

健康診断データを用いた生活習慣改善行動における 価値推定手法の提案

濱谷 尚志¹ 田中 茂樹¹ 深澤 佑介¹

概要：

昨今のデータ収集、活用の機運の高まりにより、健康診断のデータ化が進められている。本研究では健康診断のデータを用い、健康の維持や向上のための行動変容のための技術として、生活習慣改善行動における価値推定手法を提案する。提案手法では健康診断における生活習慣に関する問診への回答結果、および医師による受診者の総合所見のデータを用い、健康状態予測モデル、価値計算フレームワークおよびアドバイス生成機構の組み合わせによりユーザが取りうる生活習慣改善行動の価値を定量化する。実際に会社従業員のべ 5000 人超の健康診断データに基づき提案手法による予測モデルを構築した結果、各データにおける翌年の医師による総合所見の予測精度が約 64.1%であることを確認した。さらに、提案手法により実際に 99.6%の従業員に対し生活習慣改善のアドバイスを提示できることを確認し、生活習慣の改善において重要な問診項目を複数確認した。

1. はじめに

昨今のデータ収集・活用の機運の高まりにより、あらゆる情報がデータとして収集される時代が到来しつつある。健康に関するデータも例外ではなく、労働安全衛生法に基づき事業者により労働者に対して実施される、医師による健康診断もデータ化が進められている。同時に、我が国は長寿高齢化という課題に直面しており、超高齢社会において予期される医療費や介護費増大といった課題に対処していくために、生産年齢人口における生活習慣病の予防、すべての世代の健康の維持・向上により健康寿命を伸ばすことが不可欠である。

こういった背景の中で、生活習慣病の予防という観点で、これまでに様々な研究者らにより蓄積した健康診断のデータを用いて高血圧、糖尿病など特定の疾患の発症予測を試みるアプローチがなされてきた [1]。こういった手法では、大規模なデータ分析により健康診断項目と疾病の因果関係を学習することで、各ユーザの健康診断データに対してリスクを判断することが可能となる。一方で、実際にユーザに運動のタイミングなどを通知しユーザの生活改善のために介入するためには人の複数の認知判断プロセスを経る必要があるが、最終的な介入成功率が全体の 3 割未満にとどまるという実験報告もなされており [2]、実際にユーザの行動変容を促すことは容易ではないことが示されている。し

たがって、蓄積した健康診断データにより疾患リスク自体の予測はできるようになりつつあるが、リスクの予測結果を踏まえてリスクを回避するためにユーザに対して生活習慣改善などの介入を成功させることが課題となっている。

そこで本研究では健康の維持、向上のための行動変容技術を検討する。ヘルスケアの分野においては、健康増進のためにアドバイスを実施する際に、実際にユーザがアドバイスに従って行動してくれる割合を高めなければならない。ユーザの行動への動機づけとしてよく用いられるのが金銭的なインセンティブである。インセンティブは単純で効果のある方法と考えられる一方で、インセンティブ目的の行動はユーザの納得感を伴わない可能性があり、必ずしも持続可能な手段であるとは言い難い。健康の維持という観点では、長期的な取り組みが求められるため、本研究では納得感に注目し、ユーザ自身が納得感を持って自発的、かつ持続的に健康増進に取り組むためにアドバイスに対する納得感を高めるアプローチを採用する。

そのための具体的な手段の一つとして、本稿では健康診断における生活習慣に関する問診への回答結果、および医師による受診者の総合所見のデータを用いてユーザが取りうる生活習慣の改善行動に対し将来の健康状態に対する行動の価値を定量的に計算する手法を提案する。提案手法では、ある一年でのユーザの問診回答、および総合所見のデータに基づいて、翌年の総合所見を予測するモデルを機械学習アルゴリズムを用いて構築する。この予測モデル

¹ (株) NTT ドコモ

を用い、ユーザが生活習慣を改善した場合における翌年度の総合所見の予測値、および生活習慣を現状から改善しなかった場合における翌年度の総合所見の予測値を比較することで、生活習慣改善による総合所見の予測値改善効果を算出する。この比較を、ユーザが取りうるあらゆる生活習慣改善の組み合わせに対して行うことで、ユーザに対して最も効果的な生活習慣改善のアドバイスを行うことが可能になる。

本手法により、食生活の改善・飲酒習慣の改善・運動習慣の改善など、健康維持・推進につながる生活習慣の改善行動のそれぞれの行動に伴う価値を定量化することができる。健康診断における生活習慣の問診はユーザ自身が回答する項目であり、健康診断で診察する項目の中でもユーザの努力によって制御しやすい項目であることから、各生活習慣改善に伴うユーザの負担と、健康増進におけるそれぞれの行動の価値を照らし合わせることで、ユーザがどの行動を改善するかを決定するための支援が可能になる。

実際に会社従業員のべ 5000 人超の健康診断データに基づき、提案手法による予測モデルを構築した結果、各データにおける翌年の医師による総合所見の予測精度が約 64.1%であることを確認した。さらに、提案手法により実際に 99.6%の従業員に対し生活習慣改善のアドバイスを提示できることを確認し、生活習慣の改善において重要な問診項目を複数確認した。

本研究の貢献は以下の通りである。

- 健康状態予測モデルを活用した価値計算方式の提案により、生活習慣の改善という価値の定量化が難しいものに対し価値を定義し計算する手法を提案した。
- 既存の健康診断の問診項目において、問診項目の回答を掛け合わせることで新しい意味を持つ特徴量（掛け合わせ特徴量）を提案し、実際に予測モデルにおいて元の項目よりも重要な説明変数となることを確認した。
- 提案手法により生成可能なアドバイスの傾向を分析し、データに基づいて価値の高い行動を明らかにした。

