

COVID-19に関する頻出医療指示パターンの時期による差異と 差異発生時期の可視化

趙 子泰[†] Le Hieu Hanh[†] 松尾 亮輔^{††} 山崎 友義[†] 荒木 賢二^{†††}
横田 治夫[†]

[†] 東京工業大学 情報理工学院 〒152-8552 東京都目黒区大岡山 2-12-1

^{††} 一般社団法人ライフデータイニシアティブ 〒606-0805 京都市左京区下鴨森本町 15

^{†††} 宮崎大学 医学部附属病院 医療情報部 〒889-1692 宮崎県宮崎市清武町木原 5200

E-mail: [†]{kotai,hanh,hlh}@de.cs.titech.ac.jp, [†]yokota@cs.titech.ac.jp,

^{††}matsuo@ldi.or.jp,

^{†††}{yamazaki.cp,araki6925}@gmail.com

あらまし 電子カルテの普及に伴い、蓄積された医療情報を解析することによる医療支援が期待されている。例としては、電子カルテデータから頻出医療指示パターンを抽出する手法が提案されている。本研究では、疾病の異なる時期の医療指示パターンの差異を認識するために、最長共通サブシーケンスバリエーション (LCSV) と、それを用いた併合シーケンスバリエーション (MSV) を COVID-19 感染波の第一波から第五波までの医療指示シーケンスに適用し、時期毎の医療指示パターンの差異を可視化した。また、感染波間の医療指示パターンの差異発生時期を把握するために、スライディングウィンドウを実装し、ウィンドウサイズとスライド単位を調整することで、時期的に異なる特徴を持つ頻出パターンを抽出し、頻出パターン間の類似度を評価し、差異発生時期を可視化した。

キーワード 電子カルテ, 医療支援, シーケンシャルパターンマイニング, シーケンスバリエーション, 異なる時期の医療指示の比較, COVID-19

1 序論

1.1 研究背景

近年、紙のカルテに代わり電子カルテの普及に伴い、今後さらに国内外の電子カルテの普及率が増加していくことが予想される。これに伴い、蓄積された医療情報の二次利用が期待されている [1]。また、国として医療情報の二次利用の必要性が認識され、電子カルテなどの医療・健康に関する記録を全国規模で一元的に集める「千年カルテプロジェクト」[2] が始まっている。二次利用の例としては、特定の病気の患者に対して疾病毎の典型的な医療指示の流れであるクリニカルパスの作成や改善支援のため、電子カルテデータから頻出の医療指示パターンを抽出する手法が提案されている。従来、クリニカルパスの作成は医療関係者自身の医学的経験に基づいて行われており、人力でクリニカルパスを収集・分析して改善するのは容易ではなかった。そのため、計算機によって電子カルテをデータ工学の観点から電子カルテにを解析することで、医療行為改善の支援を行う研究が進められている。

一方、COVID-19 の感染拡大に伴い、変異株の出現や、新たな治療方法・薬剤の開発が行われることにより、時期による医療指示パターンが変わっている。時期による医療指示パターンの差異とその差異の発生時期を把握することは、今後の感染症対策において有用であり、医療の改善に貢献する可能性がある。

既に COVID-19 の治療に関する電子カルテデータは蓄積されており、それらを解析することにより、時期による治療パターンの違いと違いの発生時期を示すことが医療支援につながることを望まれる。

1.2 研究の目的

本稿では、異なる時期の COVID-19 に対する医療指示シーケンスの違い及び違いの発生時期を可視化することで、変異株や治療方法、利用薬剤の違いの医療指示シーケンスに及ぼす影響を把握し、医療支援を行うことを目指す。ここで示す異なる時期の医療指示シーケンス解析のアプローチの対象は、COVID-19 に限らず、新たな感染症にも適用可能である。

実際に入院し、治療が受けられた COVID-19 の患者の電子カルテデータを用い、複数の医療機関における第一波から第五波までの頻出医療指示シーケンスをそれぞれ抽出し、時期による違い及び違いの発生時期を可視化する。可視化のアプローチとして、異なる医療機関でのバリエーションを含む頻出医療指示シーケンスの違いを示すために提案している [14]、最長共通サブシーケンスバリエーション (LCSV) と、それを用いた併合シーケンスバリエーション (MSV) を、時期の異なる COVID-19 の電子カルテデータに適用し、時期による頻出の医療指示パターンの違いを示す。また、スライディングウィンドウのアルゴリズムを実装し、ウィンドウサイズとスライド単位を調整することで、時期的に異なる特徴を持つ頻出パターンを抽出し、頻出パ

ターン間の類似度を LSCV を用いて評価し,MSV を用いて差異発生時期を示す. MSV による医療機関間での頻出医療指示パターンの差異の提示は我々の研究グループが提案する独自の手法で, その手法を医療機関内での同一の疾病に対する異なる時期の間の差異及び差異発生時期に適用する本研究でのアプローチも, 他では行われていない新たな試みである.

1.3 本稿の構成

本稿は以下の通り構成される. 2 節では本研究の関連研究について説明する. 3 節では提案手法である異なる時期による治療パターンの差異と差異発生時期の可視化アプローチについて述べる.4 節では, 3 節の手法を用いて実際の医療指示データを解析することと異なる時期による治療パターンと差異発生時期の MSV の作成を行う. 5 節では, 提案手法に関する考察を述べる. 最後に 6 節でまとめと今後の課題について述べる.

2 関連研究

本節では, 本研究に関連するシーケンシャルバリエーション (SV), シーケンシャルパターンマイニング (SPM), LCSV と MSV について説明する.

