

予防・健康領域に対するデータ利活用による効果的なアプローチ

田中 博典 田尻 俊宗

要旨

日本の少子高齢化が今後も加速し、2040年には国民医療費は66.7兆円になると試算されています。医療費の増大、働き手の減少が社会課題となるなか、予防・健康管理は特に重要視されています。本稿では、AIを用いたデータ分析により、予防・健康管理に対してやみくもに対策を打つのではなく、過去データに基づいた科学的かつ効果的なアプローチへの挑戦を紹介します。主に、従来の統計的手法とAI（異種混合学習）を組み合わせることで、生活習慣と検査値の因果関係をとらえる手法を説明します。



ヘルスケア/ビッグデータ/異種混合学習/機械学習/データマイニング/因果推論

1. はじめに

近年、日本は国内の高齢化率が27.7%（2017年）と、超高齢社会を迎えています。今後は更に進展し、2036年には、33.3%となり国民の3人に1人が65歳以上となる社会が到来すると言われてしています¹⁾。国民医療費も、2018年度の39.2兆円から、2040年度には約1.7倍の66.7兆円まで膨れ上がると試算されています²⁾。特に生活習慣病は、2004年度時点で医療費全体の3割、死亡原因の6割を占めており³⁾、医療費増大の要因の1つとなっています。生活習慣病は完治が難しいこともあり、予防・健康管理が特に重要とされています。

厚生労働省は2015年度から、すべての健康保険組合に対して、「データヘルス計画」の作成と実施を求めています。やみくもに保健事業を実施するのではなく、データを活用して科学的にアプローチすることで、事業の実効性を高めて医療費削減などにつなげることがデータヘルス計画の狙いとされています⁴⁾。

本稿では、NEC独自のAI技術を用いて過去の健診データから傾向を学習し、効果的な保健事業のためのデータ利活用取り組み事例を紹介します。

2. 健診データを活用した予防アプローチ

NECソリューションイノベータでは、図1に示すように、蓄積された健診データをビッグデータ分析し、将来の健診結果の予測に取り組んでいます。

2.1 予測の目的

健診データに含まれる生活習慣や検査値を合わせて分析し、そのままの生活習慣を続けた場合の予測や、生活習慣を改善することで変化する予測をシミュレーション化することで、個人の行動変容促進や、改善効果の高い取り組み提案を実現することが目的となります。

2.2 予測技術

予測のためのデータ分析には、NEC独自技術の異種混合学習⁵⁾を用いています。多種多様なデータのなかから精度の高い規則性を自動で発見し、その規則に基づいて、状況に応じた最適な予測が可能です。また、なぜその予測に至ったか、予測の根拠を示すことが可能です。

2.3 ホワイトボックス型AI

昨今、AI活用において「説明責任」が求められるように

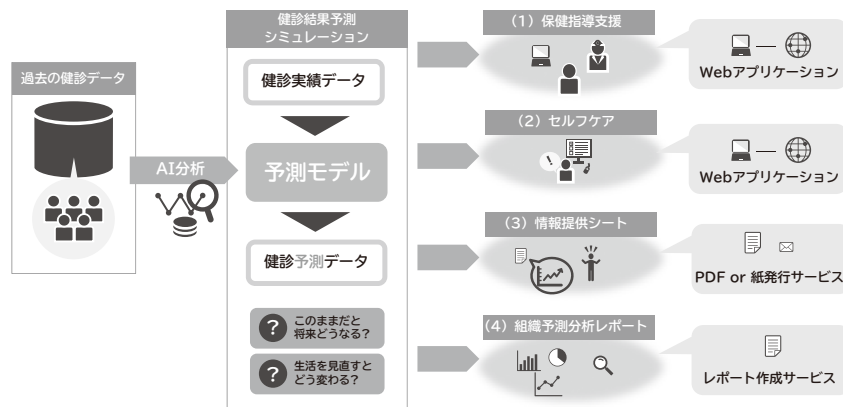


図1 将来の健診結果予測概要

なっています。AIの判断結果が人に影響を与える場合、なぜそう判断したのかが分からなければ、納得感、公平性など、さまざまな課題が浮上します。近年、日本や欧州ではAIの公正や倫理に関する原則がまとめられており、2018年12月に日本政府が策定したAIの原則案では、個人情報、セキュリティの保護などとともに、「AIを利用した企業に決定過程の説明責任」という文言が含まれています。