2. 関連研究

2.1 健康リスク予測に関する研究

近年、健康診断に関するデータの電子化、およびデータを活用した疾病のリスク予測の取り組みが進められている。Ushida らは fuzzy neural network を用いてメタボリックシンドロームのリスク因子を推定する取り組みを行っている [3]。さらに、具体的な疾患のリスク予測として、機械学習アルゴリズムにより高血圧 [4]、心血管疾患 [5]、糖尿病 [6]、[7] を予測するアプローチがなされている。これらのアプローチでは、neural network や boosting などの手法により、蓄積したデータに基づいて予測モデルを構築する。また、予測モデルにおける特徴量の工夫として、医師が記載するコメントに対し自然言語処理を適用し特徴を抽

出するアプローチもなされている [5]。また、近年様々な分野で活用されている深層学習により、糖尿病、高血圧、脂肪肝の 3 種類の疾患を同時に予測する取り組みも行われている [1]。以上のように、蓄積した診断データに対し深層学習を含む機械学習アルゴリズムを適用することで、現在の診断結果に基づいた将来の糖尿病などの生活習慣病の発症リスク予測が行われている。

一方、医療分野における AI 技術の別のトレンドとして、MRI や CT 画像に対し、画像処理分野で大きな成果を収めている深層学習を適用することによる画像中の器官のセグメンテーション、特定の被写体の検知、健康状態の診断などの研究も盛んに行われている [8]。医療用途の画像を用いた認識等では一般に高い精度を期待できる一方で、医療機関で専用の設備を用いて撮影を実施する必要がありデータを収集しづらいという課題がある。また、MRI や CT 検査は何らかの疑義により実施することが主であり、健康な状態の人のデータを得る目的では行われないと考えられる。対照的に健康診断のデータは年次ごとに実施することが多く、健康な状態の人のデータも数多く含まれる一方で、得られる情報は生活習慣の問診結果、身長や体重、血圧など計測が容易な一般的な医療データに留まる。以上のように医療画像と健康診断データでは健康に関して異なる範囲をカバーしており、それぞれのデータを活用した健康状態の診断や予測といった研究開発が今後も重要であると考えられる。

2.2 行動変容に関する研究

昨今のセンサの小型化により、様々なデバイスに搭載されたセンサを用いて実世界の多様なデータを収集することが可能になりつつある。センシングしたデータからいかに現象を読み解くかという観点での分析が現在の AI 技術開発の主流であるが、一方でセンシングにより得られた知見を実世界にフィードバックすることも重要である。ヘルスケアの分野においては、これまで様々なセンサを用いてユーザーの身体状態を計測することに主眼が置かれてきたが、今後は収集したデータに基づいてユーザーの健康状態の改善を支援することが重要となってくる [9]。

Choi らは、行動変容に至るまでの人の受容モデルを提案している。健康支援のための介入策に対して、人は perception, availability, adherence という三種類の判断を経て介入作の実行に至るという実験的研究がなされている [2]。この実験結果では、最終的に介入の成功に至る事例は全体の 3 割未満と報告されている。さらに、タスクを実行可能な状態である (available) にも関わらず介入策に従わなかったユーザが 6 割を超えていることがわかっており、ユーザーの行動変容を促すことは容易ではないことが示されている。

2.3 本研究の位置づけ

以上の研究により、機械学習アルゴリズムを用いて膨大なデータを活用することで健康状態を推定・予測するモデルの構築が可能になりつつある一方で、実際にユーザに対して習慣的な運動の実施など行動変容を促すことは容易ではないことがわかってきている。行動変容を促すことが困難である理由の一つとして、タスクに従うことの意義や必要性が十分に理解されず、タスクの実行が必要と判断されないことが考えられる。そこで、本研究では行動変容につながる技術開発を目的とし、タスクに従うことの価値の解釈性を高めるため、自身が取りうる行動の価値を推定する手法を提案する。具体的には、健康診断データにおける医師の総合所見判定、および生活習慣の問診項目のみを利用して価値推定モデルを構築する。これにより、推定モデルの出力となる健康の状態が可能な限り簡便に表現され、かつ入力となる変数がユーザ自身ですぐに改善可能な生活習慣となるため、提案手法により推定した行動価値がユーザにとって解釈が容易になるという利点がある。これまで、健康診断データを活用した研究において、医師の総合所見判定を予測する取り組みはこれまでになされておらず、本研究のように総合的な健康状態を考慮した予測モデルの構築は新しいアプローチである。

3. 問題設定

3.1 本研究で対象とする課題

本研究では健康の維持、向上のための行動変容技術を検討する。行動変容における課題として、先行研究 [2] により adherence の成功率を高めることが重要であると示唆されている。ヘルスケアの分野においては、健康増進のためにアドバイスを実施する際に、実際にユーザがアドバイスに従って行動してくれる割合を高めなければならない。

その手段としては、金銭的なインセンティブなどが最も単純で効果のある方法であると考えられるが、インセンティブ目的での行動はユーザの納得感を伴わない可能性があり、必ずしも持続可能な手段であるとは言い難い。そこで、本研究ではユーザ自身で納得感を持って自発的、かつ持続的に健康増進に取り組むことが重要であると考え、アドバイスに対する納得感を高めるアプローチを採用する。

そのための具体的な手段の一つとして、本論文ではユーザが取りうる生活習慣の改善行動に対し将来の健康状態に対する行動の価値を定量的に示すための手法を検討する。ユーザが取りうる数多くの選択肢の中から健康増進における価値の高い行動をピックアップしてアドバイスすることにより、納得感の高いフィードバックが行えるものと考えられる。本論文ではその基盤となる技術として、生活習慣改善行動における価値推定手法の提案を行う。提案システムを用いて行うアドバイスに対するユーザの納得感については、実際にシステムを運用してアドバイスを実施した結果

