

複数医療機関の電子カルテからの 統計情報付き頻出クリニカルパスの抽出と可視化

杉谷 美和 (指導教員：文化情報工学科・Le Hieu Hanh)

1 はじめに

近年、電子カルテの普及により医療データの集約や二次利用が進み [2]、診療プロセスの標準化が期待されている。一方で、患者個々の特性に応じた診療支援や、医療機関間でのデータ共有の必要性が指摘されている [3]。従来は単一機関のデータ解析が中心で、医療機関間の比較や疾患横断的な分析は不十分である。本研究では、複数医療機関の電子カルテデータを統合し、疾患ごとの頻出医パターン抽出・可視化するとともに、診療プロセスの詳細や検査結果の異常値発生率を示すための統計情報の算出手法を提案する。

2 提案手法

本節では、提案手法について述べる。電子カルテデータを活用して統計情報付き頻出医療指示パターンを抽出し、統計情報や異常値発生情報を付与することで、診療プロセスの標準化および医療従事者の意思決定を支援することを目的とする。提案手法は以下の4つのステップで構成される。

2.1 クリニカルパスの作成

クリニカルパスとは、特定の疾患や治療における標準的な診療プロセスを時系列で体系化した計画書である。本研究では、電子カルテデータから診療履歴を整理し、重要な医療行為 (例: 手術) を行なった日 (Day 0) を基準に経過日数を算出してクリニカルパスを構築する。

定義 2.1 (同日の医療行為の並び替えルール)

医療行為の実行時刻がないという前提で、同一日に記録された医療行為は次の順序で並び替える。

手術 → 投薬 → 検査 → 診療行為
同一カテゴリ内では、辞書順で整列する。

定義 2.2 (経過日数付き医療指示列)

患者 p_i に対する医療指示シーケンスは、基準日 t_0 (Day 0) からの経過日数 d_j を含む以下の形式で定義する。

$$S_{p_i} = \langle (d_1, a_1), (d_2, a_2), \dots, (d_n, a_n) \rangle$$

$d_j = t_j - t_0$ は経過日数を表し、 a_j は医療行為の種類を表す。なお、同じ d_j に複数の医療行為が含まれる場合もある。

このように整理したデータは、患者ごとに識別子 SID を付与して新しいデータテーブルに格納する。

2.2 統計情報付き頻出パターンの抽出

T-PrefixSpan アルゴリズム [1] を用いて疾患ごとの頻出医療指示パターンを抽出する。このアルゴリズムは、アイテム間の時間間隔を考慮しながら順序情報を保持することで、より高精度なパターン抽出を可能にする。さらに、頻出医療指示パターンの抽出と同時に検査結果や投薬量などの統計情報の抽出も行う。

定義 2.3 (統計情報付き頻出パターン)

検査結果と投薬量は、疾患の進行や治療効果を評価し、診断支援や治療の適正化に役立つ重要な指標である。頻出パターン f_s に対応する検査結果や投薬量を $X(\alpha)$ とし、観測値を x_j ($j = 1, 2, \dots, l$) とする。 k は観測値の総数とし、以下の統計情報を付与する。

$$\text{平均値: Mean}(X(\alpha)) = \frac{1}{k} \sum_{j=1}^k x_j$$

$$\text{中央値: Median}(X(\alpha)) = x_{\lceil \frac{k}{2} \rceil}$$

$$\text{最大値: Max}(X(\alpha)) = \max(X(\alpha))$$

$$\text{最小値: Min}(X(\alpha)) = \min(X(\alpha))$$

2.3 検査結果の異常値発生率の算出

頻出パターン f_s ごとに、対象疾患に関連するクリニカルパスを分析し、検査結果を対象に異常値発生率を算出する。異常値発生率は、特定のパターンにおける検査異常の傾向を把握し、診断や治療方針の改善に役立つ。異常値発生率は、検査結果が正常範囲 $[B, T]$ を逸脱する割合で定義される。

定義 2.4 (異常値発生率)

まず、対象疾患のクリニカルパス集合 $D = \{S_{p_1}, S_{p_2}, \dots, S_{p_m}\}$ を参照し、頻出パターン f_s を含むクリニカルパス群 D_{in} および含まない群 D_{out} に分割する。

$$D_{in} = \{S_{p_i} \in D | f_s \subseteq S_{p_i}\}, \quad D_{out} = D \setminus D_{in}$$