2.1 SPM

Agrawal らによって提案 [3] されたシーケンシャルパターンマイニング (SPM) は, シーケンスデータベース (SDB) から頻出するパターンを抽出する手法であり, 医療,E コマース, インターネットなどの領域で注目されている [1]. アイテムの順列をシーケンスと呼び, SDB は あるシーケンス集合に属するシーケンスと, シーケンスの識別子であるシーケンス ID を組みとする要素からなる. SPM では, 設定された最小支持度 (minsup) よりも出現頻度が高いシーケンスを頻出パターンとして抽出する. 例えば,minsup が 0.2 の場合, 全シーケンスのうち 20% 以上のシーケンスに含まれているパターンが抽出する.minsup を下げるとより多くのパターンが抽出されるが, 下げすぎると有意義な情報が多数のパターンに埋もれてしまう可能性がある.SPM を抽出するアルゴリズムとしては,Apriori [4],PrefixSpan [10],CSpan [8] などが提案されている.

2.1.1 SPM を用いた医療情報の解析

SPM は, 医療情報の特に医療指示シーケンスに適用される研究で多く取り上げられている. 佐々木らは, 実際の電子カルテデータにタイムインターバル SPM [5] を適用し, 時間間隔を考慮した頻出する医療指示パターンを抽出した. Le らは, 効率的な SPM である CSpan を拡張し, タイムインターバルの統計情報を収集する T-CSPan を提案し, 実際の電子カルテデータに適用した [9]. また, 医療指示パターンの抽出方法に加えて, 電子カルテ解析におけるプライバシー保護に関して,Le らは差分プライバシーの概念を適用する手法を提案している [6].

2.2 SV の解析

上記の SPM で抽出された頻出医療指示パターンには, パターンの一部が異なるバリエーションが存在することがある. このよう

なバリエーションは, シーケンスバリエーション (SV) と呼ばれ, 患者の状態等によって生じる医療に対応する. 例えば, 医療指示列において分岐が発生すると, シーケンスバリエーションが生じる. 例えば, (入院 → 手術 → 検体検査 → 退院) と (入院 → 看護タスク → 手術 → 処置 → 退院) のようなシーケンスの中で, 検体検査と処置がバリエーションと呼ばれる. このように, 複数のシーケンスをまとめてシーケンスバリエーションと考えることができる.SV を解析結果として提示することは, 医療支援において重要であるとされている.

本田らは, 実際の電子カルテから医療指示の SV を抽出し, それらをグラフで表すことで, 医療従事者向けに視認性の高い可視化ツールを提供した [7]. また, 本田らは,SV に対して, 患者の性別や年齢などの静的情報と, 血圧や検査結果などの動的情報に対する多変量解析を行うことで, バリエーションにおける分岐の発生要因を分析する手法を提案し, 実際の電子カルテを用いて効果を示した [11].

一方, 山田らは,SPM を適用する際に, シーケンスの識別子であるシーケンス ID (SID) を保持しつつ, 頻出シーケンスがどのシーケンスと対応関係にあるかの情報を取得することで, 頻出シーケンスの抽出と同時に SID のセットを得ることができる. そのため,SV の頻出シーケンスごとの安全性や効率性を指標化し,SV の評価がスムーズかつ正確に行えるような可視化手法を提案し, 実際の電子カルテを用いて推薦結果の評価を行った [6].

坂本らは, 分岐先頻度と実施時刻情報を用いたグループ化による適正分岐候補の選択を行い, 頻出シーケンスの分岐先の適正候補の選択による併合シーケンスの生成を行った. また, 分岐要因である患者情報の拡充や, 併合シーケンスの分岐への要因適用による医療指示推薦を提案した [12].

2.3 複数医療機関の SV の比較

それまでの解析は, 単一の医療機関を対象としていたが, Li らは, 異なる医療機関において異なるシーケンスバリエーションを含む頻出の医療指示シーケンスの違いを示すために, 二つの SV の最長共通サブシーケンスバリエーション (LCSV) を抽出し, それを使用した併合シーケンスバリエーション (MSV) の概念と導出アルゴリズムを提案した. さらに, 実際の 2 つの医療機関の電子カルテデータにこれらを適用することで, 医療機関間の治療パターンの違いを示した.

LCSV は, 従来の最長共通サブシーケンス (LCS) では SV を扱えないために,SV 間で最も長い共通のシーケンスバリエーションを導出するものである. さらに,LCSV を使用して SV のシーケンスをラベル (色) 付きで含む MSV を導出することで, 医療機関間の SV の違いを可視化することができる.LCSV および MSV の正確な定義と導出アルゴリズムについては, 参考文献 [14] を参照されたい.

