

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

安光 夕輝[†] 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: [†]{yasumitsu,hanhhlh}@de.cs.titech.ac.jp, [†]yamazaki.cp@gmail.com, [†]yokota@cs.titech.ac.jp,
^{††}matsuo@ldi.or.jp,
^{†††}araki6925@gmail.com

あらまし 頻出医療指示パターンの病院間の共通点や相違点を明らかにするために、複数病院間での頻出医療指示パターンの比較を行った。しかし、多くの病院を一度に比較すると病院間の共通点や相違点を正確に把握することが難しい。そこで、頻出医療指示パターンの距離を定義してクラスタリングを行い、対象となる病院群をいくつかのクラスタに分類し、そのクラスタ内で比較を行った。その際に、多くのシーケンスに出現し、シーケンス内での出現回数が多い医療指示があると、重要な投薬などの処置項目を抽出できない。そこで、出現頻度の高い検査項目を除外してマイニングを行うことで、投薬内容の抽出を可能にした。実際に 23 病院の COVID-19 のデータを用いて提案手法を適用することで、病院ごとの頻出医療指示パターンの特徴を明らかにし、病院間の共通点と相違点を容易に把握することができた。

キーワード 電子カルテ, 医療支援, 医療機関間比較, 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

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が進み、今後さらに日本全国の医療機関で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]。二次利用の例として、ある疾病の患者に対して頻出医療指示シーケンスの抽出を行い、典型的な医療行為の流れ「クリニカルパス」を生成するというものがある。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、過去の膨大な事例を収集・分析してクリニカルパスの改善を行うことは人力では困難であった。そのため、様々なデータ工学の手法を用いて電子カルテを解析し、医療行為改善の支援を行う研究が進められている。

また、国家レベルでの医療情報管理及び二次利用の必要性が認識され、電子カルテなどの健康・診療データを全国規模で収集し、活用を進める「千年カルテプロジェクト」[2] が始まっている。「千年カルテプロジェクト」に参加する医療機関のデータが 1 つに集約されるため、複数の医療機関を跨いだ研究が期待されている。医療機関ごとの頻出治療パターンを比較し、その多様性を認識することは非常に有用であり、医療行為のさらなる改善につながるが見込まれる。

1.2 本研究の目的とアプローチ

本研究は、複数の医療機関の電子カルテを解析し、医療機関間

の比較を行い、共通点と相違点を可視化することで、医療行為の支援を行うことを目的とする。アプローチとして、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を行い、実際の電子カルテデータベースから頻出医療指示パターンの抽出を行う。この際、出現頻度の高い検査項目を除外することで各医療機関における投薬内容の抽出を可能にした。また、頻出医療指示パターンをシーケンスバリエーション (Sequence Variant, 以下 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, 以下 LCSV) を算出する。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 (Sequence Variant Distance, 以下 SVD) を定義し、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, 以下 MSV) を作成することで、医療機関間の頻出医療指示パターンの共通点と相違点を可視化する。

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可視化は我々の研究グループが提案する独自の手法で、この手法を 3 つ以上の医療機関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 節では背景知識と関連研究について述べる。3 節では提案手法である医療指示の除外、医療機関群のクラスタリング及び 3 つ以上の医療機関の LCSV と MSV を求めるアルゴリズムについて述べる。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し、医療機関群のクラスタリング及び MSV の作成を行う。最後に 5 節でまとめと今後の課題について述べる。

2 背景知識と関連研究

2.1 背景知識

李ら [14] は 2 つの SV を比較するために最長共通サブシーケンスバリエーション (LCSV) と併合シーケンスバリエーション (MSV) を提案した。以下に関連する概念の定義と LSCV 及び MSV の定義を述べる。

2.1.1 シーケンス

定義 1 (アイテムセット). アイテムセット I を以下のように定義する。

$$I = \{i_1, i_2, \dots, i_n\}$$

$i_j \in I$ がアイテムとなる。

定義 2 (シーケンス). アイテムセット I に対し、シーケンス S を以下のように定義する。

$$S = (\{s_1, s_2, \dots, s_m\}, \prec_S)$$

$s_j = (id, i), i \in I$. id は S の中でユニークなインデックスとなる。 \prec_S は S 上の全順序関係である。

2.1.2 SV

定義 3 (シーケンスバリエーション). アイテムセット I に対し、束 SV をシーケンスバリエーションと定義し、以下のように表す。

$$SV = (\{sv_1, sv_2, \dots, sv_l\}, \prec_{SV})$$

$sv_j = (id, i), i \in I$. id は SV の中でユニークなインデックスとなる。 \prec_{SV} は SV 上の半順序関係である。また、 SV が上限および下限を持っている。

例として、分岐のある医療指示列がシーケンスバリエーションとして表される。シーケンス (入院 → 検査 → 手術 → 退院) と (入院 → 検査 → 投薬 → 退院) に対して、手術と投薬を分岐と捉え、この 2 つ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。

以下、 A, B, C のような大文字をアイテムとする場合、 $A \prec B, B \prec C, B \prec D, C \prec E, D \prec E$ を、 $\langle A, B, (C, D), E \rangle$ と表す。

定義 4 (シーケンスバリエーショングラフ). シーケンスバリエーション SV に対し、有向非巡回グラフ G をシーケンスバリエーショングラフと定義し、以下のように表す。

$$G = (V, E)$$

頂点集合 V が SV のアイテム集合となる。また、 $\forall e = (sv_{out}, sv_{in}) \in E, sv_{out} \prec_{SV} sv_{in}$ である。

シーケンスバリエーショングラフを可視化することで、簡単にシーケンスバリエーションを示すことができる。例として、 A, B, C, D, E

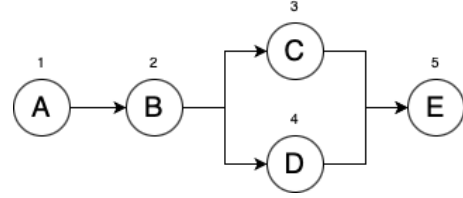


図 1 SV のシーケンスバリエーショングラフ

をアイテムとすると、 $SV = \langle A, B, (C, D), E \rangle$ のシーケンスバリエーショングラフは図 1 のように示される。

シーケンスバリエーションとシーケンスバリエーショングラフは互いに変換可能であるため、以下両者を同じものとして扱う。

2.1.3 LCS

定義 5 (共通部分列 (CS)). シーケンス S_α と S_β に対し、共通部分列 CS を下記の条件を満たすシーケンスと定義する。
 $\forall cs_i, cs_j \in CS, cs_i, cs_j \in S_\alpha \cap S_\beta$ かつ $cs_i \prec_{CS} cs_j \iff cs_i \prec_\alpha cs_j \wedge cs_i \prec_\beta cs_j$.

