

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

小林 莉華[†] 坂本 任駿[†] Le HieuHanh[†] 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

^{††} 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: rika.kobayashi@de.cs.titech.ac.jp

あらまし 病院での電子カルテの普及により、二次利用による医療の効率化や質の向上が期待されている。先行研究では、電子カルテのデータからクリニカルパスと呼ばれる医療指示列の典型例を抽出し、また、それをグラフで表現することで医療従事者の視認性を高める可視化ツールを作成してきた。本研究では、患者の動的な情報を用いた医療行為の推薦を支援するための可視化ツールを導入する。動的な患者情報を入力として次に行うべき医療行為の推薦を行う際、医療従事者がその内容を納得できるように、視覚的に分かりやすく示すことが望まれる。推薦の前提となる医療シーケンスのグラフ表現中の分岐等の情報が、動的な患者情報によって変化する様子を可視化することにより、推薦の理由となる医療情報や他の医療行為の選択の可能性を、医療従事者にとって把握しやすく提示できる方法を提案する。

キーワード 電子カルテ、シーケンシャルパターンマイニング、クリニカルパス可視化

1 はじめに

1.1 研究背景

近年、大規模病院を中心に電子カルテが普及している。電子カルテは、対象患者の検索や、患者への医療指示履歴の閲覧、共有を高速化することを可能にしており、医療指示全体の標準化、改善に貢献している。また、電子カルテは診療記録に用いられるだけでなく、二次利用による情報活用が期待されている。

二次利用の例としては、特定の病気の患者に対しての典型的な医療指示の流れ「クリニカルパス」を抽出することが挙げられる。従来の典型クリニカルパスは、医療従事者の医学的経験に基づいて作成されていたが、膨大な過去の事例から人力で作成することは容易ではなかった。このような背景から、電子カルテをデータ工学の観点から計算機を用いて分析し、医療行為改善の支援を行う研究が始まった。

電子カルテのデータを分析することでクリニカルパスを抽出し、またそれをグラフ表現した医療従事者にとって視認性の高い可視化ツールを作成することは、医療の質や効率の向上に繋がると考えられる。

1.2 関連研究

牧原らの研究 [1] では、電子カルテの履歴から、ある患者に対して行った医療指示をアイテム、医療指示の流れをシーケンス、全ての患者の医療指示の流れをデータベースと考えることで、シーケンシャルパターンマイニング（以下、SPM）を適用し、頻出シーケンスの抽出を行った。しかし、牧原らの手法には基準となる医療指示の後の部分のマイニング結果が膨大になってしまうこと、医療指示間の時間間隔を考慮していないことの

二つの課題が残った。

佐々木ら [3] は、この二つの課題解決のため、飽和オーダー列と呼ばれる概念 [4] を導入することで、出力の情報量を損なわずに出力数を削減し、またタイムインターバル SPM [5] (以下、TI-SPM) を PrefixSpan [6] に用いることで、二つのアイテム間の時間間隔を考慮した抽出を行った。しかし、TI-SPM はタイムインターバル（以下、TI）を人為的に定めるため、最適な時間間隔を定めることが困難であるという課題があった。また、これらの研究では注射のような薬剤情報を含む医療指示において、投与薬剤の種類の情報が含まれないという課題もあった。

浦垣ら [7] は、これらの課題の解決のために、アイテムを（大別 Type, 詳しい説明 Explain, 薬効コード Code, 薬剤名 Name）の四つ組によって構成することで、薬剤情報を取り入れた抽出を行った。さらに、アイテム間の時間間隔を統計処理によって定めることで、予め最適な TI セットを人為的に求めなければならないという問題に課題解決を図った。保坂ら [8] は、三つのスコアリング手法を導入することで最適な TI セットを事前に決定し、頻出なクリニカルパスの抽出を行った。しかし、これらの先行研究には医学的に解析の価値が高いクリニカルパスバリエーションに対しての処理がなされていない、出力結果が頻出シーケンスの文字列であるため、クリニカルパスバリエーションを含む場合の視認性が悪いという二つの課題があった。

本田ら [9] は、この二つの課題解決のため、シーケンス同士の共通部分を医療指示が出される日の情報を考慮して検出していくことによって、共通部分を持つシーケンスの差異「シーケンスバリエーション（以下、SV）」を検出した。そしてバリエーションを含むクリニカルパスをグラフで表現して、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。また、山田ら [10] はシーケンスや SV の評価が円滑かつ正確に行えるような可視化