3 提案手法

本節では, 関連研究の異なる医療機関間の差異を示すために提案された LCSV と MSV を医療機関内の同一の疾病に対し

て、異なる時期に行われた医療指示間の差異の可視化、スライディングウィンドウを適用した差異発生時期の可視化について提案手法の概要を説明する。

3.1 医療従事者の情報に基づくデータの事前処理による適正化

医療従事者の情報に基づいて前処理によって以下の医療指示を除外する

- 医学的に有用ではない医療指示を除外する
「調剤料計算」、「食事療養」など治療行為とは関係が薄い医療指示を除外する

- 治療中に高頻度行われる医療指示を除外する
出現頻度が高く、シーケンス中の出現回数が多い医療指示がある場合、重要な医療指示が隠されて出てこない可能性がある(例: 末梢血液一般検査、経皮的動脈血酸素飽和度測定)。そのため、出現頻度が高いかつシーケンス中の出現回数が多い医療指示を除外する必要がある。 TO/S (term occurrence per sequence)はある医療指示がシーケンスに出現する平均回数を表せる指標であり、これを用いて除外する医療指示を決定する [15]。

$$TO/S = \frac{\text{その医療指示の総出現回数}}{\text{シーケンスの総数}}$$

3.2 最小支持度の設定

有意な LCSV と MSV を得るためには、対象期間に対する最小支持度 (minsup) をどのように設定するかが問題となる。適切な minsup の選択は重要な課題であるが、まずは minsup を変化させながら抽出された SV を観察し、最も有意と判断できる SV が得られる minsup を用いることにする。

3.3 感染波毎の頻出医療指示パターンの抽出

頻出医療指示パターンの時期による差異を示すために、まず、各医療機関の対象疾病に対する異なる期間の電子カルテデータのセットに対して SPM を適用し、期間毎の SV を導出する。次に、期間間の差異を示すために、それらの SV 間の共通部分である LCSV を抽出する。さらに、求めた LCSV と元の SV を用いて期間毎のラベル (色) の付いた MSV を導出する。

3.4 スライディングウィンドウを適用した時期的な SV の抽出

スライディングウィンドウを移動させる際にウィンドウ間の比較について説明する。

時期的に特徴を持つ頻出パターンを抽出するために、本論文ではスライディングウィンドウを実装する。ここで、適切なウィンドウの種類、ウィンドウサイズ、スライド単位をどう設定するかが問題である。

本論文では、時間間隔ウィンドウ、シーケンス数ウィンドウと時間・シーケンス数ウィンドウの3種類のスライディングウィンドウを提案する。

3.4.1 時間間隔ウィンドウ

ウィンドウのサイズをある時間範囲に設定し、その時間範囲内のシーケンスに SPM を適用し、頻出パターンを抽出する。その後、設定したスライド単位でウィンドウを移動させ、新しいウィンドウ内のデータに対して再度 SPM を適用する。しかし、

時間間隔ウィンドウでは、シーケンス数が少ない期間に対して、SPM を適用しても何も抽出できない可能性がある。そのため、本論文では、シーケンス数ウィンドウを提案する。

3.4.2 シーケンス数ウィンドウ

データベースのシーケンスデータを時間順でソートし、ウィンドウサイズ分のシーケンスに SPM を適用し、頻出パターンを抽出する。

感染波の中でピークとなる時期の前後のシーケンスは、最も時期的な特徴を有していると考えられる。これらのシーケンスを用いることで、感染波の中でピークとなる時期における特徴的なパターンを把握することができる。そのため、時期的に異なる特徴を持つ頻出パターンを抽出するためには、感染波の中でピークとなる時期の前後のデータから標準偏差 σ を算出し、片側のデータの約 95 % を含む 2 倍標準偏差を採用する。その後、ピークの幅となる 4 倍の標準偏差をその時期のウィンドウサイズにすることで、その時期の特徴を持つ頻出パターンを抽出することができると考えられる。ここで、ウィンドウサイズは整数でなければならないため、標準偏差から算出したウィンドウサイズの値は、最も近い整数に丸める。

また、各感染波の最初及び最後において、入院患者数が収束しているため、対応する期間のシーケンスデータも減少することになる。これにより、スライド単位が大きい場合、次の感染波の先頭のシーケンスをスキップすることになるため、先頭時期において時期的に異なる特徴を持つ医療指示パターンを抽出できない可能性がある。

そのため、本論文では、スライド単位を動的に変更するための時間・シーケンス数ウィンドウを提案する。

3.4.3 時間・シーケンス数ウィンドウ

時間・シーケンス数ウィンドウでは、シーケンス数ウィンドウをベースに、後続期間のシーケンス数が考慮する。以下、スライド単位を動的に変更するアプローチについて説明する。

(1) まず、デフォルトのスライド単位を設定する必要がある。スライド単位を初期化する方法はいろいろあるが、本稿では評価実験として、ウィンドウサイズの半分に設定する。ここで、計算されたスライド単位が整数でない場合は、最も近い整数に丸めることになる。

(2) ウィンドウの先頭から設定された後続時期範囲のシーケンス数を調べる

(3) シーケンス数がデフォルトのスライド単位より小さい場合のみ、スライド単位をシーケンス数と同じように取る。

(4) ウィンドウをスライド単位分で移動し、新たなウィンドウ内のデータに再度 SPM を適用する。

3.5 ウィンドウ間の頻出パターンの比較

隣接のウィンドウを一つずつ比較することで、治療パターンが切り替わる時期を検出可能となる。LCSV は、LCS の拡張であり、LCS の間の比較方法を同様に拡張し、LCSV の比較が可能になる。これにより、以下の SVS (Sequence Variant Similarity) の式で、LCSV 間の比較を行うことができるようになる。

$$SVS(SV1, SV2) = \frac{2|LCSV|}{|SV1| + |SV2|}, (0 \leq SVS \leq 1)$$

ここで、 $|LCSV|$ 、 $|SV1|$ と $|SV2|$ は、それぞれ LCSV のノード数、 $SV1$ のノード数と $SV2$ のノード数である。

SVS は、2つのグループ間の治療パターンの類似度を示す指標である。 SVS 値が大きい場合、2つのグループの治療パターンはより類似しており、逆に SVS 値が小さい場合、2つのグループの治療パターンは異なっている。

