

### Boosted Regression Tree の概要と疫学における使用例

Boosted Regression Tree は機械学習の手法の一つである。Regression Tree (回帰木) と Boosting (ブースティング) を組み合わせて使用するもので、データマイニングにおいて強力な手法として知られている<sup>1</sup>。

Tree(木)に基づく方法は特徴空間を長方形の集まりに分割し、各長方形に単純なモデルを当てはめる<sup>2</sup>。非線形な問題や3変数以上の複雑な交互作用に対応でき、(分類問題の場合) 結果が解釈しやすいというメリットがある<sup>3</sup>。一方で、予測性能が低いという弱点がある。

この問題を解決するために、Boosting が使用される。Boosting はアンサンブル学習の一種で、大量の弱い分類器を組み合わせて強力な合議を形成する<sup>2</sup>という発想のもと開発された。間違っって学習されたデータに大きな重みをつけて分類器の作成を繰り返すことで、最終的な分類器の予測性能を高めるのが特徴である。

抄読会では、Tree と Boosting について学習したことをまとめ、発表する。加えて、Boosted Regression Tree の疫学における使用例として、デング熱発生リスクの地理的な分布を推定した論文を紹介する。

#### 参考文献

1. Bhatt, Samir, et al. "The global distribution and burden of dengue." *Nature* 496.7446 (2013): 504-507.
2. 杉山, ほか “統計的学習の基礎 データマイニング・推論・予測” 共立出版(2014)
3. Bi, Qifang, et al. "What is Machine Learning? A Primer for the Epidemiologist." *American journal of epidemiology* (2019).