6N-07

# 複数医療機関の電子カルテデータを用いた頻出医療指示パターンの分岐要因推定

杉谷 美和 † 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 † 城西大学 ‡†

## 1 はじめに

近年，電子カルテデータの二次利用による医療プロセスの解析が進んでいる．同データから抽出された頻出医療指示パターンの分岐要因が，患者病態か施設特性によるかを定量的に推定することは，実態に即した診療プロセスの理解において不可欠である．先行研究 [1] では，多重共線性による因子の不安定さが課題であった．例えば WBC（白血球数）と CRP（C 反応性タンパク）のような相関の強い項目は適正な因子評価を妨げる．また，施設属性と個別病態を定量的に分離・評価する手法も未確立であった．

本研究では，抽出された頻出医療指示パターンを対象に，分岐要因を統計的に推定する手法を提案する．提案手法では，患者属性や検査値等を抽出し，臨床的指標に基づき離散化する．これに L2 正則化 [2] ロジスティック回帰を適用することで，多重共線性の影響を抑制し，安定的な寄与度の算出を実現する．評価実験では，27 医療機関の胃の悪性腫瘍患者データを対象に本手法を適用し，医療機関や患者の状態が医療指示の選択に与える影響を推定することで，提案手法の有効性を検証する．

## 2 提案手法

抽出された頻出医療指示パターンの分岐要因を統計的に推定する手法を述べる．

Factor Estimation of Branching in Frequent Medical Order Patterns using Electronic Health Record Data from Multiple Medical Institutions

† Miwa Sugitani†, Ryosuke Matsuo†, Tomoyoshi Yamazaki†, Kenji Araki†, Masato Oguchi†, Haruo Yokota‡, Hieu Hanh Le†
†Ochanomizu University ‡Josai University

### 2.1 対象選定と変数抽出

先行研究 [3] に基づき，抽出された頻出医療指示パターンのうち，医療現場で作成された標準クリニカルパスとの比較で算出された F 値の上位・下位各 5 件を選定する．これは，標準パターンと例外パターンの両面から分岐要因を網羅的に推定するためである．各パターンから手術等の共通指示を除外した検査・薬剤項目を，分岐要因推定の目的変数 $y$ として定義する．説明変数には，静的情報（患者の入院期間，年齢，性別，ボディマス指数（BMI），医療機関コード）および動的情報（分岐直前のバイタル，検査値）を用いる．

### 2.2 前処理と特徴量化

連続値（バイタル，検査値）の欠損に対し，同一疾患群の患者において前日のデータがある場合は LOCF 法 [4]，無い場合は当該疾患患者全体の平均値による補完を適用する．過学習抑制と臨床的解釈性向上のため，連続値は「低・正常・高」の 3 値に離散化する．年齢は，小児（18 歳未満），成人（18 歳以上 65 歳未満），高齢者（65 歳以上）の 3 区分とする．これらカテゴリ変数および施設属性を One-Hot Encoding によりダミー変数化する．

### 2.3 要因推定モデルの構築

特定項目の実施（$y \in \{0,1\}$）を目的変数とし，項目ごとに独立した二項ロジスティック回帰モデルを構築する．多重共線性や施設間差による不安定化を抑制するため，以下の目的関数 $J(\theta)$ を最小化する L2 正則化 [2] を適用する．

$$J(\theta) = -\sum_{i=1}^{n}[y_i \log(h_\theta(x_i)) + (1-y_i)\log(1-h_\theta(x_i))] + \lambda \sum_{j=1}^{m} \theta_j^2$$

ここで，$n$：サンプル数，$m$：特徴量数，$y_i$：正解ラベル，$x_i$：説明変数ベクトル，$\theta$：回帰係数ベクトル，$h_\theta(x_i)$：予測確率，$\lambda$：正則化パラメータを表す．各因子の寄与は，回帰係数から算出される



Copyright ©2026 Information Processing Society of Japan.
All Rights Reserved.

