

医療におけるデータマイニング の活用

津本周作

島根大学医学部医療情報学

tsumoto@med.shimane-u.ac.jp

自己紹介

昭和58年 4月(1983)大阪大学医学部医学科入学

平成元年 3月(1989)大阪大学医学部医学科卒業

平成元年 6月 千葉大学医学部附属病院医員(研修医)

平成 2年10月 松戸市立病院救急部医員

平成 3年10月 千葉大学医学部附属病院医員(医療情報部)

平成 5年 5月 東京医科歯科大学助手

(難治疾患研究所医薬情報)

平成11年 5月(1999) 島根医科大学助教授

(医学部医学科医療情報学講座)

平成12年 7月 島根医科大学教授

平成15年10月 島根大学教授

平成19年10月 統計数理研究所 教授(客員)

現在に至る

学位: 平成9年 3月 博士(工学) 東工大

概要

- ・ ここ20年の医療の変化
 - EBMの普及
 - 診療情報の電子的蓄積が進む
 - FDAによる認可プロセス
- ・ データマイニングの応用例
 - 髄膜脳炎データベース (1995)
 - インシデントレポートの解析 (1) (2000)
 - インシデントレポートの解析 (2) (2017)
 - 検査データの解析 (2005)
- ・ まとめ

ここ20年で変わったこと

- 医療の側
 - EBMの浸透
 - 診療ガイドライン: 診療のアルゴリズム化
 - 診療情報の電子化
- 情報の側
 - 計算機の速度の向上, 並列処理の深化
 - 記憶容量の爆発的増加
 - インターネットの進展
 - 電子化された情報の利活用の技術の深化
- FDAがAI技術を含む医療機器を認可しはじめた

