統計数理(2018)
第 66 巻 第 2 号 267–278

特集「サービス科学の今」
［研究ノート］

地域健康政策へのベイジアンネットワークの応用

鳥海 航[1]・生方 裕一[1]・久野 譜也[2,4]・岡田 幸彦[3,4]

(受付 2017 年 12 月 31 日；改訂 2018 年 3 月 15 日；採択 3 月 16 日)

要　　旨

本稿は，データ中心科学としてのサービス科学の新たな展開である地域健康政策のためのサービス科学のあり方について，国立研究開発法人日本医療研究開発機構の「AI を活用した保健指導システム研究推進事業」として採択された筑波大学の取り組み事例をもとに議論している．自治体が行う地域健康政策では説明責任が強く求められるため，担当職員にとって説明容易性の高い統計手法を用いる必要がある．また，どの自治体，どの疾病に対しても応用可能な分析方法論を確立する必要がある．本稿では，これらの必要性を満たす統計手法として，制約ベースアプローチで条件付き独立性を χ^2 検定によって行い，より効率的な構造学習が可能な Local to Global アプローチのアルゴリズムを採用したベイジアンネットワークが有用であることを主張する．そして，自治体 A の実際の健康関連ビッグデータを用いて，どの自治体，どの疾病に対しても応用可能な疾病発症ベイジアンネットワークの試行を行っている．そして，本稿で紹介した取り組み事例をふまえ，地域健康政策におけるサービス科学のあり方と今後の研究課題について議論している．

キーワード：地域健康政策，人工知能，説明責任，説明容易性，ベイジアンネットワーク．

1. はじめに

データ中心科学としてのサービス科学は，わが国が抱える社会課題への貢献も期待されている．わが国は，人口減少・少子高齢化の先進国である．すでに 2013 年度にわが国の国民医療費は 40 兆円を超え，今後さらに増大することが見込まれている．さらに，地域間の健康格差の問題が指摘され始めている．これからの超高齢時代において日本国憲法第 25 条が規定する国民の健康で文化的な最低限度の生活を営む権利をどのように保障していくかは，わが国が直面する最も重要な社会課題の 1 つである．

この社会課題に対してわが国は，総務省の支援のもと「健幸長寿社会を創造するスマートウエルネスシティ総合特区」として先導的に構築された健幸クラウドや，厚生労働省の支援のもと構築された国民健康保険中央会の KDB(国保データベース)のように，健康関連ビッグデータ(個人の国民健康保険，介護保険，特定健診等のデータ)の蓄積を政策的に進めてきた．そして，これらの健康関連ビッグデータを基礎として，厚生労働省はデータヘルス計画の策定・実

[1] 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
[2] 筑波大学 体育系：〒 305–8573 茨城県つくば市天王台 1–1–1
[3] 筑波大学 システム情報系：〒 305–8573 茨城県つくば市天王台 1–1–1
[4] 筑波大学 人工知能科学センター：〒 305–8573 茨城県つくば市天王台 1–1–1

行を，総務省は証拠に基づく政策立案(EBPM; Evidence-Based Policy Making)を推進し，地域の健康課題に即した健康政策の実施が促されている．さらに，国立研究開発法人日本医療研究開発機構(AMED)は，2017 年度に「AI を活用した保健指導システム研究推進事業」を公募し，健康関連ビッグデータと人工知能を活用した健康・医療戦略の推進のための基盤研究を開始した．

健康関連ビッグデータの蓄積・活用と人工知能への注目は，データ中心科学としてのサービス科学の新たな地平を示す．それは，社会課題解決型の応用統計数理についての「学」への期待とも言えよう．そこで本稿では，地域健康政策のためのサービス科学のあり方について，上述した AMED の基盤研究として採択された筑波大学の取り組み事例[1)]をもとに議論したい．ここでの重要な論点は，地域健康政策の立案・実行を担う自治体職員(以下，担当職員)の知識・スキル水準との整合性である．なぜなら，どんなに高度な統計手法を用いたとしても，合意形成や説明責任を果たすために担当職員が説明・説得することができなければ，人工知能は無用の長物となってしまうのである．

そこでまず本稿では，地域健康政策における人工知能の使い手である担当職員の目線から，「説明・説得のしやすさ」としての説明容易性を重視した統計手法の選定を考える．その結果，地域健康政策における制約ベースアプローチのベイジアンネットワークの有用性が導かれる．次いで，健康関連ビッグデータを所与として，どの自治体，どの疾病にでも広く応用可能なベイジアンネットワークの構築方法について考察する．そして，自治体 A の実際の医療レセプトおよび特定健診データを用いた疾病発症ベイジアンネットワークの構築を行う．なお本稿では，紙幅の関係から，全ての疾病の中で代表的な高血圧と糖尿病の結果のみを取り上げる．最後に，本稿での取り組み事例をふまえ，地域健康政策におけるサービス科学のあり方と今後の研究課題について議論したい．

