動的治療レジメ推定におけるQ学習とその周辺

京都大学大学院医学研究科　　大前　勝弘

1　はじめに

個々の患者の特徴や遺伝情報に合わせて治療選択を考慮したいという潮流は，個別化医療という標語と共に近年非常に高い注目を浴びている．特に，個々の患者の臨床背景や治療歴，臨床転帰（以下，これらをまとめて履歴と呼ぶ）に基づいて適応的に患者の治療を選択するというパラダイムは動的治療レジメ（Dynamic Treatment Regime, DTR）と呼ばれる．近年においては，医療技術やデータマネジメント手法の発展により，これまでにない規模でDTRを評価するためのデータを継続的に収集することが可能になり始めており，これらのデータから最適な治療レジメを推定したいという関心がより高まっている．

2　Q学習とその問題点

DTR評価における統計的な問題設定は以下の通りである．治療ステージ $k \in \{1, 2, \cdots, K\}$ において，アウトカム $o_k \in \mathcal{O}_k$ が観測された患者に，治療 $a_k \in \mathcal{A}_k$ を，ある治療方針 d_k に基づいて施したい．この治療方針 d_k は，履歴 $h_k = (\bar{o}_k, \bar{a}_{k-1})$ をもとに決定され，その治療方針の善し悪しは，方針 $d = (d_1, \cdots, d_K)$ と履歴 h_k を与えたもとでの，その後の経過アウトカム $Y_{k+1}(h_{k+2}), Y_{k+2}(h_{k+3}), \cdots, Y_K(h_{K+1})$ の総和 $\sum_{j=k}^{K} Y_j(h_{j+1})$ の期待値で与えられるとする．ただし，$\bar{o}_k = (o_1, o_2, \cdots, o_k)$ はステージ k 以前に観測されたアウトカムの組，$\bar{a}_{k-1} = (a_1, a_2, \cdots, a_{k-1})$ はステージ k 前に行われた治療の組で，経過アウトカムはこれらの既知関数である．このように，ステージ k で観測しうる履歴の集合を $\mathcal{H}_k$ とした場合に，治療方針 $d_k : \mathcal{H}_k \to \mathcal{A}_k$ のうち $V_k(h_k) = \mathrm{E}_d[\sum_{j=k}^{K} Y_j(H_j, A_j, O_{j+1}) | H_k = h_k]$ の期待値をもっとも大きくするような各ステージにおける治療方針 d_k の列 $(d_1, d_2, \cdots, d_K)$ を推定する問題として，最適なDTR推定の枠組みが定式化される．

上記設定は強化学習 (Reinforcement Learning) の問題と見ることができ，「Q関数」と呼ばれる関数の最適化を段階的に解くQ学習のアイデアをもとにした最適治療方針の推定が目指される．ただし，強化学習での典型的な仮定と異なり，治療の遅れ効果などを依拠した非マルコフ決定過程を考える必要が生じる．この場合には，計算コストがステージ数や治療選択数に応じて爆発的に増加してしまうため，何かしらのモデル近似を伴うことになる．医療データのサンプルサイズが比較的小さいことから，Q関数を履歴の線形関数で近似した上でのQ学習が有望であり，これを基にしたアイデアがMurphy(2005)を起点として数多く提案されている．しかし，このように単純な線形Q学習の場合にでさえ，逐次的な最適化に伴う推論のnon-regularityが生じてしまうことが知られている (Chakraborty 2010)．これにより，Q関数のモデルパラメータの推定量にバイアスが生じたり，正確なType-I errorや名目上の被覆確率の算出を困難にさせてしまう．本発表では，DTR推論におけるこのようなnon-regularityの問題及びこれを解決するために提案されているアイデアの一連とそれらの問題点を共有し，新たな方向性について議論したい．

参考文献

[1] Murphy, S. A. (2005). A generalization error for Q-learning. *Journal of Machine Learning Research,* 6, 1073–1097.

[2] Chakraborty, B., Murphy, S. A., Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19, 317–343.