

2024年度第4回城西大学公開講座

医療・健康・介護のデータとその利用

2024年9月25日

城西大学理学部数学科 特任教授
東京工業大学 名誉教授
横田 治夫

hyokota@josai.ac.jp

2024/9/25

©横田治夫

1

自己紹介

- 1957年（昭和32年）生まれ（西年）67歳
- 長野県長野市出身
- 東京工業大学卒業・修士課程修了
- 富士通株式会社就職
- 第五世代コンピュータ研究開発プロジェクトに出向（4年間）
- 富士通研究所 帰任（出向中も含めると富士通に10年間在籍）
- 北陸先端科学技術大学院大学 助教授（6年間：石川県）
- 東京工業大学大学院情報理工学研究科 准教授・教授（25年間）
 - 2023年 定年退職 東京工業大学名誉教授
- 2023年から城西大学理学部数学科 特任教授（1年間）



2024/9/25

©横田治夫

<https://www.josai.ac.jp/~hyokota/>

2

研究分野

- 学生の頃から**情報工学**（コンピュータ）分野の研究に携わる
- 特に「**データ工学**」と呼ばれる分野が専門
 - データベース、データの蓄積、データの解析等
 - 最近「データサイエンス」で注目されるようになってきました
- 現在「日本データベース学会」の会長
- ここ10年程、**医療データ解析**の研究に注力
 - 電子カルテで集められた医療データの解析
 - 2022年「電子カルテデータ解析」の書籍上梓 ⇨
 - 医療支援のためのエビデンス・ベースド・アプローチ
 - **2023年度大川出版賞受賞**
- 本日：右の書籍に関連した話題
 - より詳しく知りたい場合は、書籍をご参照ください



共立出版

2024/9/25

©横田治夫

3

本題に入る前に（アナウンス）

- 城西大学に「**情報数理学科**」が誕生（来年2025年度から）
 - 城西大学**紀尾井町キャンパス**の数学科が情報数理学科に！
 - 時代の要求に合致した学生の育成
 - **数学と共に情報の理論を身につけ、社会の課題解決のできる学生を育てる**



約45km 南東



©横田治夫

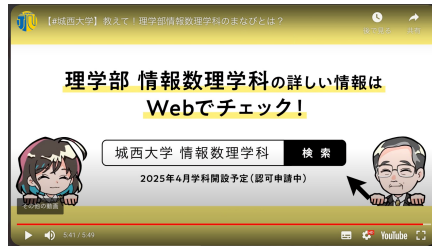
紀尾井町キャンパス3号棟

2024/9/25

4

新学科紹介動画

- 新学科紹介の YouTube 動画もあります
 - ミライちゃんと一緒に出演しています
 - お時間があったらご笑覧ください



<https://www.youtube.com/watch?v=r7vVtl-x5iQ>

2024/9/25

©横田治夫

5

本題：本日の予定

- 時代的背景
- 医療ビッグデータ
- 医療・健康・介護データの中身
- 休憩
- プライバシー保護と情報セキュリティ
- 医療・健康・介護データの利用
- 今後の展望
- Q&A

2024/9/25

©横田治夫

6

時代の流れ

- 情報技術の発達に伴い、**社会の情報化**が進んでいます
- 情報化が進むことで**蓄積されるデータ量**も増えていきます
 - 「**ビッグデータ**」という言葉聞いたことがある方もおられるのでは
- 世の中の（世界中の）**総データ量**を推定している人達がいま
 - 色々な推定がされていますが、その一つによると2025年の総データ量は

175 ZB（ゼットバイトと読みます）
 $1ZB = 10^{21} B$

と言われてもピンと来ませんよね

2024/9/25

©横田治夫

7

もう少し分かり易い（？）イメージ

- **スマホのデータ容量**を例に考えてみましょう
 - 皆さん、ご自分のスマホのデータ容量はどの程度かご存じでしょうか
 - 一般的に、128GB（ギガバイト）～1TB（テラバイト）程度
 - $1GB = 10^9 B$, $1TB = 10^{12} B$
 - 175ZB は、128GB のスマホの **1.37 兆倍** (1.37×10^{12} 倍)
 - これでもピンときませんよね
 - 世界の人口：2024年 80 億人 (80×10^8) を超えたとされています
 - 赤ちゃんからお年寄りまで **一人170台を超えるスマホ分**のデータを持つイメージ
- 実際は、企業等の組織が多くのデータを持っています
 - 特にプラットフォーム（GAFAM）と呼ばれる企業の**データセンター**

