々は既に、糖尿病関連や経皮的冠動脈インターベンションのデータについて強化学習や逆強化学習を用いた解析を行っているが [Asoh 12-1, Asoh 13]、今回、心不全関連のデータについても強化学習による解析を始めたので報告する。以下に、我々が取り扱うデータとそのモデル化、分析の結果、課題および展開の可能性について述べる。

### 2. 分析対象データと方法

今回分析対象とするデータは、東京大学医学部附属病院循環器内科を受診または入院した患者約 2,000 名分の、検査データ及び投薬データである。患者の状態の観測変数としては、下記の検査 (主に血液検査) の実施日と結果 (数値) が与えられている：

- BNP: 心臓に対する負担度の指標

連絡先: shiro@ni.aist.go.jp

- BUN, CRE, eGFR: 腎機能の指標
- CK: 筋肉に含まれる酵素量
- CKMB: 心筋に含まれる酵素量
- T.CH, T.G,d HDL-C, LDL: 血中コレステロール値
- HB-A1C, GLU: 血糖値
- K: 血中カリウム値
- U.A: 尿酸値
- CRP: 感染等による炎症の指標

患者からの痛み等の訴えや医師の所見に関する情報は含まれていない。医師の措置としては、投薬 (薬名、分量) についての情報が与えられている。各レコードは、検査データでは (患者 ID, 日付, 検査名, 検査結果数値)、投薬データでは (患者 ID, 日付, 薬剤名, 用法, 1 日の分量, 日数, 総処方量) のような構造をしている。投与されている薬の種類はデータ全体では約 320 種類あり、それらが異なる容量の組合せで投与される。従って、検査値、投薬パターンともに種類が非常に多く、それらをそのままの形で状態、行動として強化学習を適用することは不可能である。患者の状態および医師の行動を何らかの形で情報縮約することが必要になる。

そのため、今回は、1996 年 9 月以降 2010 年 10 月までのデータから、心不全の通院治療に関連するレコードを抽出して分析した。このように疾患を限定することで、関連する検査の種類と薬物の種類をかなり減らすことができる。それぞれの患者ごとのデータは、一定の治療方策に従ったひとつのエピソードと考えられる。関連する検査としては、表 1 に示す 7 つをそれぞれ 2 つの閾値にて normal, medium, severe の 3 値に離散化した。ただし、K については、3.5 から 5 の範囲にある場合だけを normal とし、その他は severe とした。

表 1: 利用した検査種と離散化のための閾値

検査名	単位	閾値 1	閾値 2
BNP	pg/ml	100	300
BUN	mg/dl	20	30
CRE	mg/dl	1	1.5
CRP	mg/dl	0.3	2
eGFR	ml/min/1.73m <sup>2</sup>	60	30
K	mEq/l	3.5	5
U.A	mg/dl	7	9

心不全の治療に関係する薬物として、利尿剤、 $\beta$ -遮断薬としてカルベジロール、ビソプロロール、ARB、ACE 阻害剤、ピモベンダン、ジゴキシンを選択し、それぞれの有効成分名を行動とした。なお、ARB、ACE 阻害剤、 $\beta$ -遮断薬は降圧薬として、 $\beta$ -遮断薬は狭心症治療薬として用いることもあるが、今回はすべて解析対象とした。投与量は無視して組合せのみを考慮した。

こうした実データを強化学習の枠組みで分析するにあたり、まずデータの欠損や整合性のチェックを行い、別に用意した対応表により、薬剤名を有効成分名に変換した。さらに、1 イベ

ントだけのデータを除き、各患者について 63 日以下の時間間隔をもつエピソードだけを抽出した。こうして作成したエピソードを用いて off-policy Q 学習を実行した。今回の検討では探索用の方策として医師の方策を利用したことになる。報酬関数は、複数の検査値が同時に normal となる場合を 1 とし、その他を 0 とした。強化学習の際の時間間隔は検査間隔とした。