例：認知症ガイドライン

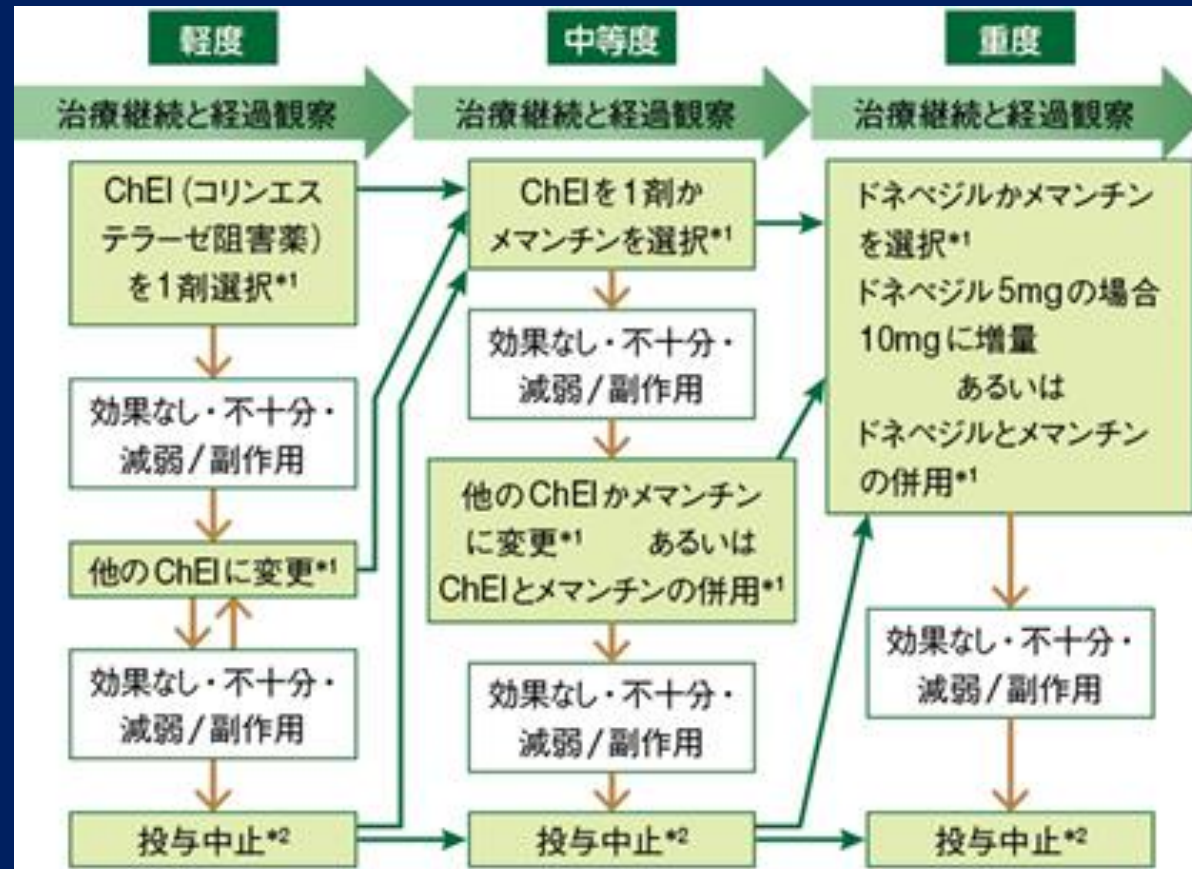


図 アルツハイマー型認知症・病期別の治療薬剤選択のアルゴリズム

* 1: 薬剤の特徴と使用歴を考慮して選択

* 2: 急速に認知機能低下進行例があり、投与中止の判断は慎重に

(「認知症疾患診療ガイドライン2017」より)

ここ20年で変わったこと

- 医療の側
 - EBMの浸透
 - 診療ガイドライン: 診療のアルゴリズム化
 - 診療情報の電子化
- 情報の側
 - 計算機の速度の向上, 並列処理の深化
 - 記憶容量の爆発的増加
 - インターネットの進展
 - 電子化された情報の利活用の技術の深化
- FDAがAI技術を含む医療機器を認可しはじめた

医療現場において

- 入力の手間
 - 電子カルテの普及
 - 検査機器のネットワーク対応, IoT,
- 診療支援
 - インターネット資源の充実
 - DBとしての電子カルテ
 - AIを用いた医療機器の登場
- AI技術の導入環境ははじまっている。

医療応用 (Nature, 2019)

- 画像診断がほとんど

領域	画像
放射線科/神経内科	CT 頭部急性病変
	CT 脳出血
	CT 頭部外傷
	CXR (転移性肺腫瘍)
	CXR 肺病変
	マンモグラフィ
	手首単純X線
病理	乳がん
	肺がん
	脳腫瘍
	乳がん転移
	乳がん転移
皮膚科	皮膚がん
	メラノーマ
	皮膚病変

医療応用(2) (Nature, 2019)

- 画像診断がほとんど

領域	画像
眼科	糖尿病性網膜症
	糖尿病性網膜症
	糖尿病性網膜症
	先天的白内障
	網膜疾患
	加齢黄斑変性症
	未熟児網膜症
	AMDと糖尿病性網膜症
	消化器
	大腸内視鏡でのポリープ
循環器	心電図
	心電図

医療応用(3) (Nature, 2019)

- 画像以外での機械学習/深層学習の適用例
 - ほとんど深層学習が使われていない

項目	N	AUC	深層学習
3-12ヶ月以内の死亡率	221284	0.93	○
再入院	1068	0.78	特徴選択モデル
敗血症	230936	0.67	サポートベクターマシン
敗血症ショック	16234	0.83	ベイズ解析
クロストリジウム感染	256732	0.82	ロジスティック回帰 (GLM?)
認知症	76367	0.91	クラスタリング
アルツハイマー病 (+アミロイド像)	273	0.91	ロジスティック回帰
133疾患でのオンセット	298000	0.7-0.9	○
自殺	5543	0.84	ロジスティック回帰
せん妄	18223	0.68	ロジスティック回帰

医療応用(4) (Nature, 2019)