CS の部分列も CS となるので、CS は複数存在する。そのため、最長共通部分列を定義する。

定義 6 (最長共通部分列 (LCS)). シーケンス S_α と S_β に対し、最長共通部分列 LCS を下記の条件を満たす共通部分列と定義する。 S_α と S_β に対する任意の共通部分列 CS に対し、 $|CS| \leq |LCS|$. $|CS|$ は CS のノード数を表す。

入力列の個数が任意である一般の場合については、この問題は NP 困難である [19]. 入力列の個数が一定のときには、この問題は動的計画法によって多項式時間で解く事ができる。また、3 つ以上の LCS を求めるための高速なアルゴリズムが提案されている [20].

2.1.4 LCSV

SV の共通部分である共通サブシーケンスバリエーション (CSV) を定義する。

定義 7 (共通サブシーケンスバリエーション (CSV)). シーケンスバリエーション SV_α と SV_β に対し、共通サブシーケンスバリエーション CSV を下記の条件を満たすシーケンスバリエーションと定義する。
 $\forall csv_i, csv_j \in CSV, csv_i, csv_j \in SV_\alpha \cap SV_\beta$ かつ $csv_i \prec_{CSV} csv_j \iff csv_i \prec_\alpha csv_j \wedge csv_i \prec_\beta csv_j$.

CSV のサブシーケンスバリエーションも CSV となるので、CSV は複数存在する。そのため、最長共通サブシーケンスバリエーション (LCSV) を定義する。

定義 8 (最長共通サブシーケンスバリエーション (LCSV)). シーケンス SV_α と SV_β に対し、最長共通サブシーケンスバリエーション $LCSV$ を下記の条件を満たす共通サブシーケンスバリエーションと定義する。 SV_α と SV_β に対する任意の共通サブシーケンスバリエーション CSV に対し、 $|CSV| \leq |LCSV|$. $|CSV|$ は CSV のノード数を表す。

LCSV の例を図 2 に示す。

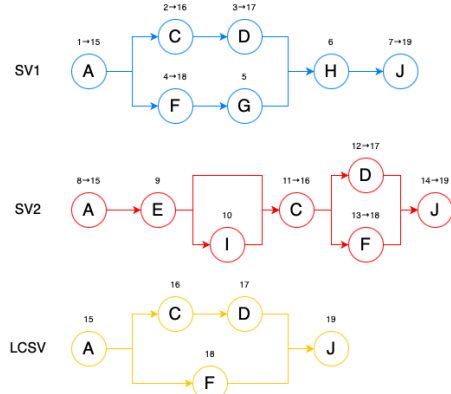


図2 SV_1 と SV_2 の LCSV

2.1.5 MSV

2つのSVをLCSVを基に併合したMSVを定義する。

定義 9 (併合シーケンスバリエント (MSV)). シーケンスバリエント SV_α, SV_β とその LCSV に対し, 併合シーケンスバリエント MSV を下記の条件を満たすシーケンスバリエントと定義する. $\forall sv_i, sv_j \in SV_\alpha \cup SV_\beta, sv_i, sv_j \in MSV$ かつ $sv_i \prec_{MSV} sv_j \iff sv_i \prec_\alpha sv_j \wedge sv_i \prec_\beta sv_j$. $\forall msv_i, msv_j \in MSV, msv_i \prec_{MSV} msv_j \implies (msv_i, msv_j \in SV_\alpha \wedge msv_i \prec_\alpha msv_j) \vee (msv_i, msv_j \in SV_\beta \wedge msv_i \prec_\beta msv_j)$. また, LCSV に出現したノードに label $\{\alpha, \beta\}$ を与え, 出現していないノードに属しているSVにより, label α か β を与える。

MSV の例を図3に示す。

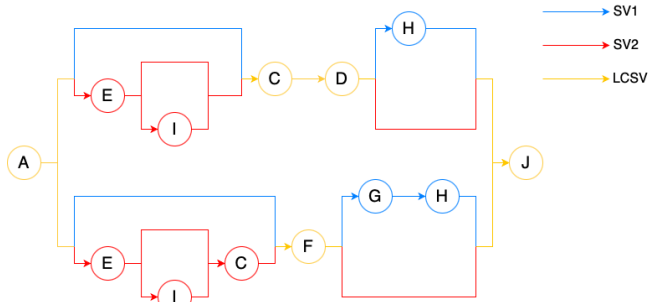


図3 SV_1 と SV_2 の MSV

2.2 関連研究

2.2.1 SPM

Agrawal らによって提案された SPM はシーケンシャルデータベース (以下, SDB) から頻出シーケンスを抽出する手法であり [3], 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 [21]. 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. SDB から, 出現頻度の最小値として設定した最小支持度 (minsup) より出現頻度が高いシーケンスを抽出する. SPM を行うアルゴリズムは Apriori [3], PrefixSpan [4], CSpan [5] 等, 多数提案されている。

2.2.2 SPM による医療情報の解析

佐々木ら [6] は, タイムインターバル SPM [7] (以下 TI-SPM) を PrefixSpan に用いて, 実電子カルテデータから, 時間間隔を考慮した頻出シーケンスの抽出を行った. Le ら [8] は, TI-SPM を効率的な SPM である CSpan [5] に用いた T-CSPan を提案し, 実電子カルテデータに適用した. さらに Le ら [9] は, 電子カルテ解析におけるプライバシー保護に関して差分プライバシーの概念を適用する手法を提案した。

2.2.3 SV の解析

本田ら [10] は, 実電子カルテデータから医療指示の SV を抽出し, グラフで表現して医療従事者への視認性の高い可視化ツールの提供を行った. さらに本田ら [11] は, SV に対して, 患者の性別や年齢といった静的情報と, 血圧や検査結果といった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の要因分析を行う手法を提案し, 実電子カルテデータを用いて効果を示した. 山田ら [12] は, 頻出シーケンスとシーケンスの対応関係を, シーケンスの識別子であるシーケンス ID を保持しながらマイニングすることによって, 頻出シーケンス抽出と同時に取得し, 頻出シーケンスごとの安全性や効率性の指標を算出し, SV 評価が円滑かつ正確に行える可視化手法を提案し, 実電子カルテデータを用いて推薦結果の評価を行った. 坂本ら [13] 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選択を行い, 頻出シーケンス分岐先の適正候補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を拡充し, 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。

2.2.4 SV の比較

李ら [14] [15] は 2つのSVの共通部分であるLCSVと, 2つのSVを併合したMSVを提案し, 2医療機関間の頻出医療指示パターンの差異を示した. 趙ら [16] は COVID-19 の実電子カルテデータから, 異なる時期の頻出医療指示パターンを抽出し, MSV を用いてその差異を示した. また, 出現頻度の高い検査項目を除外することで投薬内容の抽出を可能にした. さらに, 趙ら [17] は SV 間の類似度を定義し, スライディングウィンドウを用いて頻出医療指示パターンが変化する時期を推定した. 黒川ら [18] は MSV を用いて基礎疾患の有無による頻出医療指示パターンの差異を示した。