次に、各群の経過日数 d_j ごとの検査結果 x_j を以下のように分類する。

$$C(x_j) = \begin{cases} -1 & (x_j < B) \quad (\text{低値}) \\ 0 & (x_j \in [B, T]) \quad (\text{正常}) \\ 1 & (x_j > T) \quad (\text{高値}) \end{cases}$$

経過日数 d_j ごとの高異常率 (HR), 低異常率 (LR), 正常率 (NR) は以下の式で定義される。なお、 $X(\alpha, d_j)$ は経過日数 d_j における検査結果の集合を表す。

$$HR(\alpha, d_j) = \frac{|x_i \in X(\alpha, d_j) | C(x_i) = 1|}{|X(\alpha, d_j)|} \quad (\text{高異常率})$$

$$LR(\alpha, d_j) = \frac{|x_i \in X(\alpha, d_j) | C(x_i) = -1|}{|X(\alpha, d_j)|} \quad (\text{低異常率})$$

$$NR(\alpha, d_j) = 1 - (HR(\alpha, d_j) + LR(\alpha, d_j)) \quad (\text{正常率})$$

2.4 可視化

本研究では、頻出パターンと異常値発生率の解析結果を可視化し、診療プロセスの特性や分岐点を明確化する手法を提案する。頻出パターンでは、医療指示を表す各ノードがサポート値 $Sup(\alpha)$ に基づいて配置され、重要なパターンを強調する。異常値発生率は棒グラフ形式で表示し、ツールチップによる詳細情報や検査項目ごとの絞り込み機能も実装する。

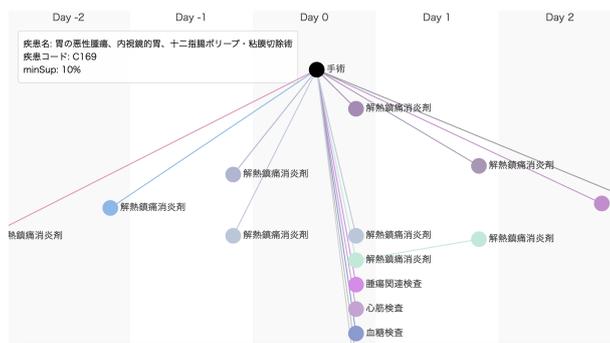


図 1: 頻出パターンの可視化

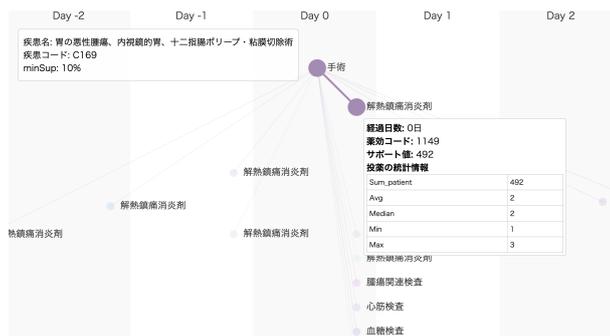


図 2: ノードクリック時の統計情報表示と強調機能の例

3 評価実験

本節では、提案手法に基づき抽出した頻出医療指示パターンおよび異常値発生率の可視化結果を示す。評価は「パターン間の差異(バリエーション)の可視化」、「統計情報付き頻出医療指示パターンの抽出」、「検査結果の異常値発生率の算出」を基準に行った。

3.1 実験環境

本研究では、頻出パターンの抽出に PrefixSpan ライブラリを改良し、T-PrefixSpan [1] と同様の機能を実装した。Python でコードを記述し、データベースは PostgreSQL を使用した。複数の医療機関から提供された電子カルテデータを基に人工データセットを作成し、ランダム性を持たせて異常値を含むパターンを再現して評価を行った。人工データセットには医療機関の識別情報は含まれていないが、将来的には識別情報を用いて施設間の診療プロセス比較を目指す。

3.2 実験結果

図 1 は、「胃の悪性腫瘍、内視鏡的胃、十二指腸ポリプ・粘膜切除術」に関する医療指示パターンを可視化したものである。閾値 (minSup) を 10% に設定し、抽出されたクローズドパターンを反映している。各ノードは医療指示、エッジはその流れを示し、サポート値が高いパターンほど上部に配置される。

図 2 はノードをクリックした際の様子を示す。ツールチップには、経過日数、薬剤コード、サポート値、統計情報(患者数、平均値、中央値、最小値、最大値)が表示される。クリックしたノードを含む頻出シーケンス全体が強調され、頻出パターンと異常値の関連性が視覚的に把握できると確認できた。

