

生活習慣病改善のための問診票と健康診断結果を組み合わせた ベイジアンネットワークモデルの提案

Proposal of a Bayesian Network Model Combining a Medical Questionnaire and Medical Checkup Results for Improving Lifestyle Diseases

井口 拓己¹ 吉野 孝¹ 高木 伴幸² 小池 廣昭³
Takumi Iguchi Takashi Yoshino Tomoyuki Takagi Hiroaki Koike

1. はじめに

生活習慣病は、食事や運動、睡眠、飲酒、喫煙、ストレスなど、日々の生活習慣が深く関与し発症する病気の総称である¹。生活習慣病は特に、「がん・脳卒中・心筋梗塞」といった3大疾病や、3大疾病に「高血圧症・糖尿病・腎疾患・肝疾患」を加えた7大疾病と、「肥満・脂質異常症」などが代表的な病気として挙げられ、日本人の死因の上位を占める病気も含まれる。厚生労働省が公開する令和3年版厚生労働白書によると、生活習慣病に含まれる「悪性新生物・糖尿病・高血圧性疾患・心疾患・脳血管疾患」の総患者数は1,785万人で、死亡数は約71万人であった²。

生活習慣病は名前の通り、各個人の生活習慣が大きく関係する病気であるため、生活習慣を改善することにより、発病を防ぐことや発症後の治療が可能である。そのため、日々の生活習慣に気を配ることによる予防や、発病後に自らの生活習慣を見直し、生活習慣を改めることが大切である。しかし、生活習慣が原因で、生活習慣病になることを理解していても、面倒であることを理由に生活習慣改善に取り組むことができない人がある³。

そこで本研究では、生活習慣を問診票として本人に記載させ、健康診断の結果と組み合わせ、ベイジアンネットワークを用いることで、生活習慣を改善するきっかけ作りを目的としたモデルを作成する。

2. 関連研究

生活習慣病に着目し、ベイジアンネットワークモデルによる分析を行った研究がある [1][2]。清らは、メタボリックシンドロームを対象に、特定健康診査から、ベイジアンネットワークを用いてモデルを作成した [1]。過去5年間の経過推移により、メタボリックシンドロームを改善できたかどうかに着目し、メタボリックシンドロームを脱出するための生活習慣改善を提示する目的で、モデルを作成した。モデルの作成には、飲酒や喫煙、ストレス、運動など、生活習慣に関する質問が記載された問診票と、健康診断に

よる検査結果を使用した。本研究とは、生活習慣に関する質問項目と健康診断による検査結果を、ベイジアンネットワークを用いて分析している点で類似しているが、本研究では、メタボリックシンドロームだけに着目するのではなく、複数の病気を含んだ生活習慣病についてベイジアンネットワークのモデル作成を行う。鈴木らは、2型糖尿病患者の「生活習慣の改善意思」に影響を与える因子および地域性の特徴を可視化し、生活習慣の改善によって期待される行動変容の効果について検討した [2]。糖尿病患者の生活習慣の改善への効果的な支援を行うこと、またその支援の効果指標の提案を目的とし、特定健診データから、地域別のベイジアンネットワークモデルを作成した。モデルの構築には、医学的知見を用いず、データから学習したデータ学習型モデルによる分析を行った。本研究とは、生活習慣の改善に着目している点で類似しているが、データ学習型モデルを利用した分析ではなく、医学的知見を用いて作成されたベイジアンネットワークモデルを提案する。

生活習慣病以外の病気を対象としたベイジアンネットワークモデルによる分析に、Fahmiらの関節リウマチ (RA) の診断精度向上を目的とした研究がある [3]。リウマチ専門医の知識と、血液検査などのデータを組み合わせ、ベイジアンネットワークを用いることで、知識ベース型のモデルを作成した。このモデルを用いることで、患者がRAである確率を予測し、RAとその他の炎症性関節炎 (IA) を分類した。結果として、AUCは0.86であり、診断精度は、RAが82%、IAが79%であった。また、知識ベース型モデルとは別に、データのみを用いて学習した、データ学習型モデルを作成し、両者を比較したところ、知識ベース型の方が予測・分類のどちらにおいても良い性能であったことを示した。本研究では、関節リウマチを診断の対象とするのではなく、生活習慣病を対象にベイジアンネットワークを用いてモデルを作成する。