手法の提案を行った。しかし、これらの可視化は一般的な医療指示の流れを示すクリニカルパスやそこから得られる分岐の評価のみを対象としており、個々の患者における性別や年齢、過去に行った医療指示履歴などの情報を考慮していない。

1.3 本研究の目的

本研究は、医療従事者に向けて患者に対する適切な医療行為の選択を支援をすることを目的とする。そのために、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を推薦の結果や、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する。電子カルテデータの解析結果の文字列による出力は、プログラムに対する知識のない人にとって理解することは容易ではないが、出力を可視化することで、推薦の理由となる医療情報や他の医療行為の選択の可能性を、医療従事者にとって把握しやすく提示する方法を提案する。

1.4 本稿の構成

本稿は以下の通り構成される。2 節では、本研究で利用する頻出医療指示列の抽出と次に行うべき医療指示の推薦の手法を背景知識として説明する。3 節では、得られたクリニカルパスを動的な患者情報を用いて可視化していく手法を提案手法として述べる。最後に、4 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 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る [2] 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, sid を組みとする要素からなる。

2.2 TI-SPM

当初 Agrawal らの提案手法 [2] は、アイテム間の時間間隔を考慮していない。例えば、2021 年 1 月 1 日に注射を行い、その翌日に手術を行うようなシーケンスと 2021 年 1 月 1 日に注射を行い、その 1 カ月後に手術を行うシーケンスを同一のシーケンスとみなしていた。この 2 つを異なるものとして扱うべきである医療指示のよう時間間隔が重要なシーケンスに対するマイニング手法として、Chen らは TI-SPM と呼ばれる手法を提案した [5]。TI-SPM は、時間間隔を含んだ SDB D 、最小支持度 $MinSup$ ($0 \leq MinSup \leq 1$)、事前に設定した時間間隔 TI-セットを入力とすることで、TI-頻出シーケンスを出力する。

2.3 T-PrefixSpan

TI-SPM である I-PrefixSpan [5] はアイテム間の時間間隔が固定となりやすいデータに対して用いられるアルゴリズムであり、TI-セットとして事前に設定した入力が必要となる。このためアイテム間の時間間隔は定めた TI-セットによって変化して

しまい、正確なものにならないという問題点が存在した。浦垣らは TI-SPM の問題点を解決するべく、T-PrefixSpan [7] を導入した。T-PrefixSpan は 2 アイテム間の時間間隔を外れ値処理を含む統計情報を用いて表現を行った。これにより、TI-セットを事前に決定する必要なく、時間間隔情報を含んだパターンの抽出を可能とした。

2.4 T-CSpan

T-PrefixSpan は PrefixSpan を元に行っているため、計算時間が長いという問題がある。Le らはその解決のため、CSpan を元にした、T-CSpan [11] を導入した。T-CSpan も T-PrefixSpan と同様に時間間隔は統計情報を用いているため、TI-セットを必要としない。以下に T-CSpan に関する概念の定義と、T-CSpan の説明を行う。

定義 1. タイムアイテム (i, t)

アイテム集合 I が与えられ、アイテム $i \in I$ の発生時刻が t であるとき、 i と t の組 (i, t) をタイムアイテムと定義する。

定義 2. タイムシーケンス s

タイムアイテムからなる順列 s をタイムシーケンスと定義し、以下のように表す。

$$s = \langle (i_1, t_1), (i_2, t_2), \dots, (i_n, t_n) \rangle$$

また、タイムシーケンス s の長さ $length(s)$ を $length(s) \equiv n$ とし、シーケンス $O_s = \langle i_1, i_2, \dots, i_n \rangle$ を s のオリジナルシーケンスと呼ぶ。

定義 3. 時間間隔 TI_k

タイムシーケンス $s = \langle (i_1, t_1), (i_2, t_2), \dots, (i_n, t_n) \rangle$ において、時間間隔 TI_k を以下のように定義する。

$$TI_k \equiv t_{k+1} - t_k (k = 1, 2, \dots, n-2, n-1)$$

定義 4. タイム SDB D

タイムシーケンス集合 S が与えられたとき、タイム SDB D を以下のように定義する。

$$D \equiv \{(sid, s) \mid sid \text{ は識別子}, s \in S\}$$

ただし、任意の 2 要素の識別子 sid は異なる値を持つ。

タイム SDB に含まれる全てのタイムシーケンスのオリジナルシーケンスからなる SDB をオリジナル SDB と定義したとき、タイム SDB から抽出されるタイム頻出シーケンス及びを飽和タイム頻出シーケンスを以下のように定義する。