4 実 験

本節では、提案手法の有効性を確認するため、COVID-19の第一波から第五波までの実電子カルテデータを用いて、それぞれの感染波の間の差異及び差異発生時期を MSV として示す実験を行う。なお、感染波の間の差異に関しては、既に第 42 回医療情報学連合大会にて発表を行っているため、詳しくはそちらを参照されたい [13]。

4.1 対象データ

今回の実験では、「COVID-19 感染の診療への影響調査と予測モデル開発」に関する研究の研究協力機関である 2 医療機関 (医療機関 A、B) から提供された、COVID-19 に対する実電子カルテデータ中の医療指示データを対象とした。本実験では、両医療機関の電子カルテデータの中で COVID-19 を特定するために、ICD-10 コードを用いている。COVID-19 の場合は、ICD-10 コードが U071 の患者のシーケンスを抽出する。また、COVID-19 の各波に対応する入院期間によってデータを分ける。医療機関 A と医療機関 B の第一波から第五波までの各波における患者数とシーケンス数を表 1 に示す。同一患者が複数のシーケンスを構成することがあるため、シーケンス数が患者数より増えている。

表 1 COVID-19 の波毎の患者数とシーケンス数

	医療機関 A		医療機関 B	
	患者	Seq	患者	Seq
第一波 (2020.04.01-2021.06.30)	42	49	10	10
第二波 (2020.07.01-2020.10.31)	52	63	35	35
第三波 (2020.11.01-2021.02.28)	96	108	55	57
第四波 (2021.03.01-2021.06.30)	89	99	24	24
第五波 (2021.07.01-2021.09.30)	98	104	35	35

表 2 本稿で使用する医療指示名称の略号と薬効

略号	医療指示名称	略号	医療指示名称
AT	アセトアミノフェン錠 (解熱鎮痛剤)	MT	メジコン錠 (咳止め剤)
BIL	ビリルビン検査 (血液検査)	MTT	メトクロプラミド錠 (消化器機能異常治療剤)
CBC	末梢血液一般検査 (血液検査)	OT	酸素吸入
CM	カモスタットメルシ 酸塩錠 (経口蛋白分解酵素阻害剤)	PCT	PCT 半定量 (血液検査)
CRP	C-リアクティブ・ プロテイン (血液検査)	PSL	プレドニゾロン錠 (ステロイド剤)
CTD	コンピューター断層診断	PT	プロトロンビン時間検査 (血液検査)
D-D	D-ダイマー検査 (血液検査)	RO	リスベリドン内用液 (抗精神病剤)
DT	デカドロン (ステロイド剤)	ST	センノシド錠 (緩下剤)
FRTN	フェリチン定量 (血液検査)	VWFa _g	VWF 抗原 (血液検査)
HBsAG	HBs 抗原 (血液検査)	XP	単純撮影
LT	ルネスタ錠 (不眠症治療薬)	LS	ロキソプロフェン ナトリウム (鎮痛・抗炎症・解熱剤)
TMR	ツムラ大建中 湯エキス (腹が冷えて痛み、 腹部膨満感のあるもの)	ST	ダイフェン (合成抗菌剤)
MM	マグミット錠 (制酸・緩下剤)	AS	バイアスピリン (抗血小板剤)
LL	リリカ OD 錠 (疼痛治療剤)	OB	オルベスコ (吸入ステロイド 喘息治療剤)
SM	酸化マグネシウム錠 (制酸・緩下剤)	IZ	イグザレルト錠 (選択的直接作用型 第 Xa 因子阻害剤)
AA	アセトアミノフェン錠 (解熱鎮痛剤)	CT	CT 撮影

4.2 実験環境と設定

個人情報保護のために、利用者を限定した実験環境として、宮崎大学医学部附属病院が管轄するクラウド環境を用いる。解析を行う東京工業大学のメンバーは、匿名化された電子カルテデータをクラウドから取り出して解析するのではなく、解析プログラムの Docker Image を作成して、その Docker Image をクラウド上で実行し、個別の患者情報を含まない頻出医療指示パターンのみ含む結果を取り出すことで、情報セキュリティとプライバシー保護を確保している。この研究は、宮崎大学の倫理委員会および東京工業大学の人を対象とする研究倫理審査委員会の承認を得ている。

3 の提案手法で述べたように、有意な SV を抽出するために

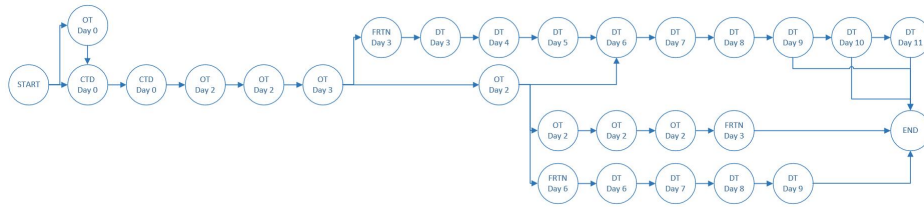


図 1 医療機関 A の第四波 (2021.03.01-2021.06.30) の SV (minsup=0.2)

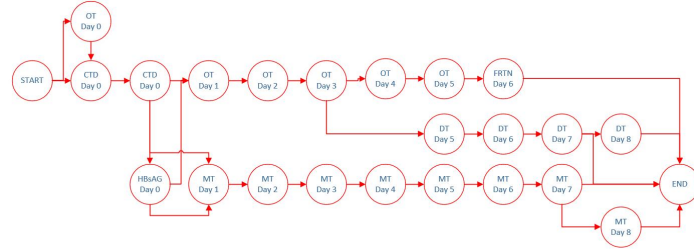


図 2 医療機関 A の第五波 (2021.07.01-2021.09.30) の SV (minsup=0.2)