3 提案手法

本節では頻出な医療指示の除外, 医療機関群のクラスタリング, LCSV と MSV を求めるアルゴリズムについて述べる。

3.1 頻出な医療指示の除外

電子カルテデータに含まれる医療指示は検査項目と処置項目に大別される. 処置項目には主に投薬などが含まれ, ある疾病に対する治療パターンを解析する際に, 処置項目は重要な医療指示である. しかし, 医療指示シーケンスには多くの検査項目が含まれているため, SPM を行って得られる頻出医療指示パターンは出現頻度の高い検査項目を多く含んでおり, 処置項目

が抽出できないという問題がある。この問題は、多くのシーケンスに出現し、シーケンス内での出現回数が多い医療指示によって引き起こされる。これは SPM で得られた最長の頻出パターンを用いて SV を作成しているためである。そこで、処置項目を抽出するために、これらの検査項目を事前に除外してから SPM を行う。

除外する医療指示を決定するにあたり、医療指示の出現頻度を評価する指標が必要である。そこで、ある医療指示がシーケンスに出現する平均回数、 TO/S (*Term Occurrence per Sequence*) を提案する。

定義 10 (TO/S).

$$TO/S = \frac{\text{その医療指示が全シーケンス中に出現する回数}}{\text{全シーケンス数}}$$

TO/S はある医療指示について医療機関ごとに計算される値である。複数の医療機関を対象に実験を行うにあたり、医療機関ごとの TO/S を参照して除外する医療指示を一元的に決定する必要がある。そこで、対象となる医療機関群での TO/S の平均値 $Ave(TO/S)$ と最大値 $Max(TO/S)$ を算出する、

$Ave(TO/S)$ が高い医療指示は多くの医療機関で共通して出現頻度が高い医療指示であるため、これを算出すること自体に意味はあるが、医療指示の除外を行う上で適切な指標とは言えない。なぜなら、一部の医療機関でのみ TO/S が高い医療指示の場合、 $Ave(TO/S)$ は低くなってしまいうからである。対して、 $Max(TO/S)$ は特定の医療機関でのみ出現頻度の高い医療指示についても考慮することができる。そのため、 $Max(TO/S)$ が低い医療指示はどの病院でも出現頻度が低く、SPM 時に処置項目が抽出できないという問題を引き起こさない。

$Max(TO/S)$ が高い検査項目を除外するにあたり、その閾値を定める必要がある。処置項目の抽出を目的としているので、 $Max(TO/S)$ が高い検査項目から順に除外し、ある値よりも $Max(TO/S)$ が小さい検査項目について、除外しなくても抽出される処置項目が変化しないような値の最大値を閾値とする。

3.2 医療機関群のクラスタリング

対象となる医療機関群についてそれぞれの医療機関の SV を抽出した後、これらの比較を MSV を用いて行う。対象となる医療機関の数が増えると SV の共通部分である LCSV が存在しなくなることが懸念される。MSV を用いる理由は SV の共通点と相違点の双方を示すことであり、共通部分が存在しない SV 群から MSV を作成しても、SV を単に並べたものと大差無く、SV をうまく比較できているとは言えない。

この問題を解決するためにクラスタリングを用いて医療機関群をいくつかのクラスタに分類し、そのクラスタ内で MSV を作成する。クラスタリングに際して、SV 間の距離を定義する必要がある。SVD(Sequence Variant Distance) を提案する。

定義 11 (SVD). 2 つの $SV(SV_1, SV_2)$ の距離 SVD を以下のように定義する。 $|SV|$ は SV のノード数を表す。

$$SVD = 1 - \frac{2|LCSV(SV_1, SV_2)|}{|SV_1| + |SV_2|}$$

クラスタリングの手法は階層的クラスタリングを用いる。DBSCAN も検討したが、要素数 1 のクラスタが多くできてしまった。階層的クラスタリングによって作成されたデンドログラムを基にクラスタを作成する。クラスタを分割する距離の閾値は、クラスタ内の SV について LCSV が存在し、かつクラスタ数が最小になるようなものとする。階層的クラスタリングの手法は最短距離法、最長距離法、重心法、群平均法、ウォード法の 5 つを行う。3 つ以上の医療機関について MSV を作成する本研究の提案手法をより多くの医療機関に適用するために、要素数が 3 以上のクラスタに含まれる医療機関が最多となる手法を採用する。

3.3 3 つ以上の SV について LCSV を求めるアルゴリズム

3 つ以上の SV について LCSV を求めるにあたり、李ら [14] が提案した 2 つの SV について LCSV を求めるアルゴリズムを拡張した。以下は 3 つの SV についての例である。

Algorithm 1 LCSV

```

1: input:  $SV_1, SV_2$  and  $SV_3$ 
2: output:  $LCSV$ 
3: number  $SV_1$  from 1 to  $s$ 
4: number  $SV_2$  from  $s + 1$  to  $t$ 
5: number  $SV_3$  from  $t + 1$  to  $u$ 
6:  $r_1, r_2, r_3 \leftarrow$  the number of paths of  $SV_1, SV_2, SV_3$ 
7:  $LCSs \leftarrow$  a list contains  $LCS$ 
8: for  $i = 1, 2, \dots, r_1$  do
9:   for  $j = 1, 2, \dots, r_2$  do
10:    for  $k = 1, 2, \dots, r_3$  do
11:       $LCS(i, j, k) \leftarrow$  the LCS of path  $i$  of  $SV_1$ , path  $j$  of  $SV_2$  and path  $k$  of  $SV_3$ 
12:      add  $LCS$  to  $LCSs$ 
13:    end for
14:  end for
15: end for
16: remove duplicate of  $LCS$  in  $LCSs$ 
17:  $LCSV \leftarrow$  combination of  $LCS$  in  $LCSs$ 
18: re-number nodes in  $LCSV$  from  $u + 1$ , also re-number them in  $SV_1, SV_2$  and  $SV_3$ 

```

10 行目から始まる 3 重の for ループを N 重にすることで、 N 個の SV について LCSV を求めることができる。

3.4 3 つ以上の SV について MSV を求めるアルゴリズム

李ら [14] が提案した 2 つの SV について MSV を求めるアルゴリズムは共通部分 LCSV に 2 つの SV を併合するというもので $MSV(SV_1, SV_2, LCSV)$ と表すことができる。これを用いて 3 つ以上の MSV を求めるアルゴリズムを 2 つ提案する。

3.4.1 直接併合法

N 個の SV について LCSV を求め、それに N 個の SV を直接併合するものである。

3.4.2 距離ベース併合法

クラスタリングの際に得られたデンドログラム上で距離が近いものから順に併合するものである。

Algorithm 2 MSV(直接併合法)

```
1: input:  $SV_1, SV_2, \dots, SV_N$ 
2: output:  $MSV$ 
3:  $LCSV = LCSV(SV_1, SV_2, \dots, SV_N)$ 
4:  $MSV = MSV(SV_1, SV_2, LCSV)$ 
5: for  $i = 3, \dots, N$  do
6:    $MSV = MSV(SV_i, MSV, LCSV)$ 
7: end for
```