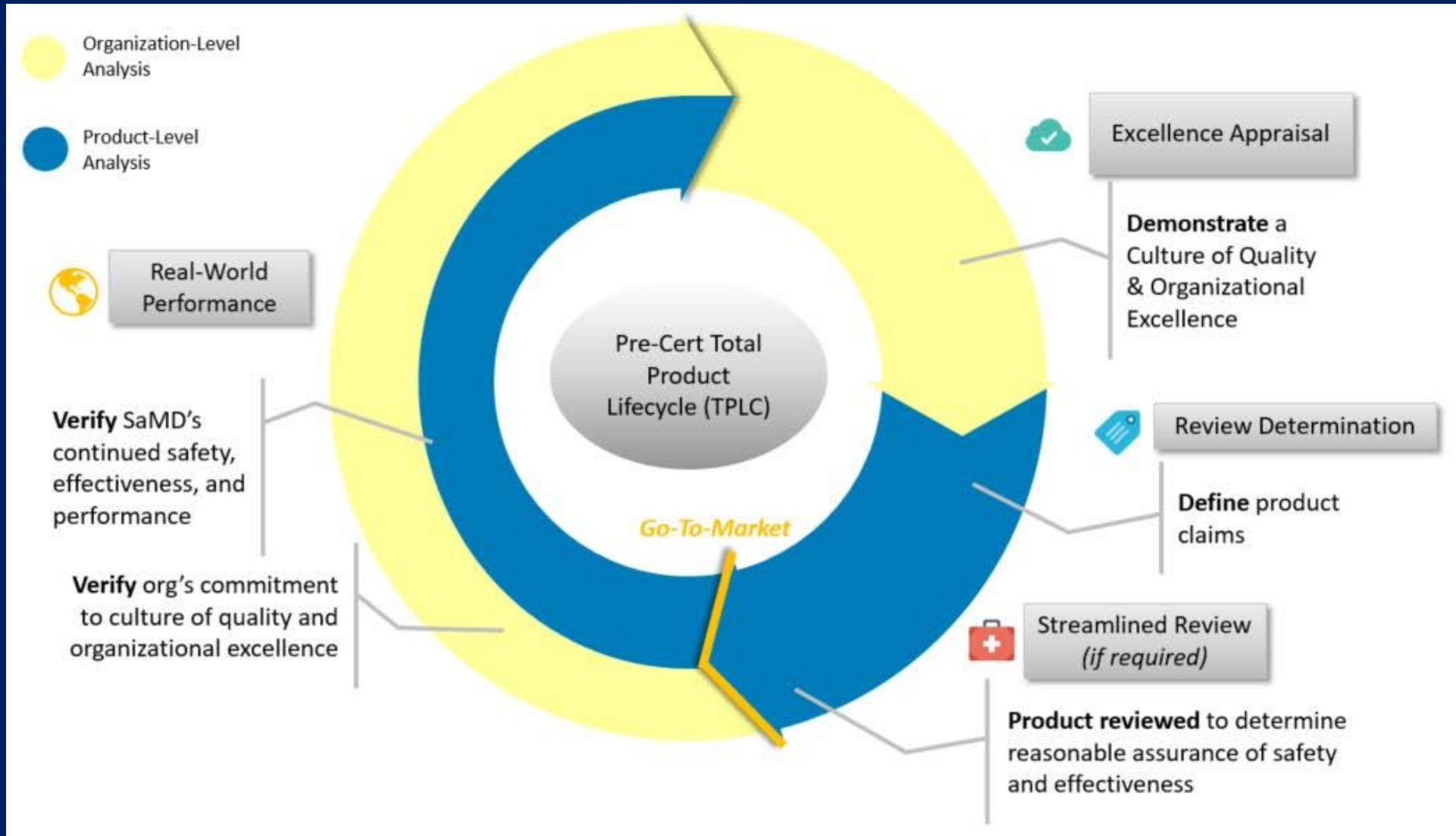
- 昨年からFDAで医療機器への認可が激増。

会社	認可時期	内容
Apple	2018年9月	心房細動検知
Aidoc	2018年8月	診断
iCAD	2018年8月	マンモグラフィでの診断
Zebra Medical	2018年7月	冠動脈石灰化スコアリング
Bay Labs	2018年6月	心エコーEF値決定
Neural Analytics	2018年5月	パラメディカルによる脳卒中診断装置
IDx	2018年4月	糖尿病性網膜症診断
Icometrix	2018年4月	頭部MRI解析
Imagen	2018年3月	単純X線での手首骨折
Viz.ai	2018年2月	CTでの脳卒中診断
Arterys	2018年2月	肝・肺癌診断(MRI,CT)
MaxQ-AI	2018年1月	CT 脳出血診断
Alivector	2017年11月	Apple Watchでの心房細動検知
Arterys	2017年1月	心臓MRI解析

