

令和 5 年 6 月 19 日現在

機関番号：14201

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K10620

研究課題名（和文）特定保健指導のアウトカム最適化を目指すAIを活用した大規模テキストデータ解析

研究課題名（英文）AI-based large-scale text data analysis for optimizing outcomes of specific health guidance.

研究代表者

池之上 辰義（Ikenoue, Tatsuyoshi）

滋賀大学・データサイエンス学系・講師

研究者番号：70761443

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究では、保健指導記録と健診情報を突合することによって、特定保健指導で行われている指導内容とその結果についての縦断的な評価を行った。機械学習を用いて保健指導記録の分類と解析をおこない、4つないし6つに保健指導内容が分類できることを可視化した。さらに、ディープラーニングを用いて保健指導記録から1年後の健診での腹囲5cm減少を予測するモデルの構築を行い、精度62%のモデルと構築した。一方で既存の解析法を用いて「特定保健指導」についての被指導者反応性を検討した。肥満に対する意識と行動は、保健指導による体重減少と関連していることを明らかにした。

研究成果の学術的意義や社会的意義

本研究はこれまでブラックボックスとなっていた特定保健指導がその結果にどのように影響するかを機械学習を用いて検討した。今後、より精緻なモデルの作成により保健指導の質的な向上に寄与することが期待される。

研究成果の概要（英文）：In this study, a longitudinal evaluation of the content and results of specific health guidance was conducted by comparing health guidance records with health check-up information. The health guidance records were classified and analysed using machine learning, and it was visualised that the content of health guidance could be classified into four or six categories. Furthermore, a model for predicting a 5 cm reduction in abdominal circumference at the health check-up one year later was constructed from the health guidance records using deep learning, and a model with an accuracy of 62% was constructed. On the other hand, existing analysis methods were used to examine the recipients' reactivity to 'specific health guidance.' It was found that attitudes and behaviours towards obesity were associated with weight loss due to health guidance.

研究分野：衛生学および公衆衛生学分野関連：実験系を含まない

キーワード：特定保健指導 特定健診 機械学習 テキスト解析 ディープラーニング

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

特定保健指導は特定健診でメタボリック症候群が指摘された人を対象に生活習慣に関する指導を行い、生活習慣病の重症化予防を目的に 2008 年 4 月より始まった。特定保健指導は「保健指導員による対面指導と電話指導」がその中核であり、指導の質が指導効果の成否を決定する。しかしながら、特定保健指導制度は指導のボリューム(頻度とかける時間)を定めているのみであり「標準的な健診・保健指導プログラム」で一定の推奨指導内容が示されているが、指導要素(運動、食事、医療機関への受診など)のバランスや指導内容の伝え方は、指導員や保健指導実施業者ごとに異なっている。「質の高い保健指導」においては、この指導要素の伝え方が重要である。一方で、保健指導記録の中にテキストデータとして指導内容が記録されているが、これらテキストデータの活用がされておらず、指導の実態はブラックボックスとなっている。また、指導効果の検証は多くなされているものの、指導内容が無視されたり、選択バイアスにより過大な評価を行ったりしている。これら指導の実態を把握し、保健指導の効果を最大化するための指導内容に対する科学的な検討が必要とされている。

保健指導記録などテキストデータとして記録されるヘルスデータは多岐にわたる。しかし、これらのデータは従来まで報告書としての入力を目的にシステムへ蓄積されるのみで、本来の用途外である集計や解析の対象とするには煩雑な処理作業と高度な解析環境を必要とした。また、情報の重要度も様々な多量のテキストデータが混在し、解析に必要な情報のみを抽出する技術が不足していた。日本語特有の接続詞や外字体の使用による字体のバラツキは、テキストデータ解析を困難にしていた。さらに、テキストデータの経時的な突合、レセプト情報など関連するヘルスデータとの突合は共通 ID が必要であり、大規模な保健指導記録の保有者とレセプト、特定健診記録の保有者が異なることからその突合は困難なものとなっていた。また、健診の影響を評価するにも、比較対照の選択が困難で、選択バイアスを含んだ結果の提示しかできなかった。