Algorithm 3 MSV(距離ベース併合法)

```
1: input:  $Tree(SV_1, SV_2, \dots, SV_N)$ 
2: output:  $MSV$ 
3:  $MSV = MSV\_TREE(Tree(SV_1, SV_2, \dots, SV_N))$ 
```

Algorithm 4 MSV_TREE

```
1: input:  $Tree$ 
2: output:  $MSV$ 
3:  $Tree_{left}$  is left subtree
4:  $Tree_{right}$  is right subtree
5: if  $Tree_{left} \neq Null$  then
6:    $MSV_{left} = MSV\_TREE(Tree_{left})$ 
7:    $MSV_{right} = MSV\_TREE(Tree_{right})$ 
8:    $LCSV_{Tree} = LCSV(SV \text{ in } Tree)$ 
9:   return  $MSV(MSV_{left}, MSV_{right}, LCSV_{Tree})$ 
10: else
11:    $Tree$  has one  $SV$ ,  $SV_{leaf}$ 
12:   return  $SV_{leaf}$ 
13: end if
```

4 実験

23 の医療機関について、COVID-19 の第 5 波に対する実電子カルテデータを用いて提案手法を適用し、有用性を確認する。

4.1 データセット

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 23 医療機関 (医療機関 A,B,...,W) から提供された、COVID-19 の第 5 波 (2021.07.01-2021.09.30) 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。それぞれの医療機関の医療指示シーケンスの情報を表 1 に示す。

4.2 実験環境

個人情報保護のために、利用者が限定された宮崎大学医学部附属病院が管轄するクラウド環境 (Amazon Work Space) を実験環境として用いる。解析者が電子カルテデータをクラウドから直接取り出さずに、解析プログラムの Docker Image を作成し、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を結果として取り出すことで、情報セキュリティの担保とプライバシー保護を行っている。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。

4.3 医療指示の同一視

同一内容だが表記が異なる医療指示、同一の成分だが名称が異なる薬、同じ目的だが手法が異なる検査について正しく出現頻度を評価したり、SPM で頻出パターンとして抽出するために、これらの医療指示を同一のものとして扱う前処理を行う。これらの医療指示の選定には、診療報酬情報提供サービス [22] から得られるレセプト電算コードの区分番号を参考にした。

4.4 実験結果

4.4.1 医療指示の除外

頻出医療指示パターン中に出現する医療指示について $Ave(TO/S)$, $Max(TO/S)$ を求め、出現頻度の高い検査項目及び COVID-19 の治療と関係が薄い医療指示を除外した。その結果を検査項目と処置項目に分けてそれぞれ表 2 と表 3 に示す。

出現頻度の高い検査項目を除外するにあたり、除外する医療指示の $Max(TO/S)$ の閾値は肝炎ウイルス検査とした。肝炎ウイルス検査よりも $Max(TO/S)$ が低い検査項目は除外しなくても処置項目が抽出できなくなることとはなかった。また COVID-19 の治療と関係が薄い医療指示の選定には医療従事者からの意見を参考にした。

4.4.2 SV の抽出

23 の医療機関について前項で説明した医療指示の同一視と除外を行った後、SPM を適用し、頻出医療指示パターンを SV として抽出した。医療機関 A,B,...,V では $minsup = 0.2$ とした。医療機関 W では $minsup = 0.2$ とすると SV が得られなかったので $minsup = 0.1$ とした。

4.4.3 医療機関群のクラスタリング

抽出した SV について SVD を距離関数として階層的クラスタリングを行った。最短距離法、最長距離法、重心法、群平均法、ワード法の結果をデンドログラムとしてそれぞれ図 4, 図 5, 図 6, 図 7, 図 8 に示す。群平均法とワード法では要素数が 3 以上のクラスタに含まれる医療機関が最多なのでこの 2 つの手法を採用し、クラスタ内での LCSV を図 9 と図 10 に示す。また、医療指示の略号を表 4 に示す。各クラスタの共通部分は酸素吸入 (OT), インスリン (IN), アセトアミノフェン (AA), デキサメタゾン (DM), ヘパリン (HP), AA と CT 撮影 (CT) となっており、各医療機関の特徴的な処置項目によって分類されている。肺の炎症を抑えるステロイド剤であるデキサメタゾン、血栓症を予防するヘパリン、酸素を投与する酸素吸入は COVID-19 の治療において特に重要な処置項目である。これらの処置項目の出現の有無は患者の重症度と関連しており、医療機関の特徴を表している。

また、群平均法とワード法の違いは医療機関 B,N,V が DM を共通部分とするクラスタに分類されるか、HP を共通部分とするクラスタに分類されるかの違いのみである。医療機関 B,N,V は SV に DM と HP が両方含まれているため、クラスタリングの手法の違いによって結果が変わったと考えられる。

また、異なるクラスタに含まれる SV は特徴が異なり、クラスタが併合される距離が遠いほど違いも大きく、実際に LCSV が存在しないものが多い。

表 1 対象となる医療機関の COVID-19(第 5 波) のシーケンス情報

医療機関	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
シーケンス数	130	23	11	102	89	79	126	36	29	104	104	30	179	228	182	29	57	283	55	69	22	28	22
平均シーケンス長(除外前)	181.7	256.0	45.3	153.7	242.6	284.3	90.2	127.4	211.4	317.5	112.2	275.7	282.4	272.5	181.0	73.7	310.0	139.2	174.5	189.4	340.0	220.3	137.1
平均シーケンス長(除外後)	18.7	56.6	5.3	22.0	47.2	31.8	23.2	15.6	37.2	51.3	35.3	57.8	40.6	46.2	23.0	16.4	66.0	13.3	21.3	36.2	51.1	38.6	18.4

表 2 Ave(TO/S) と Max(TO/S) 及び除外する医療指示 (検査項目)

