

多医療機関間における頻出医療行為の利便で視認性の高い可視化手法の提案

澤村 今日子[†] 松尾 亮輔[†] 山崎 友義[†] 荒木 賢二[†] 小口 正人[†]

横田 治夫^{††} Le Hieu Hanh^{†††}

[†] お茶の水女子大学理学部 〒112-8610 東京都文京区大塚 2-1-1

^{††} 城西大学理学部 〒102-0093 東京都千代田区平河町 2-3-20

^{†††} お茶の水女子大学共創工学部 〒112-8610 東京都文京区大塚 2-1-1

E-mail: [†]g2120518@is.ocha.ac.jp

あらまし 近年、多数医療機関の電子カルテデータを跨いで頻出医療指示パターンを比較し、医療機関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、医療行為の改善等に繋げることへの研究が進んでいる。しかし、現状では分析結果の可視化処理が手動で行われており、人的コストも多くかかる上にミスが起きる可能性が高まり、医療従事者は分析結果を正確に把握することが困難である。そこで、本論文は多数医療機関における電子カルテデータの解析結果の可視化処理を自動で行う web アプリケーションを開発する。先行研究ではノード数が多い医療シーケンスの可視化手法や、医療行為の詳細な情報を図から把握することが困難である。そこでクリックでデータの詳細が出る機能や、同医療機関は同じ色のノードで同医療行為は同じ色の文字にする機能などを実現する。それにより、医療機関が増えノード数が増えても、整然かつ詳細なグラフを実現することができる。開発アプリケーションは医療従事者に評価されることによって有効性を確認する。

キーワード 可視化, クリニカルパス, web アプリケーション

1 はじめに

1.1 研究背景

近年、医療行為の効率性・安全性向上のために、単一の医療機関の電子カルテデータに分析が進んでいる [1] [2] [3] [4] [5] [6] [7]。また、特定疾病に対する医療行為の比較のために、複数医療機関の電子カルテデータに分析も行われている。安光らの研究では、複数医療機関の電子カルテデータにシーケンシャルパターンマイニングを適用し、それぞれの医療機関のシーケンスバリエーション (以下 SV) を抽出し、それらに対してクラスタリングを行い、同じグループに所属された SV に対して生成された併合シーケンスバリエーション (以下 MSV) のテキストファイルが算出された [8]。しかし、この研究ではこの MSV の可視化がまだ手動で行われており、人的コストも上がり、ミスの危険性も高まるという問題点がある。これまで、他に可視化 web アプリケーション [7] が提案されたが、データ形式が MSV のデータと互換性がなく、加えて複雑な分岐に対応していないためそのまま適用できない。

そこで、本研究は MSV の自動可視化 web アプリケーションを提案する。本提案により、より手軽に医療従事者が他医療機関の医療行為を比較できるため、医療行為のミス削減が期待できる。また、ある疾病に対する複数医療機関で共通の医療行為が一目でわかるため、まだ歴の浅い新人医療従事者の勉強にも使用でき、国内全体の医療技術向上に繋がることを期待できる。

1.2 関連研究

小林らの研究では [7]、動的な患者情報を入力として与えた際に得られる次に行うべき医療行為の推薦の結果や頻出医療指示列中のノードの出現率や分岐の要因などといった様々な情報をグラフ表現した可視化ツールを web アプリケーションとして開発した。本アプリケーションはノードの距離で時間間隔を示したり、ノードの大きさで出現率を示したりなど視覚的にわかりやすい機能を持つ。しかし単一医療機関の可視化の自動化しかできていないこと、頻出医療指示列をステージごとに切り分けて表示ができないこと、検査タイプの種類ごとに色表現を変更することの 3 つの課題が残った。

また、既存の研究では [8], [9]、多医療機関の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、そのクラスタ内で分類を行った。そこで最終結果の MSV をグラフとして可視化したのだが、その処理は手動で行われており正確性や迅速性に課題が残った。