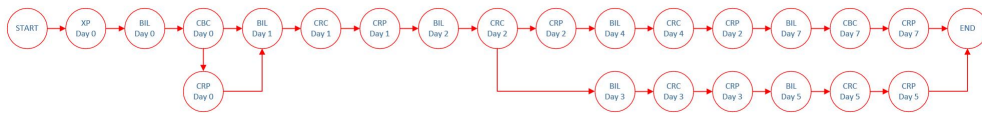


図 3 医療機関 A の第五波 (2021.07.01-2021.09.30) の SV (TO/S 値が高い指示除外せず, minsup=0.2)

治療に関係の薄い医療指示や出現頻度が高く、シーケンス中の出現回数が多い医療指示を除く必要がある。今回の実験では、医療関係者からのアドバイスのによって、医療機関 A で TO/S 値が 1.17 以上、医療機関 B で 1.04 以上の医療指示を除外した。minsup の設定に関しては、実際にいくつかの値に設定して SV を抽出して内容を確認した結果、感染波毎の頻出パターン抽出については、医療機関 A では 0.2 を、医療機関 B では 0.25 を採用することにした。差異発生時期を探すために適用したスライディングウィンド内の minsup については、医療機関 A と医療機関 B では両方 0.2 を採用することにした。実験結果の図に使用した医療指示の名称の略称と薬効を表 2 に示す。

4.3 実験結果

4.3.1 期間毎の SV の抽出

医療機関 A と医療機関 B の電子カルテデータを用い、表 1 に示す期間毎に COVID-19 の患者の医療指示データから TO/S 値が高い医療指示を除いた後、医療指示シーケンスに SPM を適用し、SV を求めた。ここで両医療機関の全ての期間の SV を示すのはページ数の関係で困難なことから、医療機関 A で minsup を 0.2 とした場合の第四波、第五波の SV を、それぞれ図 1、図 2 に示す。TO/S 値が高い医療指示の除外の効果を示すため、除外しなかった場合の第五波の SV を図 3 に示す。図 2 と図 3 を比較すると分かるように、除外しなかった場合には同じ minsup では、特徴的と思われる MT (メジコン錠) の投薬が SV に出現しないことが分かる。

4.3.2 異なる期間の LCSV と MSV の導出

医療機関 A の第四波と第五波の SV に対して、隣接している

第四波第五波間（黄色）の LCSV を図 4 に示す。さらに、導出した LCSV を基に、隣接しているそれぞれの期間の MSV を図 5 に示す。図 5 の MSV では、第四波第五波間の LCSV を黄色で、第四波を青色、第五波を赤色で示している。このように色でラベル付けし、LCSV とそれぞれの期間を対応する色でラベル付けすることで、隣接する期間の頻出医療指示パターンの違いを可視化することができる。

図 5 を詳細に見てみると、第 4 波では DT が 9 回 (Day3-Day11) 出現しているのに対し、第 5 波では DT が 4 回 (Day5-Day8) しか出現していないことがわかる。したがって、日本の第 5 波の DT は、第 4 波に比べて減少していることがわかる。医療関係者からのフィードバックによって、これは、高齢者へのワクチン効果と中和抗体薬の投与により、COVID-19 による重篤な肺障害症例が少なくなったことが原因として考えられる [18]。また、第 5 波では OT ありとなしの分岐があるが、第 4 波ではそのような分岐はない。医療関係者からのフィードバックによって、OT を必要とするバリエーションではステロイド剤 DT と経過観察を実施する必要がある。酸素吸入の必要のないとするバリエーションでは咳止め薬である MT を投薬するという頻出パターンとなっている。

4.3.3 差異発生時期の可視化

医療機関 A の第 1 波から第 5 波までのデータに時間間隔ウィンドウを適用し、遷移図は図 7 に示す。ウィンドウの大きさは一か月、スライド単位は一週に設定した。

