



医療ビッグデータから生活習慣病発症リスクの高精度予測を可能にする機械学習モデルの開発

キャンパスライフ健康支援・相談センター

特任助教 大山 飛鳥

 Researchmap <https://researchmap.jp/a.ooyama>

招へい研究員 瀬戸 ひろえ

 Researchmap <https://researchmap.jp/SetoHiroe>

教授 山本 陵平

 Researchmap <https://researchmap.jp/yamamotor>

研究の概要

生活習慣病は、個人の生活習慣の改善により発症を抑制できる一方で、健康的な生活習慣の改善や継続は容易ではない。生活習慣の改善のため、自身の生活習慣病のリスクを正しく認識し健康意識の向上に務めることが非常に重要である。これまで、個人の生活習慣病の発症リスクの推定を目指し、統計モデルや機械学習モデルを用いた様々な発症予測モデルが開発されてきた。しかし、開発時に利用可能なデータ規模がボトルネックとなり、十分な精度や信頼性が得られず、本格的な社会実装は困難を極めていた。こうした背景に対し、大阪府国民健康保険の特定健診ビッグデータを活用し、糖尿病・高血圧・脂質異常症の3年以内の発症確率を高精度で予測するAIの開発を目的として研究を実施した。

研究の背景と結果

将来の生活習慣病の発症リスクを予測するモデルを開発するにあたって、これまでの研究では、疾病の発症予測において古典的統計モデルと機械学習を用いたAIモデルの予測精度の性能に差異はないと考えられており、発症リスクの予測に対する機械学習の有用性について十分に解明されていなかった。その背景には、疾病の発症予測を精度良く行うために不可欠となる大規模リアルワールドデータの入手や活用が困難であったことから、機械学習モデルの能力を最大限に引き出すためのデータ数が不足していたという課題が挙げられる。我々の研究グループでは、大阪府国保連合会および大阪府保険者協議会の協力が得られたことで、毎年60万人規模の特定健診ビッグデータを利用することが可能となった。このうち、糖尿病歴の無い約28万人の国保被保険者を解析対象とし、モデルの学習に利用する人数を変化させたときに糖尿病発症の予測確率をどれほど正確に推定できるかを検証した。統計モデルには、従来より活用されてきたロジスティック回帰モデルを、機械学習モデルには、ビッグデータを効率よく最大限に活かすため、モデルの柔軟性や学習速度に優れた勾配ブースティング決定木モデル(LightGBM)を用いた。

図1に示すように、千人規模のデータではロジスティック回帰モデルとLightGBMで予測精度にほとんど差がなかったものの、学習に利用できる人数が増加するにつれてLightGBMの予測精度は向上し、1万人規模のデータでは、LightGBMがロジスティック回帰の予測精度を有意に上回った。これは、従来の小規模のデータによる予測能の比較では両モデルの性能にほとんど差がない一方で、ビッグデータの活用際には機械学習モデルが統計モデルより優れていることを示している。今日、医療機関や自治体が保有する大規模なリアルワールドデータの活用が注目を集めている。本研究は、これら医療ビッグデータ解析における機械学習モデルの有用性を示した。

研究の意義と将来展望

本研究により、ビッグデータにおける機械学習の優位性が明らかとなり、生活習慣病の発症確率を高精度で予測するAIモデルの開発に成功した(図1)。現在は、本モデルを大阪府営のヘルスケアアプリ「アスマイル」に実装し、本アプリを通じて自身の生活習慣病の発症リスクを知ることが可能となった(図2)。これまで、専門知識を持たない住民が健診の項目値を正確に理解することは困難であったが、生活習慣病の発症確率を提示することで、自身の健康意識の向上と自律的な健康行動を促すための重要なツールとして活用されることが期待される。

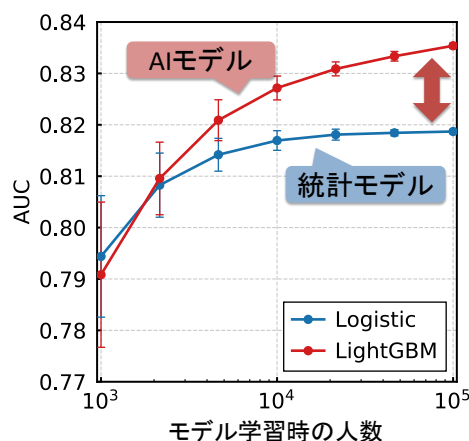


図1. 学習に使用したデータ数とAUCとの関係。



図2. ヘルスケアアプリ「アスマイル」搭載の健康予測AI。

特許

論文

Seto, Hiroe; Oyama, Asuka; Kitora, Shuji et al. Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data. Scientific Reports. 2022, 12, 15889. doi: 10.1038/s41598-022-20149-z
Seto, Hiroe; Toki, Hiroshi; Kitora, Shuji et al. Seasonal variations of the prevalence of metabolic syndrome and its markers using big-data of health check-ups. Environmental Health and Preventive Medicine. 2024, 29, 2. doi: 10.1265/ehpm.23-00216
Kotoku, Jun'ichi; Oyama, Asuka; Kitazumi, Kanako et al. Causal relations of health indices inferred statistically using the DirectLiNGAM algorithm from big data of Osaka prefecture health checkups. PLoS ONE. 2020, 15 (12), e0243229. doi: 10.1371/journal.pone.0243229
大山 飛鳥; 古徳 純一; 土岐 博 ほか. 医療ビッグデータで見る多死社会 : 3. 大阪府のKDBビッグデータから見る超高齢化社会. 情報処理. 2024, 65 (4), e15-e23. doi: 10.20729/00233422

参考URL

キーワード 特定健診、リアルワールドデータ、機械学習、生活習慣病、国民健康保険