ベイジアンネットワークによる地域健康予測

佐々木　健佑，久野　譜也，岡田　幸彦

本稿では，地域健康政策のための疾病の新規発症予測を行う．予測対象は，重症化による財政への影響と，早期の特定・予防の有用性の観点から，3 年以内の 2 型糖尿病新規発症とする．まず，地方自治体がもつ医療レセプト・特定検診のデータから，予測モデルを構築する．この際に本稿では，自治体職員による説明容易性を重視し，条件付き独立性検定としてカイ二乗検定を用いる HITON-PC アルゴリズムを用いたベイジアンネットワークを採用する．次に，提案手法の有用性を検証するため，Nanri et al. [1] で提案されている糖尿病リスクスコアによる予測との精度比較を行い，地域健康政策への応用可能性について述べる．

キーワード：地域健康政策，国保データベース，健幸クラウド，ベイジアンネットワーク，人工知能

1. はじめに

地域健康政策では，地域住民の健康推進を目的とし，自治体の担当部局がさまざまな健康関連事務事業を実施・運営している．地域健康政策の運営主体である自治体医療保険者は，医療レセプトなどのデータを分析して，地域健康政策・保健事業を進めていく必要がある．費用対効果の高い保健事業を行うためには，施策の対象とすべき被保険者を適切に選別し，その集団に合わせた施策を集中的に実施していくことが重要である [2]．地域健康政策において自治体が取り組むべき健康課題は多岐にわたる．中でも糖尿病は，わが国において生活習慣と社会環境の変化によって急激に患者数が増加している疾病であり，発症後，網膜症・腎症・神経障害などの合併症を引き起こしうる疾病であることから [3]，地域健康政策において非常に重要な疾病であると認識されている．2015 年に日本医師会，日本糖尿病対策推進会議，厚生労働省により策定された「糖尿病性腎症重症化予防プログラム」では，プログラムにおける自治体の役割として，「健診データやレセプトデータなどを用いて，被保険者の疾病構造や健康問題などを分析し，地域の関係団体とともに問題意識の共有を行う」こと，「対策の立案にあたり，地域の医療機関における連携体制のあり方，ハイリスク者を抽出するための健診項目や健診実施方法，食生活の改善や運動対策などのポピュレーションアプローチなど，様々な観点から総合的に検討した上で，保健指導や受診勧奨の内容について検討する」ことが明記されている [4]．このことから自治体職員は，1. 被保険者の疾病構造や健康問題を分析すること，2. ハイリスク者の抽出方法などを検討することの二つの役割が期待されていることがわかる．2 に関連して，医療レセプトデータ・健診データを用いた糖尿病の発症予測モデルの構築とハイリスク者の抽出に関する研究は数多く行われてきた．代表的なものとして，Doi et al. [5] や，Nanri et al. [1] によって開発された糖尿病リスクスコアがある．糖尿病リスクスコアでは，性別，年齢，BMI，腹部肥満，喫煙の有無，高血圧の有無，空腹時血糖，HbA1c[1] を糖尿病発症の危険因子とし，それぞれの危険因子のカテゴリに割り振られた点数を合計することによって，以降 3 年間に糖尿病を新たに発症する確率を算出することができる．このようなツールは，主に健康診断後の保健指導の場面において，保健師が被保険者にリスクの大きさを説明するために利用されることが想定されており [2]，実際に保健指導の現場で活用されている [6]．このように，糖尿病リスクスコアは，被保険者のリスクを手軽にかつ正確に算出することができるため，保健指導の場面において非常に有用なツールである．一方で地域健康政策の立案という観点において自治体職員は，被保険者のリスクを正確に予測しハイリスク者を特定することだけでなく，その地域特有の疾病構造や健康問題の把握までを行うことが期待されている．つまり，高い精度で将来のリスクを予測し，同時にそのリスクを引き起こす要因を示すことができる方法論が必要である．これらの要件を満たす疾病発症予測モデ

ささき　けんすけ
筑波大学大学院システム情報工学研究科
〒 305–8573 茨城県つくば市天王台 1–1–1
s1820549@s.tsukuba.ac.jp
くの　しんや
筑波大学体育系，筑波大学人工知能科学センター
おかだ　ゆきひこ
筑波大学システム情報系，筑波大学人工知能科学センター