2. 地域健康政策で有用な統計手法の考察

2.1 行政運営に求められる説明容易性

わが国の地方行政は，日本国憲法および地方自治法・関連諸規則に従い，都道府県・市町村(以下，自治体)による地域別の行政執行が基本となっている．自治体の執行機関としての長は住民による直接選挙で選ばれ，同じく直接選挙で選ばれた議事機関としての議会との相互牽制のもと，予算を編成し，政策および事務事業を実行する．予算編成，政策および事務事業の実行等を所管するのは，自治体の職員である．一般的に，自治体の職員は部局別の組織構造に配属され，当該組織は部から課，そして係へと細分化されている．

地域健康政策に限定すると，住民の健康生活を保障することを主目的とし，自治体には健康推進課や保健推進課といった名称の担当部局が存在する．そして，それらの中に設置された係単位で予算案の作成や事務事業の実施等が行われることが一般的である．つまり，地域健康政策の立案と実行は，自治体の基本計画や首長の公約等を基礎として，主として健康推進関係の係の担当職員が事務事業を所管する．そのため，担当職員にとって，内的には上長，財政課のような予算査定担当，首長，そして議会に対する逐次的な合意形成を行うことが，外的には住民に対して説明責任を果たすことが，地域健康政策上で不可欠となる．

以上のわが国地域健康行政の特徴は，前述した国が推進する健康関連ビッグデータの蓄積・活用に際し，応用統計数理の観点で重要な注意点を喚起する．統計学の初歩的知識しか有さない担当職員が，統計学の知識がない多様な利害関係者に説明・説得しなければならない姿を想定する必要があるのである．そして前述のとおり，この文脈で人工知能の活用が期待されている．ここでの避けられない利害関係者からの問いは，「なぜその健康事務事業を行う必要があ

るのか？」である．この問いに対し，健康関連ビッグデータから当該地域の健康課題の因果メカニズムを導いたうえで，その重要な原因候補を特定し，それらを証拠として健康事務事業を立案する，という EBPM が求められる時代になった．

地域健康政策における EBPM において，人工知能の役割は大きい．なぜなら，「健康関連ビッグデータを用いた因果推論 ⇒ 重要な原因の特定 ⇒ 政策立案」という一連の流れの大部分を，人工知能が代替できる可能性が高いからである．一方で，内的な説明・説得プロセスを勝ち抜くとともに，住民への説明責任を果たすためには，「どうやってその証拠が生み出されたのか？」という追加的な利害関係者からの問いに担当職員が答える準備をしておく必要がある．つまり，地域健康政策における EBPM では，証拠の科学性や可読性だけでなく，証拠とその生成過程の説明容易性が強く求められるのである．

説明容易性は，データ中心科学としてのサービス科学がこれまで見すごしてきた重要な論点であると考えられる．内的な説明・説得プロセスも含む広い意味での説明責任が求められない場合には，予測・判別精度のみを重視し，ブラックボックス型とも呼称される計算論的機械学習を用いることは有用となろう．しかしながら，説明責任が強く求められる場合には，注意を要する．なぜなら，証拠とその生成過程の説明をも求められるからである．以上を鑑みると，地域健康政策の文脈では説明責任が強く求められることから，人工知能の利用者の知識・スキル水準と整合する説明容易性の高さを基準とした統計手法の選定が望ましいと考えられる．

ここで，地域健康課題の因果メカニズムを推論でき，説明容易性が相対的に高い手法として，地域健康政策におけるベイジアンネットワークの有用性を主張できよう．ベイジアンネットワークはグラフィカルモデルの一種であり，有向グラフと条件付き確率表によって因果メカニズムを表現することができるため，不確実性の伴う複雑な社会現象を柔軟にモデル化することが可能である（本村, 2000; Pearl, 2003; Kalisch et al., 2010）．また，変数間の確率的な関連性を生かしたソフトウェア開発や病気における要因分析（Park and Kim, 2013; Harris et al., 2017），疾病の診断・管理（Velikova et al., 2014），映画のレコメンドシステム（Ono et al., 2007）など，様々な用途での活用が提案されてきた．

そして近年，ベイジアンネットワークは政策立案現場での応用可能性が議論されるようになった．例えば，環境政策の効果分析を行った Carriger et al.（2016）は，ベイジアンネットワークを用いた確率的推論を行うことで，不確実性が高い状況下でも EBPM が容易になることを示している．また上野（2010）は，ベイジアンネットワークを用いて過疎地域での人口減少要因分析を行い，政策提案を行っている．ここで上野（2010）は，ベイジアンネットワークは独立変数を変化させたときの目的変数の変化を確率として予測できるため，政策立案の際に有用性が高いことを指摘している．そして，ベイジアンネットワークを用いることで社会現象全体の構造が可視的に明らかになり，より深い洞察を得ることが可能となる（鶴田・寒河江, 2015）．