2024/9/25

©横田治夫

8

なぜ大量のデータが貯まるのか

- データは企業の戦略立案に欠かせない**武器**
- 例：スーパー、コンビニ、ネットストア等の商品販売
 - 皆さんの**購入履歴**は全て蓄積されています！
 - スーパーのレジで入力されたデータはコンピュータに記憶
 - 購入の履歴を解析することで、**顧客の動向を把握**
 - 買い物籠の中を解析するので、「バスケット解析」などとも呼ばれます
 - 例えば、よく売れる組み合わせで**商品を配架や推薦**
 - ビールの近くにつまみ類、紙おむつの近くに粉ミルク、等々
 - ネットストアで、「**これを買われた方はこちらを買われています**」の推薦
- 非常に簡単な見積もりをしてみましょう
 - 顧客一人のデータ（レシート）：1KB（10³B）、1秒間に100人
 - 100KB/s → 360MB/h → 10GB/day → 3TB/year → **30TB/10years**

2024/9/25

©横田治夫

9

医療ビッグデータ

- **医療・健康・介護**に関する分野でも大量のデータが蓄積
- **病院の電子化**が進んでいることを実感しませんか？
 - 健康で病院に普段は行かない人でも、健康診断や人間ドックなどで
- 具体的な例
 - 医者は紙の**カルテ**の代わりに端末に入力
 - **検査結果**も全部電子的に入力
 - 様々な**医用画像**も電子化
 - 普段の**体重、血圧、脈拍**等も電子的に集計可能
 - 介護施設あるいは健康維持のためのデータ
- もう少し具体的に見てみましょう

2024/9/25

©横田治夫

10

そもそもカルテとは

- 個人の医療に関する経過等を記載したもの
 - **Karte** はドイツ語（一般的によく使われる）
 - 英語では **Medical Record**
 - 日本語では、正式には**診療録**と表記
- 医師法 第二十四条
 - 医師は、診療をしたときは、遅滞なく診療に関する事項を**診療録**に記載しなければならない
 - 二 前項の診療録であつて、病院又は診療所に勤務する医師のした診療に関するものは、その病院又は診療所の管理者において、その他の診療に関するものは、その医師において、**五年間これを保存**しなければならない
- 電子カルテ：カルテを電子化したもの
 - 英語では **Electronic Medical Record (EMR)** → **永年保存**

2024/9/25

©横田治夫

11

カルテ電子化の効用

- **1次利用**：カルテの**管理コスト**の大幅低減
 - **スペース効率、アクセスコスト、廃棄コスト**
 - コンパクトなストレージデバイスに記録
 - 当該患者の情報は簡単に検索（アクセス）可能
 - スペースを要しないので、5年毎の廃棄不要
 - 各種検査結果、医用画像もすべて電子化
 - 関連する**情報をリンク**することが可能
 - 医療従事者の前の端末に必要な情報を容易に表示
- **2次利用**：折角集まった電子カルテを**解析**して有効利用
 - 様々な解析による医療支援
 - いくつかについては後述
 - 複数の**医療機関**の電子カルテを**集めて解析**することも可能に
 - いくつかのプロジェクトが動いている

2024/9/25

©横田治夫

12

医療オーダーとクリニカルパス

- **医療オーダー（医療指示、Medical Order）**
 - 医療関係者からの検査の指示、処置の指示、投薬の指示等
- **オーダーリングシステム**
 - 電子的に医療オーダーを伝え記録するシステム
 - 電子カルテシステムの一部ともいえるが、独立したシステムの場合も
 - 電子カルテシステムより導入率は高い
- **クリニカルパス（標準診療計画、Clinical Pathway）**
 - 疾病に対する施術毎の典型的医療指示の順番
 - 例：血液検査 → CTスキャン → 麻酔 → 手術 → 心電図 → 投薬 → 血液検査

2024/9/25

©横田治夫

13

医用画像

- **医用画像（Medical Image）**
 - 病気の診断の為に人体の様々な部位を放射線、磁気共鳴画像、核磁気共鳴技術、陽電子検出技術、超音波、内視鏡等を用いて画像化したものを指す（レントゲン、CT、MRI、エコー、胃カメラ、…）
 - 病名を判定し、次に行う医療行為を決定するために重要であり、特に病変の部位や範囲を特定するのに有効
- **画像解析**
 - 機械学習あるいは深層学習による**パターン識別**技術の進化は驚異的
 - 疾患部位の**自動判別**を行い、医療従事者のサポートを行うことは既に実際の医療現場でも行われるようになってきている