[1] ヘモグロビン A1c，血糖コントロールの目安となる指標のこと．

ルとして，ベイジアンネットワークによる予測モデルが考えられる．ベイジアンネットワークによる予測モデルでは，被保険者の疾病新規発症の予測に加え，医療レセプトや健診データから作成された原因系確率変数と結果系確率変数の間の確率的な依存関係を同時に示すことが可能である．これにより，疾病の発症を引き起こす背景要因は何か，どの要因に介入するべきか，などといった政策立案上重要となる事項について情報を提供し，EBPM（Evidence-based Policy Making，エビデンスに基づく政策立案）を支援することが可能となる．

本稿では，ベイジアンネットワークを用いた糖尿病新規発症予測モデルを紹介し，その応用可能性について議論をする．本モデルでは，まず医療レセプトデータ，健診における問診票データ，血糖検査値の各項目を確率変数によって表現し，ベイジアンネットワークの投入変数とする．次に，制約ベース構造推定アルゴリズムであり，条件付き独立性検定にカイ二乗検定を用いる HITON-PC を用いて確率変数間の因果構造を推定する．そして，構築した因果モデルを用いて，個人別の糖尿病新規発症を予測する．なお，この予測モデルの構築と，予測精度の実証的評価の研究は，国立研究開発法人日本医療研究開発機構 (AMED) による「AI を活用した保健指導システム研究推進事業」の一部として行われたものである．

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，確率的なグラフィカルモデルの一種である．確率変数をノードとし，ノード間の確率的な依存関係を DAG (Directed Acyclic Graph) と条件付き確率表によって表現する．ベイジアンネットワークを用いた予測のプロセスは，三つの段階で構成される．まず第 1 段階として，データからノード間の確率的依存関係を学習する構造学習が行われる．次に第 2 段階において，パラメータ推定による条件付き確率の計算を行う．最後に第 3 段階において，確率推論による予測値の出力を行う．ベイジアンネットワークの確率推論では，任意のノードの観測情報をエビデンスとして与え，目的とする確率変数の事後確率を求める．ベイジアンネットワークの詳細については，文献 [7] を参照されたい．

ベイジアンネットワークはさまざまな分野で活用されている手法である．本稿の主題から注目すべきは，政策立案におけるベイジアンネットワークの応用と，保健医療分野におけるベイジアンネットワークの応用であろう．政策立案の観点から，ベイジアンネットワークを諸問題に応用している研究として，上野 [8] や鶴田と寒河江 [9] がある．上野 [8] では，どのような政策的手段が，人口減少に対してどの程度効果的であるかについて，ベイジアンネットワークを用いて検証している．鶴田と寒河江 [9] では，少子化の因果関係の分析にベイジアンネットワークを利用し，出生率に関わる社会経済的要因を明らかにすることによって，少子化対策に関する政策について議論している．また保健医療の分野で，特に医療レセプトデータや健診データを用いたベイジアンネットワークの応用に関する研究は複数行われている．宮内 [10] では，医療レセプトデータ，特定健診データを利用し，メタボリックシンドロームのマネジメントを主目的としたベイジアンネットワークの構築を行っている．ここでは，保健指導レベルのリスク評価に着目し，各ノードの値を変化させた場合の保健指導レベルのリスク変化を観測するなどの方法によるベイジアンネットワークの活用を提案している．三好ら [11] では，医療レセプトデータと特定健診のデータを用いて，生活習慣病による医療費の予測を目的としたベイジアンネットワークの構造学習方法を提案している．さらに，構築した予測モデルを実際の保健指導の現場で利用し，医療費抑制を実現したことが報告されている [12]．鳥海ら [13] では，同様のデータを用いて高血圧症，糖尿病をもつ被保険者ともたない被保険者の要因を，ベイジアンネットワークの構造学習によって特定し，地域健康政策上重要な要因を把握する方法論としての HITON-PC アルゴリズムの有用性を主張している．これらの研究は，保健医療の文脈におけるベイジアンネットワークの有用性を主張する研究である．一方で，被保険者の疾病発症に対して，疾病発症あり，なしの 2 値分類問題として扱い，ベイジアンネットワークによる予測モデルの構築と，予測性能の評価までを行っている研究は蓄積が少ない．