その他、趙らの研究 [10] では、複数の医療機関における COVID-19 の診療プロセスを比較し、頻出治療パターンの違いを分析している。これにより、各施設の特徴やバリエーションが明確化され、診療の標準化および医療連携の促進が期待されている。

1.3 本研究の目的とアプローチ

本研究では、多医療機関間のクリニカルパスの分析結果を自動で可視化処理することで迅速かつ正確な結果を作成し、医療

従事者の医療行為選択支援をすることを目的とする。

同じラベル名は同じ色にする機能やノードには詳細な情報を表示する機能一目で分析結果が容易にわかるように工夫する。それによって他の作業で忙しい医療従事者も、時間をかけすぎずに分析結果を確かめることができ、より医療現場に寄り添った web アプリケーションとし、実際に医療現場で使用されることを目的とする。

1.4 本稿の構成

本稿は以下の通り構成される。第 2 節では背景知識について述べ、第 3 節では提案手法である自動化の可視化処理のアルゴリズム、搭載機能について説明する。第 4 節では実際に提案手法を用いて実際の複数医療機関の併合シーケンスバリエーションの可視化結果を報告し、最後に第 5 節でまとめと今後の課題について述べる。

2 背景知識

2.1 クリニカルパス

日本クリニカルパス学会の定義によると [11], 患者状態と診療行為の目標, および評価・記録を含む標準診療計画であり, 標準からの偏位を分析することで医療の質を改善する手法のことである。また, それとは別に電子クリニカルパスというものもある。電子クリニカルパスとは情報通信技術を用いて標準診療計画を作成し, 標準診療計画に基づく診療の実施を支援し, 患者個別の診療状況とその評価を記録し, 逸脱事例の集計と分析などを処理する医療管理手法のことである。

2.2 LCSV と MSV

李ら [12] [13] [14] によって, 最長共通サブシーケンスバリエーション (以下 LCSV) と併合シーケンスバリエーション (以下 MSV) が定義された。以下それらについて述べる。

2.2.1 シーケンス

アイテムセット I を以下のように定義する。

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

そしてアイテムセット I に対しシーケンス S を以下のように定義する。

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

ここで $\prec_s$ とは S 上の順序関係である。

2.2.2 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, 以下のようにシーケンスバリエーションと定義する。

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

$sv_j = (id, i), i \in I$. 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, 医療指示列において分岐が発生した場合, シーケンスバリエーションになる。シーケンス (入院 → 検査 → 手術 → 注

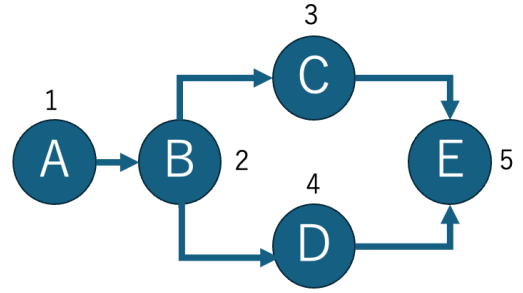


図 1 SV のシーケンスバリエーショングラフ

射 → 退院) と (入院 → 投薬 → 手術 → 処方 → 退院) に対して, 注射と処方がバリエーションと呼ばれ, この二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。SV のバリエーションに対する安全性と効率性を定量的に比較できるため, SV を抽出することは非常に有用である。また, バリエーションが現れる要因を調査することで, 特定の患者に対する治療のような医療行為の改善支援が可能となる。

以下 A, B, C のような大文字をアイテムとして説明する場合 A, B, C は id 付きアイテムの $name$ を指す。また, $A \prec B, B \prec C, B \prec D, C \prec E, D \prec E$ を $\prec A, B, (C, D), E$ と表す。

SV を可視化する方法として, シーケンスバリエーショングラフを定義する。

2.2.3 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,

$$\forall e = (sv_{out}, sv_{in}) \in E, sv_{out} \prec_{SV} sv_{in}$$

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 がアイテムで, $SV = \prec A, B, (C, D), E$ のシーケンスバリエーショングラフが図 1 のように示されている例として A, B, C, D, E がアイテムで, $SV = \prec A, B, (C, D), E$ のシーケンスバリエーショングラフが図 1 のように示されている。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扱う。