説明容易性が必要とされる地域健康政策では，地域健康課題の因果メカニズムを可視的かつ確率的に洞察できるベイジアンネットワークの有用性が高いものと考えられる．

2.2 説明容易性と制約ベースアプローチ

ベイジアンネットワークの構造学習は，スコアベースアプローチと制約ベースアプローチという 2 つのアプローチが存在する（Koller and Friedman, 2009）．スコアベースアプローチは，データセット D から導出される構造 G のスコアを $\mathrm{Score}(G|D)$ と定義すると，式(2.1)のようにスコア関数が最大となる構造を探索するようなアプローチである．

$$G^* = \mathrm{argmax}_G \, \mathrm{Score}(G|D) \tag{2.1}$$

ここで使用されるスコア関数は，情報理論アプローチとベイジアンアプローチが存在し，

基本的にいずれのスコア関数も対数尤度を基礎としている．式(2.2)は最も素朴な対数尤度によるスコア関数であるが，この方法は過学習の問題が指摘されてきた（Liu et al., 2012）．そこで，情報理論アプローチでは BIC（Schwartz, 1978)等を，ベイジアンアプローチでは BDeu（Heckerman et al., 1995)等を，式(2.2)に対する罰則項として付け加えることで，よりよい構造推定を目指すスコア関数が式(2.3)と式(2.4)である．ここで式(2.2)の D_{ij} は，データセット D における $i(i=1,\ldots,n)$ 番目の変数の $j(j=1,\ldots,N)$ 番目のデータサンプルを表している．また，PA_{ij} は，i 番目の変数の親ノード集合を表している．式(2.3)におけるは，i 番目の変数における状態数である．式(2.4)における r_i と q_i は，i 番目の変数の状態数とその親ノード集合の状態数である．同式中の D_{ij} は，データセット D における i 番目の変数の親ノード集合の状態が j となるデータサンプルを表している．その中でも D_{ijk} は，i 番目の変数の状態が k となるデータサンプルを表している．α はモデル構築者が決めるパラメータである．

$$LL(D|G)=\sum_{j=1}^{N}\log P(D_j|G)=\sum_{i=1}^{n}\sum_{j=1}^{N}\log P(D_{ij}|PA_{ij}) \tag{2.2}$$

$$BIC(D|G)=LL(D|G)-\sum_{i=1}^{n}\frac{\log N*p_i}{2} \tag{2.3}$$

$$BDeu(D|G)=LL(D|G)-\sum_{i=1}^{n}\sum_{j}^{q_i}\sum_{k}^{r_i}\log\frac{P(D_{ijk}|D_{ij})}{P(D_{ijk}|D_{ij},\alpha_{ij})} \tag{2.4}$$

このスコアベースアプローチの構造探索は，変数が増加すると指数関数的に計算量が増加する欠点が指摘されてきた(Chickering et al., 2004)．一方で，後述する制約ベースアプローチと比較して，サンプル数が少ない場合でも構造推定の精度を高く維持できるメリットがある（Tsamardinos et al., 2006)．つまり，変数が比較的少なく，サンプル数も比較的少ない場合には，スコアベースアプローチが技術的に優位であると考えられる．しかし前述のとおり，わが国ではすでに多変量大規模サンプルの健康関連ビッグデータが蓄積されていることから，低サンプル数の問題は重要ではなく，むしろ多変量に係る計算量の問題が重大となるため，わが国地域健康政策におけるスコアベースアプローチの優位性があるとは言い難い．むしろ，地域健康政策で求められる説明容易性の観点からすると，後述の制約ベースアプローチと比較して，大きな課題があることを指摘せざるを得ない．統計学の初歩的知識しか有さない担当職員が，統計学の知識がない多様な利害関係者に対して，スコア関数にもとづく構造推定について説明するのは容易ではないのである．

相対的に説明容易性が高いと考えられる制約ベースアプローチは，条件付き独立性に注目した構造推定を行う．ここで，有限個の要素からなる確率変数集合 V に対して，$P(V)$ を V の同時確率分布とし，X，Y，Z を V の部分集合としよう．この時，$P(y,z)>0$ に対して式(2.5)が成り立つ場合，X と Y は Z を所与とした条件付き独立であると考える．

$$P(x|y,z)=P(x|z) \tag{2.5}$$

制約ベースアプローチは，この条件付き独立性を χ^2 検定，G^2 検定，相互情報量の大小などで判定する点が特徴的である．その中で最も古典的な方法は，PC アルゴリズム（Spirtes et al., 2000; Madsen et al., 2017)や TPDA アルゴリズム（Cheng et al., 2002)のように，無向完全グラフや全域木から条件付き独立性を基準として構造を絞り，方向付けを行う Global アプローチである．Global アプローチは，条件付き独立性を基準として構造を絞ることと，方向付けを