3. 糖尿病新規発症予測モデル

3.1 データセット

本稿では，わが国のほぼすべての自治体が登録し活用している国保データベース（以下，KDB）システムに収録されたデータを前提とし，株式会社つくばウエルネスリサーチの健康関連ビッグデータ基盤である健幸クラウドに蓄積された，40 歳から 74 歳の国民健康保険（以下，国保）被保険者の医療レセプトデータ，特定健診問診票データ，検査値データを用いる．また，本

表 1　確率変数の定義

系	変数名	定義	データの種類
–	outcome	2013-2015年に，以下のいずれかに該当 ： 1，それ以外 ： 0 1．空腹時血糖 >= 126mg/dl 2．HbA1c >= 6.5% 3．疾病分類番号0402(糖尿病)のレセプト点数が1以上	血糖検査値， 医療レセプト
基本属性	性別_男性	男性 ： 1，それ以外 ： 0	基本属性
	年齢_40_44	ベースライン年度に40-44歳に該当 ： 1，それ以外 ： 0	基本属性
	年齢_45_49	ベースライン年度に45-49歳に該当 ： 1，それ以外 ： 0	基本属性
	年齢_50_54	ベースライン年度に50-54歳に該当 ： 1，それ以外 ： 0	基本属性
	年齢_55_59	ベースライン年度に55-59歳に該当 ： 1，それ以外 ： 0	基本属性
	年齢_60_64	ベースライン年度に60-64歳に該当 ： 1，それ以外 ： 0	基本属性
	年齢_65_69	ベースライン年度に65-69歳に該当 ： 1，それ以外 ： 0	基本属性
	年齢_70_74	ベースライン年度に70-75歳に該当 ： 1，それ以外 ： 0	基本属性
運動機能	ほぼ同じ年齢の同性と比較して歩く速度が速い	該当 ： 1，それ以外 ： 0	問診票
身体活動	運動習慣あり	該当 ： 1，それ以外 ： 0	問診票
	一時間以上の身体活動の実施あり	該当 ： 1，それ以外 ： 0	問診票
栄養・食生活	食べる速度が速い	該当 ： 1，それ以外 ： 0	問診票
	週3回以上就寝前に夕食をとっている	該当 ： 1，それ以外 ： 0	問診票
	週3回以上朝食を欠食している	該当 ： 1，それ以外 ： 0	問診票
アルコール	飲酒頻度が多い	該当 ： 1，それ以外 ： 0	問診票
煙草	喫煙している	該当 ： 1，それ以外 ： 0	問診票
貧血	貧血である	該当 ： 1，それ以外 ： 0	問診票
行動変容ステージ	生活習慣_無関心層	「運動や食生活等の生活習慣を改善してみようと思う」 改善するつもりはない，改善するつもりである(概ね6か月以内) に該当 ： 1，それ以外 ： 0	問診票
	生活習慣_改善意識あり	「運動や食生活等の生活習慣を改善してみようと思う」 近いうちに(概ね1か月以内) 改善するつもりであり，少しずつ始めている に該当 ： 1，それ以外 ： 0	問診票
	生活習慣_行動変容実施者	「運動や食生活等の生活習慣を改善してみようと思う」 既に取り組んでいる(6か月未満)，既に取り組んでいる(6か月以上) に該当 ： 1，それ以外 ： 0	問診票
休養・こころの健康	十分な睡眠がとれている	該当 ： 1，それ以外 ： 0	問診票
体組成計	二十歳の時の体重から10kg以上増加している	該当 ： 1，それ以外 ： 0	問診票
	この一年間で体重の増減が3kg以上あった	該当 ： 1，それ以外 ： 0	問診票
	低栄養	低栄養 (BMI18.5未満) ： 1，それ以外 ： 0	身体計測
	標準体重	標準体重 (BMI18.5-25.0未満) ： 1，それ以外 ： 0	身体計測
	肥満_軽度	肥満_軽度 (BMI25.0以上-30.0未満) ： 1，それ以外 ： 0	身体計測
	肥満_中程度以上	肥満_中程度以上 (BMI30.0以上) ： 1，それ以外 ： 0	身体計測
服用・病歴	血圧を下げる薬を使用している	該当 ： 1，それ以外 ： 0	問診票
	コレステロールを下げる薬を使用している	該当 ： 1，それ以外 ： 0	問診票
	高血圧の罹患者	該当 ： 1，それ以外 ： 0	医療レセプト
	脂質異常症の罹患者	該当 ： 1，それ以外 ： 0	医療レセプト
地域特性	居住地域_A小学校区	該当 ： 1，それ以外 ： 0	基本情報
血糖検査値	FPG_under_100	空腹時血糖 ≦ 100mg/dl に該当 ： 1，それ以外 ： 0	血糖検査値
	FPG_100_110	空腹時血糖 > 100mg/dl かつ 空腹時血糖 < 100mg/dl に該当 ： 1，それ以外 ： 0	血糖検査値
	FPG_over_110	空腹時血糖 ≧ 110mg/dl に該当 ： 1，それ以外 ： 0	血糖検査値
	HbA1c_under_5.6	HbA1c ≦ 5.6% に該当 ： 1，それ以外 ： 0	血糖検査値
	HbA1c_5.6_6	HbA1c > 5.6% かつ HbA1c < 6.0% に該当 ： 1，それ以外 ： 0	血糖検査値
	HbA1c_over_6	HbA1c ≧ 5.6% に該当 ： 1，それ以外 ： 0	血糖検査値