3. 分析概要

3.1 分析対象データ

本研究では、以下の2つのデータを組み合わせて、各個人のデータとして扱う。データは2つの施設⁴で収集され、合計273人であった。

問診票: 食事や運動、睡眠の質、飲酒頻度や量、喫煙の有無、ストレス度合いなどが問診票形式で回答された個人の生活習慣に関するデータである。各項目に対

¹ 和歌山大学 システム工学部, Faculty of Systems Engineering, Wakayama University

² 和歌山市医師会成人病センター, Wakayama City Medical Association Center

³ 小池クリニック, Koike Clinic Co.,Ltd

⁴ 「生活習慣病とは?」e-ヘルスネット, 入手先:
<https://www.e-healthnet.mhlw.go.jp/information/metabolic/m-05-001.html>

² 「令和3年版厚生労働白書 資料編 2 保険医療」厚生労働省, 入手先: <https://www.mhlw.go.jp/wp/hakusyo/kousei/20-2/dl/02.pdf>

³ 「3 調査結果の詳細 (4. 生活習慣病の予防や改善に関する食意識や実践について)」農林水産省, 入手先: <https://www.maff.go.jp/j/syokuiku/ishiki/h30/3-4.html>

⁴ データの匿名性を考慮し、データが収集された施設名は非公開とする。

表 1: 各判定結果の分布

	異常あり	異常なし	計
血圧	54 人 (20 %)	219 人	273 人
脂質	122 人 (45 %)	151 人	273 人
糖尿	27 人 (10 %)	246 人	273 人

応じた選択肢があらかじめ用意されており、患者はその選択肢の中から回答する。

健康診断結果：身長や体重などの基本項目に加え、血液検査によって判断される「高血圧症」や「脂質異常症」などの生活習慣病の判定が含まれたデータである。各病気の人数と、「異常あり」の割合を表 1 に記載する。血圧、脂質、糖尿に異常がある人数は、それぞれ 54 人、122 人、27 人であり、本研究で使用したデータにおいては、生活習慣病の中でも病気によって分布が異なる。

本研究で用いる問診票と健康診断結果は、匿名加工されたものを使用した。また、本研究は和歌山大学倫理審査委員会による承認を得て行った（承認番号：令 04-00-02J）。

3.1.1 データの前処理

問診票と健康診断結果は、個人 ID で紐づけを行い、同一人物のデータとして扱う。問診票には、生活習慣に関する項目が複数個用意されており、本研究ではその中からモデルの構築に 13 項目を利用した。各項目に対して、あらかじめ用意されている回答選択肢があるが、モデルを構築する際に、さらにその選択肢の分類を行った。

ベイジアンネットワークのモデル作成に用いた項目と、分類された各項目の状態を、表 2 に記載する。健康診断結果には、3.1 節で述べたように、様々な健康診断結果が含まれているが、本研究では、「BMI」「判定 脂質」「判定 血圧」「判定 糖尿」の 4 項目を利用する。「判定 ○○」のデータには、異常あり・異常なし・ほぼ正常・軽度異常・要治療・要精密検査など、様々な結果が記載されているが、本研究では、「異常なし」と「ほぼ正常」以外は、すべて「異常あり」とした。さらに、BMI は数値データであるため、定められた閾値⁵によって離散化し、「肥満」「普通体重」「低体重」に分類を行い、肥満度とした。

3.2 ベイジアンネットワーク

ベイジアンネットワークとは、複数の確率変数の間の定性的な依存関係をグラフ構造によって表し、個々の変数の間の定量的な関係を条件付確率で表した確率モデルである [4]。グラフを構成する個々の事象はノードと呼ばれ、各ノード間の確率的な依存関係を矢印で表現する。ベイジアンネットワークのグラフは有向非巡回グラフ (Directed Acyclic Graph: DAG) と呼ばれ、あるノードから他のノードを通過して元に戻ることはない。また、各ノードの条件付確率分布は、各ノードがとる具体的な値ごとに割り当てられた条件付確率表 (conditional probability table: CPT) で表される。