シーケンス数ウィンドウでは、時期的に特徴を持つ頻出パターンを完全に抽出できないため、本論文ではシーケンス数ウ

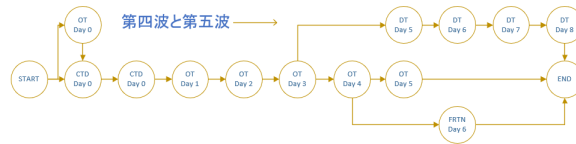


図 4 医療機関 A の第四波第五波間の LCSV

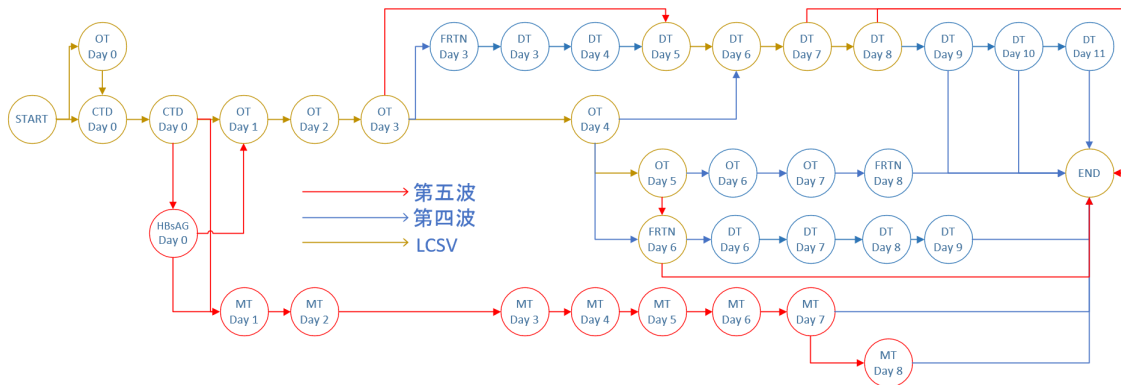


図 5 医療機関 A の第四波と第五波の MSV

インドウの実験結果は割愛する。

医療機関 A の第 1 波から第 5 波までのデータに時間・シーケンス数ウィンドウを適用し、遷移図は図 8 に示す。時間的特徴の異なる頻出パターンを抽出するために、ピークとなる時期の前後のシーケンス数の 4 倍標準偏差をウィンドウサイズとし、評価実験としてデフォルトスライド単位をウィンドウサイズの半分に設定した。

遷移図では、第一波は紫色で、第二波を水色、第三波を緑色、第四波を青色、第五波を赤色で示している。また、COVID-19 に関する診療手引きにより、全五波の感染人数のピークとなる時期及び第 1,2 回のワクチン接種開始日を図 8 に括弧と矢印を付けている。さらに、投薬の増減、OT の増減、検査のみのパターンへの変更及び投薬の変化を図に示している。

図 8 の時間間隔ウィンドウの遷移図を詳細に見てみると、ピーク 1 では CM が出ているが、ピーク 1 の時期を過ぎると、CM が使用されなくなったことがわかる。しかし、ピーク 2 の時期では CM を再びメイン投薬として投与し始めたことになった。第三波のピークの後においては、第一、二回目のワクチン接種開始日の後に、CM が出てこないことが分かる。その代わりとして、第四波からステロイド剤の DT が出てきたことが分かる。医療関係者からのフィードバックによって、第一波から第三波の時期に CM を使用する理由はカモスタットがコロナウイルスの治療に有効だとの報告 [20] があり、使用されていたと考えられる。しかし、2021 年 6 月に効果ない判定 [16] を受けていたため、投薬を停止した。一方、ステロイド剤の使用は、呼吸器学会で示されている大規模なランダム化比較試験での結果 [17] に基づいていると思われる。ピーク 4 とピーク 5 では SM と MT が出てきたと分かる。また、波の途中でも他の投薬パターンに切り替わることもあるため、同じ感染波内でもパターンが変わることがあると分かる。

4.3.4 医療機関間の差異

同じ期間であっても、医療機関によって頻出医療指示パターンが異なる。紙面の関係で全ての期間での比較を示すことはできないが、第三波における医療機関 A と医療機関 B の MSV を図 6 に示す。

図 6 の MSV では、LCSV のノード数が 2 つしかなく、他の医療指示がほぼ全部異なることがわかる。そのため、同じ第三波における両医療機関の頻出医療指示パターンが異なっていることが分かる。特に、医療機関 A では CM の使用に対して、医療機関 B では IZ が使用されていた。また、医療機関 A のみステロイド剤の DT を使用していた。一方、医療機関 B では消化器機能異常治療剤である MTT 投与の医療指示が頻出医療指示パターンとして出現していた。

差異発生時期の可視化について、医療機関 B ではシーケンス数が少ない時期が多いため、時間間隔ウィンドウを用いて SPM を適用すると何も抽出できない場合がある。そのため、医療機関 B では時間・シーケンス数ウィンドウだけ実験を行った。医療機関 B の時間・シーケンス数ウィンドウによって得た遷移図は図 9 に示す。

図 9 の遷移図により、医療機関 B が第二波から第五波まで PSL を使用していたことがわかる。医療関係者からのフィードバックによって、早期にプレドニゾロンの投与を開始した根拠として、診療の手引き 2 版 [19] を参考にしたと考えられる。また、PSL の長期使用については、医療機関 B の医師が PSL の使用に慣れている可能性がある。

4.3.5 時間間隔ウィンドウと時間・シーケンス数ウィンドウの違い

同じ感染波であっても、ウィンドウの種類によって遷移図が異なることがわかる。例えば、時間間隔ウィンドウの第一波と第二波では、時間・シーケンス数ウィンドウで出てこない投薬 TMR、MM、LL、OB が出ていることがわかる。時