稿で用いるデータは，匿名で本稿にご協力くださった自治体 A，B，C，D，E，F，G の 2011 年 4 月から 2016 年 3 月までの 5 年度分のデータである．

表 1 は，ベイジアンネットワークによる予測モデル構築に用いる原因系変数と，結果系変数の一覧を示している．原因系変数については，鳥海ら [13] を参考に，政策的な介入が可能な原因候補としての身体活動，栄養・食生活，アルコール，煙草，貧血，行動変容，休養・こころの健康，運動機能，体組成の九つの系と，ターゲットセグメント化の観点で有用である原因候補群としての基本属性，服用・病歴，地域特性（自治体小学校区レベル）の三つの系の計 12 個の系に分類する．結果系変数には，本稿の予測対象である 2 型糖尿病新規発症の事象を表す確率変数として Nanri et al. [1] の定義に従い，以下の条件を用いて定義する．ベースライン年度を t 年とし，まず t 年に以下の 1，2，3 の条件にいずれも該当しない個人を抽出する．その中から，$t+1$ 年から $t+3$ 年の間に以下のいずれかの条件を満たした個人は，2 型糖尿病を新たに発症したとして確率変数の値に 1 を，そうでない場合には 0 を設定する．

1．空腹時血糖 $\geqq 126$ mg/dl
2．HbA1c $\geqq 6.5\%$
3．疾病分類番号 0402（糖尿病）のレセプト点数が 1 点以上（入院・外来を含む）

また，先行研究において糖尿病発症予測の重要な予測変数であることが指摘されている空腹時血糖 (FPG)，HbA1c についても，血糖検査値系として確率変数に加える．分析に際して，リストワイズ法により欠損値を除外し，利用するサンプルを抽出する．なお，疾病分類番号 0402 には，2 型糖尿病以外に 1 型糖尿病などの疾病も含んでいるが，本稿で用いている国保被保険者が 40 歳から 74 歳の範囲内であること，1 型糖尿病の発症時期のピークが思春期にあること [3] を踏まえ，本稿では 2 型糖尿病を対象として扱うこととする．

3.2　ベイジアンネットワークモデリング

上記の定義に従って作成したデータから，全体の 2/3 を用いた学習データ，1/3 を用いた検証データを作成する．学習データから，条件付き独立性にカイ二乗検定を用いる HITON-PC アルゴリズムによる構造学習を行う．ベイジアンネットワークの構造学習では，モデルの仮定や明白な知識を事前情報として反映させることで，分析の精度を上げることができる [9]．そこで，原因系変数と結果系変数の関係を政策立案の場面で解釈可能な形で表現するために，以下の制約条件を