行うこととを再帰的に行うことで，より効率的に構造学習を行う RAI アルゴリズム（Yahezkel and Lerner, 2009）が提唱されるに至っている．

一方で，制約ベースアプローチには，GS アルゴリズム（Margaritis and Thrun, 2000）のような，まず部分的に局所グラフを作成し，それぞれの局所グラフをつなげることにより全体の無向グラフを作成し，方向づけを行う Local to Global アプローチ（Gao et al., 2017）も存在する．さらに，Local to Global アプローチを構成する要素技術の発展（Aliferis et al., 2010b）として，MMPC アルゴリズム（Tsamardinos et al., 2006）や HITON-PC アルゴリズム（Aliferis et al., 2003; Aliferis et al., 2010a; Aliferis et al., 2010b）のように，ターゲットノードに対する辺候補を親子関係の蓋然性から導出し，局所構造を探索する方法も提案されている．

一般的に Local to Global アプローチは，Global アプローチと比較して，条件付き独立性検定の試行回数について，効率的に構造学習をすることが可能であるとされる（Tsamardinos et al., 2006）．この検定における効率性は，次の 2 点において効果的である．まず Global アプローチで必要となる高次の条件付き独立性検定を必要としないため，高次の検定結果による信頼性や効率性の低下を防ぐことができる．次に，検定の試行回数自体を少なくすることで，構造推定の精度を維持できる（Koller and Friedman, 2009）．このように効率性が高く，信頼性も維持できる点が，Local to Global アプローチの利点としてあげられる．一方で，Local to Global アプローチは，関係の強い部分的な変数群に関する局所構造の推定を行い，探索空間を制限する（Gao et al., 2017）．つまり，Global アプローチと比較して，Local to Global アプローチには全てのグラフを構造推定の候補としない欠点が存在している．

ここで注意すべきは，地域健康政策のためにベイジアンネットワークを用いる文脈である．統計学の初歩的知識しか有さない担当職員にとって，最もなじみがあり，説明容易性が高いのは，条件付き独立性を χ^2 検定で判断することであると考えられる．さらに，自治体は国の政策によって健康関連ビッグデータをすでに有し，今後もデータがさらに大規模に蓄積されていくことを考えると，多変量大規模データからいかに効率的に構造推定を行い，地域健康政策上の情報ニーズに迅速に応えていくかが重要となろう．こうした理由から，わが国の地域健康政策においてベイジアンネットワークを適切に応用するためには，説明容易性の観点から制約ベースアプローチで条件付き独立性を χ^2 検定によって行い，より効率的な構造学習が可能な Local to Global アプローチのアルゴリズムを採用することが望ましいと考えられる．

3. 疾病発症ベイジアンネットワークの基本設計と使用データ

本稿の以降では，前章で導いたわが国地域健康政策において有用だと考えられる統計手法の条件に従い，実際の健康関連ビッグデータを用いた地域健康課題の因果メカニズムのモデリングを試行したい．この試行では，前述の条件を満たす HITON-PC アルゴリズム（Aliferis et al., 2003; Aliferis et al., 2010a; Aliferis et al. 2010b）を用い，条件付き独立性は χ^2 検定で判断し，Verma and Pearl（1990）の Inductive Causation アルゴリズムによる因果モデルの構築を行う．このとき，非巡回制約を満たす方向付けを，Meek（1995）の伝統的なオリエンテーションルールによって行う．本稿で試行する因果モデリングは，国が政策的に蓄積を進めてきた健康関連ビッグデータの活用を想定し，わが国全ての自治体で応用可能な方法を追求する．具体的には，ほぼ全ての自治体が登録し，活用していると言われる KDB を前提とし，KDB に収録されている住民の国民健康保険（以下，国保）の医療レセプトデータと特定健診の結果データを用いる．それらのデータの概要は，以下のとおりである．

わが国では，被用者保険と国保のいずれかの医療保険に加入することが義務付けられている．自治体を保険者とする国保への加入者数は 3,303 万人であり，その平均年齢は 51.5 歳であ

る(厚生労働省, 2016)．KDB では，被保険者である住民個人の疾病区分ごとの医療レセプトが含まれており，ここでの疾病区分は厚生労働省が定める「社会保険表章用疾病分類」の区分に従い，糖尿病や高血圧などの 121 疾病が存在している．この国保の医療レセプトデータを用いれば，ある住民が特定の疾病を発症しているか否かの判定が可能となる．そして，こうして判定された住民別・疾病別の発症の有無のデータは，個別自治体における健康課題についての因果メカニズムを特定する際の，欠かせない結果変数となる．