2. 研究の目的

指導が「いつ」、「どのように」行われたのかを明らかにするために、詳細に記載された保健指導記録をもとにテキストマイニングにより保健指導を分類し、保健指導の実態を明らかにする。これにより、保健指導内容のバリエーションを明らかにする。(保健指導記録の分類と解析)

さらに、その指導内容が被指導者の「アウトカム」に関連性を持つのか明らかにするために、保健指導記録にリンクする指導後の健診およびレセプト記録をもとにアウトカムとの関連性を明らかにする。さらに、指導を受けた対象者内でのアウトカムの比較を行い、効果判定で必要となる介入試験を行うにあたって必要不可欠な対象者の選定、効果量の推定に重要な知見を与えるとともに、暗黙知として扱われてきた指導要素のバランスや指導内容の伝え方とアウトカムの関連性を明らかにする。(アウトカムとの関連性解析)

3. 研究の方法

【保健指導記録の分類と解析】

使用したデータセットでは、各指導後に指導員が選択する目標カテゴリー(適正受診、活動量、栄養バランスなど 19 のカテゴリー)と目標カテゴリーに対応した自由記載が存在している。これらの関連性を「ラベル付きテキスト」として AI に学習させ、自由記載記録の持つ特徴を数値化し、可視化をした。自由記載部分についてクラウドシステムである Google Cloud Platform (GCP) Natural Language API で日本語形態素解析を行い、形態素解析の結果から名詞、動詞を抽出し、順序を加味した単語ベクトルを作成した。作成した単語ベクトルを Doc2vec (Lau, J. H., & Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. arXiv preprint arXiv:1607.05368.) の手法を用いて、それぞれの文章の持つ類似性を目標カテゴリーを特徴量として表示させた(ポートフォリオの作成)。さらに、この特徴量をベクトルとして距離を計算し、距離の近いもの同士を同じグループとするように、k-means 法を用いて分類した。分類された各指導における特徴量の偏差値を求め、その平均値を算出し、レーダーチャートを作成し、研究者の合議にて臨床的に意味づけが可能なグループを選択した。

【アウトカムとの関連性解析】

テキスト解析手法の急激な進歩に応じて、当初予定していた文脈による分散を用いた集団への分類である Doc2vec からディープラーニングを用いる Transformer を基盤とした解析に解析手法の変更を行う必要があると考えた。そこで、Transformer に基づいた BERT モデルを用いて保健指導記録から 1 年後の健診での腹囲 5 cm 減少を予測するモデルの構築を行った。まず、モデル 5062 人の保健指導記録のみを用いてモデルの構築を行った。ついで、特定保健指導の対象となった健診における健診情報(性別、ALT、HbA1c、中性脂肪)を保健指導記録にテキストとして追加しモデルの構築を行った。モデル作成とは別にランダムに選択した 613 人の指導記録についてバリエーションセットとしてモデルの当てはめを行い、精度について検討を行った。

【「特定保健指導」についての反応性を考察のための解析】

保健指導時の体重減少と肥満の自覚・対策の関連性を検討するために保有コホートにおいて「特定保健指導」についての反応性を検討した。本コホート研究では、2013年4月から2020年3月までに日本全国で実施された保健指導介入に参加し、1年間の体重変化を評価した40~74歳の個人を対象とした。エクスポージャーは、ベースラインの質問票で定義された過食や運動不足の意識と、患者が肥満に対して行動を起こすかどうかの組み合わせとして定義した。性別、年齢、過去の指導経験などの交絡因子は、重回帰モデルで調整した。

4. 研究成果

【保健指導記録の分類と解析】

18,006 あった特定健診指導記録を Dog2Vec を用いて特徴量に変換した。特徴量のもととなる情報は、指導員の登録した指導内容とした。右の例はテキスト解析により得られたポートフォリオの例である。特徴量は1を最大とし、0を最低としている。本例では受診に関連する項目、自身の状況把握に関する項目、間食などの生活習慣に関する特徴量が高いことがわかる。

STEP1: テキスト解析により保健指導のポートフォリオを作成 (ポートフォリオの例)