表 2: モデル構築に用いた項目

項目	項目の状態
睡眠の質	(熟眠) or (熟眠ではない)
喫煙	(喫煙者) or (禁煙者) or (吸ったことがない)
ストレス度	(重度) or (普通) or (軽度)
朝食	(毎日食べる) or (時々食べない) or (全く食べない)
間食頻度	(ほとんど毎日) or (たまに食べる) or (食べない)
飲酒頻度	(ほぼ毎日) or (たまに飲む) or (飲まない)
ウォーキング	(する) or (ほぼしない)
ウォーキング以外の運動	(する) or (しない)
肥満度	(肥満) or (普通体重) or (低体重)
体重変化	(増えた) or (減った) or (変わらない)
判定 脂質	(異常あり) or (異常なし)
判定 血圧	(異常あり) or (異常なし)
判定 糖尿	(異常あり) or (異常なし)

ベイジアンネットワークの特徴は、分析や意思決定に用いられる人間の知見とデータを組み合わせることにより、要因を分析できることであり、気象予報や医療的意思決定支援、マーケティングなど様々な分野で利用・研究が行われている⁶。

また、ベイジアンネットワークを用いる利点としては、確率推論を行うことで、任意の変数に関する確率分布を求め、様々な条件における定量的な評価ができる点である。特に、他のデータ分析手法と比較すると、モデルを構築した後に、そのモデル上で確率推論を実行できる点がベイジアンネットワーク独自の利点である [5]。確率推論では、ある観測された変数を事実 (エビデンス) として、モデルに入力することで、その他の変数の確率分布がどう変化するか (事後確率) を求めることができ、確率による予測が可能である。

3.2.1 ベイジアンネットワークのモデル構築

本研究では、株式会社 NTT データ数理システムが提供する「BayoLinkS (ver.9.0)」⁷を用いて、ベイジアンネットワークのモデルを構築する。モデルは、医学的知見に基づき、各ノード間にどのような因果関係があるかをリンクでつなげて作成した。また、各ノードが持つ CPT は 3.1 節で述べた 273 人のデータから算出した。

3.2.2 モデルの評価方法

作成したベイジアンネットワークモデルは、BayoLinkS のモデル検証機能によって評価する。BayoLinkS のモデル検証機能では、適合率、再現率、F 値を算出でき、これらによってモデルを評価する。それぞれの値は、混同行列における真陽性 (TP)、偽陽性 (FP)、偽陰性 (FN) を用い

⁵ 「BMI」 e-ヘルスネット、入手先：
<https://www.e-healthnet.mhlw.go.jp/information/dictionary/metabolic/yml-002.html>

⁶ 「依存関係・因果関係の視覚化に使われるベイジアンネットワークとは？」 NTT データ数理システム、入手先：
<https://www.msiism.jp/article/what-is-bayesian-network.html>

⁷ <https://www.msi.co.jp/bayolink/>

表 3: モデル検証結果

判定結果	適合率	再現率	F 値	
血圧	異常あり	0.79	0.85	0.82
	異常なし	0.96	0.95	0.95
脂質	異常あり	0.86	0.88	0.87
	異常なし	0.91	0.90	0.90
糖尿	異常あり	0.70	0.93	0.80
	異常なし	0.99	0.95	0.97

て、以下の数式によって算出される。

$$\text{適合率} = \frac{TP}{TP + FP}$$

$$\text{再現率} = \frac{TP}{TP + FN}$$

$$\text{F 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

4. ベイジアンネットワークモデル

作成したモデルを図 1 に記載する。各ノードは楕円で表され、矢印が各ノード間の因果関係を表す。緑色の楕円は、表 2 の「判定 脂質」「判定 血圧」「判定 糖尿」であり、その他の項目が黄色の楕円で表される。

4.1 モデルの評価

本研究では、モデル構築に用いたデータからランダムに 50 % (137 人分) 選び、テストデータとして使用し、モデル検証を行った⁸。

モデル検証の結果を表 3 に記載する。再現率は、血圧・脂質・糖尿の「異常あり」「異常なし」の両方において、85 % を超えていた。適合率は、血圧と糖尿の「異常あり」が 80 % を下回ったが、表 1 の分布にも示したように、血圧と糖尿の「異常あり」である割合が全体の 20 % と 10 % であり、テストデータを作った際に含まれる「異常あり」の数が少なかったからであると考えられる⁹。

4.2 確率推論

4.2.1 モデルの妥当性確認

確率推論では、実際に存在し得るエビデンスをいくつか想定し、そのエビデンスをモデルに与えることで、事後確率が算出される。この事後確率によってモデルの妥当性を確認する。モデルに与えた各エビデンスを表 4 に記載する。また、モデルにエビデンスを与えた際の、確率推論の結果を表 5 に記載する。