定義 5. タイム頻出シーケンス P

最小支持度 $MinSup$ ($0 \leq MinSup \leq 1$)、タイム SDB D が与えられたとき、 $P = \langle i_1, X_1, i_2, \dots, i_{n-1}, X_{n-1}, i_n \rangle$ ($\forall j \ i_j$ はアイテム, $\forall k \ X_k$ は 5 つの値の組 $(min_k, mod_k, ave_k, med_k, max_k)$) の $O_P = \langle i_1, i_2, \dots, i_{n-1}, i_n \rangle$ を考えた時、 O_P が D のオリジナル SDB の $MinSup$ において頻出シーケンスであれば、タイム頻出シー

ケースとする。ただし、 $\min_k, \text{mod}_k, \text{ave}_k, \text{med}_k, \text{max}_k$ は以下に示す。

オリジナルシーケンスを構成したとき、 O_P をサブシーケンスとするような D に存在する全てのタイムシーケンス $S = \langle i'_1, t_1, i'_2, t_2, \dots, i'_{m-1}, t_{m-1}, i'_m \rangle$ において、 $i_k = i'_{j_k}, i_{k+1} = i'_{j_{k+1}}$ を満たす $k = 1, 2, \dots, n-1, 1 \leq j_1 < j_2 < \dots < j_{n-1} < j_n \leq m$ を考えたとき、時間間隔 $TI_k = t'_{j_{k+1}} - t'_{j_k}$ の集合 Set_{TI_k} を構成できる。このとき、 $X_k = (\min_k, \text{mod}_k, \text{ave}_k, \text{med}_k, \text{max}_k)$ において、 \min_k を Set_{TI_k} における最小値、 mod_k を Set_{TI_k} における最頻値、 ave_k を Set_{TI_k} における平均値、 med_k を Set_{TI_k} における中央値、 max_k を Set_{TI_k} における最大値とする。ここで、時間間隔 $X_j = (\min_j, \text{mod}_j, \text{ave}_j, \text{med}_j, \text{max}_j)$ ($1 \leq j < n$) に対して、 $\min_j = \text{max}_j$ が成り立つとき、アイテム i_j 及び i_{j+1} の時間間隔は一定となる。特に $\min_j = \text{max}_j = 0$ の場合は、同日に起こるとする。また、 O_P を P のオリジナルパターンとする。

定義 6. 飽和タイム頻出シーケンス

タイム SDB D から抽出したタイム頻出シーケンス集合 Σ に属する A に対し、以下の条件を満たす $B \in \Sigma \setminus A$ が存在しないとき、 A を飽和タイム頻出シーケンスと定義する。

(1) A, B のオリジナルパターンを A', B' としたとき、 $A' \subseteq B'$ が成り立つ。

(2) (1) が成り立つとき、 $A = \langle a_1, T_1, a_2, \dots, a_{n-1}, T_{n-1}, a_n \rangle, B = \langle b_1, T'_1, b_2, \dots, b_{m-1}, T'_{m-1}, b_m \rangle$ としたとき、 $a_k = b_{j_k}, a_{k+1} = b_{j_{k+1}}$ となる $k = 1, 2, \dots, n-1, 1 \leq j_1 < j_2 < \dots < j_n \leq m$ が存在する。このとき、全ての $T_k = (\min_k, \text{mod}_k, \text{ave}_k, \text{med}_k, \text{max}_k)$ 、 $T'_{j_k} = (\min'_{j_k}, \text{mod}'_{j_k}, \text{ave}'_{j_k}, \text{med}'_{j_k}, \text{max}'_{j_k})$ に対して、 $\min_k \geq \min'_{j_k}$ かつ $\text{max}_k \leq \text{max}'_{j_k}$ が成立する。

(3) $\text{Sup}(A) \leq \text{Sup}(B)$

ここでタイム頻出シーケンスのサポート値 $\text{Sup}(A)$ を $\text{Sup}(A) \equiv |\{s \mid s \subseteq S, (sid, S) \in D, sid \text{ は } S \text{ の識別子}\}|$ と定義する。

T-CSpan は SDBD, 最低支持度 MinSup を入力とする。まず D 中の全ての頻出単一タイムシーケンスを決定する。次に、各 k -タイムシーケンスにおいて、射影データベースを構成し、射影データベース内で頻出アイテムを見つけ、タイムシーケンスを生成し、最後に生成された頻出タイムシーケンスに対して、各アイテム間の時間間隔の結果を計算する。T-CSpan は、発生チェックを利用して計算結果に飽和タイム頻出シーケンシャルパターンのみを追加するため、効率的なアルゴリズムとなっている。