として健康増進に具体的に寄与したかどうかの情報を収集する必要があり、今後の研究における取り組みとする。

3.2 利用する健康診断データ

本研究では株式会社 NTT ドコモグループで勤務する従業員より同意を得て収集した 1994 年から 2016 年までの合計 109,272 件の健康診断データを用いる。健康診断のフォーマットは厚生労働省による標準的な質問票 [10] に従うものであり、健康診断データには受信日や受診場所、身体測定値など要配慮の個人情報も含まれるが、本研究では以下の情報を用いる。

説明変数となるデータ：表 1 に示す、生活習慣に関する合計 14 個の問診項目への回答であり、回答は 2 から 5 通りの選択肢のうちから選択される。なお、表の回答選択肢の数値コードを説明変数として用いるため、一般的に健康によいとされる回答により大きな値を付与するようにしており、実際の回答用紙と記載の順序が異なる項目が存在することに留意されたい。本研究では後述の価値推定アルゴリズムにおいて価値推定の対象となる説明変数をユーザが解釈・改善可能な変数とするために、表 1 に示すとおり、ユーザ自身が回答する問診回答結果のみを用いる。

目的変数および説明変数となるデータ：表 2 に示す医師の総合初見判定であり、各医療機関によって符号での表記や日本語での表記など若干の揺れがあるが、各医療機関で共通している 7 段階の分類（異常なし、日常生活問題なし、要注意、経過観察、要治療、要精密検査、治療中）を含むデータを用いる。この分類のうち、本研究では表 2 で示す 3 種類のコードを目的変数として用いる。この理由として、符号 E（要治療）や F（要精密検査）などの治療が必要な状態を分類する際には血圧などのバイタルデータの計測が重要となると考えられるが、前述の通り本研究では解釈に重きを置き、表 1 に示す生活習慣に対する問診回答結果に基づいて健康状態を予測するためである。すなわち、本研究では健康上問題のない状態、観察が必要な状態、および治療が必要な状態という 3 種類の大局的な健康状態を予測するモデルを構築する。したがって本来収集しているデータより低い解像度での健康状態の予測となるが、ユーザに行動指針を提示するという目的においては、3 種類の大局的な健康状態を考慮することにも依然として効果があると考えられる。また、健康診断データは毎年更新される性質のデータであるため、ある一年の総合所見判定を予測するために前年の総合所見判定を利用することが可能である。したがって、ある一年における総合所見判定の診断結果の情報も翌年の総合所見の予測に用いる説明変数とする。ある時点での健康状態の確定値、およびそこから一年間の生活習慣に基づいて、翌年の健康状態を予測することにより、ユーザによって異なる健康状態を考慮した将来の健康状態予測モデルを構築できる。

表 1 説明変数となるデータ

ID	問診項目	回答選択肢
1	現在、たばこを習慣的に吸っている。	2 (1: はい, 2: いいえ)
2	20 歳の時の体重から 10kg 以上増加している。	2 (1: はい, 2: いいえ)
3	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	2 (1: いいえ, 2: はい)
4	日常生活において歩行又は同等の身体活動を 1 日 1 時間以上実施	2 (1: いいえ, 2: はい)
5	ほぼ同じ年齢の同性と比較して歩く速度が速い。	2 (1: いいえ, 2: はい)
6	この 1 年間で体重の増減が ± 3 kg 以上あった。	2 (1: はい, 2: いいえ)
7	人と比較して食べる速度が速い。	3 (1: 速い, 2: ふつう, 3: 遅い)
8	就寝前の 2 時間以内に夕食をとることが週に 3 回以上ある。	2 (1: はい, 2: いいえ)
9	夕食後に間食 (3 食以外の夜食) をとることが週に 3 回以上ある。	2 (1: はい, 2: いいえ)
10	朝食を抜くことが週に 3 回以上ある。	2 (1: はい, 2: いいえ)
11	お酒 (清酒, 焼酎, ビール, 洋酒など) を飲む頻度	3 (1: 毎日, 2: 時々, 3: ほとんど飲まない)
12	飲酒日の 1 日当たりの飲酒量	4 (1: 3 合以上, 2: 2-3 合未満, 3: 1-2 合未満, 4: 1 合未満)
13	睡眠で休養が十分とれている。	2 (1: いいえ, 2: はい)
14	運動や食生活等の生活習慣を改善してみようと思いますか	5 (1: 改善するつもりはない, 2: 改善するつもりである (概ね 6 ヶ月以内), 3: 近いうちに改善するつもりであり (概ね 1 ヶ月以内), 4: 既に改善に取り組んでいる (6 ヶ月未満), 5: 既に改善に取り組んでいる (6 ヶ月以上))

表 2 目的変数となるデータ

符号	総合所見	コード	説明
A	異常なし	2	健康上問題がない状態
B	日常生活問題なし		
C	要注意	1	観察が必要な状態
D	経過観察		
E	要治療	0	治療が必要な状態
F	要精密検査		
G	治療中		

4. 提案手法

本章では、3.2 節で述べた健康診断データを用い、生活習慣改善の行動の価値を推定する手法を提案する。提案する手法により、ユーザが取りうる生活習慣改善の選択肢に対して価値を判断し、定量的な価値を含めてアドバイスを生成することが可能になる。

4.1 提案手法の概要

提案手法の概要を図 1 に示す。提案手法を構成する要素は 3 つあり、それぞれ健康状態予測モデル、価値計算フレームワーク、アドバイス生成機構である。提案手法では健康診断のデータが入力として与えられると、その問診回答を参照し、ユーザが取りうる生活習慣改善の選択肢の組み合わせを網羅的に生成する。次に、生成した選択肢のそれぞれを健康状態予測モデルに入力し、翌年の健康状態の予測結果を得る。この際に、予測結果を表 2 に示す 3 段階のコード (クラス) としてだけでなく、3 種類の各クラスと予測される確率を合わせて計算する。続けて、価値計算フレームワークは生活習慣改善の各選択肢に対して予測されたクラスとその確率について、生活習慣を改善しない場合