表 4 のエビデンス A は、明らかに健康状態にある人を想定し、エビデンス B~D は、エビデンス A に比べて悪い生活習慣を想定した。また、エビデンス E~G は運動習慣を考慮したエビデンスを想定した。

⁸本研究では利用可能なデータ数が少ないため、モデル構築用データとは別のテストデータを作成した際に含まれる各病気の「異常あり」の数が少なく、モデルの検証が行えなかった。そのため、モデル構築用データからテスト用データをランダムに選び、モデル検証を行った。また [1] においても、モデル構築用データとテストデータは同じデータを使用していたため、参考にした。

⁹テストデータとした 137 人中「異常あり」の人数は、血圧が 27 人、脂質が 58 人、糖尿が 15 人であった。

表 4: エビデンス

	各ノードの状態
A	(肥満度：普通体重) (喫煙：吸ったことがない) (飲酒頻度：飲まない) (朝食：毎日食べる) (体重変化：変わらない)
B	(肥満度：普通体重) (喫煙：吸ったことがない) (飲酒頻度：ほぼ毎日)
C	(肥満度：普通体重) (喫煙：喫煙者) (飲酒頻度：飲まない)
D	(肥満度：肥満) (喫煙：喫煙者) (飲酒頻度：ほぼ毎日) (間食：ほとんど毎日) (体重変化：増えた)
E	(肥満度：肥満) (間食頻度：ほとんど毎日) (ウォーキング：ほぼしない) (ウォーキング以外の運動：しない) (体重変化：増えた)
F	(肥満度：肥満) (間食頻度：ほとんど毎日) (ウォーキング：ほぼしない) (ウォーキング以外の運動：する)
G	(肥満度：普通体重) (間食頻度：たまに食べる) (ウォーキング：する) (ウォーキング以外の運動：しない)

表 5: 推論結果

エビデンス	血圧 異常あり	脂質 異常あり	糖尿 異常あり
A	0.17	0.29	0.08
B	0.33	0.61	0.38
C	0.28	0.31	0.29
D	0.49	0.56	0.52
E	0.51	0.54	0.40
F	0.46	0.45	0.40
G	0.33	0.28	0.43

表 5 から、喫煙や飲酒の習慣があるエビデンス B, C, D や、間食の頻度が多いエビデンス D, E, F は、明らかに健康状態であるエビデンス A に比べて、各病気の「異常あり」である事後確率が高いことが示された。また、運動習慣のあるエビデンス F と G に比べ、運動習慣のないエビデンス E は各病気の「異常あり」である事後確率が高くなった。

これらの結果から、作成したモデルの妥当性が確認できたと考えられる¹⁰。

4.2.2 確率推論による生活習慣の改善

本研究で作成したモデルを用いてどのように生活習慣の改善が可能であるかを、表 4 と表 5 から考察する。

エビデンス D のような人が、エビデンス C のように喫煙習慣があっても、推論結果から、肥満度を下げ、飲酒の習慣を減らすことができれば、生活習慣病である確率を現在より下げることが可能であると分かる。