Max(TO/S)で除外 Covid19の治療と関係が薄いので除外																									
医療指示名	Ave(TO/S)	医療指示名	Ave(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)	医療指示名	Max(TO/S)
末梢血液	3.68726	PCT	0.32413	経皮的酸素飽和度測定	10.955	SARS-CoV-2抗原検出	2.304	末梢血液	3.68726	糖試験紙法	10.222	VWF抗原	2.232	呼吸心拍監視	9.528	肝炎ウイルス検査	2.103	尿一般	1.867	梅毒トレポネマ	1.646	心電図	1.591	CT撮影	1.467
呼吸心拍監視	3.30809	BNP	0.30109	糖試験紙法	10.222	肺炎抗原	1.867	呼吸心拍監視	3.30809	呼吸心拍監視	9.528	梅毒トレポネマ	1.646	呼吸心拍監視	3.30809	肺炎抗原	1.867	呼吸心拍監視	3.30809	肺炎抗原	1.867	呼吸心拍監視	3.30809	肺炎抗原	1.867
CRP	3.29974	終末呼吸気道ガス濃度測定	0.29057	呼吸心拍監視	9.528	肺炎抗原	1.867	CRP	3.29974	呼吸心拍監視	9.528	梅毒トレポネマ	1.646	CRP	3.29974	呼吸心拍監視	9.528	CRP	3.29974	呼吸心拍監視	9.528	CRP	3.29974	呼吸心拍監視	9.528
経皮的酸素飽和度測定	3.25183	乳酸	0.28552	生食注シリンジ	8.267	肺炎抗原	1.867	経皮的酸素飽和度測定	3.25183	生食注シリンジ	8.267	肺炎抗原	1.867	経皮的酸素飽和度測定	3.25183	生食注シリンジ	8.267	経皮的酸素飽和度測定	3.25183	生食注シリンジ	8.267	経皮的酸素飽和度測定	3.25183	生食注シリンジ	8.267
単純撮影	2.99991	ABO血液型	0.27461	末梢血液	7.455	梅毒トレポネマ	1.646	単純撮影	2.99991	末梢血液	7.455	梅毒トレポネマ	1.646	単純撮影	2.99991	末梢血液	7.455	単純撮影	2.99991	末梢血液	7.455	単純撮影	2.99991	末梢血液	7.455
糖試験紙法	2.95887	鼻腔吸引液採取	0.2373	グルコース	6.828	心電図	1.591	糖試験紙法	2.95887	鼻腔吸引液採取	0.2373	心電図	1.591	糖試験紙法	2.95887	鼻腔吸引液採取	0.2373	糖試験紙法	2.95887	鼻腔吸引液採取	0.2373	糖試験紙法	2.95887	鼻腔吸引液採取	0.2373
Dダイマー	2.64187	梅毒トレポネマ	0.2297	総ビリルビン	6.545	CT撮影	1.467	Dダイマー	2.64187	梅毒トレポネマ	0.2297	CT撮影	1.467	Dダイマー	2.64187	梅毒トレポネマ	0.2297	Dダイマー	2.64187	梅毒トレポネマ	0.2297	Dダイマー	2.64187	梅毒トレポネマ	0.2297
点滴注射	2.3433	中心静脈圧測定	0.21991	グルコース	6.828	CT撮影	1.467	点滴注射	2.3433	中心静脈圧測定	0.21991	CT撮影	1.467	点滴注射	2.3433	中心静脈圧測定	0.21991	点滴注射	2.3433	中心静脈圧測定	0.21991	点滴注射	2.3433	中心静脈圧測定	0.21991
総ビリルビン	2.01943	APTT	0.20565	総ビリルビン	6.545	BUN	1.414	総ビリルビン	2.01943	APTT	0.20565	BUN	1.414	総ビリルビン	2.01943	APTT	0.20565	総ビリルビン	2.01943	APTT	0.20565	総ビリルビン	2.01943	APTT	0.20565
血液ガス分析	1.19974	留置カテーテル	0.14204	CRP	6.364	クレアチニン	1.379	血液ガス分析	1.19974	留置カテーテル	0.14204	クレアチニン	1.379	血液ガス分析	1.19974	留置カテーテル	0.14204	血液ガス分析	1.19974	留置カテーテル	0.14204	血液ガス分析	1.19974	留置カテーテル	0.14204
液体酸素測定装置化酸素貯槽	1.19283	HbA1c	0.14143	点滴注射	6.203	AST	1.353	液体酸素測定装置化酸素貯槽	1.19283	HbA1c	0.14143	AST	1.353	液体酸素測定装置化酸素貯槽	1.19283	HbA1c	0.14143	液体酸素測定装置化酸素貯槽	1.19283	HbA1c	0.14143	液体酸素測定装置化酸素貯槽	1.19283	HbA1c	0.14143
生食注シリンジ	1.06165	KL6	0.12965	フェリチン	5.611	鼻腔吸引液採取	1.342	生食注シリンジ	1.06165	KL6	0.12965	鼻腔吸引液採取	1.342	生食注シリンジ	1.06165	KL6	0.12965	生食注シリンジ	1.06165	KL6	0.12965	生食注シリンジ	1.06165	KL6	0.12965
フェリチン	0.99843	クレアチニン	0.12517	血液ガス分析	5.227	KL6	0.962	フェリチン	0.99843	クレアチニン	0.12517	KL6	0.962	フェリチン	0.99843	クレアチニン	0.12517	フェリチン	0.99843	クレアチニン	0.12517	フェリチン	0.99843	クレアチニン	0.12517
グルコース	0.96587	R-吸気分布	0.10578	Dダイマー	5.091	BNP	0.933	グルコース	0.96587	R-吸気分布	0.10578	BNP	0.933	グルコース	0.96587	R-吸気分布	0.10578	グルコース	0.96587	R-吸気分布	0.10578	グルコース	0.96587	R-吸気分布	0.10578
麻薬等の鎮痛	0.78752	BUN	0.10443	PT	4.043	インテロイキン	0.904	麻薬等の鎮痛	0.78752	BUN	0.10443	インテロイキン	0.904	麻薬等の鎮痛	0.78752	BUN	0.10443	麻薬等の鎮痛	0.78752	BUN	0.10443	麻薬等の鎮痛	0.78752	BUN	0.10443
観血的動脈圧測定	0.75796	VWF抗原	0.09704	ナトリウム及びクロール	3.722	ABO血液型	0.888	観血的動脈圧測定	0.75796	VWF抗原	0.09704	ABO血液型	0.888	観血的動脈圧測定	0.75796	VWF抗原	0.09704	観血的動脈圧測定	0.75796	VWF抗原	0.09704	観血的動脈圧測定	0.75796	VWF抗原	0.09704
中心静脈注射	0.73761	CK-MB	0.09604	観血的動脈圧測定	3.633	HbA1c	0.873	中心静脈注射	0.73761	CK-MB	0.09604	HbA1c	0.873	中心静脈注射	0.73761	CK-MB	0.09604	中心静脈注射	0.73761	CK-MB	0.09604	中心静脈注射	0.73761	CK-MB	0.09604
CT撮影	0.736	TSH	0.09487	動脈血採取	3.333	FT4	0.768	CT撮影	0.736	TSH	0.09487	FT4	0.768	CT撮影	0.736	TSH	0.09487	CT撮影	0.736	TSH	0.09487	FT4	0.768	CT撮影	0.736
肝炎ウイルス検査	0.72717	FT4	0.07696	APTT	3.31	TSH	0.768	肝炎ウイルス検査	0.72717	FT4	0.07696	TSH	0.768	肝炎ウイルス検査	0.72717	FT4	0.07696	肝炎ウイルス検査	0.72717	FT4	0.07696	肝炎ウイルス検査	0.72717	FT4	0.07696
尿一般	0.58487	結核菌特異的インターフェロン-γ産生能	0.07574	乳酸	3.202	TARC	0.745	尿一般	0.58487	結核菌特異的インターフェロン-γ産生能	0.07574	TARC	0.745	尿一般	0.58487	結核菌特異的インターフェロン-γ産生能	0.07574	尿一般	0.58487	結核菌特異的インターフェロン-γ産生能	0.07574	尿一般	0.58487	結核菌特異的インターフェロン-γ産生能	0.07574
動脈血採取	0.53887	AST	0.0627	総蛋白	3.152	LDLコレステロール	0.621	動脈血採取	0.53887	AST	0.0627	LDLコレステロール	0.621	動脈血採取	0.53887	AST	0.0627	動脈血採取	0.53887	AST	0.0627	動脈血採取	0.53887	AST	0.0627
総蛋白	0.50504	肺サーファクタント蛋白	0.04535	中心静脈注射	3.045	CK-MB	0.607	総蛋白	0.50504	肺サーファクタント蛋白	0.04535	CK-MB	0.607	総蛋白	0.50504	肺サーファクタント蛋白	0.04535	総蛋白	0.50504	肺サーファクタント蛋白	0.04535	総蛋白	0.50504	肺サーファクタント蛋白	0.04535
PT	0.46717	インターロイキン	0.043	中心静脈圧測定	3	グロブリンクラス別ウイルス抗体	0.552	PT	0.46717	インターロイキン	0.043	グロブリンクラス別ウイルス抗体	0.552	PT	0.46717	インターロイキン	0.043	PT	0.46717	インターロイキン	0.043	PT	0.46717	インターロイキン	0.043
細菌顕微鏡検査	0.44274	TARC	0.033778	麻薬等の鎮痛	2.87	結核菌特異的インターフェロン-γ産生能	0.517	細菌顕微鏡検査	0.44274	TARC	0.033778	結核菌特異的インターフェロン-γ産生能	0.517	細菌顕微鏡検査	0.44274	TARC	0.033778	細菌顕微鏡検査	0.44274	TARC	0.033778	細菌顕微鏡検査	0.44274	TARC	0.033778
細菌培養同定検査	0.40674	SIL2	0.02852	細菌培養同定検査	2.652	肺サーファクタント蛋白	0.517	細菌培養同定検査	0.40674	SIL2	0.02852	肺サーファクタント蛋白	0.517	細菌培養同定検査	0.40674	SIL2	0.02852	細菌培養同定検査	0.40674	SIL2	0.02852	細菌培養同定検査	0.40674	SIL2	0.02852
SARS-CoV-2抗原検出	0.4023	グロブリンクラス別ウイルス抗体	0.02787	終末呼吸気道ガス濃度測定	2.6	留置カテーテル	0.478	SARS-CoV-2抗原検出	0.4023	グロブリンクラス別ウイルス抗体	0.02787	留置カテーテル	0.478	SARS-CoV-2抗原検出	0.4023	グロブリンクラス別ウイルス抗体	0.02787	SARS-CoV-2抗原検出	0.4023	グロブリンクラス別ウイルス抗体	0.02787	SARS-CoV-2抗原検出	0.4023	グロブリンクラス別ウイルス抗体	0.02787
心電図	0.37822	LDLコレステロール	0.02743	R-吸気分布	2.433	SIL2	0.455	心電図	0.37822	LDLコレステロール	0.02743	SIL2	0.455	心電図	0.37822	LDLコレステロール	0.02743	心電図	0.37822	LDLコレステロール	0.02743	心電図	0.37822	LDLコレステロール	0.02743
ナトリウム及びクロール	0.35891	フィブリノゲン	0.01578	細菌顕微鏡検査	2.364	フィブリノゲン	0.21	ナトリウム及びクロール	0.35891	フィブリノゲン	0.01578	細菌顕微鏡検査	2.364	ナトリウム及びクロール	0.35891	フィブリノゲン	0.01578	ナトリウム及びクロール	0.35891	フィブリノゲン	0.01578	ナトリウム及びクロール	0.35891	フィブリノゲン	0.01578