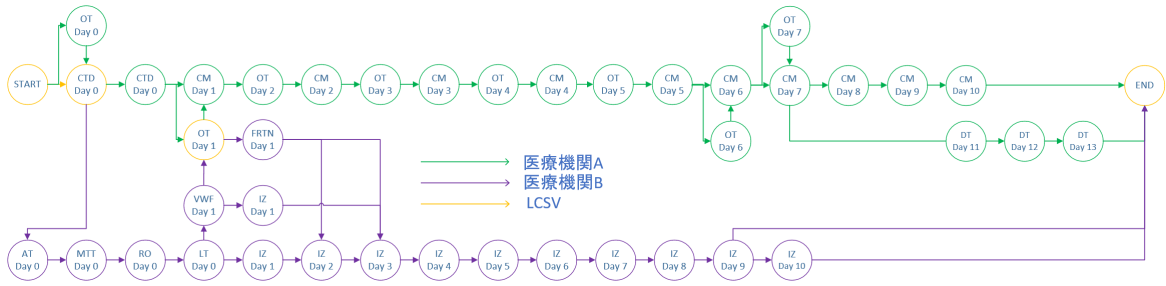


図 6 第五波の医療機関 A(minsup=0.2) と医療機関 B(minsup=0.25) の MSV

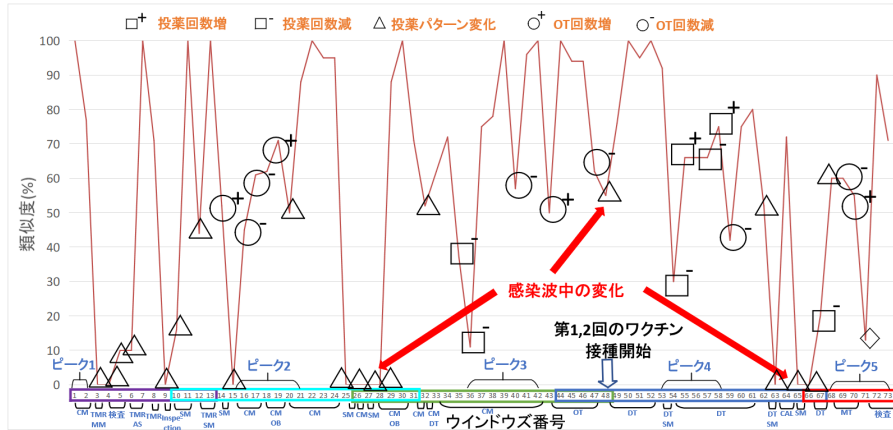


図 7 医療機関 A のパターン類似度の遷移図 (時間間隔ウィンドウ)



図 8 医療機関 A のパターン類似度の遷移図 (時間・シーケンス数ウィンドウ)

間・シーケンス数ウィンドウでは、主要な投薬の遷移状況は時間間隔ウィンドウと同じであるが、投薬量の増減や OT の増減については時間間隔ウィンドウよりも変化の頻度が少ないことがわかる。また、出現する薬の種類も時間間隔ウィンドウより少ないことがわかる。

時間間隔ウィンドウでは、治療パターンが頻繁に変化し、投薬の種類も多く出現している。一方、時間・シーケンス数ウィンドウでは、変化が少ないため、各時期の主要な投薬の使用状況の変化がよりわかりやすくなる。したがって、時間間隔ウィンドウは、各時期の投薬量の変化を捉え、より詳細な投薬の使用状況を把握する場合、時間間隔ウィンドウを使うべきである。また、各時期の主要な投薬の遷移状況を捉える場合、時間・シーケンス

数ウィンドウを使うべきである。

5 考察

3 の提案手法で述べたように、適切な minsup の設定は課題として残っている。例として、minsup を実験で用いた値より小さくすると、SV の長さが大きく増加してしまう。また、今回は minsup を少しずつ変更しながら SV を抽出し、有意であるかどうかで判断しているが、この判断自体が自明ではない。実験において、頻出する医療指示を除外することで、特徴的な投薬の医療指示が SV に出現するようになるが、重症度リスク因子を判定する CRP などの医療指示を除外し、抽出した SV にそれらが含まれないのは、要因分析の面からも、必ずしも適切

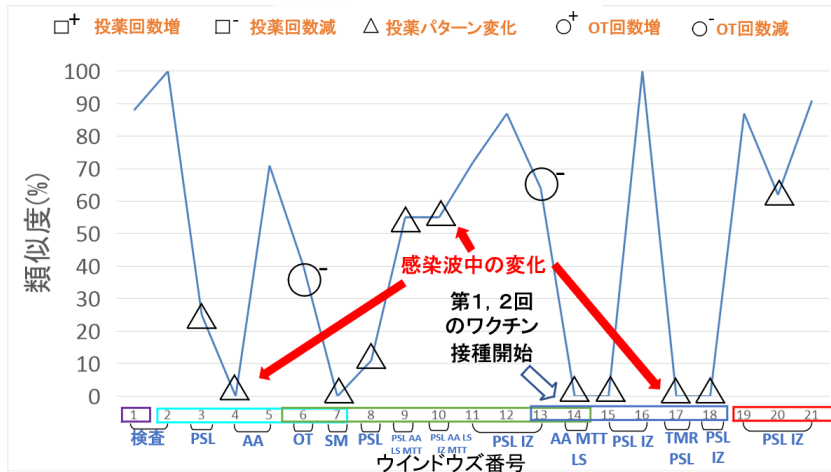


図9 医療機関Bのパターン類似度の遷移図(時間・シーケンス数ウィンドウ)

とは言えない。医学的な面を考慮しながら適切な minsup と除外する医療指示の選択は、今後の重要な課題である。

一方、適切なウィンドウの種類及びウィンドウサイズの設定も課題の一つだと考えられる。ウィンドウサイズとスライド単位を変更すると、抽出されるパターンの細部が若干異なることがある。例として、シーケンスウィンドウのウィンドウサイズを実験で用いた値より小さくすると、投薬の量が変わることになる。どの組み合わせが最も医学的に意味があるかを決めるのは困難だが、試した組み合わせを医療従事者に示し、有用な組み合わせを選んでもらうことができる。