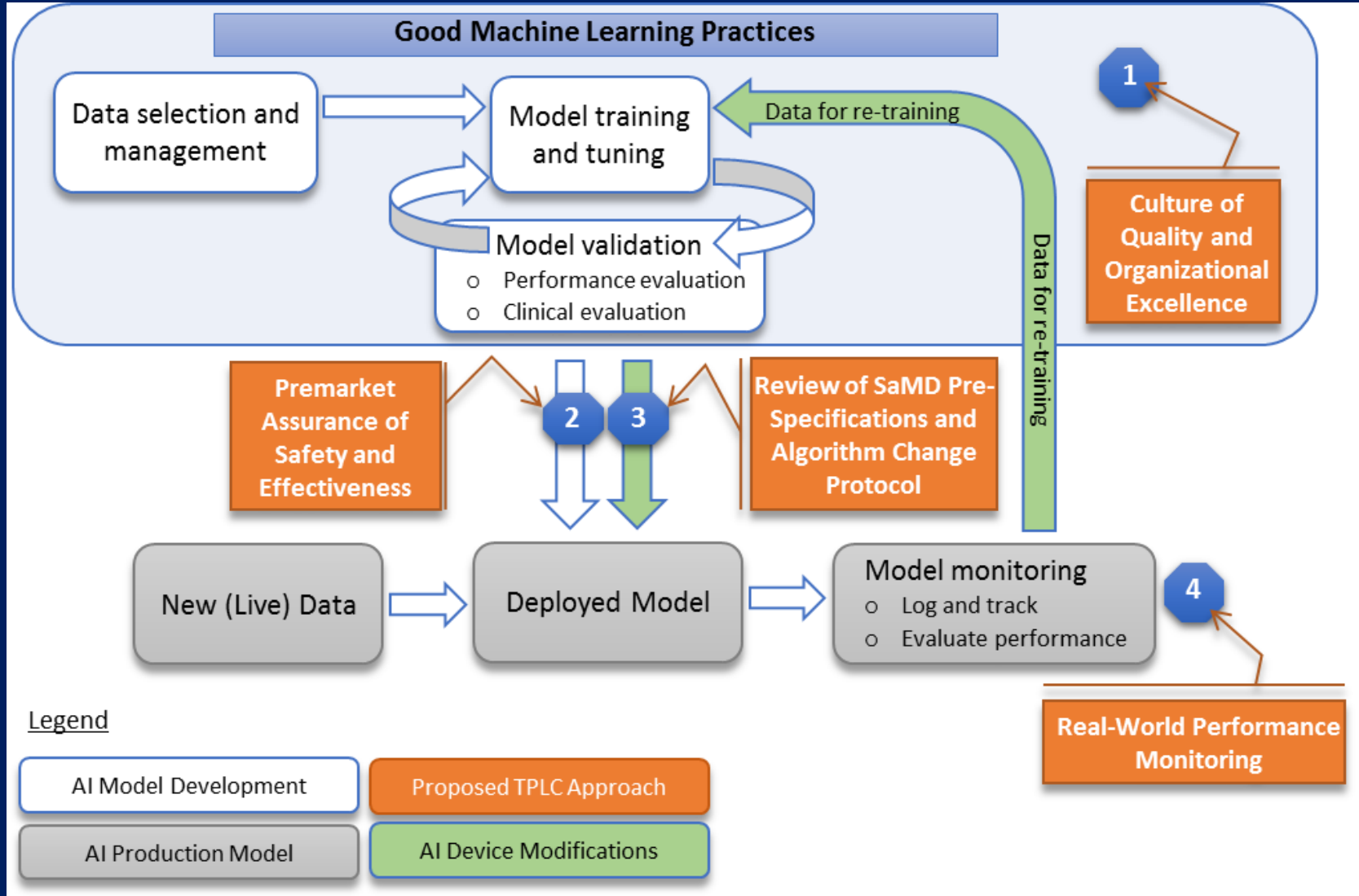
FDAの医療機器認可

- これまでは、ハードウェアが前提
 - 精度は限りなく100%に近いことが要求。
- **ソフトウェア: プログラム医療機器**
 - 精度が高いことが求められるが、学習によって精度が上がることで予想されるので、100%に近くなくてよい。
 - ハードウェアと同様、事前仕様の厳格化で対応。
 - **整ったソフトウェアライフサイクルが導入されているか?**