企業のAI活用において「公平性」「説明責任」や「透明性」を求められている状況です。予防を目的とした将来健診予測を実現するにあたり、予測の根拠を確認し妥当性を評価することは重要なプロセスとなります。

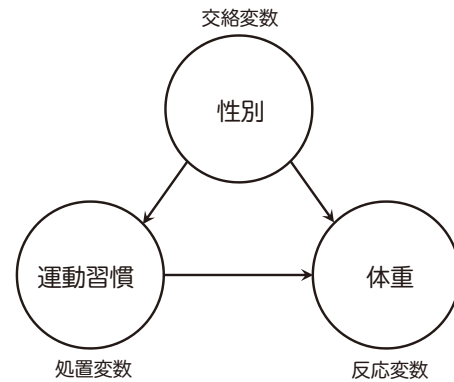


図2 交絡変数

3. 生活習慣の改善が検査値に与える影響の推定

第2章1節で述べた目的を達成するためには、生活習慣を変えたときに検査値にどの程度影響を与えるか（以降、因果効果）をとらえたうえで、予測を行う必要があります。

本章では、運動習慣の変化が体重に与える因果効果を推定することを試みます。

検証には、倉敷中央病院の2011年度から2017年度の健診データ（計297,409件、85,991人分）を用います。

運動習慣の有無は、特定健診における標準的な質問票⁶⁾において定められている問診項目「1回30分以上の軽く汗をかく運動を週2日以上、1年以上実施（はい、いいえの2択で回答）」から判断します。

3.1 検証1 運動習慣の有無による体重の層別比較

まず、2017年度受診者（N=29,436）を「運動習慣：

はい」と回答した群（A群、N=7,392）と、「運動習慣：いいえ」と回答した群（B群、N=22,044）に分割し、各群の同年度の体重の平均値を比較してみます。このとき、各群の体重の平均値はA群：61.87kg、B群：61.05kgとなりました。この結果から、運動習慣を未実施から実施に変化させたときの体重への因果効果は両群の体重平均値の差0.82kgである、と推定してよいでしょうか。これは、大きく以下2点の問題があり、不適切な推定と言えます。

第1に、運動習慣と体重の両方への因果関係を持つ変数（交絡変数）の影響を考慮していません。例えば、「性別」変数について、男性の方が女性よりも平均体重が大きく、かつ運動習慣を実施している傾向にある場合、2群間の体重平均値の差は、運動習慣による効果だけでなく、単に両群の男女比率の差によって生じる効果が含まれている可能性があります（図2）。実際、A群の男性比率は0.590に対し、B群の男性比率は0.449となっており、データか

らも性別が交絡変数である可能性が示唆されます。

第2の問題として、比較対象の数値が個人「内」の変化に関する因果効果を示すものになっていないことが挙げられます。前述の検証では、同年度における個人「間」の運動習慣の違いによって2群に分割しています。そのため、2群間の体重平均値の差は、あくまで「運動習慣のない群とある群の体重平均値の差」でしかなく、本来知りたかった「運動習慣のない個人が運動習慣を身に付けた場合の体重変化値」を示すものにはなっていません。