事前情報として与え，構造学習を行う．

1. 糖尿病新規発症のノードからほかの原因系変数への矢印を引かない
2. 可変変数から不変変数への矢印を引かない
3. 同じ系に属する変数同士で矢印を引かない

構造学習によって生成された無向グラフに対して，その方向づけを Meek [14] によるオリエンテーションルールによって行う．次に，ネットワークと学習データから条件付き確率を計算し，予測モデルを構築する．この際，ベイジアンネットワークの確率推論では，着目するノードの取りうる値の確率値が出力される．そのため，疾病新規発症の 2 値分類を行うためには，確率変数の値である 0 と 1 を分ける閾値を決定する必要がある．本モデルでは，学習データによって構築した予測モデルと，学習データの正解ラベルから ROC 曲線を描き，感度と特異度の和が最大となる点を cutoff ポイントとして採用し閾値とする．構築したモデルについて，5 節で詳細を述べる．

3.3 予測モデルの評価指標

構築する予測モデルの妥当性を示すために，Nanri et al. [1] に従って三つの指標でその予測性能を測る．一つ目は，実際に疾病を新たに発症する人の中で，予測モデルによって正しく発症すると予測された人の割合を意味する感度 (Sensitivity) である．感度は以下の式で計算される．

$$Sensitivity = \frac{TP}{TP + FN}$$

TP: True positive，予測モデルの予測値が陽性を示し，実測値が実際に陽性であったケースの数
FN: False negative，予測モデルの予測値が陰性を示し，実測値が実際には陽性であったケースの数

二つ目は，実際に疾病を発症しなかった人の中で，予測モデルによって正しく発症しないと予測された人の割合を示す特異度 (Specificity) である．

$$Specificity = \frac{TN}{TN + FP}$$

TN: True negative，予測モデルの予測値が陰性を示し，実測値が実際に陰性であったケースの数
FP: False positive，予測モデルの予測値が陽性を示し，実測値が実際には陰性であったケースの数

三つ目は，cutoff ポイントに依存せずに，予測モデルの全体的な予測性能を評価する Area under the ROC curve（以下，AUC）である．AUC は，予測モデルの閾値を変えながら感度と特異度を同一平面上にプロットした曲線である ROC 曲線の積分値として計算される．AUC については，ベイジアンネットワークによる予測モデルと，ベンチマークとしての糖尿病リスクスコアを比較するため，Delong による AUC の差の検定を行う．Delong による AUC の差の検定の帰無仮説 H_0 と対立仮説 H_1 は以下のとおりである．

$$H_0 : \mathrm{AUC_{BN}} = \mathrm{AUC_{Riskscore}}$$
$$H_1 : \mathrm{AUC_{BN}} \neq \mathrm{AUC_{Riskscore}}$$

すべての分析は，R version 3.5.1 を用いて行った．

表 2　自治体ごと糖尿病新規発症の分布

市町村	データ	outcome: 0	outcome: 1	割合
自治体 A	学習データ	3,051	125	0.039
	検証データ	1,524	62	0.039
自治体 B	学習データ	1,611	104	0.061
	検証データ	805	51	0.060
自治体 C	学習データ	1,679	78	0.044
	検証データ	838	38	0.043
自治体 D	学習データ	291	12	0.040
	検証データ	96	3	0.030
自治体 E	学習データ	1,085	76	0.065
	検証データ	541	37	0.064
自治体 F	学習データ	690	31	0.043
	検証データ	344	14	0.039
自治体 G	学習データ	3,098	109	0.034
	検証データ	1,548	54	0.034

4. 予測精度の評価

自治体ごとの 3 年以内糖尿病新規発症 (outcome) の分布は表 2 のとおりである．

はじめに，自治体ごとに構築した予測モデルと，糖尿病リスクスコアによる予測の結果を表 3 に示す．表中の BN がベイジアンネットワークを用いた予測モデル，Riskscore が糖尿病リスクスコアを表す．また，図 1 に AUC の箱ひげ図を示している．表 3 と図 1 から，ベイジアンネットワークを用いた予測モデルによる予測性能と糖尿病リスクスコアによる予測性能には，AUC における大きな差は見られず，いずれの自治体においても同程度の予測性能を示していることがわかる．AUC についての差の検定を行った結果についても，すべての自治体において有意水準 5% で帰無仮説を棄却しないという結果であった．感度については，七つの自治体の中で，4 自治体（自治体 B，C，E，F）においてベイジアンネットワークを用いた予測モデルが糖尿病リスクスコアを上回った．特異度については，5 自治体（自治体 A，B，D，E，G）においてベイジアンネットワークを用いた予測モデルが上回った．感度，特異度