一方，特定健診は「高齢者の医療の確保に関する法律」に従い，「医療保険者(国保・被用者保険)が，40〜74 歳の加入者(被保険者・被扶養者)を対象として，毎年度，計画的に(特定健康診査等実施計画に定めた内容に基づき)実施する，メタボリックシンドロームに着目した検査項目」(厚生労働省, 2013, p.5)による健康診査である．特定健診における項目は，必須項目と詳細な健診項目から構成される．必須項目は，質問表(服薬歴，喫煙歴等)，身体測定(身長，体重，BMI，腹囲)，血圧測定，理学的検査，検尿(尿糖，尿蛋白)，血液検査(脂質検査，血糖検査，肝機能検査)からなる (厚生労働省, 2009)．

これら特定健診データの中で，疾病発症ベイジアンネットワークで用いるために離散変数として取り扱うことが可能な投入変数として，少なくとも表 1 の 33 変数を作成・使用可能である．これらの変数は，KDB に参加する全ての自治体で作成・使用可能である．本稿は全自治体への応用可能性を目指した疾病発症ベイジアンネットワークの試行であり，本稿ではこの試行のために匿名でご協力くださった自治体 A の 2015 年度のデータを用いる．自治体 A はわが国における典型的な中小自治体であり，国保の医療レセプトと特定健診のデータが個人単位で揃っている全住民は 3,391 人であった．この本稿で用いる 33 個の離散変数に関する全 3,391 人の分布は，表 1 のとおりである．

ここで本稿では，疾病発症ベイジアンネットワークを構築するにあたり，表 1 の 33 個の投入変数を，政策的に改善が可能な原因候補として位置付けられる生活習慣系・食習慣系・健康意欲系・身体指標系，政策的に変化させることは難しいがターゲットセグメント化には有用である個人属性系・病歴系，という 6 つにグループ化した．これらに加えて，自治体内の地域特性を反映すべく，どの自治体でも使用可能な地区変数として，自治体 A 内の 8 つの小学校区を地域特性系として用意した．これら 7 つのグループと各グループ内の変数リストは，図 1 のとおりである．そして本稿では，疾病発症ベイジアンネットワークの構築に際し，グループ内の変数の間の関係性を考慮しない制約を設けた．加えて，疾病発症の有無が結果変数となるように，疾病発症の有無から全ての投入変数に対して矢印が引かれることがない制約を設けた．

なお，本稿の疾病発症ベイジアンネットワークは，全ての疾病区分に対して応用可能である．紙幅の関係から，本稿では，代表的な生活習慣病である高血圧と糖尿病の 2 疾病を取り上げることとする．

4. 試行結果と議論

前述の基本設計とデータによって，高血圧と糖尿病の疾病発症ベイジアンネットワークが構築された．疾病へと矢印が向かわない投入変数と矢印を省略した簡易図は，図 2(高血圧)と図 3(糖尿病)のとおりである．なお，図中の符号は関連する 2 変数(全て二値データである)についての条件付き確率表をもとに付され，+ は正の関係を，− は負の関係を示している．

図 2 と図 3 から，自治体 A では，メタボ判定基準に該当する住民が高血圧と糖尿病の発症確率が高いことがわかる．これが因果関係を示していることが医学的に確からしいのであれば，自治体 A ではメタボ対策の事務事業を重点的に行うことが有効となりそうである．そして，糖尿病発症に関しては，このメタボ対策の事務事業を，特に男性に対して重点的に実施すること

表 1. 本稿で使用する 33 個の投入変数と分布(全 3,391 人).

変数名	内容	人
性別	女性	1,864
	男性	1,527
40~44歳	該当者	90
45~49歳	該当者	85
50~54歳	該当者	103
55~59歳	該当者	181
60~64歳	該当者	664
65~69歳	該当者	1,170
70~74歳	該当者	1,098
BMI（Underweight）	該当者	269
BMI（Normal）	該当者	2,341
BMI（Pre-obese）	該当者	696
BMI（Obese）	該当者	85
メタボ判定	基準該当	606
_基準該当	基準該当ではない	2,785
メタボ判定	予備群該当	272
_予備群	予備群該当ではない	3,119
既往歴	有り	129
（脳血管）	無し	3,262
既往歴	有り	185
（心血管）	無し	3,206
既往歴	はい	13
（腎不全）	いいえ	3,378
貧血	はい	450
	いいえ	2,941
喫煙	はい	396
	いいえ	2,995

変数名	内容	人
体重変化	はい	945
	いいえ	2,446
運動習慣	はい	2,002
不十分	いいえ	1,389
歩行習慣	はい	1,318
不十分	いいえ	2,073
歩行速度	はい	1,646
	いいえ	1,745
食習慣（早食い）	早い	704
_早い	早くない	2,687
食習慣（早食い）	遅い	245
_遅い	遅くない	3,146
食習慣	はい	464
（就寝前夕食）	いいえ	2,927
食習慣	はい	338
（夜食）	いいえ	3,053
食習慣	はい	169
（朝食抜き）	いいえ	3,222
飲酒	毎日	919
_毎日	毎日ではない	2,472
飲酒	時々	787
	時々ではない	2,604
睡眠不十分	はい	590
	いいえ	2,801
改善意識	健康改善するつもりはない	1,174
	健康改善するつもりである	2,217
改善行動	健康改善行動をしていない	1,953
	健康改善行動を始めている	1,438

