

# 医療健康支援技術の研究開発と糖尿病患者の血糖コントロール不良予測への適用

NTTでは、AI（人工知能）で医療・健康等データを分析し、健康維持に資する効果的かつ効率的な介入を支援する、医療健康支援技術に取り組んでいます。本稿では、医療・健康等データの「データがまばら」という課題を解決する、欠損の影響を抑制した特徴抽出手法について説明します。また、東京大学医学部附属病院（東大病院）と共同で糖尿病治療の課題の1つである患者の血糖コントロール不良予測へこの手法を適用した事例を紹介します。

くろさわ ひさし†1 ふじの あきのり†2

倉沢 央 / 藤野 昭典

はやし かつよし†3

林 勝義

NTTサービスエボリューション研究所<sup>†1</sup>  
 NTTコミュニケーション科学基礎研究所<sup>†2</sup>  
 NTT研究企画部門<sup>†3</sup>

## 生活習慣病予防介入の重要性

生活習慣病とは、食生活や運動習慣、睡眠、飲酒などの生活習慣が発症や進行に大きく関与する疾患群であり、糖尿病やがんなどがこれに含まれます。生活習慣病は、医療費の約3割、死亡者数の約6割を占めており、その発症・進行予防は健康寿命の延伸に向けた重要課題となっています。生活習慣病の発症・進行予防には未病状態や発

症初期の患者への積極的な介入が効果的なことが知られており、特定健診・特定保健指導などの介入施策が実施されています。しかし、このような介入施策は国や自治体、健康保険組合への多大な費用負担や医療従事者への多大な稼働を必要とするため、ICT活用による生活習慣病予防介入の効率化・効果増進が求められています。

## NTTの取り組む医療健康支援技術

NTTでは、AI（人工知能）を活用して医療・健康等データを分析し、健康維持に資する効果的かつ効率的な介入を支援する、医療健康支援技術を研究開発しています。医療健康支援技術の概要を図1に示します。本技術は、個人の将来の疾病リスクや治療行動、治療効果を予測します。そして、予測結果に基づき個人に効果的な介入支援



図1 医療健康支援技術の概要

プランを策定し、医療従事者を通じた介入や、身の回りのさまざまなデバイスやロボットを通じた介入に役立れます。

医療健康支援技術の研究開発には、NTTがこれまで培ってきたデータ分析に関する知見が数多く活かされています。一例を挙げると、人の行動分析の知見を糖尿病治療の課題の1つである患者の「受診中断」という治療行動の分析に活かし、東京大学医学部附属病院（東大病院）と共同で電子カルテデータを基に受診中断を個人レベルで予測するモデルを構築しました<sup>(1)</sup>。さまざまな領域に共通する課題の解決をめざして研究開発したデータ分析技術を医療健康領域に展開し、研究開発を推進しています。