2.5 多変量解析によるシーケンスの分岐要因推定

本田ら [12] は、SPM で抽出された頻出医療指示シーケンスに存在するバリエーションが、シーケンスを生成した際に生じる分岐にどのように影響したか要因を推定する手法を提案した。この手法は患者の年齢や入院時期などの静的情報と体温や血圧などの動的情報の両方を取り入れた多変量解析を行い、分岐に影響した要因を推定するものとなっている。

響した要因を推定するものとなっている。

2.6 SID を保持する SPM

山田ら [10] は、SPM によって得られたデータを分析する際に抽出元のデータベースの情報が必要である場合があることに着目し、SPM 時に SID を保持しながら行う手法を提案した。例として、表 1 のようなタイム SDB D において、最小支持度 $\text{MinSup} = 0.4$ におけるマイニングを考える。

表 1 タイム SDB D

sid	タイムシーケンス
1	$\langle (A, 1), (B, 3), (C, 7), (E, 10) \rangle$
2	$\langle (A, 1), (B, 4), (E, 7) \rangle$
3	$\langle (A, 2), (B, 6), (B, 9) \rangle$
4	$\langle (A, 2), (B, 5), (F, 10) \rangle$
5	$\langle (A, 2), (B, 7) \rangle$

タイム頻出シーケンスは、 $\{\langle A \rangle, \langle 1, 2, 3, 4, 5 \rangle, \langle B \rangle, \langle 1, 2, 3, 4, 5 \rangle, \langle E \rangle, \langle 1, 2 \rangle, \langle A, (2, 3, 3, 3, 5), B \rangle, \langle 1, 2, 3, 4, 5 \rangle, \langle B, (3, 5, 5, 5, 7), E \rangle, \langle 1, 2 \rangle, \langle A, (2, 2, 2, 2, 3), B, (3, 5, 5, 5, 7), E \rangle, \langle 1, 2 \rangle\}$ となる。また、飽和タイム頻出シーケンスは $\{\langle A \rangle, \langle 1, 2, 3, 4, 5 \rangle, \langle B \rangle, \langle 1, 2, 3, 4, 5 \rangle, \langle A, (2, 3, 3, 3, 5), B \rangle, \langle 1, 2, 3, 4, 5 \rangle, \langle A, (2, 2, 2, 2, 3), B, (3, 5, 5, 5, 7), E \rangle, \langle 1, 2 \rangle\}$ となる。これにより、SID 情報をもとに電子カルテデータから必要な情報を取得することが可能となる。

2.7 患者情報と医療指示履歴に基づいた医療指示推薦

SPM により抽出された頻出シーケンス集合に含まれるアイテム i について患者数に対する頻度 ρ を算出し、頻出シーケンス集合で生じたバリエーションに含まれる $i_k \dots$ について上限頻度 ϕ 、下限頻度 ψ を閾値として、 ρ が

1. $\rho \leq \phi$ の時、 $i_k \dots$ をグループ化
2. $\psi \leq \rho < \phi$ の時、 $i_k \dots$ を分岐とする
3. $\rho < \psi$ の時、 $i_k \dots$ を除外

という処理を行う。これにより得られたシーケンスについて実施時刻情報を用いてグループ化された i の時間間隔を算出し、グループ内での時間関係を明確化することで、頻出シーケンス集合から併合シーケンスを生成する。

また、シーケンス中の情報である、シーケンスのどのアイテムを参照しても同じ値が得られる静的情報とシーケンスの参照する要素によって得られる値が異なる動的情報を用いて多変量解析を行うことで、分岐の要因を推定する。

以上により得られた併合シーケンスと分岐の要因、患者情報と医療指示履歴に基づいて、次にオーダーされる医療指示を推薦する。

3 提案手法

本章では、動的な患者情報を用いた医療行為の推薦を支援す



図 1 頻出医療指示列の可視化図

るための、頻出医療指示列の可視化の手法について説明する。

頻出医療指示列の抽出から可視化までの流れは図 2 のようになっている。本研究では四角で囲まれた部分を行っていく。ある病気についての頻出医療指示列をグラフ表現するだけでなく、年齢や入院歴、医療指示履歴といった動的な患者の情報を入力することで得られる、次に行うべき医療行為を可視化する。