が，自治体 A では有効かもしれない．

一方，図 2 の高血圧発症に関しては，45 歳から 74 歳までの喫煙する男性が，そして貧血で食事が遅い 70 歳から 74 歳までの女性が，さらに男女を問わず太り気味の住民が，自治体 A では高血圧の発症確率が高そうである．もしこれが自治体 A の実態に即していると担当職員が感じるのであれば，自治体 A ではメタボ対策だけでなく，これら 3 つの領域に特化した高血圧対策の事務事業に注力することが有効となりそうである．

これらの結果と解釈はあくまでも 2015 年度の自治体 A のデータをもとにした試行であり，実用化を目指すうえでは，医学的見地からの補強や解釈が困難な矢印への対応方法の検討など，さらなる研究開発が必要である．しかし，本稿で主張したいことは，わが国地域健康政策の文脈に即し，説明容易性を重視した統計手法の選定を行い，かつどの自治体，どの疾病にも広く応用可能な方法論を設計することで，データ中心科学としてのサービス科学の新たな地平

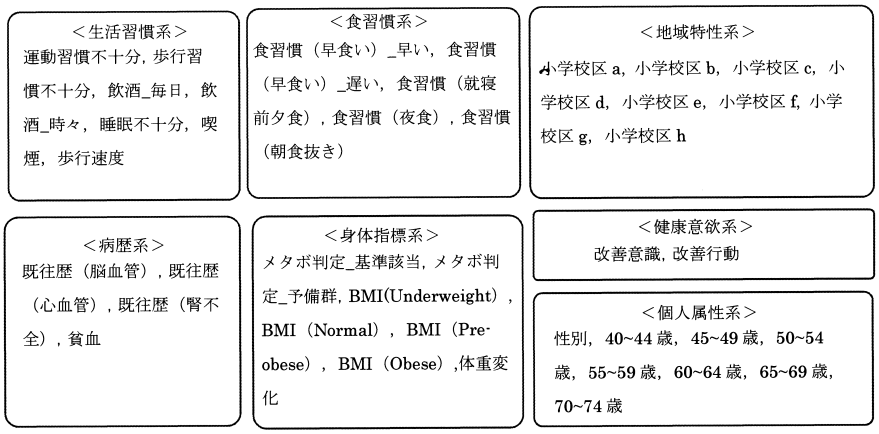

図 1．7 つのグループと変数.

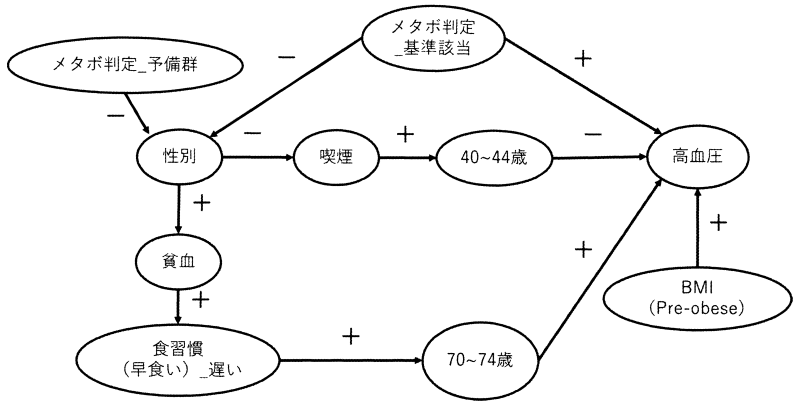

図 2．高血圧の疾病発症ベイジアンネットワーク.

が切り拓かれる可能性があることである．その中で，本稿が考察し試行した疾病発症ベイジアンネットワークの構築方法論は，今後のわが国の地域健康政策において 1 つの有用なやり方であると考えられる．

5. おわりに

本稿では，地域健康政策のためのサービス科学のあり方について，AMED の基盤研究として採択された筑波大学の取り組み事例をもとに議論した．自治体が行う地域健康政策では説明責任が強く求められるため，統計学の初歩的知識しか有さない担当職員にとって説明容易性の高い統計手法を用いる必要がある．また，国が政策的に蓄積してきた健康ビッグデータを前提とし，どの自治体，どの疾病に対しても応用可能な分析方法論を確立する必要がある．本稿では，これらの必要性を満たす統計手法として，制約ベースアプローチで条件付き独立性を χ^2 検定によって行い，より効率的な構造学習が可能な Local to Global アプローチのアルゴリズムを採用したベイジアンネットワークが有用であることを主張した．