表 3 Ave(TO/S) と Max(TO/S) 及び除外する医療指示 (処置項目)

Covid19の治療と関係が薄いので除外

医療指示名	Ave(TO/S)	医療指示名	Max(TO/S)
酸素吸入	2.25557	ヘパリン	8
ヘパリン	1.74117	酸素吸入	7
デキサメタゾン	1.29474	ネキシウムカプセル	6.686
インスリン	1.27465	ランソプラゾール	5.773
アセトアミノフェン	1.27235	ブレドニゾロン	5.406
ブレドニゾロン	0.93039	インスリン	4.667
ミヤBM細粒	0.79291	ミヤBM細粒	4.5
ランソプラゾール	0.74017	メジコン	3.462
ネキシウムカプセル	0.5767	デキサメタゾン	3.427
オルミエント	0.57665	ロゼレム	3.367
人工呼吸	0.57091	人工呼吸	2.636
カルボシステイン	0.50239	カルボシステイン	2.615
メジコン	0.47317	アセトアミノフェン	2.145
ロゼレム	0.46187	ペルソムラ	2.135
ペルソムラ	0.30648	オルミエント	2.026
メトクロプラミド	0.24291	センノシド	1.955
センノシド	0.18348	ツムラ補中益気湯エキス顆粒	1.921
ルネスタ	0.13383	ツムラ補中益気湯エキス顆粒	1.733
ツムラ補中益気湯エキス顆粒	0.08074	リスベリドン	1.072
リスベリドン	0.07583	ルネスタ	0.928
コルヒチン	0.03769	メトクロプラミド	0.754

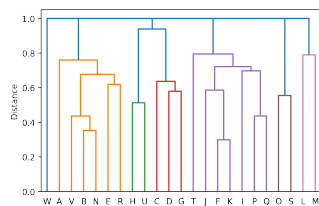
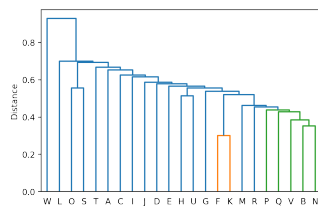


図 4 デンドログラム (最短距離法) 図 5 デンドログラム (最長距離法)

4.4.4 MSV の導出

ウォード法のクラスタリング結果のうちクラスタ 4 とクラスタ 5 について MSV を 2 種類の手法で求めた。その結果を図 11, 12, 13, 14, に示す。MSV のノードとエッジにラベルと対応した色を付けることで 3 つ以上の医療機関の SV とその共通部分を可視化できた。直接併合法は個々の SV がわかりやすいがノード数が多く冗長である。対して、距離ベース併合法は共通

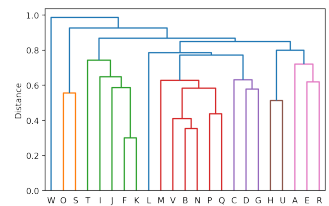
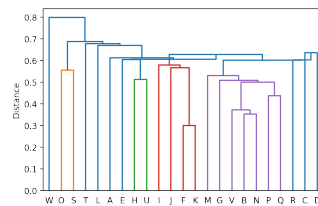


図 6 デンドログラム (重心法) 図 7 デンドログラム (群平均法)