Pre-Cert Total Product Lifecycle



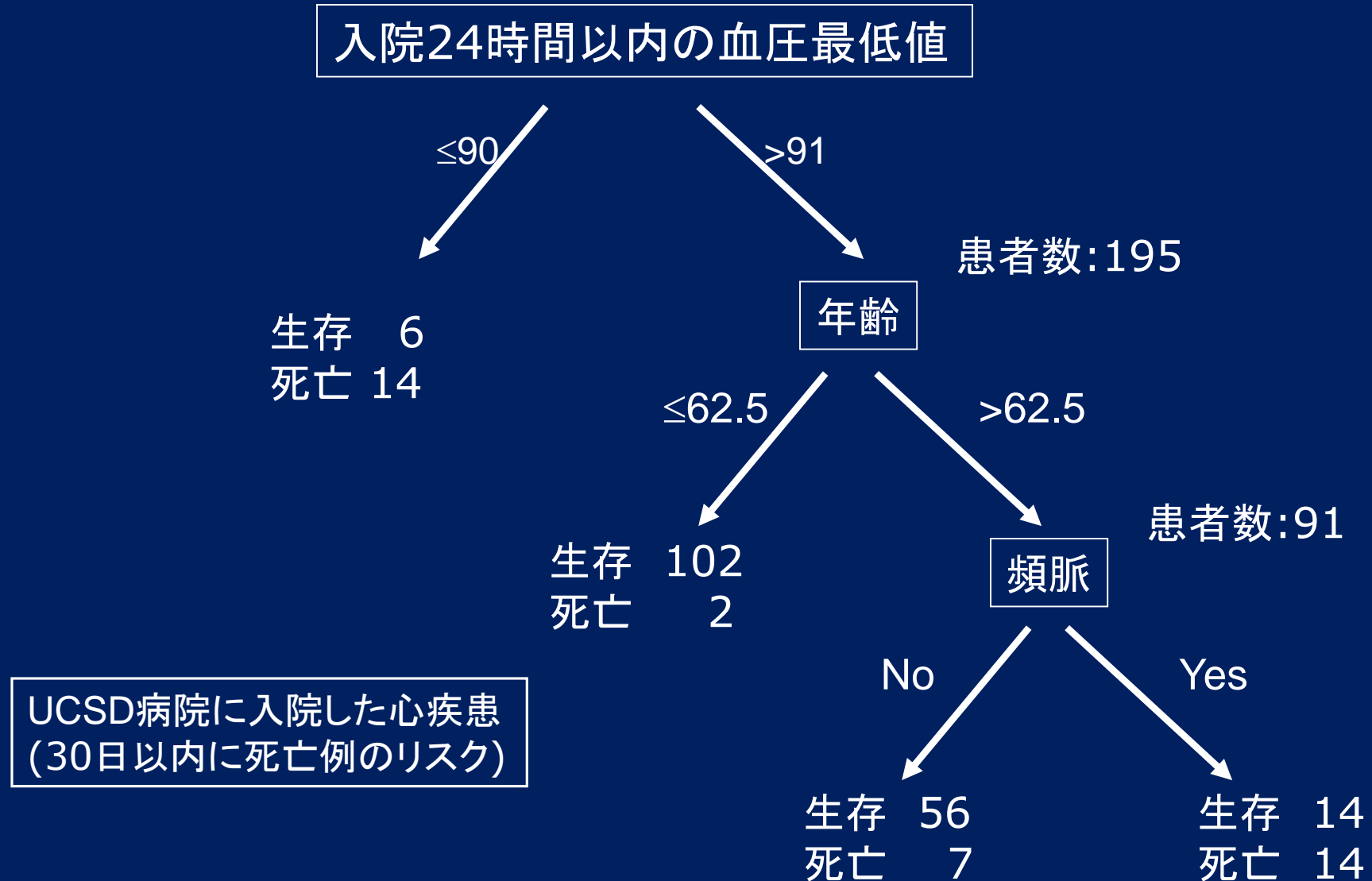
Good ML Practice



データマイニングの応用例

決定木
(CARTの応用例)

応用例1: 決定木(Breiman, 1984)



心疾患の死亡リスク

- 入院24時間以内の血圧最低値
 - 年齢
 - 頻脈
- 救急部での経験からすると、思い当たることが多い。
- 以下に示すように、得られた結果がすぐに使えるとは限らない。

データマイニングの応用例 (2)

髄膜炎のリスク因子の発見

W.Ziarko教授との共同研究

髄膜炎データベース(1995)

(Tsumoto and Ziarko, 1995)

髄膜炎データベース:

昭和54年から平成4年までの約14年間に
千葉県東北部の施設で入院治療した121人

入院時の病歴、検査所見、専門医の鑑別診断、
治療内容、治療後の経過、転帰

標本数: 121, 属性数: 34, クラス数: 2

専門家に予想されなかったルール

1. [白血球数 \leq 12000] & [性別=女性] & [CSF細胞数 \leq 1000]
⇒ ウィルス性髄膜炎
(Accuracy: 0.97, Coverage: 0.55)
2. [年齢>40] & [白血球数>8000] ⇒ 細菌性髄膜炎
(Accuracy: 0.80, Coverage: 0.58)
3. [白血球数>8000] & [性別=男性] ⇒ 細菌性髄膜炎
(Accuracy: 0.78, Coverage: 0.58)
4. [性別=男性] & [CSF細胞数>1000] ⇒ 細菌性髄膜炎
(Accuracy: 0.77, Coverage: 0.73)