2024/9/25

©横田治夫

15

検歴データ

- **検体検査の結果、医用画像、生理計測データ等**
- **検体検査**
 - 検体を用いて行われる検査
 - 検体：尿、大便、喀痰等の患者の排泄物、血液、髄液、胸・腹水、さらには手術や針等で採取した組織等の検査材料
- **検体検査の種類**
 - 血液生化学検査、血液学検査、血清学検査、病理検査、腫瘍マーカー、血液型検査、尿検査、便検査、喀痰検査、細菌検査、胸・腹水検査
- **医療検査技術の発達に伴い検体検査の項目も増加**
 - どのくらいの数の検体検査項目があると思いますか？
 - ある大学附属病院の電子カルテに含まれる検体検査の項目数：13,500 以上
- **検歴データを残すことは解析にとって重要**

2024/9/25

©横田治夫

14

遺伝子情報

- **遺伝子情報（ゲノム: Genome）**
 - 二重らせん立体構造を持つデオキシリボ核酸（DNA）によって構成
 - ヒトには**約3万個の遺伝子**があるとされている
 - 生物の設計図に相当するもので両親から引き継がれる
 - 引き継がれる部分の違いにより個人個人でパターンが異なる
- **遺伝子情報の利用**
 - **遺伝子のパターンと疾病との関連**を解析するようなアプローチも可能
 - 遺伝子情報と電子カルテを結び付けることも考えられる
 - 遺伝性疾患の**発症のリスク**を調べることも行われるようになっている
 - 遺伝的に乳癌の発生率が高い場合に乳癌の予防のために乳房を削除する事例
- **本人だけでなく両親、兄弟等の血縁者にも影響を与える情報**
 - 情報管理には最大限の注意が必要

2024/9/25

©横田治夫

16

レセプト

- 独：Rezept、英：Receipt、日（通称）：レセプト
- 医療報酬の明細書
 - 医療行為には費用が発生するため、個々の患者に対する費用を算出し、報酬を支払うための情報が必要（項目毎）
 - 医療機関が健康保険組合等に医療報酬を請求するための明細書
- 種類
 - 医師の診療に関する診療報酬明細
 - 保険薬局における調剤に関する調剤報酬明細
 - 訪問看護に関する訪問看護費明細
 - 等
- カルテ同様電子化が進む
 - 全国のレセプトを集めた NDB
 - 2008年度から開始、2022年度までに約250億件
- 検査結果、診察記録等は含まれない

2024/9/25

©横田治夫

17

生理計測データ（健康データ）

- 生理計測データ (PMD: Physiological Measurement Data)
 - 体重、体温、血圧、脈拍等：体調や健康状態を表す重要な指標
 - 介護施設、あるいは個人の自宅等での普段の生活の中で計測
- IoT (Internet of Thing) の技術の進歩
 - 計測デバイスが高機能化されてインターネットに接続
 - 医療向け：IoMT (Internet of Medical Things)
 - Apple Watch: 心拍数、心電図の情報
 - 小型のパルスオキシメーター
- 呼称
 - 電子カルテのEMR (Electronic Medical Record) に対して、電子健康記録 (EHR: Electric Health Record) と呼ぶこともある
 - 明確に区別されている訳ではなく、電子カルテをEHR と呼ぶこともある
 - 個人健康記録 (PHR: Personal Health Record) という呼び方もある

2024/9/25

©横田治夫

18

介護データ

- 医療と同様、介護に関しても多くのデータが集積されている
- 様々な情報が対象
 - 要介護認定情報
 - 身体機能、生活機能、認知機能、精神・行動障害、社会生活への適応
 - 支援・介護記録（ケアの内容）
 - 内服薬情報
 - 介護レセプト情報
 - 介護施設の情報
- オープンデータおよびデータベース
 - 匿名化オープンデータが厚生労働省老健局老人保健課から公表
 - 匿名医療保険等関連情報データベース（NDB）
 - 国民健康保険データベース（KDB）