表 3　予測結果

市町村	手法	AUC (95% CI)		cutoff ポイント	感度	特異度
自治体 A	BN	83.613	(76.872, 90.353)	0.099	0.710	0.923
	Riskscore	87.866	(83.783, 91.949)	0.028	0.806	0.816
自治体 B	BN	73.114	(64.574, 81.653)	0.065	0.627	0.839
	Riskscore	74.383	(67.181, 81.585)	0.028	0.608	0.814
自治体 C	BN	87.312	(80.073, 94.550)	0.055	0.842	0.800
	Riskscore	82.510	(75.641, 89.379)	0.055	0.684	0.882
自治体 D	BN	75.355	(41.345, 100.000)	0.099	0.667	0.854
	Riskscore	89.063	(74.222, 100.000)	0.055	0.667	0.813
自治体 E	BN	80.067	(71.041, 89.093)	0.115	0.703	0.874
	Riskscore	76.055	(66.274, 85.837)	0.055	0.568	0.867
自治体 F	BN	81.966	(68.832, 95.101)	0.079	0.714	0.892
	Riskscore	74.761	(59.713, 89.810)	0.115	0.429	0.939
自治体 G	BN	79.888	(71.638, 88.138)	0.091	0.704	0.924
	Riskscore	86.333	(81.042, 91.624)	0.055	0.815	0.786
全体	BN	$Mean = 80.188$, $SD = 4.816$		—	$Mean = 0.710$, $SD = 0.066$	$Mean = 0.872$, $SD = 0.045$
	Riskscore	$Mean = 81.567$, $SD = 6.426$		—	$Mean = 0.654$, $SD = 0.136$	$Mean = 0.845$, $SD = 0.053$

AUC: Area under the ROC curve

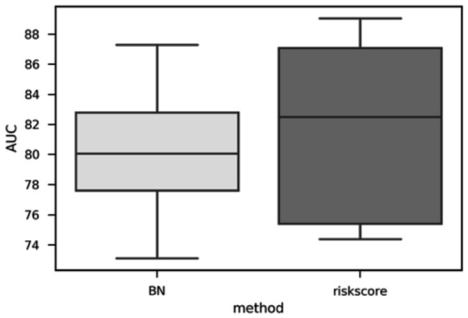

図 1　AUC の箱ひげ図

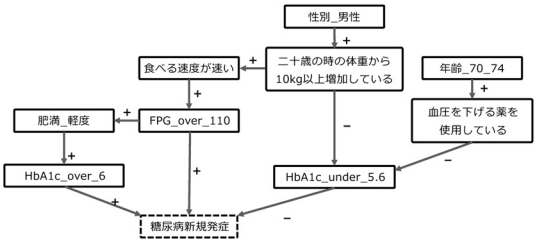

図 2　自治体 B におけるベイジアンネットワーク

の両方について，糖尿病リスクスコアを上回った自治体は 2 自治体（自治体 B，E）であった．また，自治体ごとの予測結果を平均して比較すると，感度と特異度についてはベイジアンネットワークを用いた予測モデルが上回り，AUC については糖尿病リスクスコアが上回った．なお，各指標について t 検定による平均値の差の検定を行った結果，有意水準 5%で帰無仮説を棄却しないという結果であった．

5.　地域健康政策への応用

5.1　ベイジアンネットワークの解釈

図 2 は，自治体 B における糖尿病新規発症ベイジアンネットワークから，糖尿病新規発症に矢印が引かれている投入変数のみを抜粋した簡略図である．図中の ± の符号は，矢印の親ノードと子ノードの間の条件付き確率表を元に付与しており，+，− がそれぞれ正の関係，負の関係を表している．図 2 から，自治体 B では，軽度の肥満であり，HbA1c が高い被保険者，また，二十歳のときの体重から 10 kg 以上増加しており，人と比較して食事を食べる速度が速い男性ほど空腹時血糖値が高く，糖尿病の新規発症確率が高いという関係性がわかる．このことから，血糖検査値の値が高いことに加え，食事を食べる速度が速く，二十歳のときの体重から 10 kg 以上増加している男性に特に着目した保健指導を行っていくことが有効である可能性があることがわかる．また，糖尿病の予防的な保健事業を検討する際は，男性でかつ体重増加が比較的大きい被保険者を対象にすることで，糖尿病発症の予防効果を高められる可能性があることがわかる．