男性，年齢>40 ? : 細菌性髄膜炎

一般的な知識とは一致しない



何らかのシナリオが必要



男性，年齢>40 ? → 糖尿病等，生活習慣病
→ 細菌性髄膜炎

予想されなかったルールの再評価

性別: 男性: 74/121 30例: 細菌性 (30/37)
12例: 慢性 (12/14)
2例: 糖尿病
3例: 肝硬変
7例: 慢性副鼻腔炎

年齢>40: 51/121 23例: 細菌性 (23/37)
7例: 慢性疾患 (7/14)
1例: 糖尿病
2例: 肝硬変
4例: 慢性副鼻腔炎

得られた結果からの反省

- 男性, 年齢といった項目は, およそ一般的な専門知識から外れている
 - 外れている場合の説明は?
 - 「情報の粒度」というのが大事。
 - 専門家は粒度を細かくしようとする。
 - 粒度を粗くすることで, 違ったものが見えることがある。
- **項目に関する知識の階層が大事。**

データマイニングの応用例 (3)

インシデントレポートの解析
(松岡氏, 横山氏(Kdicons)との共同研究)

医療事故防止のためのデータマイニング

看護師の内服薬事故: 2001年4月～2002年3月
(とある救急部)

- ・前半6ヶ月: 従来フォーマットでの内服薬事故データ



要因の検知



- ・後半6ヶ月: 「環境因子」=年数、勤務人員、多忙度、患者数等の追加データの収集



詳細な分析



リスク回避 (内服薬ミスの激減)

Tsumoto S and Hirano S
(2010) Fundamenta Informaticae
98: 107-121

決定木 (2)

*** 決定木 ***

- 12不穏数 B棟基準1 = 0: 準夜 (4→3)
- 12不穏数 B棟基準1 = 1以上:
 - 15-1出室検査の出室 = 出室: 深夜 (2→2)
 - 15-1出室検査の出室 = 無かった:
 - 6)集中力の欠如 = 0:
 - 経験年数分類(1:3年未満、2:3から10年未満、3:10年未満) = 1: 日勤 (4→4)
 - 経験年数分類(1:3年未満、2:3から10年未満、3:10年未満) = 2: 日勤 (3→3)
 - 経験年数分類(1:3年未満、2:3から10年未満、3:10年未満) = 3: 準夜 (2→2)
 - 6)集中力の欠如 = 1:
 - 11大量輸液療法数B棟基準1 = 1以上: 日勤 (2→2)
 - 11大量輸液療法数B棟基準1 = 0:
 - 受け持ち患者の重症度 = 軽度: 準夜 (1→1)
 - 受け持ち患者の重症度 = 重: 深夜 (3→2)
 - 受け持ち患者の重症度 = 中: 深夜 (3→2)

**不穏患者数が1以上ならば、
時間帯、経験年数等に関係なく、
高確率 90%(18/20)で、内服薬ミスが生じる**

シナリオ

内服薬ミス要因の決定木による「環境因子の解析」から、

- 1) 不穏患者数が1以上ならば、高確率90%で、
内服薬ミス。
- 2) 経験年数、時間帯、などに関わり無く発生。

現状分析： 内服薬の準備(種類,数量等)は、勤務時間帯の担当者が実施していた。

勤務の引継ぎ等で、勤務時間開始前の準備時間が30分以下の場合もあり、事前の薬の仕分けができず、勤務時間中に行うので、**不穏患者に注意をとられ**、ダブルチェックができず、ミスが起きやすい。

(シナリオ： 仮説)

シナリオからの改善

内服薬ミスの対策には、何を改善すればよいのか。

- 経験年数、時間帯等に関係の無いことが判明.
- 業務への割り込みが重要.
- 勤務が終わった者が次の時間帯の投薬準備を2名で実施(ダブルチェック)。
- 投薬実施者は、数量、種類などをチェック。
- 実施後、内服薬ミスが激減した。(10分の1に減少)

245例 → 24例

得られた結果からの反省

- インシデントは「経験年数が大きく関わっている」という「常識」
 - 得られた項目は常識から外れている。
 - 外れている場合の説明は？
 - 「情報の粒度」だけではなく、「シナリオ」も大事。
- 項目がワークフローに関するものであれば、ワークフローの流れを細かく考えることが大事。