2.2.4 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。

$$\forall cs_i, cs_j \in CS, cs_i, cs_j \in S_\alpha \cup S_\beta \text{ かつ } cs_i \prec_{CS}$$

$$cs_j \Leftrightarrow cs_i \prec_\alpha cs_j \wedge cs_i \prec_\beta cs_j.$$

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。

2.2.5 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

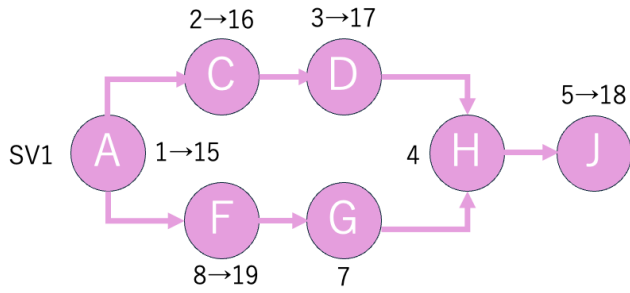


図 3 MSV の例

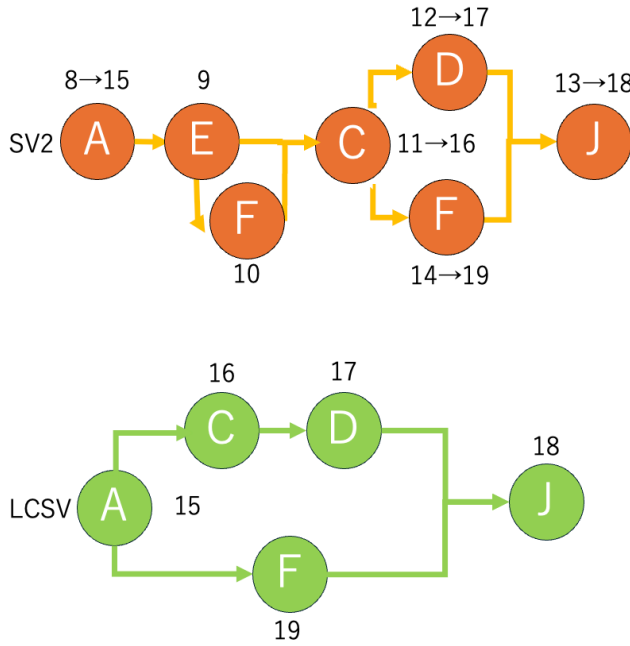


図 2 LCSV の例

意の共通部分列 CS に対し, $|CS| \leq |LCSV|$. $|CS|$ は CS のノード数を表す.

例として, シーケンス $\langle A, B, C, D \rangle$ と $\langle A, C, B, A, D \rangle$ を考える. 長さ 2 の共通部分列が 5 つ存在する: $\langle A, B \rangle$, $\langle A, C \rangle$, $\langle A, D \rangle$, $\langle B, D \rangle$, $\langle C, D \rangle$. 長さ 3 の共通部分列が 2 つ存在する: $\langle A, B, D \rangle$, $\langle A, C, D \rangle$. それ以上長い共通部分列が存在しない. 従って最長共通部分列が $\langle A, B, D \rangle$ と $\langle A, C, D \rangle$ である.

2.2.6 共通サブシーケンスバリエント (CSV)

シーケンスバリエント SV_α と SV_β に対し, 共通サブシーケンスバリエント CSV 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. $\forall csv_i, csv_j \in CSV, csv_i, csv_j \in SV_\alpha \cap SV_\beta$ かつ $csv_i \prec_{CSV} csv_j \Leftrightarrow csv_i \prec_\alpha csv_j \wedge csv_i \prec_\beta csv_j$. CSV のサブシーケンスバリエントも 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ントを定義する.