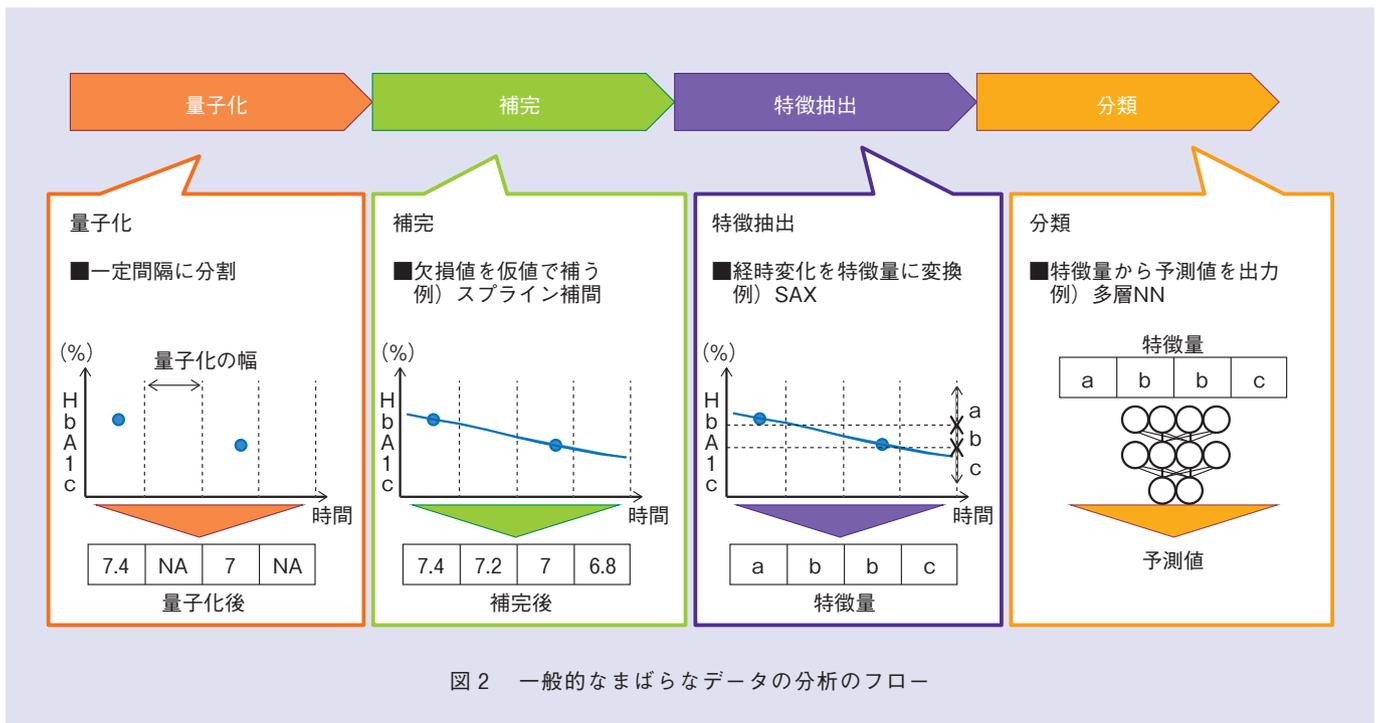
本稿では、「データがまばら」という課題を解決する、欠損の影響を抑制した特徴抽出手法について説明します。この技術は、医療健康支援技術を構成する、個人の将来の疾病リスクや治療行動、治療効果の予測精度向上に寄与し、介入支援の効果や効率を高めることに役立ちます。手法の説明の後に、東大病院と共同でこの手法を糖尿病患者の血糖コントロール不良予測へ適用し、予測精度向上の効果を確認した事例を紹介します。

### まばらな医療・健康等データの課題

医療・健康等データの多くは、不定期に計測された値で構成されます。一般に、病院や診療所では、受診日と同

一日に検査や処方が実施されます。通院間隔は病態や患者や医師の都合により変化するため、検査や処方の実施間隔もそれに応じて変化し、間隔は一定でなくなります。また、血圧値や活動量のような家庭内で計測される値は、計測忘れが生じると、間隔が一定でなくなります。このように計測間隔が一定でない値を、「まばらなデータ」と呼びます。

個人のまばらなデータから、将来の疾病リスクや治療行動、治療効果を予測する場合、従来は、「量子化」「補完」「特徴抽出」「分類」という4つのステップで処理していました（図2）。量子化は、まばらなデータを一定間隔に区切り、区切られた各期間に代表値



もしくは欠損を表す値を割り当てる処理です。例えば、1日間隔で区切った場合、計測忘れの日には欠損を表す値が埋められます。補完は、量子化されたデータの欠損値を仮の値で補う処理です。例えば、線形補完は、線形多項式で経時変化を表現できると仮定し、欠損値の前後の値から多項式の係数を算出し、仮の値を推定します。特徴抽出は、補完によって擬似的に計測間隔が一定になったデータから経時変化を表す特徴量を抽出する処理です。例えば、特徴抽出手法の1つであるSAX (Symbolic Aggregate approXimation) は、実数値の一定間隔のデータから記号列の特徴量に変換します。最後の分類は、特徴量から疾病リスクや治療行動、治療効果の予測値を出力する処理です。例えば、分類手法の1つであるSVM (Support Vector Machine) は、疾病リスク有無を見分けるのに適した特徴量空間における超平面を訓練データから学習します。