表1　要因推定結果のオッズ比（胃の悪性腫瘍）

| 変数名 | Day -1 止血機能 | Day -1 腎機能 | Day 1 血液代用剤 | Day 1 血算定 | Day 1 感染症 | Day 1 腎機能 |
|---|---|---|---|---|---|---|
| 施設規模 | | | 4.28 | 2.25 | 2.53 | |
| 疾患別患者数 | | | 3.36 | | 2.65 | 0.42 |
| 施設：大規模 | | | 0.46 | 0.48 | | |
| WBC：高 (Day 1) | | | | 2.82 | 7.72 | |
| WBC：正常 (Day 1) | | | | 2.47 | 3.46 | 6.46 |
| CRP：正常 (Day 0) | | | 2.15 | 4.19 | 6.46 | 6.38 |
| CRP：高 (Day 0) | | | | 0.15 | 0.06 | 0.10 |
| 血小板：正常 (Day 0) | | | | | | 0.47 |
| γ-GTP：高 (Day -2) | 2.07 | | | | | |
| γ-GTP：正常 (Day -2) | 0.49 | | | | | |
| γ-GTP：正常 (Day 0) | | | 0.43 | | | |
| γ-GTP：高 (Day 0) | | | | | | 5.17 |
| ALT：高 (Day -2) | | 2.03 | | | | |
| クレアチニン：高 (Day -2) | 0.48 | | | | | |
| RBC：正常 (Day -2) | | 0.48 | | | | |

オッズ比（$\exp(\theta_j)$）により評価する．

## 3　評価実験

提案手法の有効性を検証するため，27 医療機関の胃の悪性腫瘍患者を対象とした評価実験を実施した．本解析では，手術当日を Day 0 と定義し，L2 正則化ロジスティック回帰の目的関数 $J(\theta)$ を最小化する際の正則化パラメータ $\lambda$ は 1.0 に固定した．抽出された頻出医療指示パターンのうち，F 値に基づき選定された分岐項目の解析結果を表 1 に示す．なお，標準化係数の絶対値が 0.7 以上の項目を，臨床的寄与の高い主要因子として抽出した．本研究は LDI の倫理委員会の承認を得ている（No. 2024_MIL_0004_A001）．

### 3.1　手術前日（Day -1）の検査選択要因

手術前日の解析では，患者病態に基づく検査選択要因を推定できた．表 1 より，止血機能検査（Day -1）への γ-GTP 高値（Day -2）（オッズ比 2.07）や，腎機能検査（Day -1）への ALT 高値（オッズ比 2.03）の有意な寄与を確認した．これらは，臓器状態によらず麻酔医の判断が前提となる手術において，脂肪肝等に伴う止血悪化の確認や「念のため」の検査，施設規定を反映していると推測される．

### 3.2　手術翌日（Day 1）の医療指示と施設特性

手術翌日の解析では，術後の管理における臨床判断と施設特性の両面を定量的に抽出できた．まず施設特性については，血液代用剤に対し施設規模（オッズ比 4.28）および疾患別患者数（オッズ比 3.36）が高い正の寄与を示しており，手術の割合が多い病院と多くない病院の差を反映している．また，胃の手術は経口での引水開始が遅いため点滴時間が伸びる特性があり，午後の手術開始では翌日も点滴が継続される実態が示された．次に臨床判断については，感染症および血算定検査で，当日 WBC 高値（オッズ比 7.72, 2.82）と前日 CRP 正常値（オッズ比 6.46, 4.19）が強い正の相関を示した．この背景には両者の必須の関係があり，当日の CRP が正常であっても翌日に WBC 数で生体反応の程度を確認する臨床判断や，そこに生じる施設差が反映されている．また，腎機能検査において当日 γ-GTP が示す寄与（オッズ比 5.17）は，術前の機能不良を術後も継続評価する臨床判断を反映している可能性が高い．

## 4　終わりに

本研究では，抽出された頻出医療指示パターンの分岐要因を統計的に推定する手法を提案した．L2 正則化ロジスティック回帰により，医療指示の選択に関わる患者病態と施設特性をオッズ比で定量的に分離し，評価を行った．胃の悪性腫瘍患者を対象とした評価実験を通じて，臨床現場の判断基準や施設固有の傾向を反映した主要因子を適切に抽出可能であることを確認した．今後は対象疾患を順次拡大し，提案手法の網羅的な有効性と汎用性を検証する．

## 謝辞

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

## 参考文献

[1] H.H. Le et al. Methods for analyzing medical-order sequence variants in sequential pattern mining for electronic medical record systems. *ACM Healthcare*, Vol. 4, pp. 3:1–28, 2023.

[2] Hoerl and Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, Vol. 12, No. 1, pp. 55–67, 1970.

[3] 杉谷ら．複数医療機関間の電子カルテデータを用いた統計情報付き頻出医療指示パターンの抽出と可視化．In *DEIM*, pp. 6K–02, 2025.

[4] D. B. Little, R. J. A. & Rubin. *Statistical Analysis with Missing Data.* Wiley, 2019.



Copyright ©2026 Information Processing Society of Japan.
All Rights Reserved.