(テキスト) HbA1cの値が7.8パーセントと受診勧奨値を超えている為、受診の必要性を説明し、受診勧奨した。健診結果もあまりきちんとみておらず健康への関心が低い。メタボリックシンドロームの進行について説明し、自覚症状が無くても症状が進行する事があると伝え、まずは自身の状態を知る為にも受診が必要と強調し、早期受診を促した。減量の必要性を伝え、飲料や間食、飲酒習慣改善を目標とした。

ポートフォリオ

A.モニタリング	B.正しい情報	C.適正運動	D.適正採食	E.応答の自己管理	F.活動量	G.飲料習慣	H.間食習慣	I.飲酒習慣	J.夕食の時間と量
0.945	0.538	0.927	0.909	0.869	0.506	0.658	0.797	0.515	0.451
K.食事リズム	L.栄養バランス	M.飽和脂肪酸の量	N.塩分バランス	O.野菜/果物/きのこ類の量	P.糖分量	Q.喫煙	R.夜間起床	その他	
0.534	0.455	0.414	0.454	0.429	0.463	0.400	0.462	0.487	

上記の特徴量を基に k-means 法を用いて、保健指導記録の自由記載箇所(テキストデータ)を分類した。本検討では医師・保健師の合議を行い4グループおよび6グループへの分類を行った。

4グループに分類した時の平均ポートフォリオ



6グループに分類した時の平均ポートフォリオ



上図は得られたグループの平均特徴量のレーダーチャートである。

保健指導記録の自由記載箇所(テキストデータ)を人工知能で解析することによって、従来のデータでは見えなかった指導毎のパターンの違い(指導実態)が可視化することができた。

【アウトカムとの関連性解析】

健診情報はテキストに健診情報を付与する方法で行った。具体的には、「性別は男性。」もしくは「性別は女性。」という分を加えて、性別情報を付与した。同様にGPT(ALT)、HbA1c、中性脂肪についてもそれぞれのカットオフに基づく「正常」、「高値」、「極めて高値」のカテゴリー情報を文として指導記録に付与して指導記録文をBERTモデルに適応し、学習させた。バリデーションセットでは、指導記録のみで作成したモデルでは1年後の5kg以上の減少を予測する精度は62%を得た。一方で健診情報を付与して作成したモデルでは52%の精度となった。

本研究では、1回の保健指導によりアウトカムとの関連性を評価した。しかし、保健指導では初回面談からその後の継続的な指導が行われており、1回の指導での指導記録のみで評価するには限界があると考えられた。今後、今回用いたBERTモデルとともにTransformerに基づくGPTモデルにより各指導での要約を自動作成させ、それをBERTで解析するなどの手法が必要になると考えられた。また、健診情報についても同様にして各年での要約を作成し、アウトカムとの関連性を評価していくことが必要であることが示唆された。

【「特定保健指導」についての反応性を考察のための解析】

「特定保健指導」についての反応性を考察のための解析では、1,772名の参加者(平均[SD]ベースライン体重、77.1[9.8]kg)のうち、行動を起こし、過食の自覚がある参加者は、行動を起こさず、自覚もない参加者に比べて、体重が低かった(調整後差、-1.02kg; 95%CI、-1.52~-0.52)。一方で、行動を起こさず、運動を意識している参加者は、体重の増加と関連していた(調整後差、0.59kg; 95%CI、0.16~1.01)。これらのことから、肥満に対する意識と行動は、保健指導による体重減少と関連しており、保健指導の効果もこの患者特性により変化する可能性があり、考慮すべき交絡の一つと考えられた。

5. まとめ

本研究では、保健指導記録と健診情報を突合することによって、特定保健指導で行われている指導内容とその結果についての縦断的な評価を行った。縦断的評価を行う中で、機械学習を用いての検討を実施していたが技術的な変化が大きく当初の実施予定からの大幅な変更が求められた。現時点での精度が十分でないことから、Transformer をさらに利用し記録情報のより精緻な要約情報の作成および健診情報についても時系列での情報を活用することで精度をより高めることが可能であると考えます。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

現在、2本の論文について登校中。1本を作成中。

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	山田 ゆかり (Yamada Yukari) (00306846)	京都大学・医学研究科・特定講師 (14301)	
研究分担者	福間 真悟 (Fukuma Shingo) (60706703)	京都大学・医学研究科・准教授 (14301)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------