3.2 検証2 交絡変数と個人内変化を考慮した比較

次に、第3章1節で述べた問題を解消するための検証を試みます。具体的には、以下の手順によって検証を実施します。

- (1) 単年度の健診結果でなく、2年度分の連続した健診データを使用します。ただし、年度ごとの因果効果の傾向差はないと仮定し、 $N_1 \times 7$ 年度のデータ集合全体のうち、任意時点の2年度連続のデータを全抽出し、 $N_2 \times 2$ 年度 ($t=1, 2$) のデータ集合とします。
- (2) (1) のデータ集合から、 $t=1$ において「運動習慣：いいえ」のサンプルのみを抽出します。
- (3) (2) で抽出したサンプル群について、 $t=2$ において運動を継続的に実施した群（ケース群）とそうでない群（コントロール群）に分割します。ただし、この2群を分割する際、単純に $t=2$ において「運動習慣：はい」かどうかだけで分割すると、2年度間のどの時点で運動習慣を実施し、どの程度継続しているかが分からず、適切な推定値を得ることができません。その対策として、問診項目「運動や食生活などの生活習慣を改善してみようと思いませんか（意志なし、意志あり（6カ月以内）、意志あり（近いうち）、取り組み済み（6カ月未満）、取り組み済み（6カ月以上）の5択で回答）」（以降、「生活習慣の改善状況」）を併用します。具体的には、 $t=2$ において「運動習慣：はい」かつ「生活習慣の改善状況：取り組み済み（6カ月以上）」のサンプルのみケース群とし、それ以外をコントロール群とします。
- (4) (3) で分割した2群について、傾向スコアマッチング⁷⁾によって2群間の交絡変数の調整を行います。具体的には、 $t=2$ において処理変数「運動習

慣：はい」となる確率（傾向スコア）を、処理変数（運動習慣：はい）と反応変数（体重）の交絡変数によって推定します。交絡変数としては、 $t=1$ の「年齢」「性別」「生活習慣」（標準的な質問票に準拠したもの）、及び「生活習慣の改善状況」を使用します。推定モデルにはロジスティック回帰モデルを使用します。その後、推定した傾向スコアが近いサンプル同士を2群間でマッチングし、最終的な2群を生成します。マッチング方法は、キャリパーを伴う差近傍マッチングとし、キャリパーは推奨値である $0.2^8)$ とします。

- (5) (4) でマッチングした2群について、差分の差分法⁹⁾によって運動習慣の変化が体重に与える因果効果を推定します。具体的には、 $t=1 \rightarrow 2$ におけるコントロール群の平均体重変化値を、ケース群の平均体重変化値から差し引いた値を平均的な因果効果とします（図3）。

第1の問題（交絡変数の考慮）については、(4)の傾向スコアマッチングによって交絡変数の調整を行うことで解消します（ただし、未観測の交絡変数については未考慮）。

また、第2の問題（個人「内」因果効果の推定）については、(1)で2時点間を考慮し、(2)～(4)で2群を適切に抽出し、(5)の差分の差分法によって2群間の差と2時点間の差の両方を考慮することで解消します。ただし、ケース群が処置されなかった場合（反実仮想、図3の破線）の変化値が、コントロール群の変化値と同等という仮定を置いています。

前述した方法に基づいて検証した結果、(1)によって作成されたデータ集合は $N=145,814$ で、(2)のサンプル抽出により、 $N=108,688$ となります。更に、(3)により、

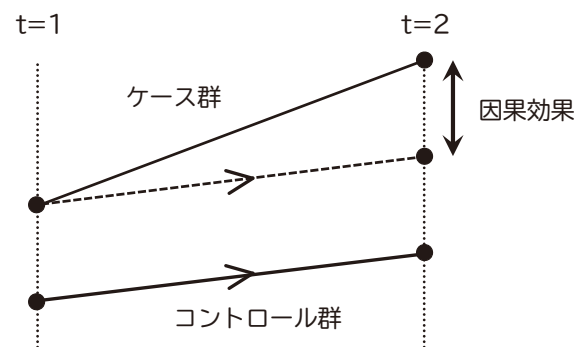


図3 差分の差分法

ケース群 (N=3,047) とコントロール群 (N=105,641) に群分割されます。(4) の傾向スコアマッチングにより、コントロール群はN=3,047に絞られ、ケース群と傾向スコア分布の近いサンプル集合となりました。傾向スコア推定について、2群の識別性能を示すROC曲線をプロットしたときのc統計量 (AUC) は 0.826となり、望ましいとされる0.8以上の値¹⁰⁾となっています。