全体のプログラムは統計解析言語 R[R 言語] で実装し、最適方策の推定には R 上の強化学習パッケージ RLearn[Asoh 12-2] を用いた。学習時の忘却係数  $\gamma$  は 0.9, 0.5 とした。

### 3. 結果と考察

一般的な心不全の治療においては、心臓の負荷を減らすために利尿剤等を利用するが、それによる副作用としての腎機能低下についても監視し、腎機能低下時には強心剤等を利用することになる。このため心臓に対する負担度の指標である BNP と腎機能の指標の中から CRE を特に重点的に検討した。以下の表 2, 3, 4 はこの二種の検査値を状態空間としたときの、 $\gamma = 0.9$  において得られた最適方策および最適方策の下での行動価値推定結果である。

表 2: BNP 検査値を状態空間としたときの最適方策および最適方策の下での行動価値推定結果 (エピソード数: 9035, エピソード系列の長さ: 2~111)

状態	行動価値	最適方策
BNP-normal	4.253	《処方なし》
	2.996	マレイン酸エナラプリル (降圧薬 [ACE 阻害剤])
	2.953	フロセミド (利尿剤)
	2.937	カルベジロール (降圧薬 [ $\alpha$ $\beta$ 遮断薬]), スピロノラクトン (利尿剤), マレイン酸エナラプリル, フロセミド
	2.665	スピロノラクトン, フロセミド
	2.070	ロサルタンカリウム (降圧薬 [ARB]), カルベジロール, スピロノラクトン
BNP-medium	2.382	《処方なし》
	1.703	スピロノラクトン, フロセミド
	1.624	フロセミド
	1.294	スピロノラクトン, ジゴキシン (強心剤 [ジギタリス]), フロセミド
	1.230	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	0.913	ジゴキシン
BNP-severe	1.508	《処方なし》
	0.960	フロセミド
	0.750	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	0.576	スピロノラクトン, フロセミド
	0.550	カルベジロール
	0.487	スピロノラクトン, ジゴキシン, フロセミド

表 3: CRE 検査値を状態空間としたときの最適方策および最適方策の下での行動価値推定結果 (エピソード数: 14800, エピソード系列の長さ: 2~794)

状態	行動価値	最適方策
CRE-normal	8.458	スピロノラクトン, フロセミド
	8.397	フロセミド
	8.390	カンデサルタン シレキセチル (降圧薬 [ARB])
	8.379	スピロノラクトン, ジゴキシン, フロセミド
	8.374	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	8.349	ジゴキシン, フロセミド
CRE-medium	4.996	《処方なし》
	4.864	スピロノラクトン, フロセミド
	4.746	フロセミド
	4.689	マレイン酸エナラプリル
	4.389	スピロノラクトン, ジゴキシン, フロセミド
	4.311	ジゴキシン
CRE-severe	2.290	フロセミド
	2.257	《処方なし》
	2.149	ロサルタンカリウム, フロセミド
	2.075	スピロノラクトン, フロセミド
	1.831	マレイン酸エナラプリル, フロセミド
	1.825	カルベジロール, フロセミド

表 4: BNP 検査値と CRE 検査値を状態空間としたときの最適方策および最適方策の下での行動価値推定結果  
(エピソード数: 8230, エピソード系列の長さ: 2~109)