このように，本稿で紹介したベイジアンネットワークによる予測モデルでは，疾病の新規発症確率を得るための予測ツールとしての用途だけでなく，疾病の新規発症に関わる要因の分析までを行うことができる．特に，医学的にすでに明らかになっている要因だけでなく，そのほかの要因と疾病の新規発症との関係に関する洞察を得ることができる点が非常に重要である．自治体 B の結果からは，日本人における横断的，経年的疫学研究からすでに明らかな危険因子である加齢，家

表 4　手法の比較

		手法の特徴	総括
ベイジアンネットワークを用いた予測モデル	利点	既知の危険因子だけでなく，ほかの要因を含めた自治体ごとの疾病構造が把握できる	市町村ごとの疾病構造を捉えることができ，政策立案の場面で有用な手法である
	欠点	すべての自治体において，示唆に富むネットワーク構造が得られるとは限らない	
糖尿病リスクスコア	利点	割り当てられた点数を合計することにより，手軽にリスク発症確率を算出できる	対面の保護指導の場面で，被保険者にわかりやすくリスクを伝えることができる手法である
	欠点	自治体ごとの疾病構造・疾病発症の背景要因について分析できない	

族歴，肥満，身体活動の低下，耐糖能異常（血糖検査値の上昇），高血圧，高脂血症の既往歴 [3] という要因以外の情報として，食事を食べる速度が間接的に糖尿病発症に関係している，などの追加的情報を得ることができる．この点が，本稿でベイジアンネットワークを用いた予測モデル構築の有用性を主張する重要なポイントである．これらの点について，次節で述べる限界を踏まえて二つの手法を比較した結果を表 4 に示している．

5.2　研究課題

最後に，本稿で紹介したベイジアンネットワークを用いた予測モデルの限界と今後の研究課題を述べる．本稿の分析では，自治体がデータ期間中に実施した糖尿病ハイリスク者への施策については考慮されていないことに留意する必要がある．そのため，実際に施策を行った個人を識別することのできるデータの蓄積と，その点を考慮したモデルの構築が必要である．また，今回ベンチマークとした Nanri et al. [1] の糖尿病リスクスコアは，比較的規模の大きい企業の労働者を研究対象として作成されたものである．これにより，ベンチマークとして用意した予測モデルが，国保被保険者とは生活習慣や食習慣などが異なる集団を想定していることによる予測性能の過小評価が発生している可能性がある．糖尿病リスクスコアを一般集団に適用した場合の懸念として Nanri et al. [1] は，糖尿病新規発症リスクの過小評価の可能性を指摘している．今後，予測モデルの有用性をより厳密に示すために，国保被保険者のデータを用いて作成された糖尿病リスクスコアを用いた追加検証が必要である．また，ベイジアンネットワークを用いた手法の限界として，すべての自治体で示唆に富むネットワーク構造が得られるとは限らない点がある．図 3 は自治体 F において構築されたネットワークの簡略図を示している．糖尿病新規発症に矢印が引かれている投入変数のみを抜粋した場合，血糖検査値と疾病発症のノードのみが抽出され，すでに明

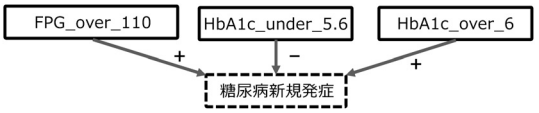

図 3　自治体 F におけるベイジアンネットワーク

らかになっている危険因子以外で，政策立案上有用な知見は得られない結果になっている．今後はこの点を考慮し，現場職員のもつ仮説や医学的な知見を事前情報として表現し，ネットワーク構造の一部に反映させたうえで構造学習を行うなどの方法を検証するなどが有効であると考えられる．