データマイニングの応用例 (4)

インシデントレポートの解析

(電通大田中教授, 坂東氏との共同研究)

応用例2: クラスタリング

- 日本医療機能評価機構のヒヤリハット事例DB
 - 検索条件
 - 報告事例区分: ヒヤリハット事例
 - 検索キーワード: 知識不足, 作業中断, 判断ミス, チェック漏れ, 操作ミス, 誤操作, ルール違反
 - 検索結果: 867件
 - 処理対象カラム: 15カラムを選択.
 - 事業区分, 事例ID, 発生年月, 発生場所, 概要, 患者の年齢・性別, 疾患名, 患者の状況, 当事者職, 事故内容, 事故詳細内容, 行動に係る要因, ヒューマンファクタ, 背景要因, 改善策

応用例2: クラスタリング

- 15カラムについて、距離を計算、距離の近さでグループ分け
- CLS5-0: 薬剤管理問題/227 件(以下「件」省略)
 - CLS10-0: 与薬, 投薬等の配薬ミス, 患者の自己管理ミスなど/211
 - CLS10-5: 内服薬, 投薬等の配薬ミス, 処方ミス/16
- CLS5-1: 医療機器に関する問題/177
 - CLS10-1: 輸血時のフィルタの扱いの問題/18
 - CLS10-4: 人工呼吸器, モニタ等医療機器の接続, 設定の問題/97
 - CLS10-6: ジャクソンリース, 酸素ボンベ等医療機器の準備に関する問題/14
 - CLS10-7: 輸液ポンプの設定, 操作に関する問題/35 CLS10-8: 放射線治療や MRI 造影時に食事やラコール等を止めなかった問題/13
- CLS5-2: 血糖値測定とインスリン注射に関する問題/73
 - CLS10-2: 食事との関連, スケール確認等の問題/73
- CLS5-3: 点滴に関する問題/357
 - CLS10-3: 点滴の投与時間ミス, 投薬ミス/357
- CLS5-4: 輸液ポンプに関する問題/33
 - CLS10-9: 滴数や投与時間に関する設定ミス, 接続ルートミス/33

得られた結果からの反省

- 全国のデータから見ても、与薬事故が多いことがわかる。
- 与薬事故が経験年数と関係なければ...
 - 与薬事故が、院内のワークフローの指標となる可能性が高い。
 - インシデントレポートの解析によって、もう少し追求できるかもしれない。
 - プロセスに関する知識をまとめておくことは大事。

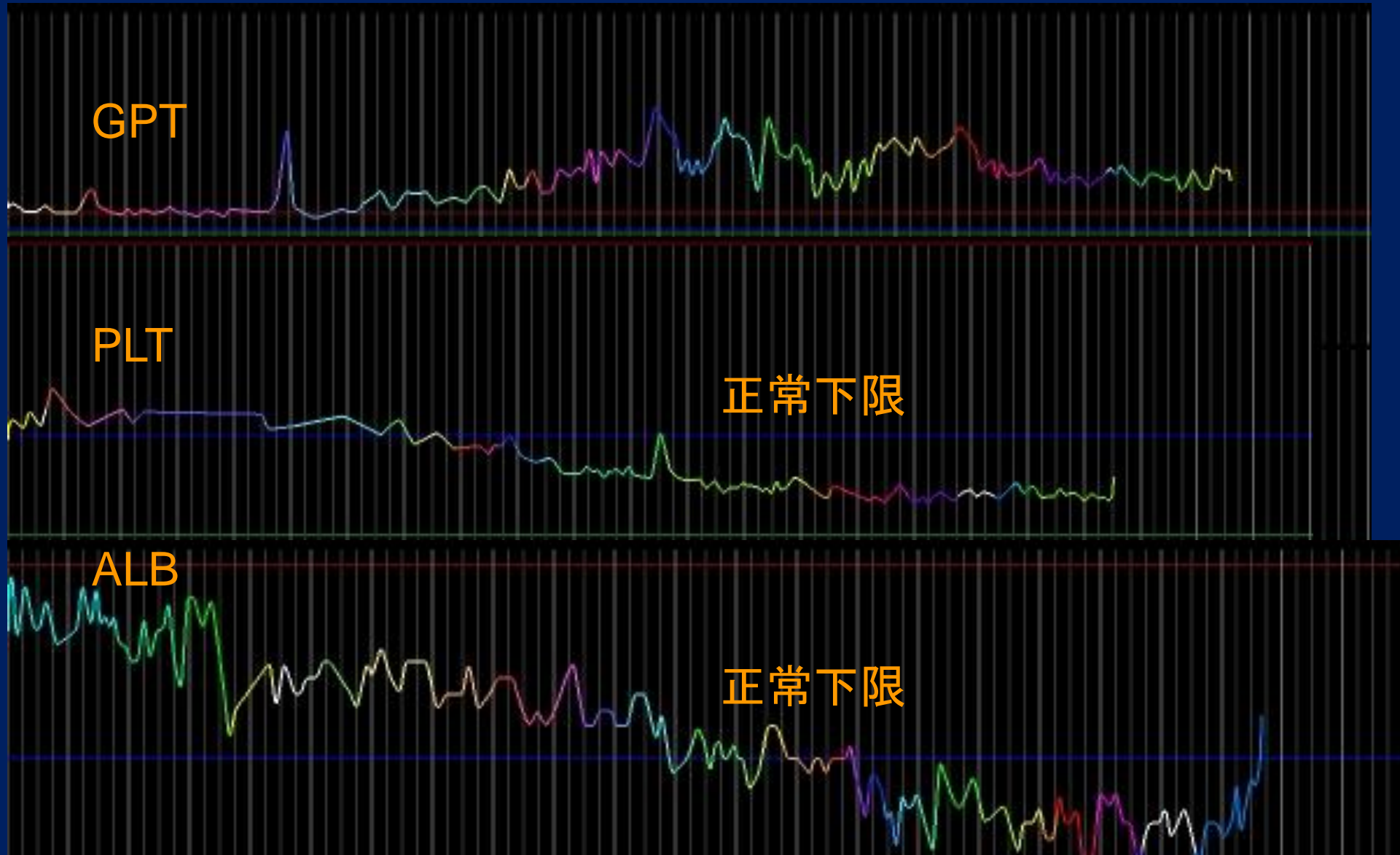
データマイニングの応用例 (5)

