

複数医療機関間の頻出医療指示パターン比較手法

横田 治夫^{*1}, Le Hieu Hanh^{*1}, 李 玉清^{*1}, 松尾 亮輔^{*2}, 山崎 友義^{*3}, 荒木 賢二^{*3}

^{*1} 東京工業大学 情報理工学院,

^{*2} 一般社団法人ライフデータイニシアティブ, ^{*3} 宮崎大学 医学部附属病院

A Method for Comparing Frequent Medical Order Sequential Patterns between Multiple Medical Institutes

Haruo Yokota^{*1}, Le Hieu Hanh^{*1}, Yuqing Li^{*1},

Ryosuke Matsuo^{*2}, Tomoyoshi Yamazaki^{*3}, Kenji Araki^{*3}

^{*1} School of Computing, Tokyo Institute of Technology,

^{*2} Life Data Initiative, ^{*3} University of Miyazaki Hospital

抄録: 医療ビッグデータの二次利用として, 複数医療機関の電子カルテに含まれる施術毎の医療指示の頻出シーケンスのパターンを比較することで, 個々の医療機関における施術の実施内容の違いを認識し, それを基に必要に応じて医療の改善の足掛かりとなりうる情報の提供を目指す. そのため, 最長共通サブシーケンスバリエーション(LCSV)と, LCSVに基づくラベル付き併合シーケンスバリエーション(MSV)の概念を提案するとともに, それらを求めるアルゴリズムを提案する. まず, 複数医療機関の電子カルテのデータ中で対象とする施術に対する医療指示の頻出シーケンスのバリエーションを求める. 次に, その間のLCSVを抽出し, 求めたLCSVに基づきラベル付きMSVを抽出することで, 医療機関間の差異を示す. 実際に, 宮崎大学医学部附属病院と宮崎市郡医師会病院の実電子カルテデータに対して提案手法を適用し, 医療機関間の差異となる部分を抽出した.

キーワード 医療指示, シーケンシャルパターンマイニング, バリエーション, 医療機関比較, 医療支援

1. はじめに

医療現場の情報化が進み, 医療に関する様々なデータが医療ビッグデータとして大量に蓄積され, その二次利用として, 多様な解析が行われるようになってきている[1]. 医療データ解析の法整備として次世代医療基盤法も施行され, その下で多数の医療機関の電子カルテを集積する千年カルテプロジェクト[2] や, レセプト情報等を集積するNDB [3] が推進されている. NDBを基に地域毎の疾病発生の統計的解析も行われている[3]が, 個々の医療機関間の比較はこれからの課題となっている.

これまで筆者らは, 電子カルテ中の医療指示にシーケンシャルパターンマイニングを適用して頻出パターンを抽出し, 医療オプションでもある医療指示シーケンスのバリエーションの抽出と可視化[4], 分岐発生要因解析やバリエーションの安全性や効率性の評価[5], 要因に基づく次の医療指示候補提示[6]等の手法を提案し, 実際の電子カルテデータを用いて評価を行ってきた. しかし, それらは一つの医療機関を対象にしたもので, 複数医療機関間の比較はできていなかった.

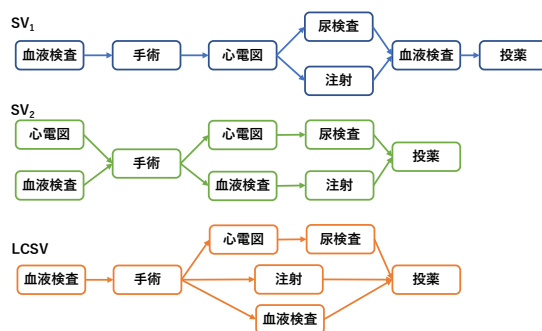


図1 LCSV の例

2. 比較のための手法

1) シーケンスバリエーションの共通部分の抽出

比較のため, 医療機関毎の頻出のシーケンスバリエーション($SV_{1,2}$)を求め, 共通部分を抽出する. 従来の最長共通部分列(LCS)ではバリエーションが扱えないため, 最長共通サブシーケンスバリエーション(LCSV)の概念とそれを求めるアルゴリズム(紙面の関係で割愛)を提案する. 例を図1に示す.