(5) の差分の差分法によって示された運動習慣を実施・継続することでの体重増減の推定値は -0.804 ($\Pr(>|t|) = 0.0649$) となり、運動習慣のない人が運動習慣を6カ月以上継続した場合、平均して0.804kg程度の体重減少が見込める、と解釈できます。

4. 将来検査値予測モデル事例

本章では、倉敷中央病院のデータを用いて異種混合学習によって構築した予測モデルのうち、1年後体重予測モデルについて紹介します。

モデルに使用され得る変数集合は、第3章の検証結果を元に、交絡変数などを考慮して決定しています。そのなかから、異種混合学習により、実際のモデルに使用される変数が自動で選択されます。

4.1 区分疎線形モデル

今回は、異種混合学習の特長である高い解釈性を特に有する区分疎線形モデルを構築しました。このモデルは、二分木構造をしており、内部ノードが条件式、外部ノード (葉ノード) が線形回帰モデルを持ちます。条件式を根ノードから葉ノードまでたどることにより、どの回帰モデルで予測するかをサンプルごとに決定します。

4.2 1年後体重予測モデル

今回構築した体重予測モデルでは、条件式によって3つの線形回帰モデルに区分される構造となりました。各線形回帰モデルは、予測対象である体重の1時点前の値への係数を持ちます。また、将来の生活習慣を固定した場合の予測を実施するためのモデルなので、予測時点の生活習慣も回帰モデルに含まれます。これは、構造化ベクトル自己回帰 (SVAR) モデルに相当します。

このモデルでは、前回健診で体重が86.5kg以上だった場合に特定の線形回帰モデルが使用されます。その線

形回帰モデルからは、前回健診時点で「夕食後の間食」習慣を持たない人が生活習慣の悪化によって「夕食後の間食」習慣を継続的に実施した場合、現状維持した場合と比較して体重が平均1.04kg増量することや、前回健診時点で「速く歩く」習慣を持たない人が生活習慣改善によって「速く歩く」習慣を継続的に実施した場合、現状維持した場合と比較して体重が平均0.89kg減量することが読み取れます。異種混合学習により、このような解釈性の高いモデルを自動で学習することが可能となっています。

5. おわりに

AIはさまざまな分野での活用が期待されていますが、今後センサーデバイスの発達などで、人間の健康に関する情報は爆発的に増えると考えられていることもあり、予防・健康領域での活用にも期待が高まっています。

これまで健診データ分析で培った技術をベースに、今後は多種多様なデータの突合分析を実現するなど、分析範囲の拡大や提供する情報の多様化を目指し、研究開発に取り組んでいきます。