4つのステップのうち、補完と特徴抽出の処理がまばらなデータを用いた予測の大きな課題でした。具体的には、補完の処理によって正確に欠損値を仮の値で補えるとは限らないため、補完後に特徴抽出して得た特徴量は、計測値の特性を正確に反映しているとは限らないという問題がありました。この問題は、特徴量の質を低下させ、予測精度低下につながります。このように、個人ごと、データの種別ごとに計測間隔が多様なまばらな医療・健康等デー

タから、質の高い特徴量を抽出することは技術的に困難でした。

### 欠損の影響を抑制した特徴抽出手法

まばらなデータから質の高い特徴量を抽出する技術を検討するうえで、私たちは質の高い特徴量を「まばらなデータの特性を反映しているだけでなく、まばらなデータの経時変化を復元できるもの」と考えました。そこで、補完と特徴抽出を同時に処理する技術を研究開発し、オートエンコーダと呼ばれる特徴抽出手法から着想を得て、欠損の影響を抑制した特徴抽出手法を考案しました。

オートエンコーダは、深層学習の一種で、次元圧縮に用いられる手法です。この手法は、入力データから特徴量に変換するエンコーダと、特徴量から入力データと同じ次元の出力データに変換するデコーダの2つで構成されます。オートエンコーダの学習では、出力データが入力データに近づくように、エンコーダとデコーダに含まれる各パラメータを収束させます。

私たちは、オートエンコーダの構造を参考に、不等間隔データから欠損補完と特徴抽出を同時に行う、「欠損を考慮したオートエンコーダ」を設計しました。欠損を考慮したオートエンコーダの概要を図3に示します。欠損を考慮したオートエンコーダは、入力データと出力データの差が最小になるようにパラメータを学習するオートエンコーダにおいて、式1に示すように、

入力データで欠損している要素以外で計算される入出力ベクトルの差 $L(w)$ が最小になるようにパラメータ $w$ を学習します。 $f$ はオートエンコーダで定義される関数を表します。

$$L(w) = X_{An} \cdot (X_{Bn} - Y_n(w))$$

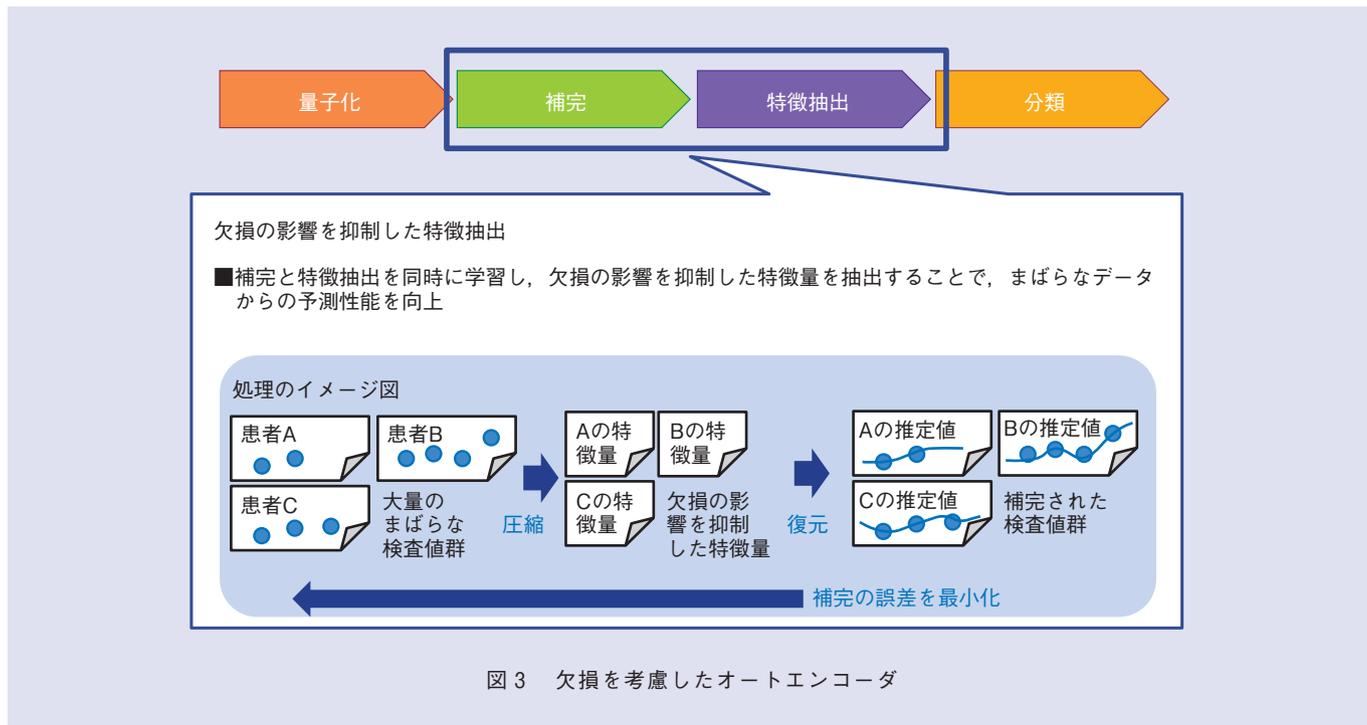
$$Y_n(w) = f(X_{An}, X_{Bn}; w) \quad (式1)$$