表 3 追加した掛け合わせ特徴量

ID	元 ID 1	元 ID 2	意味	回答選択肢
15	11	12	飲酒量の累積	1, 2, 3, 4, 6, 8, 9, 12
16	3	4	生活全体の運動習慣	1, 2, 4
17	1	11	たばことお酒の習慣	1, 2, 3, 4, 6
18	8	10	食生活の乱れ度合い	1, 2, 4

の予測クラスと確率と比較することにより、各選択肢の生活習慣改善に対する価値を定量的な値として算出する。得られた各選択肢における価値を比較することで、提案手法ではユーザに生活習慣改善のためのアドバイスを生成する。

4.2 健康状態予測モデルの構築

3.2 節で述べた健康診断データを用いて、機械学習アルゴリズムにより健康状態を予測するモデルを構築する。本研究では教師あり学習により予測モデルを構築する。具体的には、健康診断データより説明変数と目的変数のペアを多数取得し、入力として説明変数が与えられた際に、精度よく目的変数を予測できる関数を学習させる。3.2 節で述べたように、説明変数は生活習慣に関する問診項目への回答結果およびその年度の総合所見情報、目的変数は次年度の総合所見情報とする。

機械学習アルゴリズムでは、一般的に入力変数の数が多ければ多いほど複雑な表現を学習可能であるため、予測精度の高いモデルを構築することが可能である。本研究では健康診断で規定された問診回答結果を説明変数として用いる都合上、説明変数の数を増やすことは困難であるが、特徴量の中でも関連する項目において明示的に関係合わせを行うことで新しい意味合いの特徴量を作り出す工夫を行う。例えば、従来の問診項目では飲酒の頻度観点、および

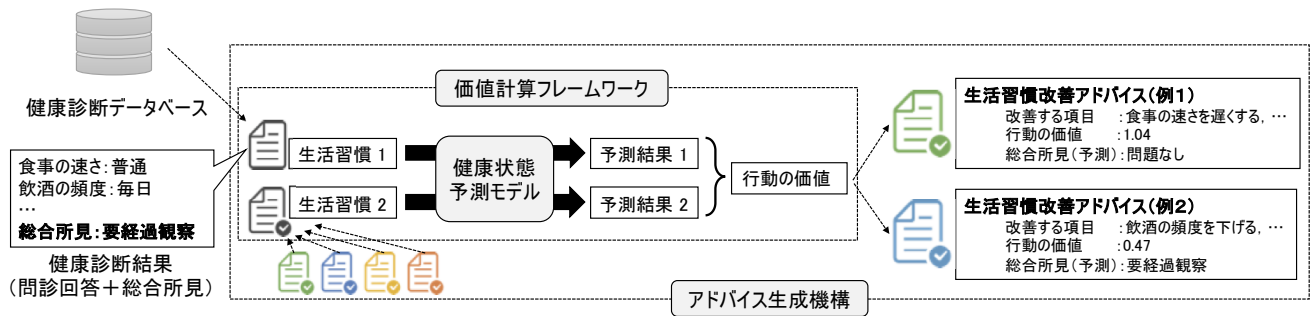


図 1 提案手法の概要

一度あたりの飲酒量観点での問診項目が存在するが、飲酒量の累積も健康上重要な観点であると考えられるため、飲酒の頻度の回答結果と、一度あたりの飲酒量の回答結果を掛け合わせることで、飲酒量の累積を考慮することを試みる。表 3 に本研究で提案する掛け合わせ特徴量を示す。本研究では合計 4 種類の掛け合わせ特徴量を提案する。上述の飲酒量の累積に関する特徴量に加え、生活全体での運動習慣（日常生活における運動 × その他の運動）、たばこ飲酒の習慣（たばこの習慣 × 飲酒習慣）、および食生活の乱れ（就寝前の夕食習慣 × 朝食を抜く習慣）を考慮する。いずれの掛け合わせ特徴量も、健康状態に関連することが期待できる。以上の 4 種類の掛け合わせ特徴量、および表 1 の 14 種類の問診結果を特徴量とし、医師の総合所見と合わせて合計 19 次元の説明変数を予測モデル構築に用いる。また、目的変数は表 2 に示す 1 次元で 3 種類のクラスを含む総合所見情報として予測モデルを構築する。

これまでに述べている通り、本研究では説明変数が予測結果に与える影響の解釈性を重視するため、予測モデルはロジスティック回帰 [11]、Extreme Gradient Boosting (XGBoost) [12]、および Light GBM [13] を用いるものとし、5 章の評価で各アルゴリズムによる予測性能を比較する。

4.3 行動における価値推定のフレームワーク

本研究で目的としている、生活習慣改善のための行動の価値の計算は、その行動の結果として将来の健康状態がどう変化したかを計ることで実現できる。しかしながら、今回利用するデータでは健康状態は表 2 に示す 3 クラスの状態であるため、3 クラスで表現される総合所見の変化を比較するだけでは細かな健康状態の変化が判断できない。そこで、本研究では先行研究 [14] で提案されているサッカープレーの系列よりゴールにつながるプレーの価値を計算する手法より着想を得て、行動による確率の変化を行動の価値とする方式を採用する。本研究における問題設定では、ある一年の健康診断の結果（問診回答および総合所見）より翌年の総合所見を予測する