2.2.7 最長共通サブシーケンスバリエント (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ント $LCSV$ を下記の条件を満たす共通サブシーケンスバ

リエントと定義する. SV_α と SV_β に対する任意の共通サブシーケンスバリエント 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.

2.2.8 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. $\forall sv_i, sv_j \in SV_\alpha \cap SV_\beta, sv_i, sv_j \in MSV$ かつ $sv_i \prec_{MSV} sv_j \Leftrightarrow sv_i \prec_\alpha sv_j \wedge sv_i \prec_\beta sv_j$. $\forall msv_i, msv_j \in MSV, msv_i \prec_{MSV} msv_j \Rightarrow (msv_i, msv_j \in SV_\alpha \wedge msv_i \prec_\alpha msv_j) \vee (msv_i, msv_j \in SV_\beta \wedge msv_i \prec_\beta msv_j)$. また, $LCSV$ に出現したノードに label α, β 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.

3 提案手法

本節は提案手法について説明する. まず, 本提案手法の想定する利用場面のついて説明したあとに, 可視化の流れを述べる. 次に, 可視化した MSV の例を紹介し, 開発した可視化ツールの特徴を説明する.

3.1 想定する利用場面

医療従事者はパソコンやタブレットなどの電子機器を使用し, 本可視化アプリケーションを使用する. 医療従事者は同じグループに分類された複数医療機関のある疾患の治療に関する MSV を確認し, 医療機関間の共通医療行為と差異医療行為を簡単に確認することができる. それによって, 対象となる疾病の特徴を瞬時に把握することができ, 医療改善につながる.

3.2 可視化の流れ

図 4 は可視化手法の流れを表示する. MSV テキストファイルを自動的に可視化しやすくするために, json ファイルに変換する. Javascript で可視化を行うため, 最終的に HTML 形式のファイルで出力される.

3.3 json 変換

MSV テキストファイルの一例の内容を図 5 に示す. このファイルには複数のノードの情報を含む. その中で, 一つノードには name, label, id, nextNodes の 4 つの要素が記載されて

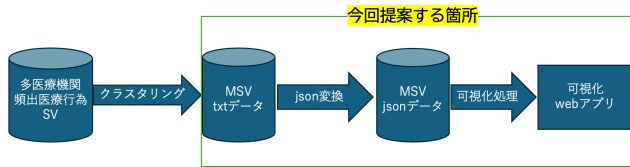


図4 可視化の流れ

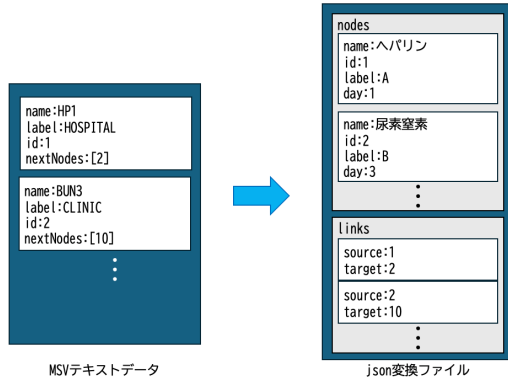


図5 MSV テキストファイルからの変換

いる。name はノードの名前と日にちの情報からなる。label はこのノードがどの SV に所属しているかを示す。id がそのノードの独自の識別子であり、nextNodes はこのノードと繋がり先の複数ノードの id を含む。この情報を用い json ファイルに変換する。変換後の json ファイルにはノード群の情報を管理する nodes とノードとノードの間の枝を管理する links の 2 種類の情報を生成する。nodes には各ノードの情報は name, id, label と day が含まれており、links にはノードとノードの間の枝の全ての情報が作成される。

図5について説明する。左が元の txt データで、右が変換した json データである。1 番上のデータを例とする。nodes には上記の 4 種のデータが含まれる。name は、元は英語の略称名「HP」と格納されている。それを json データでは日本語名「へパリン」に変換して格納する。id は 1 から順に並んでいない場合は数字を新たに割り振り直し格納する。元のデータでは「HOSPITAL」となっている label であるが、json データでは匿名処理を行ってから label に格納する。この作業により医療データの機密性が守られる。また、name の末尾の数字は day 情報を示している。そのためこの例では day が 1 となる。そして links には source と target の 2 種の情報が含まれる。source には自身の id 情報が、target には nextNodes 情報がそれぞれ格納される。