2024/9/25

©横田治夫

19

プライバシー保護

- プライバシー
 - 一般的には広範な概念：他人の干渉を許さない
 - 個人の部屋のカーテン、施錠
- 情報システムにおけるプライバシー保護
 - 意図しない情報を他人に公開しない
- 情報システムにおけるプライバシー保護のアプローチ
 - アクセス制限、匿名化（仮名化）、曖昧化
 - 具合例は後述
- プライバシー保護の指標
 - k-匿名化
 - 個人が特定される確率を 1/k 以下にする
 - 差分プライバシー (Differential Privacy)
 - 問合せ結果にノイズを加え、データベースの更新が区別できないように

2024/9/25

©横田治夫

20

要配慮データと個人情報

- **要配慮（センシティブ）データ**
 - 他人に知られたくないデータ（でも解析の対象にしたい）
 - 例：病名
- **データ解析の対象外の項目**
 - アクセス制限等で外す
 - 例：電子カルテにおける住所（地域的特性等を対象外として）
- **個人情報（氏名、個人識別符号）**
 - 生存する個人に関する情報であって、特定の個人を識別できるもの
 - **匿名加工情報**：個人を識別できないように加工
 - 復元して再識別できない
 - 第三者への提供が可能
 - **仮名加工情報**：他の情報と照合して特定の個人を識別できないように加工
 - 第三者への提供が制限（2020年の改正で導入）
- **準識別子**
 - 他の情報と組み合わせることで個人の特定が可能
 - 例：22歳の男性は佐藤一郎氏のみ（背景知識） → 特定可能
 - 曖昧化：22歳 → 20代等

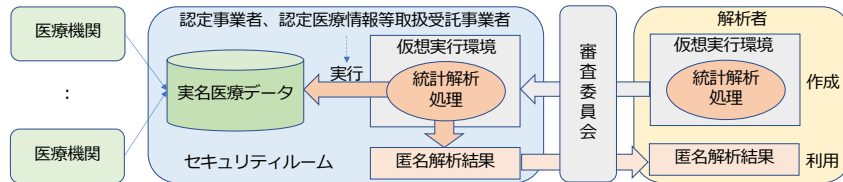
アクセス制限・匿名化・曖昧化のイメージ

| 名前 | 年齢 | 性別 | 病名 | 住所 |
|------|----|----|-------|--------------|
| 佐藤一郎 | 22 | 男 | うつ病 | 中央区銀座 1-2-3 |
| 山田花子 | 25 | 女 | バセドウ病 | 港区赤坂 4-5-6 |
| 高橋二郎 | 25 | 男 | うつ病 | 目黒区目黒 2-3-4 |
| 田中梅子 | 31 | 女 | バセドウ病 | 新宿区新宿 1-3-4 |
| 鈴木三郎 | 31 | 男 | 胃癌 | 渋谷区渋谷 2-2-5 |
| 伊藤桜子 | 36 | 女 | 胃癌 | 目黒区大岡山 1-1-1 |

| 匿名化 | 曖昧化 | アクセス制限 | | |
|--------|-----|--------|-------|----|
| 仮名 | 年齢 | 性別 | 病名 | 住所 |
| abc012 | 20代 | 男 | うつ病 | |
| def345 | 20代 | 女 | バセドウ病 | |
| ghi678 | 20代 | 男 | うつ病 | |
| jkl901 | 30代 | 女 | バセドウ病 | |
| mno234 | 30代 | 男 | 胃癌 | |
| pqr567 | 30代 | 女 | 胃癌 | |

電子カルテ解析

- **次世代医療基盤法に基づく医療データの二次利用の推進**
 - 主務大臣による**認定匿名加工医療情報作成事業者（認定業者）**の認定
 - セキュリティルーム内で認定事業者、認定医療情報等取扱受託事業者による統計解析処理
 - セキュリティルームから持ち出せるのは、統計解析処理を行い審査委員会の承認を得た情報
 - 多数の医療機関からの電子カルテ、電子レセプトデータ収集
 - **千年カルテプロジェクト**：ライフデータイニシアティブ（LDI）、NTTデータ
 - **NDB** (National Database of Health Insurance Claims and Specific Health Checkups of Japan)：日本医師会情報管理機構（J-MIMO）、ICI、日鉄ソリューションズ