また、運動習慣を考慮したエビデンス E~G では、推論

¹⁰生活習慣を改善した際の事後確率の変動が、病気ごとにどれほど差があるのかについては、医学的な知見に基づいて確認する必要があると考えている。

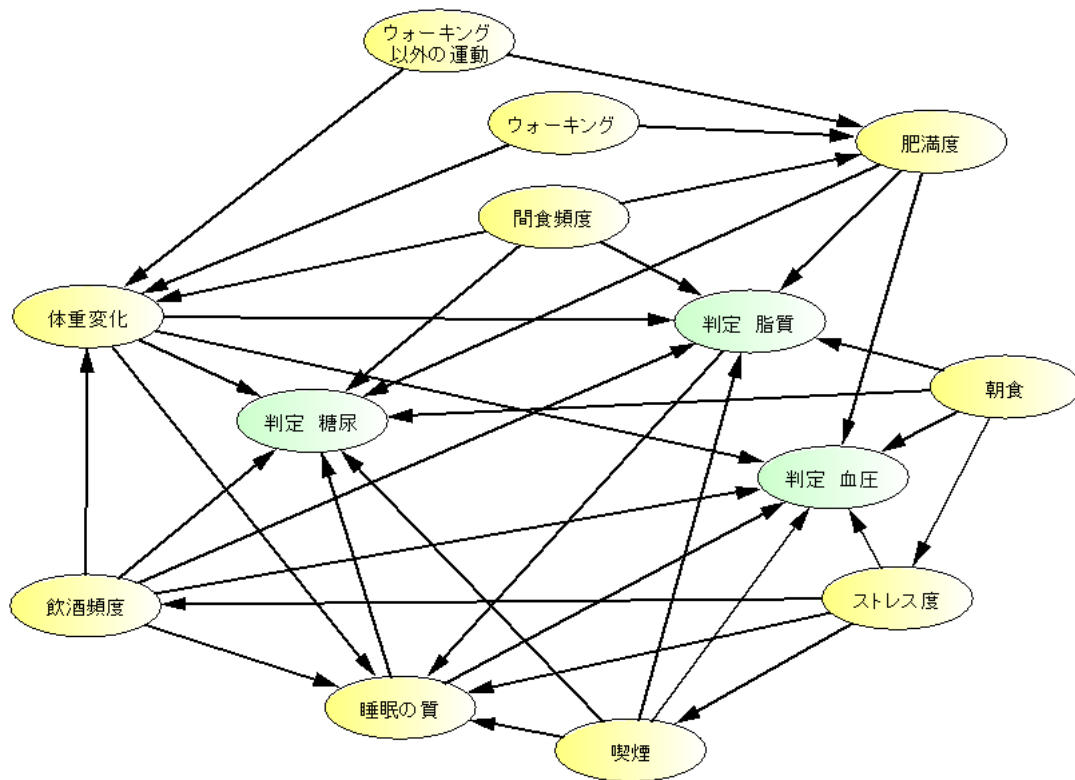


図 1: ベイジアンネットワークモデル

結果から、エビデンス E のような人が運動を生活に取り入れたり、間食の頻度を下げることで、生活習慣病の確率を現在より下げることが可能であると分かる。特にエビデンス F は、ウォーキングはしたくないが、その他の運動なら取り組むことができる人を想定しており、エビデンス G はウォーキングを生活習慣に取り入れ、普通体重になった場合を想定している。

このように、ベイジアンネットワークモデルを用いることで、取り組みやすさなどを考慮し、その人に合わせた生活習慣の改善を提案でき、生活習慣の改善が面倒だと感じる人に対しても、有効な改善の提案ができると考えられる。

5. おわりに

本研究では、生活習慣の改善を目的として、問診票と健康診断結果から、ベイジアンネットワークモデルの作成を行った。作成したモデルを用いた確率推論の結果、人それぞれに合わせた生活習慣の改善を提案できることが分かった。

本研究で作成したモデルでは、遺伝的な要因や、医療機関において病気の診断をされた人が、すでに薬や生活習慣の改善を行っている可能性を考慮することができなかった。また、本研究で使用したデータ数が少なく、各病気の分布に偏りがあったため、検証によるモデルの評価が正確に行えなかった。今後は、利用できるデータ数を増やしながら、血液検査の項目や内服薬の効果なども考慮したモデルの作成を行う。

参考文献

- [1] 清奈帆美：メタボリックシンドロームに対する効果的な保健指導のための健康診断分析モデル，慶応義塾大学大学院システムデザイン・マネジメント研究科 修士学位論文 2014 年度システムエンジニアリング学 第 168 号 (2014).
- [2] 鈴木哲平，田村菜穂美，榎本尚司，永井亘，小笠原克彦：生活習慣の改善意思に影響を与える要因および地域性の可視化-ベイジアンネットワークを用いた自治体の特定健診データ分析-，医療情報学，Vol.39, No.2, pp.85-98(2019).
- [3] Ali Fahmi, Amy MacBrayne, Evangelia Kyrimi, Scott Mclachlan, Frances Humby, William Marsh, Costantino Pitzalis: Causal Bayesian Networks for Medical Diagnosis: A Case Study in Rheumatoid Arthritis, EasyChair Preprint, No.4625 (2020).
- [4] 本村陽一，岩崎弘利：ベイジアンネットワーク技術，東京電機大学出版局 (2006).
- [5] 繁樹算男，植野真臣，本村陽一：ベイジアンネットワーク概説，培風館 (2006).