現在回答した問診結果の通りの生活習慣で一年を過ごした場合の予測結果と、より健康的な生活習慣で一年を過ご

した場合の予測結果を比較し、これらの確率的な差分を生活習慣の改善によって得られた価値と定義する。定式化すると、現在回答した問診結果 x の通りの生活習慣で一年を過ごした場合の予測結果 (l, p_0, p_1, p_2) 、より健康的な生活習慣 x' で一年を過ごした場合の予測結果 (l', p'_0, p'_1, p'_2) をそれぞれ以下の通り表せる。

$$(l, p_0, p_1, p_2) = f(x),$$

$$(l', p'_0, p'_1, p'_2) = f(x'),$$

$$l = \arg \max_{n \in \{0, 1, 2\}} (p_n).$$

ここで、 f は予測モデル、 l はモデルが予測したクラス、 p_n はクラス n と判断された確率をそれぞれ表し、目的変数は 3 クラスが含まれるため $n \in \{0, 1, 2\}$ である。これらの予測結果より、生活習慣 x' によって得られる価値 v を以下の通り定める。

$$v(x') = \Delta l + \Delta p, \quad (1)$$

$$\Delta l = l' - l, \quad (2)$$

$$\Delta p = \begin{cases} p'_{l'} - p_l, & (l' = l) \\ p'_{l'}. & (\text{otherwise}) \end{cases} \quad (3)$$

上述の式 (1) は、価値 v が習慣 x' に変わったことにより得られたクラス予測結果の差と確率の差の合計から成ることを表す。この定式化により、生活習慣改善により総合所見のクラスが改善される場合にスコアを与えつつ、総合所見のクラスが改善されない場合にも予測されたクラスの確率変化に着目することで改善の度合いを定量化できる。式 (2) は習慣 x' に変えたことにより得られたクラス予測結果の差 Δl を表す。表 2 の定義により、健康状態がより高いクラスに大きい値が割り当てられているため、 Δl の値が大きいほど行動 x' が健康改善に寄与することを表す。式 (3) は習慣 x' に変えたことにより得られるクラス予測における確率の差を表す。ここではクラス予測結果 l' および l が一致する場合と異なる場合で場合分けを行い、クラス予測結果が一致する場合にはそのクラスである確率 $p'_{l'}$ および p_l の差を習慣 x' により得られた価値 Δp とする。クラス

予測結果が一致しなかった場合には、確率 p'_l および p_l の差を比較する代わりに、新しいクラス l' と判断される確率 p'_l を価値 Δp とする。これにより、現在の状態から改善される場合にはその確率がより高い生活習慣、現在の状態と同じクラスになる場合にはそのクラスである確率がより高い生活習慣により高い価値を付与することが可能となる。

4.4 選択肢のシミュレーションによる生活習慣改善アドバイスの生成

4.3 節の価値推定方式により、現在の生活習慣 x に基づいて比較したい生活習慣 x' の価値を定量化することが可能になる。提案手法ではこの価値推定方式を用いて生活習慣改善のためのアドバイスを行うため、生活習慣改善の選択肢を網羅的にシミュレーションし、各選択肢によって得られる価値を比較することで最も効果のあるアドバイスを選択する。網羅的な選択肢のシミュレーションにおける前提として、表 1 に示される 14 個の問診回答において、一般的によいとされる項目を選ぶものとする。この理由として、学習データによっては、構築した機械学習モデルにおいてよくない生活習慣を実施した場合にも健康診断結果の改善に効果があると学習されてしまう可能性があるためである。実際に、ある問診項目で一般的によくないとされる生活習慣を回答したにも関わらず他の項目でよい生活習慣を実施したことにより健康状態が改善されるサンプルなどが存在しうが、提案手法ではユーザにとって納得感のあるアドバイスを生成することに重きを置いているため、一般的によいとされる項目だけを選択肢シミュレーションにおける対象とする。

具体的には、表 1 の各 ID で表される問診項目において、現在の回答における選択肢に対して同等かより数字の大きい選択肢を選択可能である場合にはシミュレーションの候補とする。全 14 個の項目でこの判断を実施し得られた選択肢の候補に対して、改善可能な生活習慣の組み合わせ X を生成し、4.3 節の価値推定方式に入力することで、次式の通り網羅的に生成した生活習慣の改善選択肢それぞれにおける価値をシミュレーションし、価値が最大となる生活習慣の選択肢 x_{best} を探索する。

$$x_{best} = \arg \max_{x_i \in X} (v(x_i)).$$

以上のシミュレーションに基づき、提案手法では健康状態の改善において最も効果のある生活習慣 x_{best} 、およびその時の価値 $v(x_{best})$ を算出する。実用上は生活習慣 x_{best} の内容がユーザにとって実行可能なものであるかも重要となってくるため、現在の生活習慣 x と改善後の生活習慣 x_{best} を比較し実現可能性も評価することが必要である。この方法論については、今後の研究課題としたい。

5. 評価

本章では、4.2 節で提案した健康状態予測モデル、ならびに 4.4 節で提案した生活習慣改善アドバイスの生成に関する評価を行う。

5.1 評価データおよび評価方法

3.2 節で述べた健康診断データのうち、提案手法による翌年の診断結果予測の精度検証のため同一ユーザで連続する 2 年のレコードを抽出し、さらに説明変数または目的変数に欠損が存在するデータを除き合計 5255 個のデータを用いて評価を行う。機械学習モデルの構築と精度の評価においては、アンダーサンプリングにより各クラスのサンプル数を統一した上で、3 分割交差検証により学習用データと評価用データを分離することで実施する。

5.2 健康状態予測モデルによる総合所見の予測精度

はじめに、提案手法による生活習慣改善の価値計算の前提となる健康状態予測モデルの予測精度を評価する。評価では、3 クラスで表される健康状態を正しく予測することができたかどうか集計する。性能の指標は各クラスの予測における precision (適合率), recall (再現率), および全クラスを通じての accuracy (正解率) を用いる。評価のベースラインとなる比較手法として、今年度の総合所見の結果をそのまま翌年の総合所見と見立てる方式 (baseline) を導入する。これは、総合的な健康状態が一年単位で変動することは少なく、現在の健康状態が来年そのまま継続するとしてもある程度の精度で翌年の総合所見を予測可能であるという仮説に基づく。提案手法では 4.2 節で述べたとおり、ロジスティック回帰 (LR), XGBoost (XGB), LightGBM (LGBM) をモデル構築の手法として比較する。