情報セキュリティ

- **情報セキュリティ確保は個人情報保護の基盤**
 - 確保されていなければ**漏洩、消失、改変**されることも
- **ランサムウェアって聞いたことありませんか？**
- **ランサムウェアとは**
 - 身代金目当てのサイバー攻撃
 - 典型的には、勝手にデータを暗号化して使えなくする
 - 病院は攻撃対象になりやすい
 - 電子カルテシステムが使えないと患者の生命に関わる
 - 国内も含め、世界的に多くの事例
- **対策**
 - オフライン（普段は繋がっていない）バックアップを取る
 - あるいは医療ブロックチェーンを使う（時間の関係で、説明は割愛）

蓄積された医療・健康・介護データの利用

- 医療従事者・介護者の支援
 - 医療従事者へのサポート（情報提供、推薦）
 - 介護者へのサポート（情報提供、推薦）
- 患者に対する説明
 - 医療行為・サービス内容への理解
 - セカンドオピニオンの情報提供
- 病院・施設の経営支援
 - 経営方針決定のサポート
- 国・地方自治体の支援
 - 政策立案への情報提供、医療費削減への貢献
 - 流行性疾患（例：COVID19）への対応（ワクチン接種等）
- 様々な手法で解析

2024/9/25

©横田治夫

25

相関ルールマイニング

- **マイニング（採掘）**：データの山から重要な情報を掘り出す
- 頻出する項目の組み合わせの**相関を抽出**する手法
 - 前出のお買い物履歴からの推薦にも利用
 - 「この書籍を購入した人は、こちらの書籍を購入しています」
 - 医療・健康・介護分野でも、様々なデータの相関を解析
- 相関ルールマイニングの医療への適用例（報告例）
 - 検歴データベースへの適用による**同時検査項目**の抽出
 - この検査をした人は、こちらの検査もしています
 - **医薬品による副作用報告データベースへの適用**
 - 治療パターンと**転倒インシデント**の相関の抽出
 - 体脂肪率等の健康状態と**生活習慣病発生**の相関の抽出
 - ...
- 既に様々な相関ルールマイニングによる解析が行われている

2024/9/25

©横田治夫

26

シーケンシャルパターンマイニング

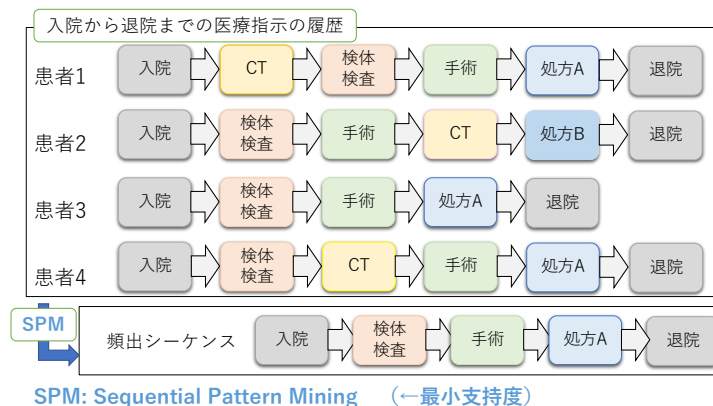
- 組み合わせだけでなく、**順番（シーケンス）**も考慮した解析
 - 医療において順番は重要
- 医療オーダーに適用することで**臨床パス**の解析が可能
 - **臨床パス**：従来は医療従事者が労力をかけて作成
 - 溜まった履歴から**頻出のシーケンスパターン**の導出が可能
- **シーケンスの分岐**の解析
 - 医療オプションに対応
 - 要因分析：どんな理由で分岐するのか、検査値、男女差、...？
- 次の検査項目の推薦
 - 検査結果のパターンから**次に実施すべき検査の項目**を推薦
- 様々な展開