$X_{An}$ は計測値があれば1、欠損していれば0の値で構成された特徴量ベクトル、 $X_{Bn}$ は欠損値を仮の値(例えば、0)で埋めた特徴量ベクトル、 $Y_n$ は $X_{Bn}$ と同じ次元で構成された出力データを表します。学習が収束すると、入力データの欠損値が補完されたデータが出力データとして得られます。同時に、欠損に影響を受けない低次元の特徴量に変換するエンコーダが得られます。

欠損の影響を抑制した特徴抽出手法は、個人ごとに計測間隔が多様であっても、一律の次元数の特徴量を抽出できます。また、データの種別が多様であっても、エンコーダとデコーダの表現力が十分であれば同一のモデル構造を転用して学習できます。

### 糖尿病患者への介入支援における技術適用

糖尿病は生活習慣病の1つであり、発症・進行予防に向けたさまざまな介入施策が実施されています。重症化予防は「血糖コントロール指標におけるコントロール不良者の割合の減少」が具体的目標に位置付けられており、血糖コントロール不良の患者には早期から治療強化の介入が実施されます。し



かし、血糖コントロールの急激な是正あるいは厳格すぎる血糖コントロールは、重篤な低血糖や最小血管症の増悪といったリスクを伴うため、患者の病態に応じて治療強化を慎重に進める必要がありました。

そこで、東大病院と共同で、糖尿病患者の将来の血糖コントロール不良を予測するモデルを構築しました<sup>(2)</sup>。血糖コントロール不良の定義は「受診日を起点に、将来64週のHbA1c（ヘモグロビンA1c）の最悪値が合併症抑制の血糖コントロール目標値7%を上回り、かつ過去64週の最悪値を上回る」としました。血糖コントロール不良を示す疑似データを図4に示します。

HbA1cは過去1、2カ月の血糖の

平均的な状態を調べるために測られます。一般に、糖尿病患者のHbA1c検査は外来受診と同一日に実施されます。通院間隔は病態や患者や医師の都合により変化するため、HbA1c検査実施間隔もそれに伴って変化し、間隔は一定ではありません。

血糖コントロール不良予測のモデルでまばらなHbA1c値から特徴抽出するにあたり、前述した欠損を考慮したオートエンコーダを用いました。予測モデルは、過去64週のHbA1c値を入力とし、1週間単位で量子化し、欠損を考慮したオートエンコーダで8次元の特徴量を抽出します。その後、8次元の特徴量を入力とした多層ニューラルネットワークの二値分類器で、将