可視化には D3.js [13] を用いることで、Web 上でグラフ表現する。

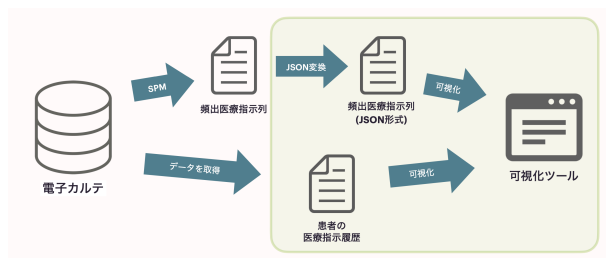


図 2 可視化までの流れ

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端末から Web アプリケーションを使用し、対象患者の症例をこのツールから選択することで、頻出医療指示列と対象患者の情報を比較し医療行為を選択していくことを想定している。特に比較的经验の浅い医者がこのような場面の利用の対象となると考えられるが、ベテランの医者に向けても、医者自身の経験に基づいて選択した医療指示とこのツールが示す頻出医療指示列を比較することで、様々な選択の可能性を確認するといった利用が考えられる。さらに、医者だけでなく看護師が患者の医療指示を確認する場面や、患者に対し医療指示の説明をする場面などでの利用も想定している。

そのため、誰もが手元の端末からアクセスし可視化グラフを確認できるよう、Web アプリケーションとして利用可能な状態になっている。

3.2 頻出医療指示列の可視化

可視化の例として、TUR-Bt（膀胱腫瘍切除手術）について頻出医療指示列を抽出した結果の全体像を図 1 に示す。

各医療指示をノードとして色のついた円形で、医療指示の流れをパスとして灰色の線で示す。図 1 より、医療指示の流れを図として視覚的に把握しやすい形で示すことができていること

がわかる。

分岐したパスなどで医療指示の差異を確認できるよう、同じ医療指示は同じ色のノードで表現している。また、分岐点ではどのような要因でその分岐が起きたのかを確認できるよう、分岐先のパス上に要因を示している。

ノードをクリックすると、図 3 のように対象医療指示における投与薬剤や薬剤コードといった詳細の表示をすることができる。

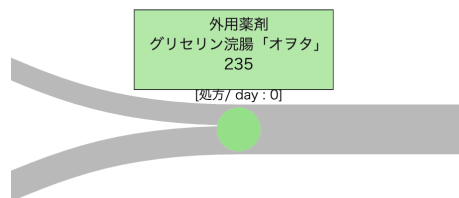


図 3 ノードの詳細

3.3 入力データの処理

クリニカルパス可視化のプログラムには、入力となる頻出クリニカルパスを JSON 形式のデータで渡している。

入力データの例として、図 4 に頻出医療指示列中に現れる手術に関する医療指示のデータを示す。頻出医療指示列はこのようなデータの配列となっている。各ノードには id を振っており、また次に現れるノードの id を next_item_id に渡すことで、分岐を含む併合シーケンスも 1 次元の配列として表すことができる。

```
{
  "id": 6,
  "order_type": "手術",
  "order_explain": "膀胱悪性腫瘍手術",
  "order_name": "",
  "code": "",
  "day": "0",
  "time_interval": [33],
  "frequency": 100,
  "pre_item_id": [5],
  "next_item_id": [7]
},
```

図 4 入力データの例

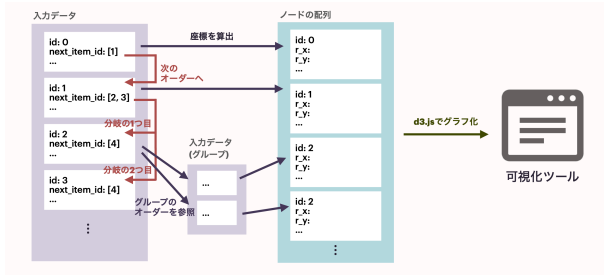


図 5 データの処理の流れ

このような入力データからグラフの生成までの流れを図5に示す。入力データ内の配列が持つ、ノードに対応する各要素に対し、グラフ内の位置を決定する x 座標と y 座標を算出し、これらの値を持った新しい要素を新たな配列に追加していく。この処理を、next_item_id に含まれる値を持つノード探していくことで、再帰的に行っていく。分岐部分では、x 座標は分岐前のノード、の隣に現れるように計算をし、また y 座標は next_item_id における配列の index により他の分岐先と位置を縦にずらしていくように計算する。これらの処理により、新しくできたノードに関する配列から、d3.js を用いてグラフ化する。

