

多施設臨床レジストリを可視化し利活用を促進するための患者状態確率モデルの研究

石井 雅通^{*1}, 美代 賢吾^{*1}, 杉山 雄大^{*2}, 大杉 満^{*2}, 植木 浩二郎^{*2}

^{*1} NCGM(国立国際医療研究センター)情報基盤センター,

^{*2} NCGM 研究所 糖尿病研究センター

Study on patient condition probability model visualizing multi-institutional clinical registry database

Masamichi Ishii^{*1}, Kengo Miyo^{*1}, Takehiro Sugiyama^{*2},

Mitsuru Ohsugi^{*2}, Kohjiro Ueki^{*2}

^{*1} NCGM(National Center for Global Health and Medicine) Healthcare Intelligence Technology Center,

^{*2} NCGM Research Institute Diabetes Research Center

抄録: 当センターでは日本糖尿病学会との共同事業「診療録直結型全国糖尿病データベース事業(Japan Diabetes compREhensive database project based on an Advanced electronic Medical record System: J-DREAMS)」を推進しており、臨床データベースの利活用が期待されている【課題】J-DREAMSで扱う臨床データは300項目以上あり、数万規模の症例データの全体を直観的に把握することは困難である【提案手法】解決策として人工知能技術のひとつである、ベイジアンネットワークを活用した患者状態確率モデルを考案した。観察項目、検査結果、処方情報の各情報をノードに設定し、データの関連を有向非循環グラフで表現するモデルである【結果】プロトタイプを構築し、糖尿病専門医による評価を実施した【考察】本モデルによりデータベース全体の俯瞰が可能となり、各ノードの事前確率を設定して確率分布の変化を探索的に確認できるため、簡易的なスクリーニングが可能となる。本研究の効果により多施設臨床レジストリの利活用促進が期待できる。

キーワード Artificial Intelligence, Bayesian Network, Clinical Registry database

1. はじめに

当センターが日本糖尿病学会と共同で推進する診療録直結型全国糖尿病データベース事業(

Japan Diabetes compREhensive database project based on an Advanced electronic Medical record System : J-DREAMS)」は10~20万人規模の症例登録をめざしている。今後、収集した多施設臨床データから創出された臨床知識を搭載した糖尿病領域における人工知能システムへの発展が期待されている。1)

臨床レジストリデータベースを人工知能で活用するにはデータの特徴を反映した“学習データ”の導出が重要である。実現には臨床ドメイン知識に加えてデータベースに蓄積された症例データ自体に対する特徴理解が必要となる。

本論では臨床レジストリデータベースの症例データの特徴理解をサポートする技術開発を目的として、臨床レジストリに登録された症例群を可

視化して患者状態の俯瞰を可能とする“患者状態モデル”の構築を試みる。

2. 方法

データベースに登録された症例の傾向や特徴を把握するため従前は記述統計を活用してきた。300以上のレジストリ項目から適切な統計を作成するには探索的に分析視点を求めて統計作成を繰り返すことになり効率が悪い。一方、BIツールを活用する場合は50以上のテーブル構造並びに300以上のデータ項目の関連についての事前理解が求められる。臨床の専門家が自ら実践するには負担が大きい。

本論では人工知能技術のひとつであるベイジアンネットワークを活用することによりテーブル構造や項目間の関連を可視化して臨床レジストリの症例群を俯瞰し、直観的かつ簡易的に取り扱うことを実現するため、以下の手法にてプロトタイプを構築する。

1) データ項目の定義抽出

J-DREAMS 糖尿病標準診療テンプレートで入力対象となっている観察項目に加えて、検査結果値、処方情報から項目を抽出し、カテゴリ化を前提としてデータ項目を定義する

2) ベイジアンネットワークモデル定義

データ項目をノードとして、データ項目間の関係を有向非循環グラフで表現する。詳細をもつ項目を親ノード、詳細項目を子ノードとして定義する

3) 症例データ抽出

J-DREAMS データベースからデータ項目に対応した実データを抽出する (約 16,000 レコード)

4) 学習データ創成

症例データを変換し、学習データを創成する

(1) データクレンジング

異常値や欠損値に対応する

(2) 特徴量データの導出

時間軸等を反映した特徴量データを導出する
例) 前回検査値からのカテゴリ変化、変動量等

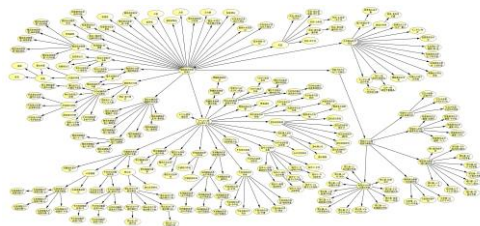


Fig.1 患者状態確率モデル図(可視化結果)



Fig.2 事前確率条件を指定した推論例

(3) 離散化

数値データをカテゴリ変数に変換する
例) 実年齢の年代カテゴリへの変換

5) ベイジアンネットワーク生成

学習データからパラメタ学習を実行し条件付き確率分布を生成する

Table.1 モデル検証例 (Creatinine)

	適合率	再現率	F値
> 正常値	0.0667	0.0021	0.0040
正常値範囲	0.7913	0.9939	0.8811
正常値>	0.0000	0.0000	0.0000

3. 結果

テンプレート項目全体の可視化が患者状態確率モデルとして実現した(Fig.1). モデルに事前確率を指定したアドホックな推論が可能となった(Fig.2). Creatinine 値(正常値範囲内外で離散化)を目的変数とした本モデルの検証結果を示す(Table.1). 異常値検出は精度がでない。

4. 考察

ベイジアンネットワークによる患者状態確率モデルを構築し、症例データベース全体可視化によりデータ分布を俯瞰可能とすることでレジストリを探索的に利用する手法を提示できた。推論モデルとしては精度に課題があるが、研究者自身による研究初期のスクリーニングに活用可能と考える。検索ログ解析により有効な検索パターンを収集し専門家の暗黙知を獲得する仕組みに発展させることでデータ活用の好循環が期待できる。

今後は時系列変化を反映した特徴量データの創出を探索し、臨床専門家の臨床的クエリに答えられるモデルへと成長させることで臨床研究に貢献していきたい。

5. 結語

臨床症例レジストリデータベースからベイジアンネットワークによる患者状態確率モデルを構築し、臨床レジストリの可視化を実現し、データベース利活用の一助となりうることを示した。

参考文献

- [1] 国立研究開発法人国立国際医療研究センター: 一般社団法人日本糖尿病学会. 診療録直結型全国糖尿病データベース事業[<http://jdreams.jp/> (cited 2018-Aug-21)], 2016