状態	行動価値	最適方策
BNP-normal, CRE-normal	2.773	《処方なし》
	1.773	マレイン酸エナラプリル
	1.755	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	1.561	フロセミド
	1.026	スピロノラクトン, フロセミド
	1.003	カルベジロール, スピロノラクトン, ジゴキシン, マレイン酸エナラプリル, フロセミド
BNP-normal, CRE-medium	0.522	《処方なし》
	0.158	カルベジロール, スピロノラクトン
	0.136	マレイン酸エナラプリル
	0.126	フロセミド
	0.124	ロサルタンカリウム, カルベジロール, スピロノラクトン
	0.105	ジゴキシン
BNP-normal, CRE-severe	0.071	《処方なし》
	0.038	カンデサルタン シレキセチル, スピロノラクトン, ジゴキシン, マレイン酸エナラプリル
	0.029	スピロノラクトン, ジゴキシン, フロセミド
	0.026	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	0.023	フロセミド, バルサルタン
	0.015	ペリンドプリルエルブミン, ジゴキシン, フロセミド, トリクロルメチアジド
BNP-medium, CRE-normal	1.025	《処方なし》
	0.538	フロセミド
	0.532	スピロノラクトン, フロセミド
	0.467	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	0.399	スピロノラクトン, ジゴキシン, フロセミド
	0.272	ロサルタンカリウム, カルベジロール, スピロノラクトン, フロセミド
BNP-medium, CRE-medium	0.302	《処方なし》
	0.102	スピロノラクトン, フロセミド
	0.075	カルベジロール, マレイン酸エナラプリル
	0.065	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	0.062	ジゴキシン
	0.059	ロサルタンカリウム, スピロノラクトン, ジゴキシン, フロセミド
BNP-medium, CRE-severe	0.038	《処方なし》
	0.010	フロセミド
	0.006	マレイン酸エナラプリル
	0.003	スピロノラクトン, フロセミド
	0.003	ピモベンダン, フロセミド
	0.003	ペリンドプリルエルブミン, カルベジロール, スピロノラクトン, フロセミド
BNP-severe, CRE-normal	0.428	《処方なし》
	0.176	フロセミド
	0.140	スピロノラクトン, フロセミド
	0.109	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	0.093	スピロノラクトン, ジゴキシン, フロセミド
	0.068	カンデサルタン シレキセチル, フロセミド
BNP-severe, CRE-medium	0.165	《処方なし》
	0.051	ピモベンダン, カルベジロール, スピロノラクトン, ジゴキシン, フロセミド
	0.042	カルベジロール, スピロノラクトン, フロセミド
	0.040	フロセミド
	0.033	マレイン酸エナラプリル, フロセミド
	0.030	カルベジロール, スピロノラクトン, ジゴキシン, マレイン酸エナラプリル, フロセミド
BNP-severe, CRE-severe	0.067	《処方なし》
	0.015	フロセミド
	0.013	カルベジロール, スピロノラクトン, マレイン酸エナラプリル, フロセミド
	0.006	カンデサルタン シレキセチル, カルベジロール, スピロノラクトン, フロセミド
	0.005	カルベジロール
	0.005	テルミサルタン

なお、計算は全検査種の組み合わせについて行ったが、三種類以上の検査値から構成する状態空間では状態数の増大に伴い、1 状態あたりのデータ数が極端に少なくなるため検討対象外とした。

まず、報酬関数として検査値が normal である場合のみを 1 として他を等しく 0 としたにも関わらず、表 2 や表 3 では、medium と severe の行動価値に差が見られた。これは、処方により患者の状態が改善したことを反映したものと考えられる。また、表 3 や表 4 の後半では状態に対する行動価値の差

があまり大きくないので、方策推定結果の順位については有効な情報となっていない可能性がある。

BNP が関係する表 2 と表 4 のすべてにおいて、処方なしが最適方策とされた点は解釈が難しい。例えば BNP に個体差があり、必ずしも検査の絶対値による状態空間の構成が適切ではない可能性も考えられる。すなわち BNP 値がもともと高値である患者の場合、より高値に変化しなければ薬剤の処方がなされない場合がありうる。別の解釈としては、医師の処方が検査値のみではなく、その背後にある何らかの状態を基準に決めら

れている可能性である。これを検討するためには真の状態を隠れ変数とするような POMDP による検討が必要である。今回利用した RLearn パッケージでは POMDP に対応した改良も検討されている。