2024/9/25

©横田治夫

27

頻出医療指示シーケンス抽出



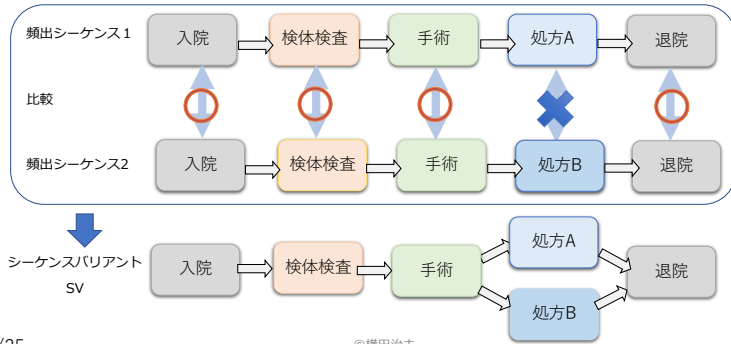
2024/9/25

©横田治夫

28

医療指示シーケンスバリエーションの存在

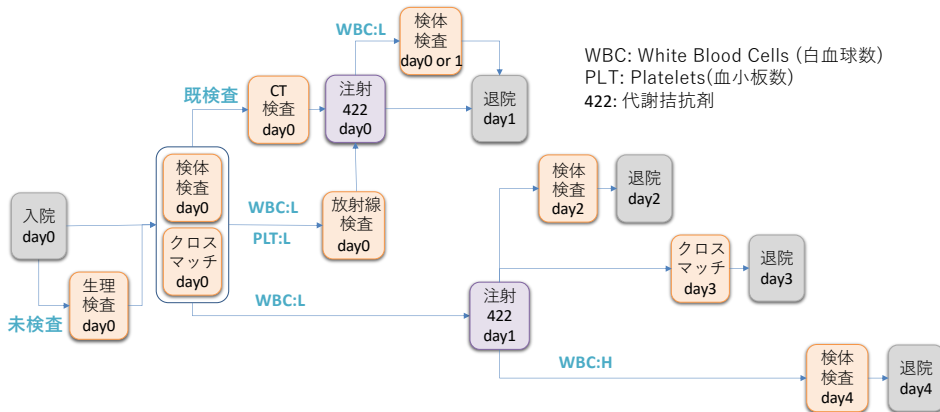
- シーケンスの分岐 (シーケンスバリエーション: SV)
 - 似た頻出医療指示シーケンスの中での違い (一部投薬する薬剤の違い等)
 - 頻出医療指示シーケンス間の比較で検出



実電子カルテを使った解析実験例

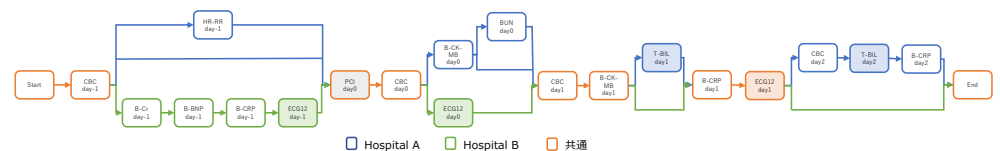
- 興味を持って頂けそうなところをいくつかピックアップ
- 対象疾病に対する治療法の**クリニカルパスの分岐要因分析**
 - 非ホジキンリンパ腫の分岐要因解析 (次のスライドで説明)
 - 検査有無、検査値によって分岐していることを可視化
 - 他の疾病でも実験 (ここでは割愛)
- 頻出シーケンスバリエーションの**複数医療機関間の差**
 - PCI (経皮的冠動脈インターベンション: カテーテルを用いた狭心症・心筋梗塞の治療法) での2病院間のシーケンスバリエーションの差
 - 病院毎の方針の差を可視化
- 頻出シーケンスバリエーションの**同一疾患の時系列における差**
 - COVID19 における感染波間の治療パターンを可視化
- これら以外にも多数の実験

非ホジキンリンパ腫の分岐要因解析



- WBC:L や PTL:L は代謝拮抗剤による骨髄抑制が疑われるために検査を行っている可能性
- 生理検査(心電図検査) は未検査の患者に, CT 検査は治療の前に再度検査を行うことが多い