3.4 可視化処理

第 3.3 節で生成された json ファイルから d3-sankey ダイアグラム [15] の形で自動的にグラフ化を行う。ノードが並べられるアルゴリズムとしては、まず name が start と end のものがあるので、それらを両端に設置する。そして次に day の値が小さい順に左から並べていく。また、SV ごとに高さを設定する。

ここでは、視覚的にわかりやすくするために、label の情報

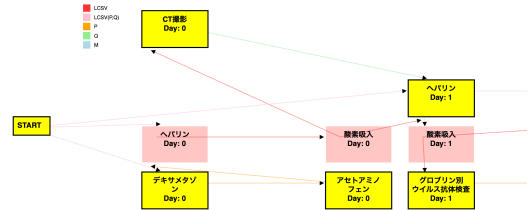


図6 ハイライトされたノード

を利用しノードと枝の色を決める。同じ SV に所属するノードと枝が原則的に同じ色にする。さらに、視認性を高めるためにノードの上に医療行為の略語ではなく、下の医療行為名を表示するように工夫する。これによって、誰でもグラフを見るだけでノードの意味をすぐ把握することができる。

また、ノードをクリックするとそのノードと同じ医療機関のノードがハイライトされる。例えば医療機関 A をクリックすると、他の全ての医療機関 A のノードと、医療機関 A が含まれた、全ての LCSV のノードがハイライトされる。

4 実験

本節では提案手法の有効性を確認するために、実データを用いた評価実験を行う。

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。その中で、同じグループに分類された医療機関 M, P, Q の MSV の結果を自動的に可視化することを目指した。本研究は一般社団法人ライフデータイニシアティブの利用目的等審査委員会による審査を受け、承認された上で研究を進めた (審査番号 2021_MIL_0007_A002)。

4.2 実験結果

図7に示した通り、医療機関 M, P, Q の MSV を自動的に可視化した結果が表示された。3 つの医療機関の共通 SV は、赤色で表示される。そして、医療機関 P と Q の共通 SV に属するノードはピンク色で区別することができる。その他、医療機関 M, P, Q のみの SV にある医療行為はそれぞれ水色、オレンジ色、薄緑色で描かれている。色に加えて、各ノードには医療行為の名前と発生した日にち情報も記載されているためグラフを見る人は誰でも SV の内容を簡単に把握することができる。

また、ノードをクリックするとそのノードと同じ医療機関の情報が含まれているノードがハイライトされた。例えば医療機関 P をクリックすると、医療機関 P のノード全てと、LCSV(P, Q) と、LCSV(P, Q, M) のノードがハイライトされる。このように、MSV テキストデータから手動でグラフ化を行った既存研究と比べて、高い視認性のグラフを自動的に生成することができることを確認できた。



図 7 自動で行なった MSV の可視化

5 おわりに

5.1 ま と め

本研究では、複数医療機関間の MSV を自動で可視化処理する web アプリケーションを開発した。手法の流れとしては、MSV のテキストファイルに含まれる医療行為情報、日付情報、ノード情報を、d3-sankey でプロットできる形に json ファイルとして変換した。その json ファイルを Javascript を用いて d3-sankey の形で可視化処理を行い、html ファイル形式で出力する。これによって、MSV の可視化も早く、正確に、わかりやすくなり医療界に対して大きく貢献できると確認できた。

5.2 今後の課題

可視化グラフにはより高度なインタラクティブな機能を開発する予定である。例えば、ユーザはある部分のみ着目したい場合に、ノードをクリックすることでより詳細な情報を提示させたりできるような機能も考えられる。また、実際に医療従事者に使用していただき評価をいただくことにより医療現場で使いやすいものになるように開発する。加えて、医療機関ごとに「その医療行為を選んだ理由」が表示される機能や、特定のノードをクリックするとそのグループだけが画面に表示される機能などを追加し、より利便かつ視認性の高いものとしていく。