次いで本稿では，自治体 A の実際の健康関連ビッグデータを用いて，どの自治体，どの疾病

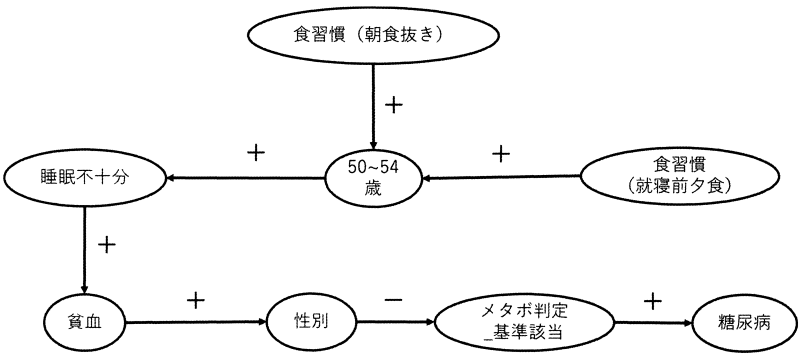

図 3. 糖尿病の疾病発症ベイジアンネットワーク.

に対しても応用可能な疾病発症ベイジアンネットワークの構築を試行した．その結果，因果とは言い難い関係性が一部含まれること，医学や公衆衛生分野の専門家による追加的な因果検証が必要であること，等の課題はあるものの，自治体 A ならではの疾病発症因果モデリングを行うことができた．そして，それを政策立案に活かす糸口を示すことができたと考える．

本稿での取り組み事例をふまえると，地域健康政策におけるデータ中心科学としてのサービス科学は，今後さらなる発展を遂げるであろう．なぜなら，本稿では個別疾病の発症の有無にしか焦点をあてていないが，合併症を含む複数の疾病の発症因果モデリング，各疾病の医療費増減の因果モデリング，社会保険・後期高齢・介護保険のデータをも包括した地域健康因果モデリング，追加的なライフスタイル・アンケートや運動データ等の取得による健康で幸せな生活のための因果モデリングなど，国民の健康長寿と国民医療費の削減との両立に貢献するデータ中心科学の発展が期待できるからである．

健康関連ビッグデータの蓄積・活用と人工知能への注目は，社会課題解決型の応用統計数理についての「学」への期待であり，本稿はその端緒を開こうとする萌芽的な研究ノートにすぎない．超高齢時代のサービス科学として，今後さらなる理論的・実証的研究が蓄積されることが求められる．

注.

[1)] AMED の「AI を活用した保健指導システム研究推進事業」は，2 つ以上の自治体との共同研究開発を条件として公募が行われた．採択結果は 2 件であり，広島大学の「自治体等保険者レセプトデータと健康情報等を基盤に AI を用いてリスク予測やターゲティングを行う保健指導システムの構築に関する研究」と，筑波大学の「自治体における保健指導の施策力に応じた最適な保健指導モデルを提示できる AI の開発研究」が採択された．いずれの取り組みも健康関連ビッグデータと人工知能の活用が想定されており，前者は住民個人への働きかけを想定したミクロ・アプローチ，後者は地域健康政策を想定したマクロ・アプローチとなっている．後者の筑波大学の取り組みは，筑波大学を代表研究機関とし，筑波大学発ベンチャーであるつくばウエルネスリサーチが自治体と連携して構築してきた 75 万人以上の健康関連ビッグデータを基盤に，筑波大学人工知能科学センターと NTT グループの人工知能技術を融合させることにより，自治体の地域健康政策を支援する AI システムを研究開発する．

参 考 文 献

Aliferis, C. F., Tsamardinos, I. and Statnikov, A. (2003). HITON: A novel Markov blanket algorithm for optimal variable selection, *AMIA Annual Symposium Proceedings*, 21–25.

Aliferis, C. F., Statnikov, A., Tsamardinos, I., Mani, S. and Koutsoukos, X. D. (2010a). Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation, *Journal of Machine Learning Research*, **11**, 171–234.

Aliferis, C. F., Statnikov, A., Tsamardinos, I., Mani, S. and Koutsoukos, X. D. (2010b). Local causal and Markov blanket induction for causal discovery and feature selection for classification part II: Analysis and extensions, *Journal of Machine Learning Research*, **11**, 235–284.

Carriger, J. F., Barron, M. G. and Newman, M. C. (2016). Bayesian networks improve causal environmental assessments for evidence-based policy, *Environmental Science & Technology*, **50**(24), 13195–13205.

Cheng, J., Greiger, R., Kelly, J., Bell, D. and Liu, W. (2002). Learning Bayesian networks from data: An information-theory based approach, *Artificial Intelligence*, **137**, 43–90.

