

令和 4 年厚生労働科学研究費 補助金
政策科学総合研究事業(臨床研究等 ICT 基盤構築・人工知能実装研究事業)
(分担)研究年度終了報告書

千年カルテの匿名加工医療情報を利用した
他医療機関の電子カルテに対するシーケンス解析に関する研究

研究分担者 横田 治夫 東京工業大学

複数医療機関における頻出医療指示パターンを比較し、医療機関間の医療指示パターンの共通点や相違点を明らかにすることで、自医療機関の特徴を確認し、他医療機関の治療パターンを参照した医療行為の改善等に繋げることが期待される。しかし、比較対象の医療機関数が多くなると共通部分が減り、医療機関の特徴を正確に把握することが難しくなる。そこで、パターンの共通部分を基にした頻出医療指示パターン間の距離を定義し、クラスタリングによって対象となる医療機関群をいくつかクラスタに分類し、そのクラスタ内で比較を行うことで特徴を把握しやすくする手法を提案する。本年度はオプトアウトされている 20 医療機関のデータに適用し、17 医療機関から頻出医療指示パターンを抽出できた。来年度はこれらのパターンに対し医療機関間の頻出医療指示パターンの共通点・差異点を示し、評価を行う。

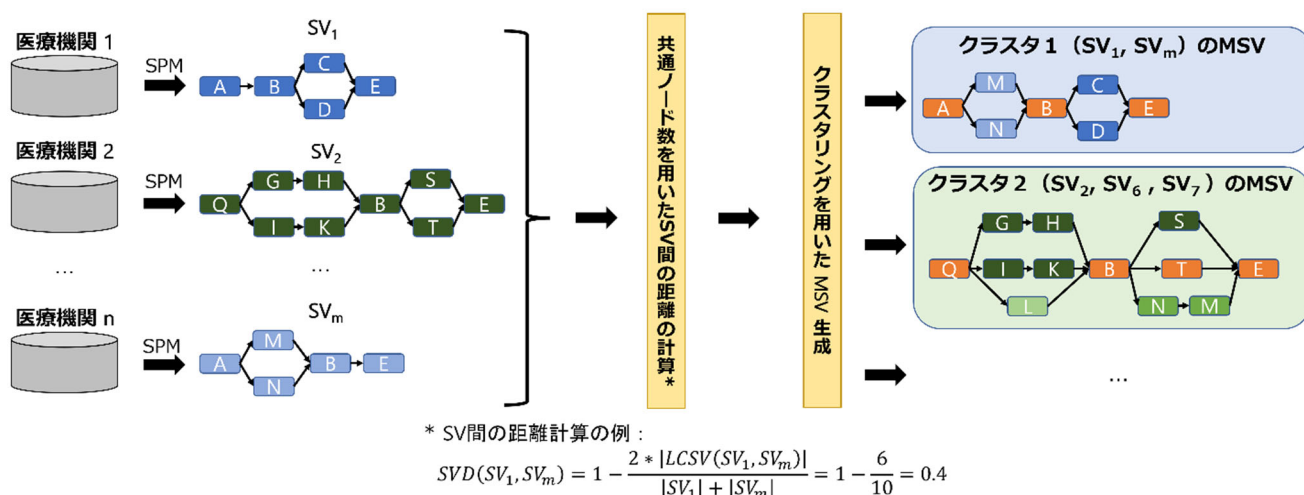


図 1: 複数医療機関の電子カルテに対する頻出医療指示パターンの解析

A. 研究目的

複数医療機関における頻出医療指示パターンを比較し、医療機関間の医療指示パターンの共通点や相違点を明らかにすることで、自医療機関の特徴を確認し、他医療機関の治療パターンを参照した医療行為の改善等に繋げることが期待される。しかし、比較対象の医療機関数が多くなると共通部分が減り、医療機関の特徴を正確に把握することが難しくなる。そこで、パターンの共通部分を基にした頻出医療指示パターン間の距離を定義し、クラスタリングによって対象となる医療機関群をいくつかクラスタに分類し、そのクラスタ内で比較を行うことで特徴を把握しやすくする。