各手法による翌年総合所見の予測精度一覧を表 4 に示す。全体の傾向として、baseline の手法ではクラス 1 (観察が必要な状態), いわば問題がない状態と治療が必要な状態の中間状態の recall が高いにも関わらず precision が低く、多くのサンプルでクラス 1 であると判断されていることが分かる。逆に、LR による予測ではクラス 1 の recall が低く、クラス 0 とクラス 2 の予測に長けていることが分かる。これらの傾向に対し、XGB および LGBM では各クラスをバランス良く予測できていることが分かる。

precision および recall の全体平均では、LR がいずれの指標においても baseline を下回っており、ロジスティック回帰のアルゴリズムがこの問題設定において効果を発揮できなかったことが示されている。一方で、baseline は提案手法のいずれと比較しても precision が高く、単純に総合所見が変わらないと予測しても多くのサンプルで予測が正しいことを表している。これに対して XGB, LGBM では

表 4 健康状態予測モデルの性能

予測手法	クラス 2 : 健康上問題がない状態		クラス 1 : 観察が必要な状態		クラス 0 : 治療が必要な状態		全体		
	precision	recall	precision	recall	precision	recall	precision (平均)	recall (平均)	accuracy
baseline	0.715	0.539	0.480	0.821	0.830	0.443	0.675	0.601	0.601
提案手法 (LR)	0.584	0.668	0.500	0.308	0.556	0.689	0.547	0.555	0.555
提案手法 (XGB)	0.654	0.588	0.542	0.637	0.683	0.632	0.627	0.619	0.619
提案手法 (LGBM)	0.638	0.611	0.560	0.578	0.726	0.733	0.641	0.641	0.641

precision と recall のバランスが取れており、予測自体の正しさを維持しながら baseline より多くのサンプルで予測できていることを示している。

また、accuracy の指標では、3 クラス全体を通じて予測がどの程度的中していたかを表しており、Light GBM による予測モデルが最も高い精度で健康状態を予測できていることが分かる。一方で、baseline、すなわち単純に総合所見が変わらないと予測する場合においても accuracy は 6 割を超えており、データの特性として過半数のサンプルで総合所見が変わらないことを示唆している。これに対し、機械学習による予測モデルを構築することで baseline に対し最大 4% の予測精度持ち上げの効果を確認できた。

以上より、予測精度の偏り、および全体的な予測精度の両観点で適切な機械学習アルゴリズムを用いて健康状態予測モデルを構築することの効果を確認できた。一方で現状の予測精度が高々 64% であり、ユーザの問診回答結果に基づいて総合所見を予測することの難しさを表している。この原因として、説明変数の次元数が少ないことや、問診回答の粒度が荒く生活習慣の細かな違いを説明変数で表現しきれないことが挙げられる。説明変数に血圧などの測定値を含めることでより精度の高い予測が可能であることが考えられるが、ユーザの解釈性とのトレードオフの関係にあり本研究では解釈性を高めることに重点を置くことから問診回答のみを用いたアプローチを実施した。今後、健康診断データ以外のデータを利用して問診回答の粒度を細くすることや、問診項目以外にユーザの生活習慣に関する情報を取得することで説明変数の量と質を充実させることにより予測精度の向上を検討したい。

5.3 健康状態予測における重要な説明変数

次に、提案手法における健康状態予測モデルにおいて、予測の際にどの説明変数が寄与しているかを分析する。図 2 に Light GBM で学習したモデルにおける各説明変数の特徴量重要度 (feature importance) の値を示す。特徴量の重要度は実数で算出されるが、図 2 では相対的に重要な特徴量を明らかにするため、特徴量重要度の値を最大値を基準とした相対値として表す。

この結果より、予測において最も重視されている項目が生活習慣改善の意識の有無であり、実際の行動よりも、改

善に対する意識の有無が翌年の総合所見の結果に反映されることが示されている。次に、本研究で提案した掛け合わせ特徴量である飲酒量の累積値が重要な説明変数となっており、単に飲酒の頻度あるいは一度あたりの飲酒量の回答よりも、累積した飲酒量が健康状態の予測において重要であることが分かる。また、翌年の総合所見の予測において、その年の総合所見の実績値は全体で 7 番目に重要であることが示されており、学習したモデルでは単純にその年の総合所見を翌年の予測に用いるだけでなく、他により重要な説明変数を鑑みた上で予測を行っていることが分かった。さらに、本研究で提案した 4 種類の掛け合わせ特徴量全てにおいて、掛け合わせ元の項目単体よりも重要度が高くなっており、掛け合わせて新しく特徴量を生成することで従来の項目よりも重要な説明変数となりうることを確認した。特筆すべき点としては、朝食習慣の乱れと夜食の習慣のそれぞれは単独では 15 番目、19 番目に重要であり、相対的に見ると重要ではないにも関わらず、それらの掛け合わせ特徴量 (食生活の乱れ度合い) は全体通じて 6 番目に重要な特徴量となっていることから、単体の項目では健康への影響をうまく考慮できていなかった一方で、それらの項目を掛け合わせることで健康への影響を考慮できるようになったことが示唆されている。