これにより、「統計情報付き頻出医療指示パターン

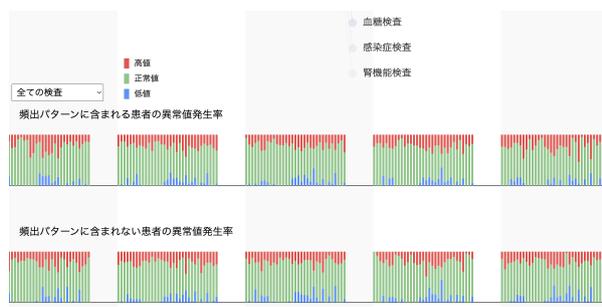


図 3: 異常値発生率の可視化

の抽出」と「パターン間の差異の可視化」の有効性が確認できる。

図 3 は、腎機能検査の異常値発生率を経過日数ごとに示した例である。各バーは高値、低値、正常値をそれぞれ赤、青、緑で表し、異常値の発生タイミングを視覚化している。バーをホバーすると、検査名、異常値の割合、サポート値などの詳細が表示される。また、関連する頻出パターンも強調表示されることで、異常値発生率と頻出パターンの関連性を視覚的に確認できる。

これらの結果は、「検査結果の異常値発生率の算出」において有効性を示している。この可視化機能により、異常値発生の分布や診療プロセス内のリスクポイントを特定し、診療の質向上やリスク低減のための適切な対策を可能にする。

4 終わりに

本研究では、複数医療機関の電子カルテデータを解析し、頻出クリニカルパスを抽出・分析する手法を提案した。本手法は、単一機関や特定疾患に限定されない包括的な解析を可能にし、SID を活用した詳細なパターン特定や異常値を含む医療指示の可視化を実現した。本手法により、診療プロセスの標準化や、個々の患者に応じた柔軟な診療方針の選択が促進されることが期待される。

今後は、大規模な実データによる検証や分岐要因の分析を進め、医療現場のフィードバックを反映して実用性と汎用性を高めたシステムの普及を目指す。

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。

参考文献

- [1] Uragaki, K. and al., et : Sequential pattern mining on electronic medical records with handling time intervals and the efficacy of medicines, in *Proc. 21st IEEE International Symposium on Computers and Communications*, pp. 20–25 (2016).
- [2] 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 (2022).
- [3] 鎌谷直之: 個別化医療とバイオインフォマティクス, *日本生体医工学学会誌*, Vol. 44, No. 3, pp. 422–428 (2006).

A 付録

表 1: 人工データセットのスキーマ

テーブル名	データスキーマ
患者情報	患者 ID, 年齢, 性別, BMI
日付情報	患者 ID, 日付, 入院日, 退院日
傷病情報	患者 ID, 入院日, ICD-10 コード
手術情報	患者 ID, 手術日, DPC コード+K コード
薬剤情報	患者 ID, 投与日, 投与の合計数, 薬効分類コード
検査情報	患者 ID, 測定日, 検査名, 検査値, 単位, 検査回数
診療行為情報	患者 ID, 実施日, レセプト電算コード
バイタル情報	患者 ID, 測定日, バイタルサイン項目名, バイタルサイン測定値, 単位

表 2: 人工データ：疾患ごとの患者数

疾患名	患者数 (名)
肺炎等	1,101
狭心症, 慢性虚血性心疾患, 経皮的冠動脈形成術等	1,100
肝・肝内胆管の悪性腫瘍, 肝悪性腫瘍ラジオ波焼灼療法	1,081
肺の悪性腫瘍手術, 肺葉切除又は1肺葉を超えるもの等	1,081
膀胱腫瘍, 膀胱悪性腫瘍手術, 経尿道的手術	1,052
椎間板変性, ヘルニア, 内視鏡下椎間板摘出 (切除) 術	1,013
急性心筋梗塞, 再発性心筋梗塞, 経皮的冠動脈形成術等	1,012
胃の悪性腫瘍, 内視鏡的胃, 十二指腸ポリープ, 粘膜切除術	1,011
脳梗塞	602
子宮頸, 体部の悪性腫瘍, 子宮悪性腫瘍手術等	484
乳房の悪性腫瘍, 乳腺悪性腫瘍手術, 乳房部分切除術	463
合計	10,000