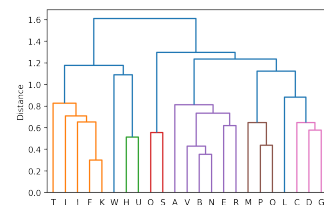


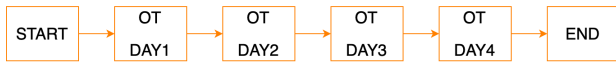
図 8 デンドログラム (ワード法)

部分をよく表しているがラベルの種類が増え MSV が複雑である。見やすさと冗長性のどちらを優先するかによって手法を選択することができる。

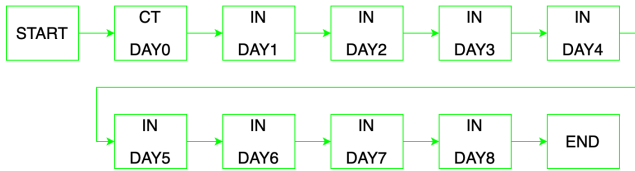
5 結 論

本研究では 3 つ以上の医療機関の頻出医療指示パターンの共通点と相違点の可視化のために、COVID-19 の第 5 波に対する電子カルテデータから各医療機関の SV を抽出し、SV 間の距離を定義して階層的クラスタリングを行い、クラスタ内で共通部分である LCSV と SV を併合した MSV を 2 つの手法で求めた。また、頻出な検査項目を除外することで処置項目の抽出を

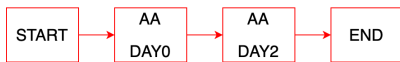
クラスタ1(F・I・J・K・T)



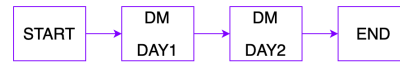
クラスタ2(H・U)



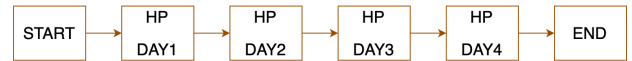
クラスタ3(O・S)



クラスタ4(A・B・E・N・R・V)



クラスタ5(M・P・Q)

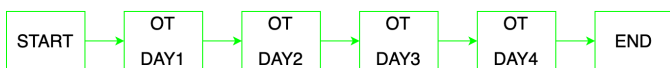


クラスタ6(C・G・D)

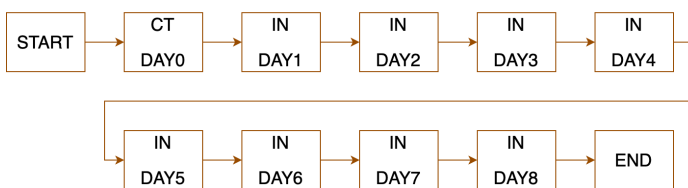


図 9 クラスタ内の LCSV(ワード法)

クラスタ1(F・I・J・K・T)



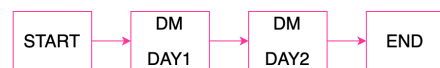
クラスタ2(H・U)



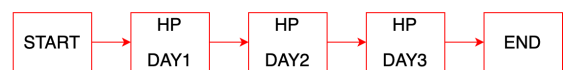
クラスタ3(O・S)



クラスタ4(A・E・R)



クラスタ5(B・N・M・P・Q・V)



クラスタ6(C・G・D)

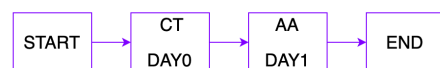


図 10 クラスタ内の LCSV(群平均法)

表 4 医療指示の略号

略号	医療指示名	概要
CT	CT 撮影	身体の内臓を撮影する検査
ECG	心電図	心臓の状態を調べる検査
UE	尿一般	尿の成分を調べる検査
TP	梅毒検査	梅毒に感染しているか調べる検査
BNP	脳性 Na 利尿ベプチド	心臓の負担を調べる検査
BUN	尿素窒素	腎臓の機能を調べる検査
GB	グロブリン別ウイルス抗体検査	様々なウイルスの抗体を調べる検査
QFT	結核菌特異的 IFN- γ 産性能	結核菌に感染しているか調べる検査
AST	アスパラギン酸アミノ基転移酵素	肝臓や心臓の異常を調べる検査
HbA1c	ヘモグロビン A1c	血糖値を評価する検査
OT	酸素吸入	酸素の投与
DM	デキサメタゾン	ステロイド剤
PD	プレドニゾロン	ステロイド剤
HP	ヘパリン	血液凝固防止剤
AA	アセトアミノフェン	解熱鎮痛剤
IN	インスリン	糖尿病治療薬
SEN	センノシド	下剤

可能にした。その結果、重要な処置項目で医療機関を分類し、クラスタ内での頻出医療指示パターンの共通点と相違点を示すことができた。

今後の課題として適切な minsup を決める指針を示す必要がある。また、クラスタの閾値については視認性を考慮してクラスタの要素数の上限を決めるなど再検討する必要がある。そして、MSV を導出するアルゴリズムについては、共通部分をよく示しつつも視認性の高い手法を検討するべきである。さらに、LCSV や MSV を自動表示するツールを開発する必要がある。

謝 辞

本研究の一部は、日本学術振興会科学研究費 (#20H04192, #21K17746) の助成及び「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 年 3 月.
- [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 Data Bases, pp. 487-499, 1994.
- [4] 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.
- [5] 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.
- [6] 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データ解析による医療行為の支援. DEIM Forum 2015, G5-1, 2015.
- [7] Y. Chen, M. Chiang, M. Ko. Discovering time-interval sequential patterns in sequence databases. Expert Systems with Applications 25, pp. 343-354, 2003.
- [8] H.H. Le, H. Edman, Y. Honda, M. Kushima, T. Yamazaki, K. 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 Science and Computational Intelligence (CSCI 2017), pp. 1726-1731, 2017.12.
- [9] H.H. Le, M. Kushima, K. Araki, H. Yokota. Differentially private sequential pattern mining considering time interval for electronic medical record systems. Proceedings of the 23rd International Database Engineering and Applications Symposium, pp. 95-103, 2019.
- [10] 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.
- [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向の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] 山田達大, 本田祐一, 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. 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クリニカルパスバリエーション分析. DEIM Forum D1-1, 2019.
- [13] 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.
- [14] 李玉清, Le Hieu Hanh, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫. シーケンシャルパターンマイニングに基づく多病院間の頻出治療パターンの比較. DEIM Forum A21-4, 2022.
- [15] Y. Li, H.H. Le, R. Matsuo, T. Yamazaki, K. Araki, H. Yokota. Comparison of sequence variants and the application in electronic medical records. Proceeding of the 33rd International Conference on Database and Expert Systems Applications (DEXA 2022), vol. 13427, pp. 117-130, 2022.
- [16] 趙子泰, 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.
- [17] 趙子泰, Le Hieu Hanh, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫. 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. DEIM Forum 2023 5a-2-3.
- [18] 黒川健人, Le Hieu Hanh, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫. 動的に医療指示を変化させたシーケンス解析における特徴的な治療パターン抽出. DEIM Forum 2023 5a-2-1.
- [19] D. Maier. The complexity of some problems on subsequences and supersequences. J. ACM. ACM Press. 25 (2): 322-336, 1978.
- [20] Q. Wang, D. Korkin, Y. Shang. A fast multiple longest common subsequence (MLCS) algorithm. IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 3, March 2011.
- [21] P. Fournier-Viger, J. C.-W. Lin, R. U. Kiran, Y. S. Koh, R. Thomas. A survey of sequential pattern mining. Data Science and Pattern Recognition, vol. 1, no. 1, pp. 54-77, 2017.
- [22] 診療報酬情報提供サービス.
<http://shinryohoshu.mhlw.go.jp/shinryohoshu/searchMenu/>

