

患者群の状態把握を目的とした臨床データ可視化の取り組み

Attempt to visualize clinical data for grasping status of patients

石井雅通
Masamichi Ishii

美代賢吾*2
Kengo Miyo

*1 NCGM(国立国際医療研究センター)
National Center for Global Health and Medicine

This paper describes an attempt to visualize clinical data collected from multi-institutional EHR System.

1. はじめに

多施設の医療機関から患者症例レジストリとして電子カルテに蓄積された臨床データを集積して臨床データベースを構築する事業が複数始まっている。当センターでは日本糖尿病学会と連携した「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)」を平成26年度から開始しており、2017年12月現在、35施設が参加し、今後4年間で100施設、10~20万人規模での登録をめざしている。

こうした患者症状レジストリの構築では臨床医師が当該症例の観察項目を入力することが一般的に実施されており、収集されるデータ項目は専門家による正解ラベルとして取り扱い可能な項目及び既知の臨床知識から判断した重要な特徴量となる項目で構成されている。こうした医師のアセスメント結果を活用し、電子カルテに記録された患者基本情報、検査結果情報、投薬情報などの構造化データから抽出した特徴量と併せて分析することで、観察する臨床データ項目の標準化による診療の質の改善及び合併症進展リスク因子の検索、合併症進展抑制効果の期待される介入の同定、糖尿病薬の副作用についての発生頻度及びリスク因子の確認、症例の未解決課題の発見といった成果が期待されている。

2. 課題

2.1 臨床データに内在する新 NP 問題とスパースネス

集積した臨床データを解析するにあたっては多施設臨床データの特徴に制約を受ける。まずデータ件数(n)に対してパラメータ数(p)が多い($n < p$)という新 NP 問題があり、かつ医療データがスパースネスであることである。

300床以上の大規模病院であれば、採用医薬品は2,000~3,000品目、検査結果値は3,000~5,000項目、検査値以外の観察項目等はすの数倍あり、データ項目としては10000を超える。今後のゲノム医療の伸展により患者ごとの遺伝子変異がパラメータに加わると数万規模となる。

一方、患者の疾患は多様でありため特定の疾病に対する分析を目的とする場合、対象となる患者(すなわちデータ件数)が稀少疾患であれば一施設あたり数名から数十名であり、一般に

析にあたってはがん種やステージごとに患者群を絞り込む必要のため新 NP 問題からは逃れられない。

2.2 患者群の状態把握のために

臨床現場においては専門医が相対する患者の状態を把握して必要な医療を提供する判断をしているが、多施設から収集した臨床データ全体を取り扱う場合は個別の患者の状態をすべて把握して解釈するアプローチは困難である。

そこで、各症例の専門家による探索的な患者群の傾向分析を可能とすることで患者群の状態把握が可能となり、新たなクニカルクエスションの選定に貢献できると考えた。

3. 提案手法

3.1 専門家による正解ラベル選択

症例レジストリで収集しているデータ項目から専門家により患者状態を表現したデータ項目を選択する。

3.2 構造化データ抽出

正解ラベル以外の構造化データ(患者基本情報、処方情報、検査結果情報)を抽出する。

3.3 患者症状単語抽出

診療録等の非構造化データから患者症状を表現する用語を抽出する。ここでは専門用語辞書として万病辞書を使用する。

3.4 患者症状単語統計

前記抽出結果より頻度統計を作成する。

3.5 時系列データ処理

3.2及び3.4の抽出データを入力として、症例レジストリ収集時点を起点として時系列情報を考慮した特徴量データを作成する。例えば、患者ごとの前回検査値からの変化量、次回検査値の変化量等を算出する。

3.6 ベイジアンネットワーク構築

3.1及び3.5の出力データを入力として、ベイジアンネットワークを構築する。

連絡先: 石井雅通, 国立国際医

母数が大きいと考えられる“がん”についても、研究のための解

3.7 専門家による主成分探索

3.6 で構築したベジアンネットワークを利用して当該症例の専門家による探索的な患者群の傾向分析を試行する。

3.8 クリニカルクエスチョンの検討

3.7 による合併症進展リスク因子、合併症進展抑制効果の介入、医薬品の副作用についてのリスク因子等の観点において新たなクリニカルクエスチョン創出の可能性について評価する。

4. 今後の計画

プロトタイプによる評価実験を計画しており、提案手法で述べた専門医によるレジストリへの入力データを正解ラベルとして活用する手法により、専門家以外が専門家の協力を得ることが容易となり、ベジアンネットワークによる可視化により専門家の気づきや暗黙知といったフィードバックをモデルへ反映することが容易となる好循環が期待できる。

ベジアンネットワークを用いた感度分析を実施することでデータ収集事業が目的としている、(1)合併症進展リスク因子の検索、(2)合併症進展抑制効果の期待される介入の同定、(3)糖尿病薬の副作用についてのリスク因子の確認、(4)糖尿病の未解決課題の発見、への貢献が期待できる。

上記プロセスで患者状態を表現する代表的な特徴量を抽出し、患者状態モデルを構成する要素として特定していくスパイラルを継続することで既存の収集対象のデータ項目の解析での貢献度を元に精査することが期待できる。入力データの評価は効率的なデータ収集項目への見直しの機会となり、データ入力元である臨床医の負担軽減の一助となすと考える。

謝辞:この研究の一部は AMED の課題番号 JP17lk1010010 の支援を受けた

参考文献

- [J-DREAMS] 診療録直結型全国糖尿病データベース事業,
<http://jdreams.jp/>
- [PMDA] MID-NET (Medical Information Database Network),
<https://www.pmda.go.jp/safety/mid-net/0001.html>
- [田中博 15] 人工知能 (AI) とビッグデータの 医療と創薬への
 応用, <http://ngbrc.com/img/file54.pdf>, 2013
- [統計関連学会連合理事会および有志 07] 統計関連学会連
 合理事会および有志, 我が国の統計科学振興への提言,
<http://www.jfssa.jp/TokeiKagakuShinkou0702.pdf>, 2007
- [矢野 17] 矢野憲, 若宮翔子, 荒牧英治, 医療テキスト解析のため
 の事実性判定と融合した病名表現認識器, 言語処理学会
 第 23 回年次大会, 2017
- [荒牧 17] 荒牧英治, 岡久, 矢野ほか, 大規模医療コーパス開発
 に向けて, 言語処理学会 第 23 回年次大会 発表論文集,
 2017

・患者状態のモデル化