参考文献

- 1) 内閣府：平成30年版高齢社会白書（全体版）
https://www8.cao.go.jp/kourei/whitepaper/w-2018/html/zenbun/s1_1_1.html
- 2) 内閣府：2040年を見据えた社会保障の将来見通し（議論の素材），pp.1, 2018.5
https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2018/0521/shiryō_04-1.pdf
- 3) 厚生労働省：平成19年度版厚生労働白書，pp.29
<https://www.mhlw.go.jp/wp/hakusyo/kousei/07/dl/0102-a.pdf>
- 4) 厚生労働省：データヘルス計画作成の手引き，pp.12, 2014.12
<https://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000069365.pdf>
- 5) 藤巻 遼平、森永 聡：ビッグデータ時代の最先端 データマイニング，NEC技報，Vol.65 No.2, pp.81-85, 2012.9
<https://jpn.nec.com/techrep/journal/g12/n02/pdf/120219.pdf>
- 6) 厚生労働省：標準的な健診・保健指導プログラム（改訂版）第2編その2，pp.49, 2013.4
https://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryō/kenkou/seikatsu/dl/hoken-program2_02.pdf
- 7) Paul R. Rosenbaum & Donald B. Rubin: The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, Vol. 70, No. 1, pp.41-55, 1983.4
- 8) Paul R. Rosenbaum & Donald B. Rubin: Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score, *The American Statistician*, Vol. 39, No. 1, pp.33-38, 1985.2
- 9) Abadie, A.: Semiparametric difference-in-differences estimators, *The review of Economic Studies*, Vol.72, No.1, pp.1-19, 2005.1
- 10) 星野 崇宏、岡田 謙介：傾向スコアを用いた共変量調整による因果効果の推定と臨床医学・疫学・薬学・公衆衛生分野での応用について，*保健医療科学*，第55巻 第3号，pp.230-243, 2006.9

執筆者プロフィール

田中 博典

NECソリューションイノベータ
イノベーション戦略本部
プロフェッショナル

田尻 俊宗

NECソリューションイノベータ
イノベーション戦略本部

関連URL

NEC 健診結果予測シミュレーション

https://www.NEC-solutioninnovators.co.jp/sl/healthcheckup_sim/index.html

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報 (日本語)

NEC Technical Journal (英語)

Vol.72 No.1 新たな社会価値を生み出すAI特集

新たな社会価値を生み出すAI特集によせて
AIとデータ活用によるデジタライゼーションの拡大

◇ 特集論文

AIの社会実装に向けた取り組み

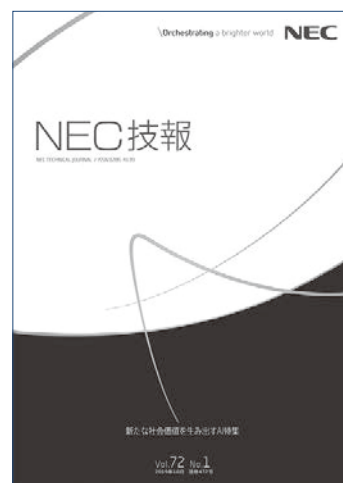
「NECグループ AIと人権に関するポリシー」とその実践に向けた取り組み
AI時代の人材育成

デジタルトランスフォーメーションを加速するAI活用サービス・ソリューション

「みんなで創るAI」を支えるNEC Advanced Analytics Platform (AAPF)
物体指紋認証技術による個人識別機能の活用
画像処理コントローラへのディープラーニング活用による外観検査ソリューション
通信予測制御技術を活用した車両の遠隔監視ソリューション
働き方改革や健康経営を支える「NEC感情分析ソリューション」
オフィスのセキュリティと利便性を向上する「顔認証ソリューション for オフィス」
業務自動化・省力化を実現する自動応答ソリューション (AIチャットボット) の概要
ビジネス創造へのワークシフトを加速するソリューション (AI for Work Shift Support) の概要と実証事例
AIを有効活用する「NEC Energy Resource Aggregation クラウドサービス」
容体変化予兆検知技術による早期退院支援の取り組み
予防・健康領域に対するデータ利活用による効果的なアプローチ
AIを活用したインサイトマーケティング事業の共創
時代のムードを味わえる「あの頃はCHOCOLATE」の開発

人とともに未来を創る最新のAI技術

あらゆる小売商品を認識可能にする多種物体認識技術
ネットワークインフラを活用して実世界を見える化する光ファイバセンシング技術
熟練者の意思決定を模倣する意図学習技術
グラフベース関係性学習 (GraphAI)
時系列データ モデルフリー分析技術
社会インフラの最適運用を支援する論理思考AI
少量データ向け深層学習技術
AIを支えるコンピューティングプラットフォーム



Vol.72 No.1
(2019年10月)

特集TOP