3.4 グルーピングされた医療指示の表示

グルーピングされた複数医療指示は、図6のようにグループ内の各医療指示を1つつつノードとして表示する。グループ内の医療指示は順序関係が明確でない場合があるため、縦に並べている。

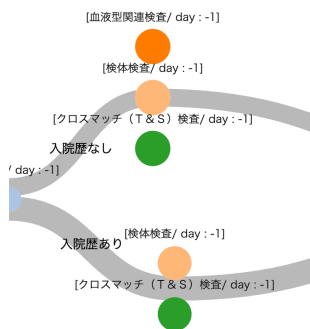


図 6 グルーピングされたノードを全て表示

グルーピングされた医療指示を、他の医療指示と同様にパスの上に乗った1つのノードとして表現したのが図7である。

グループ内の各医療指示については、ノードをクリックすることで表示される詳細表示に記載している。しかしこの場合、グループ内に存在する医療指示を把握するまでのステップが図6の場合よりも多くなる。またグループ内の医療指示数や分岐における差異を視覚的に把握しづらいという問題もある。そのため、図6のように一目でグループ内の医療指示を確認できる形で表示することを選択した。

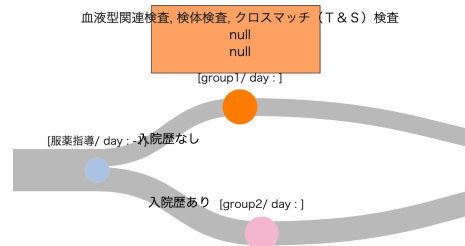


図 7 グルーピングされたノードをまとめて表示

3.5 出現率の可視化

医療従事者が頻出医療指示列から医療指示を選択する際に、その信頼性を示す一つの指標として、全患者のデータにおける出現率が有用であると考え、可視化をする。

Zhang ら [14] の手法では、SPM で得られた頻出医療指示列を、ノードの大きさやエッジの太さの異なる有向グラフとして表現することで可視化している。この際、SPM に利用した全患者のデータにおける各医療指示の出現率をノードの相対的な大きさで、ノードからノードへの遷移の頻度をエッジの太さで表している。

これを参考に、各医療指示の出現率をノードの大きさを変化させることで表現する。

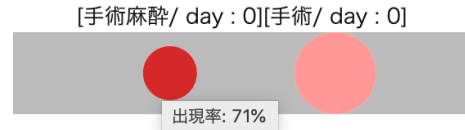


図 8 出現率の可視化

図8の右のノードは手術を表しており、出現率は100%でパスの太さいっぱいノードが表示されている一方、左のノードは出現率が71%で小さく表示されている。図1からもそれぞれのノードの大きさの違いを確認できる。また、図8のように、ノードにカーソルを合わせると対象医療指示の出現率の具体的な値が表示される。

また、分岐中の医療指示の出現率には、全患者中の出現率と分岐に対応する患者中の出現率の2種類の値を、Webアプリケーション中のボタンのクリック操作によって切り替えることができる。

3.6 時間間隔の可視化

医療指示列全体の時間関係を考慮した流れを把握したり、分岐における実施時間による差異を確認できるようにするため、各医療指示間の時間間隔についての対象患者内における平均値を、図1に示すように、パスの長さを変化させることで可視化

していく。これは画面上での配置を決定するノードの x 座標の計算に、前のノードの x 座標と時間間隔の値を用いることでノード間の距離を変化させた。また、時間間隔を反映することでパスの長さが極端に長い、または短い場合に視認性が下がることがある。そこで、出力されるパスの長さに上限と下限を設けた。

3.7 推薦の可視化

一般的な医療指示の流れを表す頻出医療指示列だけでなく、対象とする患者の情報についてもグラフに反映することで、医療指示の選択と、その根拠の把握が容易になると考えられる。可視化ツールの全体像は図 9 の通りになっており、画面上部に配置されたセレクトボックスから患者 ID を選択すると、頻出医療指示列中のどのパスを辿って医療指示を選択するべきかという推薦がグラフに反映される。

この図に示す例では、これまでに入院歴がなく、手術時間が午前だった患者の医療指示履歴を用いている。患者がこれまでに行ってきたパスは水色で、次に行うべき医療行為として推薦をしたいパスを紫色で示している。

対象患者は服薬指導の後にクロスマッチ検査、検体検査、血液型関連検査を行っており、左の分岐部分では上側のパスと合致することがわかる。患者は実際に入院歴がないため、一般的なクリニカルパスと相違なく医療指示の選択ができていないことが確認できる。また、病理診断と緊急検査まで医療指示を行っているため、そこまでのパスが水色で示されている。その先は、対象患者の手術の実施時間が 13:00 以前であったために、右の分岐部分において上側のパスを推薦したいパスとして紫色で示している。