6 結 論

6.1 ま と め

本稿では、同一疾病の異なる期間における頻出医療指示パターンの差異を可視化するために、COVID-19の第一波から第五波までの電子カルテデータに対して、期間毎のシーケンスバリエーションSVを抽出し、隣接期間間の最長の共通サブシーケンスバリエーションであるLCSVと、LCSVにそれぞれの期間のSVをラベル(色)付きで融合したMSVを求めた。また、二種類のスライディングウィンドウを適用し、時期的に異なる特徴を持つ頻出パターンを抽出し、頻出パターン間の類似度を評価することで、治療パターンが切り替わる時期を可視化した。その結果、呼吸器学会で報告されている方針等と合致する期間毎の医療指示パターンの違いを示すことができた。また、同じ時期の医療機関毎の対応の違い及び医療機関毎のパターンの遷移状況も示すことができた。

6.2 今後の課題

今後の課題として、まず適切な minsup と除外すべき医療指示に関する検討を進める必要がある。また、適切に時期的な特徴を持つ頻出パターンを抽出するために、どんなスライディングウィンドウ及びそのウィンドウのサイズとスライド単位に関しても検討を行う必要がある。一方、提案手法は他の疾病または感染波を対象にすることができるが、データベースの関係で

COVID-19 第一波から第五波までのみ解析を行ったが、今後は他の疾病と感染波に対する適用する必要がある。

謝 辞

本研究の一部は、日本学術振興会科学研究費(20H04192, 21K17746)の助成により実施された。また、本研究も「COVID-19 感染の診療への影響調査と予測モデル開発」からの支援によって行われた。データを提供頂いた研究協力機関に感謝の意を表する。

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, 2022.3.
- [2] 吉原博幸, “千年カルテプロジェクト: 本格的日本版 EHR と医療データの2次利用に向けて. 情報管理,” vol.60, no.11, pp.767-778, 2018.
- [3] R. Agrawal, R. Srikant, “Fast algorithms for mining association rules in large databases. Proceeding of the 20th International Conference on Very Large Databases,” 1994.
- [4] R. Agrawal, T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases, In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data,” pp.207-216, 1993.
- [5] 佐々木夢, 荒堀喜貴, 串間宗夫, 荒木賢二, 横田治夫, “電子カルテシステムのオーダログ解析による医療行為の支援. 日本データベース学会論文誌,” Vol. 14-J, Article No. 10, 2016.
- [6] 山田達大, 本田祐一, 菅原正彬, Le Hieu Hanh, 串間宗夫, 小川泰右, 松尾亮輔, 山崎義義, 荒木賢二, 横田治夫, “SID を保持するシーケンシャルパターンマイニングによる臨床パスバリエーション分析,” DEIM Forum D1-1, 2019.
- [7] Y. Honda, M. Kushima, T. Yamazaki, K. Araki, H. Yokota, “Detection and visualization of variants in typical medical treatment sequences. Proceeding of the 3rd VLDB Workshop on Data Management and Analytics for Medicine and Healthcare,” Springer, pp. 88-101, 2017.
- [8] V. P. Raju, G. S. Varma. “Mining Closed Sequential Patterns in Large Sequence Databases. International Journal of Database Management Systems” vol.7, no.1, pp.29-39, 2015.
- [9] H. H. Le, H. Edman, Y. Honda, M. Kushima, T. Yamazaki, Kenji Araki, H. Yokota, “Fast Generation of Clinical Pathways Including Time Intervals in Sequential Pattern Mining on Electronic Medical Record Systems”. Proceeding of the fourth International Conference on Computational Sci-

ence and Computational Intelligence (CSCI 2017), pp. 1726-1731, 2017.

- [10] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, M. Hsu. PrefixSpan, “Mining sequential patterns efficiently by prefix-projected pattern growth” Proceeding of 2001 international conference on data engineering, pp. 215-224, 2001.
- [11] 本田祐一, 山田達大, 萱原正彬, Le Hieu Hanh, 串間宗夫, 小川泰右, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫, “患者の固有情報及び動的状況を考慮したクリニカルパス分岐要因推定,” DEIM Forum D1-5, 2019.
- [12] 坂本任駿, 小林莉華, Le Hieu Hanh, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫, “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 C25-1, 2021.
- [13] Zhao Zitai, Le Hieu Hanh, 松尾 亮輔, 山崎 友義, 荒木 賢二, 横田 治夫, “COVID-19 の異なる医療機関と時期における頻出治療パターンの比較,” 第 42 回医療情報学連合大会論文集, pp. 887-892, 2022.
- [14] 横田治夫, Le Hieu Hanh, Li Yuqing, 松尾亮輔, 山崎友義, 荒木賢二, “複数医療機関間の頻出医療指示パターン比較手法,” 第 26 回日本医療情報学会春季学術大会, 2022.
- [15] 安光夕輝, Le Hieu Hanh, 松尾 亮輔, 山崎 友義, 荒木 賢二, 横田 治夫, “クラスタリングを用いた多病院間の頻出医療指示パターン比較,” DEIM Forum 5b-6-3, 2023.
- [16] 蛋白分解酵素阻害剤「フオイパン錠」新型コロナウイルス感染症 (COVID-19) を対象とした国内第相試験の結果について <https://www.ono.co.jp/news/20210611.html>
- [17] 2021/6 の呼吸器学会: <https://www.jrs.or.jp/covid19/faq/treatment/20210610105102.html>
- [18] 第 5 波の特徴 https://web.pref.hyogo.lg.jp/governor/documents/g_20210728_01_02.pdf
- [19] 診療の手引き 2 班: 第 2 版発行 日 18 月 5 年 2020 <https://www.mhlw.go.jp/content/000631552.pdf>
- [20] 東京大がカモスタットの新型コロナに対する臨床研究を計画 <https://bio.nikkeibp.co.jp/atcl/news/p1/20/03/16/06694/>