5.4 アドバイス生成内容

最後に、提案手法により生成されるアドバイスの傾向について分析する。提案手法では 4.4 節で述べたとおり、現状より一般的によいとされる項目を選択してアドバイスを生成するため、生活習慣がすでに理想的な状態、すなわち表 1 の全ての回答選択肢において最も高い数字を選択している場合、アドバイスの余地がないと判断される。実際に、5255 件のサンプルのうち、21 件で問診回答がこの状態にありアドバイス生成ができないユーザが存在したが、残りの 5234 件、すなわち約 99.6% のサンプルにおいては何らかのアドバイスを生成できることを確認した。この 5234 件のサンプルに対しては、平均して 921 通り (最大値 94084 通り、最小値 1 通り) のアドバイスを生成できることを確認した。

各サンプルで生成したアドバイスのうち、最も価値が高いと推定されたアドバイスにおける改善項目の内訳を図 3

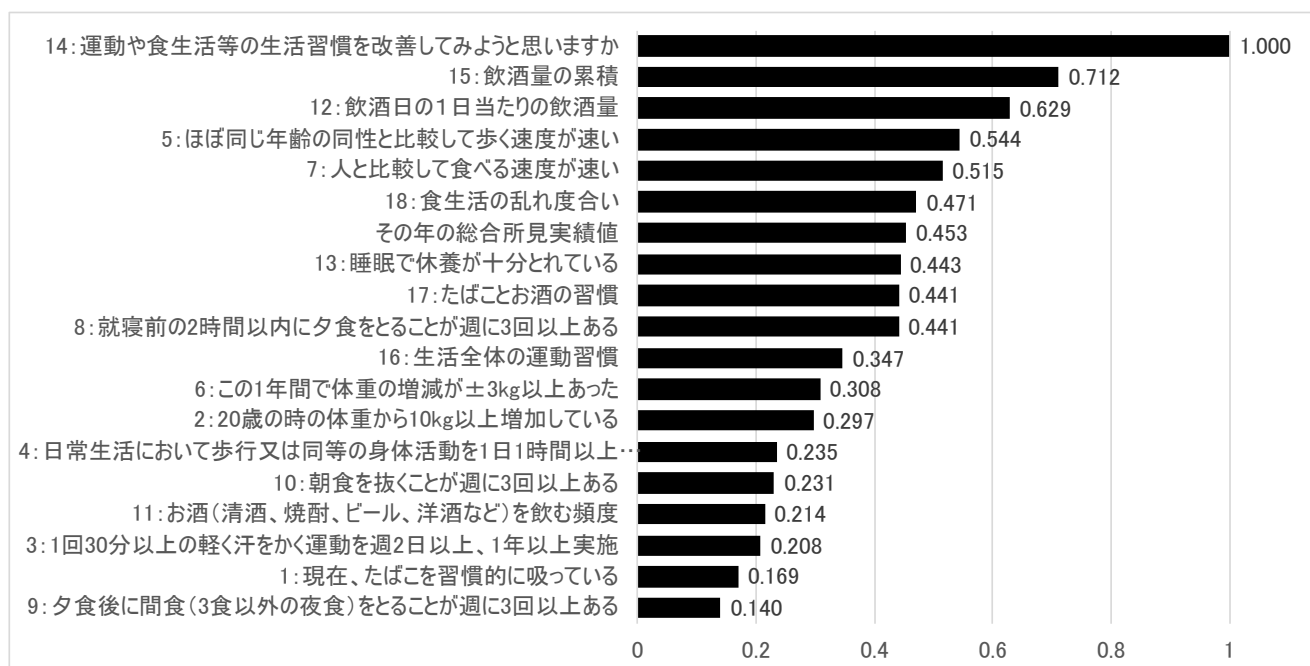


図 2 健康状態予測（提案手法：LGBM）における説明変数の相対重要度の一覧

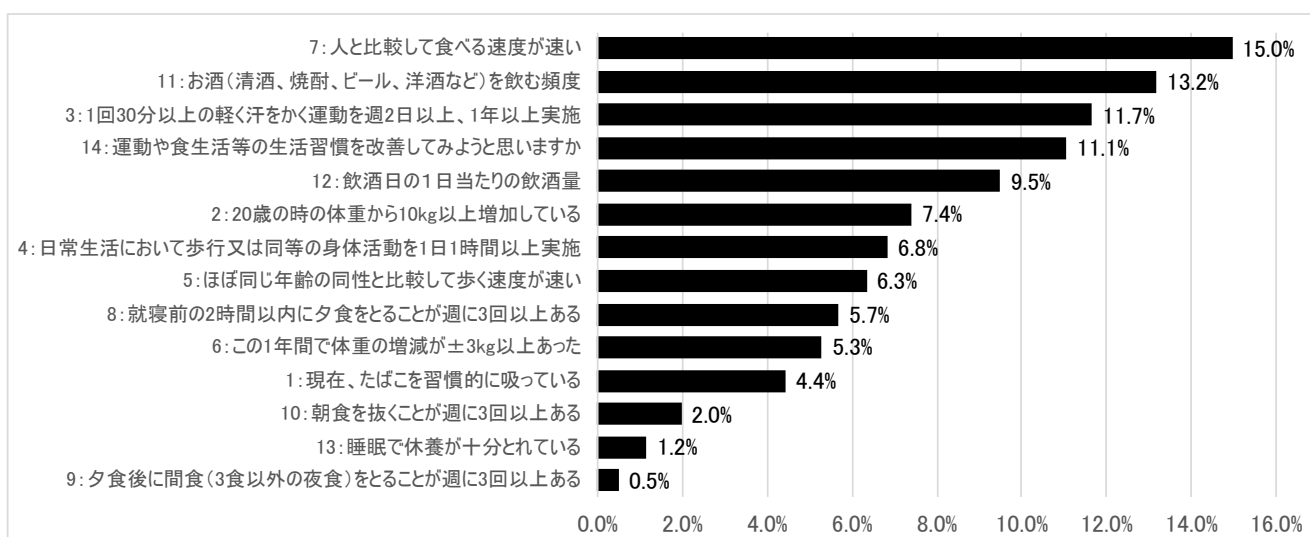


図 3 最も価値の高いアドバイスに占める各生活習慣項目の割合

に示す。なお、この結果には本研究で提案した掛け合わせ特徴量が含まれていないが、掛け合わせ特徴量は掛け合わせ元の項目の変化によって自動的に変化する値であるため、明示的にアドバイスとして生成しないことに基づく。図 3 の結果より、健康診断における総合所見の改善のためのアドバイスとして、食べる速さの改善が最も効果があると判断されたことが分かる。早食いは肥満の原因になるとも言われており、また肥満が多くの疾病の原因になりえることから、アドバイスが現実的に即していると考えられる。次に、飲酒の頻度・量、および積極的な運動の習慣が 2, 3, 5 番目に重要な項目となっており、一般的に言われているように過度な飲酒を控え、習慣的に運動することが健康

的価値が高いことが分かる。4 番目に重要な項目としては図 2 にもあったように生活習慣改善の意識を持つことが示されており、実際の行動だけでなく意識改革から実施することが重要であることが伺える。以上より、提案手法においても、一般的に効果があると言われているアドバイスが健康状態の改善において価値が高い行動であると判断されることが分かった。

既存のアドバイスの方法では、単に食事の速さを遅くすることや、飲酒を控えるようにといったアドバイスに留まるが、その効果が定量的に計れなかったことに対し、提案手法ではアドバイスに従った結果、どの程度健康状態が改善されるかを式 (1) により定量化できる。この定量化に

より、複数の実施可能なアドバイスのうち、生活習慣の改善プランだけでなくそれぞれのプランにおける価値を合わせて提示することで、従来のアドバイスよりも納得感のあるフィードバックが期待できる。

6. まとめと今後の課題

本研究では健康の維持、向上のための行動変容のための技術として、生活習慣改善行動における価値推定手法の提案を行った。提案手法では健康診断における生活習慣に関する問診への回答結果、および医師による受診者の総合所見のデータを用い、健康状態予測モデル、価値計算フレームワークおよびアドバイス生成機構の組み合わせによりユーザが取りうる生活習慣改善行動の価値を定量化する。本手法により、食生活の改善・飲酒習慣の改善・運動習慣の改善など、健康維持・推進につながる生活習慣の改善行動のそれぞれの行動に伴う価値を定量化することができる。実際に会社従業員のべ 5000 人超の健康診断データに基づき、提案手法による予測モデルを構築した結果、各データにおける翌年の医師による総合所見の予測精度が約 64.1%であることを確認した。さらに、提案手法により実際に 99.6%の従業員に対し生活習慣改善のアドバイスを提示できることを確認し、生活習慣の改善において重要な問診項目を複数確認した。

今後の課題として、提案手法で生成する生活習慣改善のためのアドバイスの実現可能性の評価軸の導入、追加のデータを用いた説明変数の質と量の両観点での拡充、および提案システムを用いて行うアドバイスに対する納得感の評価を検討している。

参考文献

- [1] Maxwell, A., Li, R., Yang, B., Weng, H., Ou, A., Hong, H., Zhou, Z., Gong, P. and Zhang, C.: Deep learning architectures for multi-label classification of intelligent health risk prediction, *BMC bioinformatics*, Vol. 18, No. 14, p. 523 (2017).
- [2] Choi, W., Park, S., Kim, D., Lim, Y.-k. and Lee, U.: Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 3, No. 2 (online), DOI: 10.1145/3328910 (2019).