このようにして、患者が一連の医療指示のうちどこまでを既に行ったのか、また今後どのような医療指示を行っていくべきなのかということが把握できる。

また、患者の医療指示履歴を、頻出医療指示列をベースにマッチしたパスの色を変えることで表現したが、実際の医療指示履歴には頻出医療指示列には現れないオーダーを含むことや、頻出医療指示列中のオーダーの一部を含まないなど過不足がある。そこで、そのような差異を確認できるようにするため、頻出医療指示列の下に患者の実際の医療指示履歴を並べて表示している。同じ医療指示のオーダーは同じ色で表現されるため、医療指示の比較が容易になっている。

3.8 症例の切り替え

可視化ツールの画面上部に配置されたセレクトボックスから症例を選択すると、別の症例についての頻出医療指示列のグラフに切り替えることができる。

本研究では TUR-Bt に加え、RFA（肝悪性腫瘍ラジオ波焼灼療法）と肝切除についての頻出医療指示列を Web アプリケーション上で表示できるようにした。

3.9 評価

宮崎大学医学部附属病院の荒木教授、山崎教授に Web アプリケーションを提示し、実際の患者に行った医療指示が頻出医

療指示列とどれほど差異があるのかといったことが視覚的にわかりやすく確認できるという点で有用性があるという意見を得た。また、医療行為のステージごとに切り分けた表示への切り替えができると良いのではないかという意見や、同じような種類の検査タイプについては同系色のノードで表現すると分かりやすいのではないかといった意見、また医者経験に基づいて作成されたクリニカルパスとの違いもわかると良いのではないかといった意見も得ることができた。

4 おわりに

4.1 まとめ

本研究では、医療従事者へ向けた医療指示の選択を支援することを目的とし、医療指示選択の根拠となる、電子カルテのデータを用いて得られたクリニカルパスの分岐要因や医療指示の出現率、時間間隔などをグラフ表現した。また、動的な患者情報を与えることで得られる次に行うべき医療行為を、グラフを変化させることで可視化した。これにより、視覚的にわかりやすくデータの解析結果を把握し、個々の患者に対する適切な医療指示の選択を容易にすることができると考えられる。

4.2 今後の課題

評価として頂いた意見を踏まえ、頻出医療指示列を医療行為のステージごとに切り分けて表示することができるようになることや、医療従事者からの助言をもとに検査タイプを種類ごとに分類するなどしてノードの色表現を変更すること、また医者経験に基づいて作成されたクリニカルパスも表示できるようにすることが必要であると考えられる。また、現状では、個々の患者に関する医療指示履歴中に含まれる医学的に有用でない医療指示も表示しているが、そのような医療指示は省略した形のグラフ表示への切り替えができるようにすることや、より細かな分岐のパターンが確認できるよう最小支持度によるグラフの切り替えができるようにすること、またグラフ表現する症例に関し当該の担当医に Web アプリケーションのページを実際に見てもらい、アンケート等の評価を行うことが必要であると考えられる。

謝 辞

本研究の一部は、日本学術振興会科学研究費補助金基盤研究(B)(#20H04192)の助成により行われた。また、本研究は宮崎大学医学部附属病院の電子カルテデータを用いている。これは宮崎大学の HP [15] に記載されており、宮崎大学の倫理審査委員会及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。関係者各位の協力に感謝する。

文 献

- [1] 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時系列分析による頻出シーケンスの抽出” DEIM Forum 2014, F6-2, 2014.
- [2] Rakesh Agrawal and Ramakrishnan 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.