表 2(BNP) については、BNP が normal である場合には見られない強心剤（ジゴキシン）が medium や severe においては方策中に出てくることが特徴的である。実際の処方においては、BNP 値が高くなり利尿剤では腎機能に影響が出る場合、強心剤が使われることがある。強心剤は表 4 においても、BNP と CRE がともに normal である場合を除き、多くの場合で使われている。

表 3(CRE) については、表 2 とは異なる傾向も見られる。強心剤は normal と medium において方策中に出現するが、severe において利尿剤の処方が最も高い行動価値をもって推定されている。これは一般的な医師の処方とは異なる。一つの解釈としては、今回の解析で処方量について考慮していないため、行動価値が正確に推定できなかった可能性がある。

忘却係数  $\gamma$  を 0.5 にした場合、行動価値の値は下がったが、最適方策および最適方策の下での行動価値推定結果の傾向は変わらなかった。

今回の検討では、off-policy Q 学習を利用したが、表 2 や表 4 において処方なしが最適方策とされたり、表 3 において利尿剤の処方が最適方策とされたりしたなど、一般的な医師の処方と異なる結果が得られた点については、off-policy Q 学習が収束していないことが原因である可能性も考えられる。この点については今後の更なる検討が必要である。

#### 4. おわりに

循環器内科における医療診療データに対し強化学習を用いた分析を行い、治療過程の改善のための情報抽出を試みた。今後、分析手法の改良を進めるとともに、より有意義な情報を抽出してゆきたい。特に、off-policy Q 学習が収束していない可能性がある点、及び報酬関数の設定については様々に検討と改良の余地があろう。今回の検討では処方量についての考慮をしていないため、今後はそれを報酬関数に組み込むべきである。BNP 値のように絶対値よりも変化量に着目するべき値についての検討も必要である。複数の検査値から状態空間を作る際に、どのようにして報酬関数を決めるかという問題も残っている。一つ

の方法として、逆強化学習 (inverse reinforcement learning) [Ng 00] の利用も考えられる。さらに、状態変数の選択、次元圧縮をデータに基づいて行う方法、述語論理式などで領域知識を表現して利用する方法 (関係強化学習 [Tadepalli 04]) など将来的に検討してゆきたいと考えている。

#### 参考文献

- [Asoh 12-1] 麻生, 城, 神寫, 赤穂, 興梧: 報酬と行動決定を伴う時系列データの強化学習を用いたオフライン分析, 電子情報通信学会技術研究報告, vol. 111, no. 419, NC2011-115, pp. 107-112, 2012
- [Asoh 12-2] 麻生, 城, 神寫, 赤穂, 興梧: エピソード時系列データ分析のための強化学習ツール RLearn, 第 15 回情報論的学習理論ワークショップ, 2012.
- [Asoh 13] 麻生, 城, 神寫, 赤穂, 興梧: 逆強化学習による医療臨床データの分析, 電子情報通信学会技術研究報告, vol. 112, no. 390, NC2012-96, pp. 13-17, 2013
- [Hauskrecht 00] M. Hauskrecht and H. Fraser: Planning treatment of ischemic heart disease with partially observable Markov decision processes, Artificial Intelligence in Medicine, vol.18, pp.221-244, 2000.
- [Ng 00] A. Y. Ng and S. Russell: Algorithms for inverse reinforcement learning, Proceedings of 17th International Conference on Machine Learning, pp.663-670, 2000.
- [R 言語] R 言語 <http://www.r-project.org/>
- [Tadepalli 04] P. Tadepalli, R. Givan, K. Driessens: Relational reinforcement learning: an overview, Proceedings of ICML-04 Workshop on Relational Reinforcement Learning, 2004.

#### 謝辞

本研究は日本学術振興会科学研究費補助金 (基盤研究 (A)) 23240043 「確率関係モデルによる医療臨床データの高度活用に関する研究」の支援を受けて行われた。