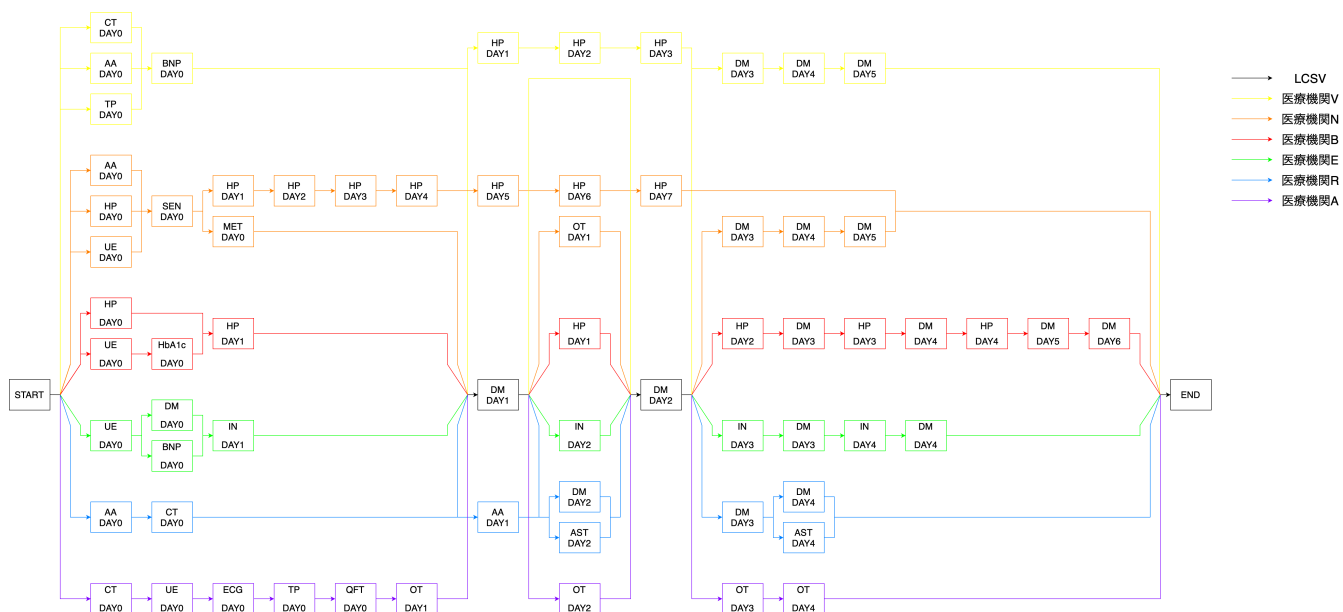


図 11 クラスタ 4 の MSV(直接併合法)

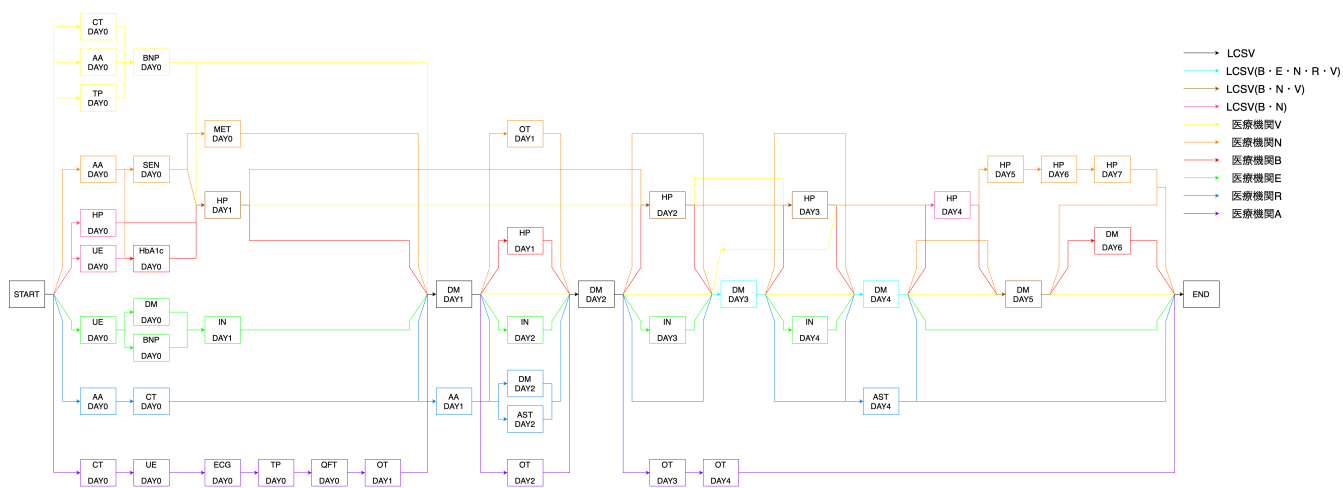


図 12 クラスタ 4 の MSV(距離ベース併合法)

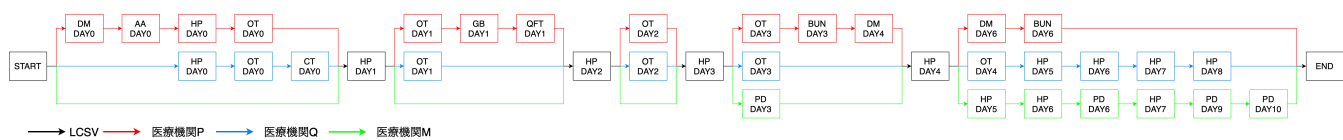


図 13 クラスタ 5 の MSV(直接併合法)

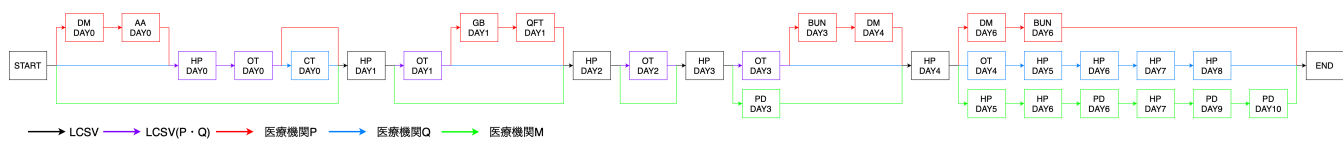


図 14 クラスタ 5 の MSV(距離ベース併合法)