来64週に血糖コントロール不良になる確率を求めます。

予測モデルの評価には、東大病院に通院する糖尿病患者7180名の電子カルテデータを医学部倫理委員会の承認を得て使用しました（承認番号10705）。対象期間は2006年11月27日から2016年1月29日とし、受診日を起点に過去と将来64週に4度以上HbA1c検査を実施したケースを対象レコードとしました。レコード総数は24万1211件、このうち血糖コントロール不良は8万6299件、血糖コントロール良好は15万4912件でした。HbA1c値を1週間単位で量子化すると、平均5.9週（標準偏差2.6週）間隔で検査が実施され、受診日を起点に過去64週のうち

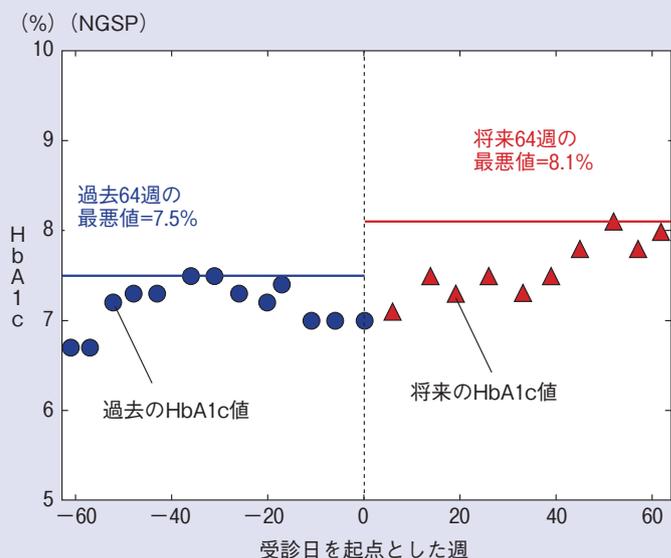


図4 血糖管理不良の一例

■参考文献

- (1) 倉沢・藤野・林：“機械学習を用いた糖尿病患者の治療行動予測,” NTT技術ジャーナル, Vol.29, No.6, pp.32-36, 2017.
- (2) 倉沢・林・藤野・芳賀・脇・野口・大江：“機械学習を用いた糖尿病外来患者の血糖管理不良予測,” 第37回医療情報学連合大会, 2017.
- (3) 山田：“デバイスをつなぐインタラクション制御技術「R-env:連舞」,” NTT技術ジャーナル, Vol.28, No.2, pp.18-21, 2016.

ち平均53.6週（83.7%）の検査値が欠損していました。

10分割交差検定\*の結果、ROC AUC (Receiver Operating Characteristic Area Under the Curve) 値が0.80の予測性能を示すことができました。これは提案モデルで予測した確率を基に糖尿病患者を比較した場合に、将来64週に実際に血糖コントロール不良になる患者を80%の正確さで見つけられることを意味します。このモデルは、糖尿病患者への血糖コントロール不良回避に向けた介入時期の絞り込みに役立つと期待されます。

\* 10分割交差検定：データセットを10分割し、そのうちの1つをテストデータ、残りの9つを訓練データとして10回検証する評価方法です。

今後の展開

ICT活用による生活習慣病予防介入の効率化・効果増進に向けて、今後もNTTは医療健康支援技術の研究開発を推進していきます。現在、標準規格に沿った電子カルテデータの普及やさまざまな生体情報を計測するデバイスの普及が進み、大量の医療・健康等データを扱える環境が整いつつあります。これらを活用しながら個人の将来の疾病リスクや治療行動、治療効果の予測モデルの改善に取り組んでいきます。また、NTTがこれまで培ってきたデバイス制御・連携技術<sup>(3)</sup>等を活用しながら、介入支援についてもさらに検討を深め、国民の健康寿命の延伸に貢献していきたいと思いをします。



(左から) 藤野 昭典/ 倉沢 央/  
林 勝義

国民の健康寿命の延伸に向けて医療健康支援技術の研究開発に継続して取り組み、人々のQoLの向上や医療費増加の抑制など社会課題の解決に貢献していきます。

◆問い合わせ先

NTTサービスエボリューション研究所  
ネットワークロボット&ガジェットプロジェクト  
TEL 046-859-3776  
E-mail hisashi.kurasawa.vx@hco.ntt.co.jp