2) 差異を示す方法

LCSV は比較対象の医療指示シーケンスの共通部分を示すが, 医療機関の間の差異は示していないため, 併合シーケンスバリエーション(MSV)の概念とそれを求めるアルゴリズム(紙面の関係で割

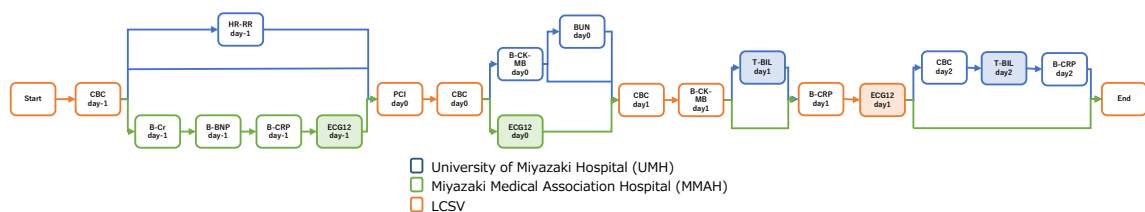


図3 UMH(438件, MiniSup=0.15)とMMAH(4,265件, MiniSup=0.25)のPCIに対するMSV



図2 SV_1 , SV_2 に対するMSV

愛)を提案する. MSV は LCSV を基に, 比較対象である SV_1 と SV_2 中に含まれるすべての頻出シーケンスを含み, 共通部分と SV_1 , SV_2 を示すラベルを持つ. ラベルを付けることで, 共通部分と, 医療機関に特有な部分を示すことができる. 図1の例に対する MSV を図2で, ラベルとして色を用い, SV_1 を青, SV_2 を緑, 共通部分をオレンジで示す. この例では, SV_1 での尿検査か注射の後の血液検査の頻出と, SV_2 での手術前の心電図の頻出が特有であること示している.

3. 実電子カルテへの適用

提案する LCSV および MSV を実際の複数医療機関の電子カルテのデータに適用を試みる. 最終的には, 前述の千年カルテプロジェクトで集積している電子カルテのデータを対象とすることを想定しているが, 現時点ではまだ準備段階であるため, 宮崎大学医学部附属病院(UMH)と宮崎市郡医師会病院(MMAH)の電子カルテを用いた. 両病院で行われたカテーテルを用いた狭心症・心筋梗塞治療法の PCI (Percutaneous Coronary Intervention) (UMH: 438 件, MMAH: 4,265 件)を対象に MSV を求め, 比較を行った. 結果の一部を図 3 に示す. 12 誘導心電図(ECG12)を入院中毎日実施するか 1 回しか実施しないかの差異が抽出できた. 術後の患者モニターによる心電情報の扱いの違いと思われる. 他, 総ビリルビン(T-BIL)の検査の頻度にも差が出たが, それらによる診療報酬点数の差は 0.5%以下であった.

4. おわりに

複数の医療機関の電子カルテに含まれる医療指示のシーケンスの頻出パターンを比較すること

で, 個々の医療機関で実施内容の違いを認識し, 場合によっては改善の足掛かりとなる情報提供を目指し, 医療機関毎の医療指示シーケンスのバリエーションを求め, それらの間の LCSV から MSV を抽出する手法を提案した. 実際に複数の医療機関の電子カルテのデータに適用し, 違いを示した.

参考文献

- [1] 横田治夫: 電子カルテデータ解析-医療支援のためのエビデンス・ベースド・アプローチ, 共立出版, 2022 年 3 月.
- [2] 千年カルテプロジェクト概要: <https://www.gehr.jp/about/index.html>.
- [3] 匿名レセプト情報・匿名特定健診等情報の提供に関するホームページ: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/iryuu-hoken/reseputo/
- [4] Y. Honda, M. Kushima, T. Yamazaki, K. Araki, and H. Yokota: Detection and Visualization of Variants in Typical Medical Treatment Sequences. In Proc. of the 3rd VLDB workshop on data management and analytics for medicine and healthcare, pp. 88–101, 2017.
- [5] H. H. Le, T. Yamada, Y. Honda, M. Kayahara, M. Kushima, K. Araki, and H. Yokota.: Analyzing Sequence Pattern Variants in Sequential Pattern Mining and Its Application to Electronic Medical Record Systems. In Proc. of the 30th Int'l Conf, on Database and Expert Systems Applications (DEXA '19), 393–408, 2019.
- [6] 坂本, 小林, Le, 松尾, 山崎, 荒木, 横田: 頻度と実施時刻によるグループ化を採り入れたシーケンス解析に基づく医療指示推薦. DEIM Forum 2021, pp. B6-5, 2020.