- [3] Ushida, Y., Kato, R., Niwa, K., Tanimura, D., Izawa, H., Yasui, K., Takase, T., Yoshida, Y., Kawase, M., Yoshida, T. et al.: Combinational risk factors of metabolic syndrome identified by fuzzy neural network analysis of health-check data, *BMC medical informatics and decision making*, Vol. 12, No. 1, p. 80 (2012).
- [4] LaFreniere, D., Zulkernine, F., Barber, D. and Martin, K.: Using machine learning to predict hypertension from a clinical dataset, *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, pp. 1–7 (2016).
- [5] Ma, S. and Chen, X.: A data mining approach to predict risk of cardiovascular, *AIP Conference Proceedings*, Vol. 2058, No. 1, AIP Publishing LLC, p. 020014 (2019).
- [6] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y. and Tang, H.: Predicting diabetes mellitus with machine learning techniques, *Frontiers in genetics*, Vol. 9, p. 515 (2018).
- [7] Chen, P. and Pan, C.: Diabetes classification model based on boosting algorithms, *BMC bioinformatics*, Vol. 19, No. 1, p. 109 (2018).
- [8] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. and Sánchez, C. I.: A survey on deep learning in medical image analysis, *Medical image analysis*, Vol. 42, pp. 60–88 (2017).
- [9] Sheeran, P., Klein, W. M. and Rothman, A. J.: Health behavior change: Moving from observation to intervention, *Annual review of psychology*, Vol. 68, pp. 573–600 (2017).
- [10] 厚生労働省: 標準的な健診・保健指導プログラム (改訂版), https://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryuu/kenkou/seikatsu/dl/hoken-program2.pdf.
- [11] Hosmer Jr, D. W., Lemeshow, S. and Sturdivant, R. X.: *Applied logistic regression*, Vol. 398, John Wiley & Sons (2013).
- [12] Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794 (2016).
- [13] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y.: Lightgbm: A highly efficient gradient boosting decision tree, *Advances in neural information processing systems*, pp. 3146–3154 (2017).
- [14] Decroos, T., Bransen, L., Van Haaren, J. and Davis, J.: Actions speak louder than goals: Valuing player actions in soccer, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1851–1861 (2019).