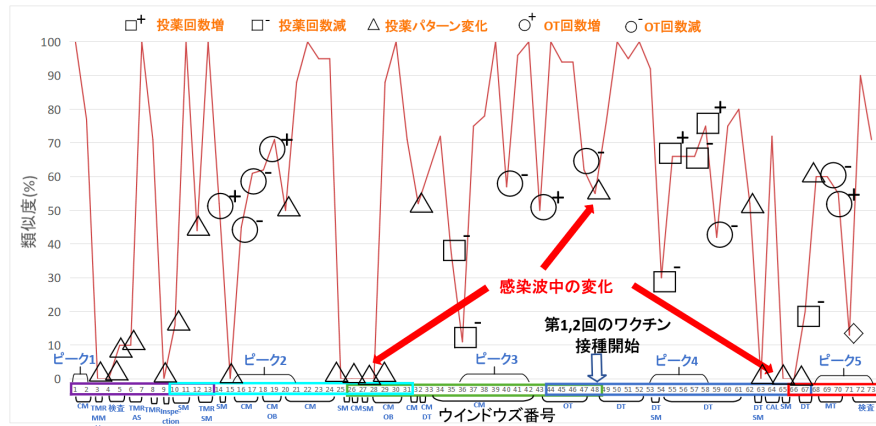
複数医療機関間の差



PCI (経皮的冠動脈インターベンション) : カテーテルを用いた狭心症・心筋梗塞の治療法の医療機関間の差

- 12点誘導心電図 (ECG12) の実施
 - 医療機関B: 毎日 (day-1, day0, day1)、医療機関A: 1日のみ (day1)
 - 術後の患者モニター (3点誘導) の心電情報の扱いの違い
- 総ビリルビン検査 (T-BIL) の頻度
 - 医療機関A の頻度が高い (検査セットの違い)
- 診療報酬点数の差
 - 僅か (PCI の点数が主、それ以外の検査等の差は小)

COVID-19 の医療指示変化時期



2024/9/25

©横田治夫

33

深層学習と医療

- 深層学習に関する技術の飛躍的進歩
 - 例：医用画像における疾患部位の自動判別
 - 既に実際の医療現場で使われている
- Chat-GPT 等の大規模言語モデル (LLM)
 - 保険診断書、紹介状等の医療文書の自動作成によるサポート
 - 医療従事者の労力削減、作成時間の短縮
 - 出力の論理性は充分ではない
 - あくまでもサポート (医療従事者による確認が前提)
- 深層学習の課題
 - 学習による変化分はブラックボックス
 - どの程度学習したら十分か不明
 - 推薦等における根拠 (エビデンス) が不十分
 - 説明可能深層学習の必要性

2024/9/25

©横田治夫

34

生理計測データ (健康データ) の解析

- 普段の生活での生理計測データの利用は重要
- 生理計測データ (IoMT) の活用イメージ
 - 健康維持
 - 体調管理
 - 介護施設においても重要
 - 医療アラート
 - 数値の変化に基づく診療のタイミングの通知
 - 診療前のデータの把握
 - COVID19 の入院前の体温、血中酸素濃度等
- 生理計測データ解析の課題
 - 生理計測データと電子カルテの付き合い合わせ
 - マイナンバーカードの利用も考えられる

2024/9/25

©横田治夫

35

介護データの解析

- 解析の対象
 - 介護保険総合データベース
 - 匿名医療保険等関連情報データベース (NDB)
 - 国民健康保険データベース (KDB)
- 国・自治体の介護関係政策決定支援
 - 実態調査、介護保険事業計画立案
 - 要介護認定情報
- 介護特有の情報の解析
 - 認知機能、精神・行動障害、支援・介護記録の利用
 - 基本は自然言語で記述：テキストマイニング技術の適用 (文章の解析)
 - 例：問題行動 (トラブル) に関連する事項の抽出
 - インシデント (問題行動発生) の予兆検出

2024/9/25

©横田治夫

36

今後の展望

- 医療の電子化が進み大量の医療データが蓄積される中
 - その二次利用であるデータ解析の重要性はますます高まっています
 - 多くのプロジェクト（含：国レベル）が動いています
- 世界的にも様々な取り組みが行われています
 - 国際的な規格も定まってきている
 - 国境を越えた連携にも期待
- 技術的な更なる発展も期待
 - 大規模言語モデルや深層学習を有効利用した実用化も近い
- 電子化された医療・健康・介護のデータ解析はさらに進む
 - 情報工学の面から医療の発展に貢献することを期待

本日のまとめ

- 医療・健康・介護に関するデータの電子化の現状
 - 電子カルテ、医療オーダー、クリニカルパス、検歴データ、医用画像、遺伝子情報、レセプト、生理計測データ（健康データ）、介護データ
- プライバシー保護と情報セキュリティ
 - 要配慮情報と匿名加工情報の扱い
 - ランサムウェア対策
- 医療・健康・介護に関するデータ解析手法の例
 - 相関ルールマイニング
 - シーケンシャルパターンマイニング
 - 実例：分岐要因分析、複数医療機関間の比較、時期による比較
 - 深層学習
 - 生理計測データ（IoMT）の解析
 - 介護データの解析
- 今後の展望