謝 辞

本研究の一部は日本学術振興会科学研究費 (#24K02943) の助成からの支援によって行われた。

文 献

- [1] Yuichi Honda, Muneo Kushima, Tomoyoshi Yamazaki, Kenji Araki, and Haruo Yokota. Detection and visualization of variants in typical medical treatment sequences. *Proceedings of the 3rd International Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH) in conjunction with the 43rd International Conference on Very Large Data Bases (VLDB '17)*, pp. 88–101, 2017.
- [2] 山田達大, 本田祐一, 荻原正彬, 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. In *DEIM Forum*, No. D1-1, 2019.
- [3] Hieu Hanh Le, Tatsuhiko Yamada, Yuichi Honda, Takatoshi Sakamoto, Tomoyoshi Yamazaki Ryosuke Matsuo, Kenji Araki, and Haruo Yokota. Methods for analyzing medical-order sequence variants in sequential pattern mining for electronic medical record systems. *ACM Transactions on Computing for Healthcare*, Vol. 4, No. 1, pp. Article No.3, 1–28, 2023.
- [4] 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. In *DEIM Forum*, No. G5-1, 2015.
- [5] Yoshitaka Arahori Muneo Kushima Tomoyoshi Yamazaki Kenji Araki Keishiro Urakaki, Tomoyuki Hosaka and Haruo Yokota. Sequential pattern mining on electronic medical records with handling time intervals and the efficacy of medicines. In *In Proceedings of the first IEEE Workshop on ICT Solutions for Health in conjunction with the 21st IEEE International Symposium on Computers and Communications*, pp. 20–25, 2016, pp. 20–25, 2016.
- [6] 牧原健太郎, 荒堀喜貴, 渡辺陽介, 串間宗夫, 荒木賢二, 横田治夫. 電子カルテシステムの操作ログデータの時系列分析による頻出シーケンスの抽出. In *DEIM Forum*, No. F6-2, 2014.
- [7] 小林梨華, 坂本任駿, Le Hieu Hanh, 荒木賢二, 横田治夫. 動的な患者情報を用いた医療行為推薦を支援するための医療シーケンスの可視化. In *DEIM Forum*, No. C25-4, 2021.
- [8] 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. In *DEIM Forum*, No. 5b-6-3, 2023.
- [9] Hieu Hanh Le, Yuki Yasumitsu, Ryosuke Matsuo, Tomoyoshi Yamazaki, and Haruo Yokota. A clustering-based sequence variants analysis method for electronic medical records of multimodal institutions. In *Proceedings of the 7th IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 653–659, IEEE, 2024., pp. 653–659. IEEE, 2024.
- [10] 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.
- [11] 日本クリニカルパス学会. <https://www.jsccp.gr.jp/index.html>. オンライン, アクセス 2025 年 1 月 8 日.
- [12] 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. In *DEIM Forum*, No. A21-4, 2022.
- [13] Yuqing Li, Hieu Hanh Le, Ryosuke Matsuo, Tomoyoshi Yamazaki, Kenji Araki, and Haruo Yokota. Comparison of sequence variants and the application in electronic medical records. In *Proceedings of the 33rd International Conference on Database and Expert Systems Applications (DEXA)*, pp. 117–130, Springer, 2022., pp. 117–130. Springer, 2022.
- [14] 李玉清, Le Hieu Hanh, 松尾亮輔, 山崎友義, 荒木賢二, 横田治夫. シーケンスバリエーションの比較と電子カルテの分析への応用. 日本データベース学会データドリブンスタディーズ, Vol. 1, No. 5, pp. 1–8, 2023.
- [15] d3/d3-sankey: Visualized flow between nodes in a directed acyclic network. <https://github.com/d3/d3-sankey>. オンライン, アクセス 2025 年 1 月 8 日.