B. 研究方法

倫理面への配慮

本研究は、千年カルテの利用目的等審査委員会¹⁾による審査を受け、承認された上で研究を進めた。

1. 背景

これまで電子カルテ中の医療指示のシーケンス解析を行い、検体検査項目をクラスタリングして検査タイプを抽出し、検査タイプと検査結果のシーケンスから次の検査項目を推薦する手法を開発してきた²⁾。また、医療オプションとなるバリエーションの安全性、効率性の評価や分岐理由の推定の提案と評価を行った³⁾。さらに、2 医療機関の医療指示のパターンの共通点・差異点

を見せる手法も提案した⁴⁾。これらの研究の知見を活かし、ライフデータイニシアティブ (LDI) が収集した複数医療機関における医療指示を解析する手法の研究を行う。

2. 開発プロセス

千年カルテの本研究用コンピューター（以下、千年カルテ環境と表記）に、Docker イメージを持ち込んでコンテナ化し、LDI と連携して試行錯誤しながら開発したプログラムを改修/・開発を行った。千年カルテ環境における開発では、以下の手順を繰り返して進めた。

- 1, プログラムと実行手順を LDI に提供
 - 2, LDI によるプログラム実行とエラー等のフィードバック
 - 3, プログラムを改修し LDI に提供
 - 4, LDI 担当者がデータセンターに入室してプログラムを実行し、データ統計やログを取得
- 最終的に開発したプログラムの評価結果を LDI よりセキュリティ便で受領した。

3. 対象データ

千年カルテに登録された 20 の医療機関の 2015～2021 年度のデータを使用した。また本研究では入院中に経皮冠動脈インターベンション (PCI) を受けた患者を対象とした。

4. 提案手法

図 1 に示すように、まず各医療機関の医療指示データを対象に、シーケンシャルパターンマイニング (SPM) を行い、頻出医療指示パターンの抽出を行う。ここで頻出医療指示パターンをシーケンスバリエーション (SV) として表し、SV 間の共通部分である最長共通サブシーケンスバリエーション (Longest Common Subsequence Variant, LCSV) を導出する。そして、医療機関間の頻出医療指示パターンの差異を表す距離を LCSV のノード数を用いて定義する。具体的に、2 つの SV の間には LCSV のノード数が多いほど、その SV らの距離が短くなるように計算する (図 1 の計算例参照)。

これを用いて階層的クラスタリングを行うことで医療機関群をいくつかのグループに分類する。さらに、グループ内で LCSV を算出し、併合シーケンスバリエーション (Merged Sequence Variant, MSV) を作成することで、医療機関の頻出医療指示パターンの共通点 (桃色) と差異点を可視化する。

C. 研究結果

本年度はオプトアウトされている 20 医療機関のデータに SPM を適用し、17 医療機関からそれぞれの頻出医療指示パターンを抽出することができた。来年度はこれらの頻出医療指示パターンに対して、SV の距離に基づいたクラスタリングを行い、医療機関間の頻出医療指示パターンの共通点と差異点を示す MSV を可視化し、評価を行う。

D. 健康危険情報

総括研究年度終了報告書にまとめて記入する。

E. 研究発表

本年度は、学会発表を行わなかった。

F. 知的財産権の出願・登録状況

本年度は、知的財産権の出願・登録を行わなかった。

参考文献

- 1) 利用目的等審査委員会の公表. 一般社団法人ライフデータイニシアティブ
[<https://www.ldi.or.jp/committee> (cited 2023-Apr-23)]
- 2) Le et al., Sequential Pattern Mining for Large Combinable Items with Values for a Set-of-items Recommendation, Proc. 34 IEEE International Symposium on Computer-based Medical Systems (CBMS2021), pp. 56-61, 2021.
- 3) Le et al., Methods for Analyzing Medical-order Sequence Variants in Sequential Pattern Mining for Electronic Medical Record, ACM Transaction on Computing for Healthcare, vol. 4, iss. 1, article No.;3, pp. 1-28, 2023.
- 4) Li et al., Comparison of Sequence Variants and the Application in Electronic Medical Records, Proc. 33rd International Conference on Database and Expert Systems Applications (DEXA 2022), pp. 117-130, vol. 13427, 2022.