6.　おわりに

本稿では，ベイジアンネットワークを用いた糖尿病新規発症予測モデルの構築とその予測性能評価，さらに，予測モデルより得られるグラフ構造から，地域健康政策への応用可能性を議論した．地域健康政策の文脈で自治体職員の意思決定を支援するための疾病新規発症予測モデルは，予測性能が十分に高いことに加えて，その予測モデルを解釈することで政策的に介入すべき要因を特定することができる方法で構築される必要がある．本稿では，この条件を満たす手法としてベイジアンネットワークを採用し，7 自治体における医療レセプト，特定健診データを用いて，3 年以内の糖尿病新規発症予測モデルを構築した．ベンチマークとして，Nanri et al. [1] による糖尿病リスクスコアを採用し，予測性能を比較した．結果的に，提案手法と既存手法は同等程度の予測性能を示した．さらに構築した予測モデルから 1 自治体を例にとり，地域健康政策の観点で，糖尿病新規発症に関わる重要な要因を示し，「なぜ」がわかる予測モデルとしてのベイジアンネットワークの可能性を示した．今後，現場のオペレーションを考慮したモデリングの方法論の確立や，より厳密なベンチマーク手法との予測性能比較，現場や医学の知見を加味した構造学習などの発展が期待される．

謝辞　本研究に際し，国立研究開発法人日本医療研究開発機構「AIを活用した保健指導システム研究推進事業」による経済的支援に心から感謝申し上げます．また，本稿執筆にあたりさまざまな助言をいただきました，株式会社つくばウエルネスリサーチ塚尾晶子様，千々木祥子様に御礼申し上げます．

参考文献

[1] A. Nanri, T. Nakagawa, K. Kuwahara, S. Yamamoto, T. Honda, H. Okazaki and M. Eguchi, "Development of risk score for predicting 3-year incidence of type 2 diabetes: Japan Epidemiology Collaboration on Occupational Health Study," *PLoS One*, **10**, e0142779, 2015.

[2] 岡山明，『基礎からわかるデータヘルス計画―保健事業の理論と実践―』，社会保険研究所，2017.

[3] 厚生労働省，「糖尿病」，https://www.mhlw.go.jp/www1/topics/kenko21_11/b7.html（2019年4月1日閲覧）

[4] 厚生労働省，「糖尿病性腎症重症化予防プログラム」，https://www.mhlw.go.jp/file/04-Houdouhappyou-12401000-Hokenkyoku-Soumuka/0000121902.pdf（2019年4月1日閲覧）

[5] Y. Doi, T. Ninomiya, J. Hata, Y. Hirakawa, N. Mukai, M. Iwase and Y. Kiyohara, "Two risk score models for predicting incident type 2 diabetes in Japan," *Diabetic Medicine*, **29**, pp. 107–114, 2012.

[6] 福岡県久山町，「ひさやま元気予報」，http://www.town.hisayama.fukuoka.jp/kenkou/kenshin/hisayamagennkiyohou.html（2019年4月1日閲覧）

[7] 植野真臣，『ベイジアンネットワーク』，コロナ社，2013.

[8] 上野眞也，"地域政策の効果を予測する―ベイジアンネットワーク分析の応用―，"熊本大学政策研究，**1**, pp. 29–40, 2010.

[9] 鶴田康人，寒河江雅彦，"ベイジアンネットワークを用いた階層型少子化因果モデルの構築，"大学ディスカッションペーパー，No. 24, 2015.

[10] 宮内義明，"メタボリックシンドロームマネジメントのための特定健診対応ベイジアンネットワークの構築，"博士論文，兵庫県立大学大学院応用情報科学研究科，2016.

[11] 三好利昇，長谷川泰隆，伴秀行，根岸正治，國近則仁，棟重卓三，"特定健診・レセプトデータを用いたベイジアンネットワークによる生活習慣病の医療費予測モデルの構築，"電子情報通信学会技術研究報告，**113**, pp. 139–144, 2014.

[12] 株式会社日立製作所，「IoT ヘルスケア（医療）事例 ウェアラブル端末で健康増進」，2016, https://www.foresight.ext.hitachi.co.jp/_ct/16970095（2019年4月1日閲覧）

[13] 鳥海航，生方裕一，久野譜也，岡田幸彦，"地域健康政策へのベイジアンネットワークの応用，"統計数理，**66**, pp. 267–278, 2018.

[14] C. Meek, "Causal inference and causal explanation with background knowledge," In *Proceedings of the 11th Conference on Uncertainty in Artificial Inteligence*, pp. 403–410, 1995.