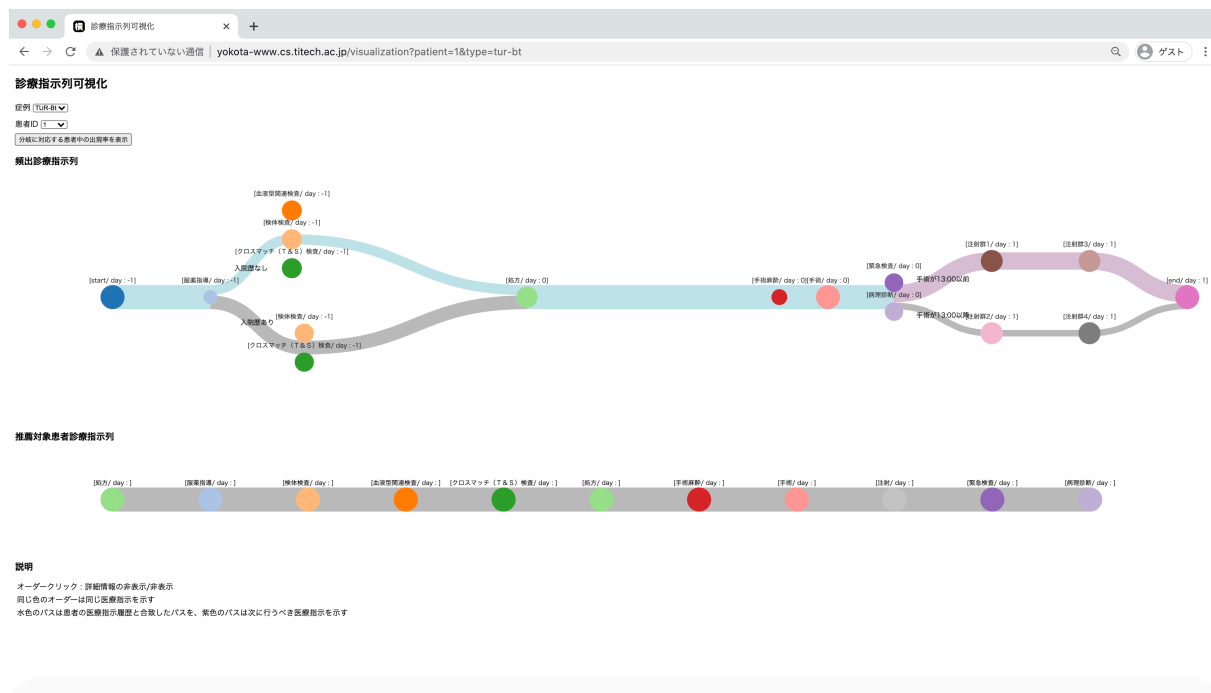


図 9 可視化ツールの全体像

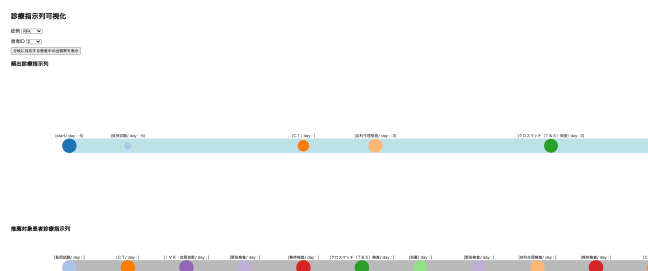


図 10 症例の切り替え

- [3] 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.
- [4] X. Yan, J. Han, R. Afshar. CloSpan: Mining closed sequential patterns in large databases. Proc.SIAM Int '1 Conf. Data Mining (SDM '03), pp. 166-177, May 2003.
- [5] Y. Chen, M. Chiang, M. Ko. Discovering time-interval sequential patterns in sequence databases. Expert Systems with Applications 25, pp. 343-354, 2003.
- [6] 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 interna
- [7] 浦垣啓志郎, 保坂智之, 荒堀喜貴, 串間宗夫, 山崎友義, 荒木賢二, 横田治夫. 電子カルテの投薬履歴における薬効に着目した医療行為パターンの抽出. DEIM Forum 2016, G7-5 2016.
- [8] 保坂智之, 浦垣啓志郎, 荒堀喜貴, 串間宗夫, 山崎友義, 荒木賢二, 横田治夫. 医療履歴の時系列解析におけるシーケンス間類似度評価による時間間隔調整の導入. DEIM Forum 2016, G7-5 2016.
- [9] 本田 祐一, 串間宗夫, 山崎友義, 荒木賢二, 横田 治夫. 典型的医療指示シーケンスの共通部分検出とその可視化. DEIM Forum 2017, I1-1, 2017
- [10] 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 2019, D1-1, 2019

- [11] Hieu Hanh Le, Henrik Edman, Yuichi Honda, Muneeo Kushima, Tomoyoshi Yamazaki, Kenji Araki, Haruo 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.
- [12] 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 2019.
- [13] D3.js., <https://d3js.org>
- [14] Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data. Yiye Zhang, Rema Padman, Nirav Patel. Journal of Biomedical Informatics 58, 2015
- [15] 宮崎大学医学部附属病院医療情報部.
<http://www.med.miyazakiu.ac.jp/home/jyoho/>