Chickering, D. M., Heckerman, D. and Meek, C. (2004). Large-sample learning of Bayesian networks is NP-hard, *Journal of Machine Learning Research*, **5**, 1287–1330.

Gao, T., Fadnis, K. and Campbell, M. (2017). Local-to-global Bayesian network structure learning, *Proceedings of the 34th International Conference on Machine Learning*, 1193–1202.

Harris, M. J., Stinson, J. and Landis, W. G. (2017). A Bayesian approach to integrated ecological and human health risk assessment for the south river, Virginia mercury contaminated site, *Risk Analysis*, **37**(7), 1341–1357.

Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*, **20**(3), 197–243.

Kalisch, M., Fellinghauer, B. A., Grill, E., Maathuis, M. H., Mansmann, U., Bühlmann, P. and Stucki, G. (2010). Understanding human functioning using graphical models, *BMC Medical Research Methodology*, **10**(1), 14–24.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models*: *Principles and Techniques*, MIT Press, Cambridge.

厚生労働省 (2009). 政策レポート(特定健康診査(いわゆるメタボ健診)・特定保健指導).

厚生労働省 (2013). 特定健康診査・特定保健指導の円滑な実施に向けた手引 Ver2.0.

厚生労働省 (2016). 我が国の医療保険について.

Liu, Z., Malone, B. and Yuan, C. (2012). Empirical evaluation of scoring functions for Bayesian network model selection, *BMC Bioinformatics*, **13**(15), 1–16.

Madsen, A. L., Jensen, F., Salmerón, A., Langseth, H. and Nielsen, T. D. (2017). A parallel algorithm for Bayesian network structure learning from large data sets, *Knowledge-Based Systems*, **117**, 46–55.

Margaritis, D. and Thrun, S. (2000). Bayesian network induction via local neighborhoods, *Advances in Neural Information Processing Systems*, **12**, 505–511.

Meek, C. (1995). Causal inference and causal explanation with background knowledge, *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, 403–410.

本村陽一 (2000). ベイジアンネットワーク, 電子情報通信学会誌, **83**(8), 645–646.

Ono, C., Kurokawa, M., Motomura, Y. and Asoh, H. (2007). A context-aware movie preference model using a Bayesian network for recommendation and promotion, *Proceedings of the 11th International Conference on User Modeling*, 247–257.

Park, H. J. and Kim, S. H. (2013). A Bayesian network approach to examining key success factors of mobile games, *Journal of Business Research*, **66**(9), 1353–1359.

Pearl, J. (2003). Causality: Models, reasoning and inference, *Econometric Theory*, **19**, 675–685.

Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), 461–464.

Spirtes, P., Glymour, C. N. and Scheines, R. (2000). *Causation, Prediction and Search*, MIT Press, Cambridge.

Tsamardinos, I., Brown, L. E. and Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm, *Machine Learning*, **65**(1), 31–78.

鶴田康人, 寒河江雅彦 (2015). 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, 金沢大学ディスカッションペーパー, No.24.

上野眞也 (2010). 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**, 29–40.

Velikova, M., Van Scheltinga, J. T., Lucas, P. J. and Spaanderman, M. (2014). Exploiting causal functional relationships in Bayesian network modelling for personalized healthcare, *International Journal of Approximate Reasoning*, **55**(1), 59–73.

Verma, T. and Pearl, J. (1990). Equivalence and synthesis of causal models, *Proceedings of the 6th Annual Conference on Uncertainty in Artificial Intelligence*, 255–270.

Yehezkel, R. and Lerner, B. (2009). Bayesian network structure learning by recursive autonomy identification, *Journal of Machine Learning Research*, **10**, 1527–1570.

Applicability of Bayesian Network to Regional Health Policy in Japan

Wataru Toriumi[1], Yuichi Ubukata[1], Shinya Kuno[2,4] and Yukihiko Okada[3,4]

[1]Graduate School of Systems and Information Engineering, University of Tsukuba
[2]Faculty of Health and Sport Sciences, University of Tsukuba
[3]Faculty of Engineering, Information, and Systems, University of Tsukuba
[4]Center for Artificial Intelligence Research, University of Tsukuba

This paper discusses the service science for regional health policy and helps to contribute to the development of service science as data centric science. Because of strong accountability towards residents, it is necessary for municipal officials to use the statistical methods when planning the regional health policy as this would make it easy to explain how and why they choose it. Also, it is necessary to establish a statistical method applicable to any municipality and any disease. In this paper, we propose that the Bayesian network adopting the algorithm of Local to Global approach which enables more efficient structural learning is useful in fulfilling these needs. This algorithm is one of the constraint-based approaches and tests conditional independence with test.

In addition, we develop a disease-causing Bayesian network applicable to any municipality and any disease.

Key words: Regional health policy, artificial intelligence, accountability, ease of explanation, Bayesian network.