

# 動的治療計画の最近の進展と今後の展望

統計数理研究所

江口 真透

## 1. はじめに

エビデンスに基づく個別化医療、精密医療の内容で被験者の個人ごとの動的治療計画 (Dynamic Treatment Regime, DTR) を統計的な観点からアプローチする研究が精力的に展開されている (Chakraborty & Moodie, 2013, Springer). 慢性病 (高血圧症, 肥満, 糖尿病, 関節リウマチ, ニコチン中毒, アルコール, ドラッグ乱用, HIV 感染症, 鬱病, 統合失調症, がん) に見られる様々な患者間の異質性, 再発, 並存疾患, 治療の時間依存の奏功性・副作用の問題がある中で, 個人ごとの最適な治療レジメを動的に行うことが目的である. 反事実仮想に基づく潜在アウトカムの推定・予測のアプローチがキーとなるが, 機械学習の強化学習について復習しながら基本的な考え方を紹介したい.

強化学習の多くはマルコフ性を仮定した決定プロセスが研究されているが, 動的治療計画では被験者の治療履歴を開始時期から最終観察期まで一つの軌跡として扱う必要がある. そのために適切な推論を得るために必要な確率の基本仮定について紹介する. このような枠組みは, James Robins, Susan Murphy たちの貢献によって合理的な体系化が成されているが非常に難解である.

## 2. 確率的フレームワーク

機械学習における強化学習は学習エージェント (例えば, ロボット) と学習する環境の相互作用によって特徴づけられる. 学習エージェントが環境のある状態を観測して可能な行動の候補の中からある行動を決定した結果, 環境が新たな状態に移り, エージェントがその変異による報酬を得る. 各ステージの状態, 行動, 報酬が3つの要素となるプロセスにおいて, 最も基本的な仮定のマルコフ性, すなわち, 「現時点の状態と行動を与えた時の新たな状態へ推移する環境の確率はそれより以前の状態と行動には依存しない」の要請を課すとき, マルコフ決定プロセスと呼ばれる. このマルコフ性は深層学習を使った最近の研究 (DQN) においても基本的な仮定である.

一方で, 典型的な DTR では診療開始時点から終了時点までのすべての共変量履歴および治療履歴が最適な DTR の構築のためには必要である. アウトカムにおよぼす原因は明確でないことが多く, 遺伝的要因, 治療従順性, 家族背景など複雑に交絡した原因が考え得るので単純なマルコフ性から記述されるモデルとは, かけ離れたところに真の確率分布が想定される. このように交絡によってさまざまな潜在的なバイアスを伴うので, 正確な推論を得るためにはランダム化されたデータを獲得することが不可欠となる. この観点から最適な DTR を得るための仮定として逐次多重ランダム化試験割り当て, Sequential Multiple Assignment Randomized Trial (SMART) が考えられた, cf. Murphy (2004) Stat. Med. このデザインは患者の初めの治療のランダム化と, それ以降のステージごとの再ランダム化が含まれる. この再ランダム化のキーはそれ以前の治療の情報に依存して決める点にある. ここで, 各ステージごとに独立にランダム化試験を行うことは良い DTR の決定には役に立たないことに注意する. ある被験者が  $K$  回の診療を受けたとき, ステージ  $j$  ( $1 \leq j \leq K$ ) の共変量  $O_j$  と治療  $A_j$  に対して履歴を  $H_j = (O_1, A_1, \dots, A_{j-1}, O_j)$  と書くとき, DTR とは  $d_j : H_j \mapsto A_j$  を決めることである. このプロセスにおいて SMART は次の2つの仮定を課す:

**A1.** 被験者のアウトカムは他の被験者の治療割り付けに影響されない.

**A2.** 履歴  $H_j$  を条件付けると治療  $A_j$  はステージ  $j+1$  以降の潜在共変量と潜在アウトカムと独立である.

この基本仮定の下では未測定交絡因子の非存在がいえるので, 再ランダム化は履歴  $H_j$  に依存させた治療  $A_j$  の割り付け確率を決めていくことが可能になる. DTR  $d = (d_1, \dots, d_K)$  の分布も扱い易い形に書き下すことができる. DTR  $d$  の価値関数  $V(d)$  は平均アウトカムによって定義され, 最適な  $d$  の全探索は後ろ向きのステージ毎の探索に帰着させた Q-学習が基本となるが, アウトカムの平均の差をモデル化した構造平均モデルを導入した G-推定が提案されている. さらに, 価値関数  $V(d)$  のパラメトリックモデルを導入して直接, 最適な  $d$  を求めるポリシー検索法が考察されている. 傾向スコアを使った逆治療確率重みづけ (IPTW) やリブリケートを発生させる拡張データによる方法 (AIPTW), また傾向スコアによらない G-計算の方法など活発な研究がある. 実際のデータ適用事例や最近の発展と今後の展望については口頭発表の中で紹介する.