検査時系列データの解析

応用例

- データ
 - 慢性肝炎の検査時系列データ (1980～2000)
 - C型肝炎自然経過例
- 方法
 - 2変量を選択, その軌跡についての類型化
 - Multiscale Matchingによる非類似度の算出
 - クラスタリング:Ward法

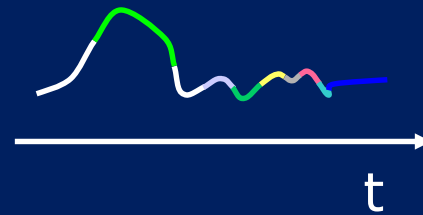
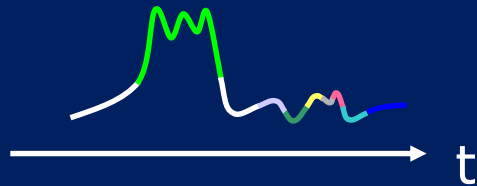
慢性C型肝炎の自然経過例



軌跡マイニング

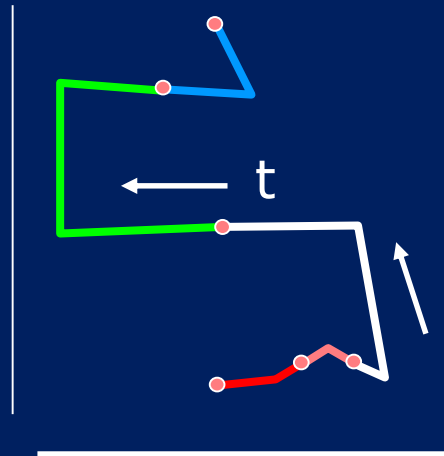
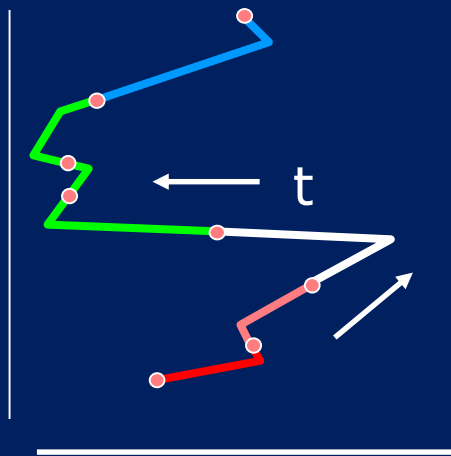
- 通常の時系列:

1変量の時間的发展



Tsumoto S and Hirano S
(2005)
IJIS 20: 737-757

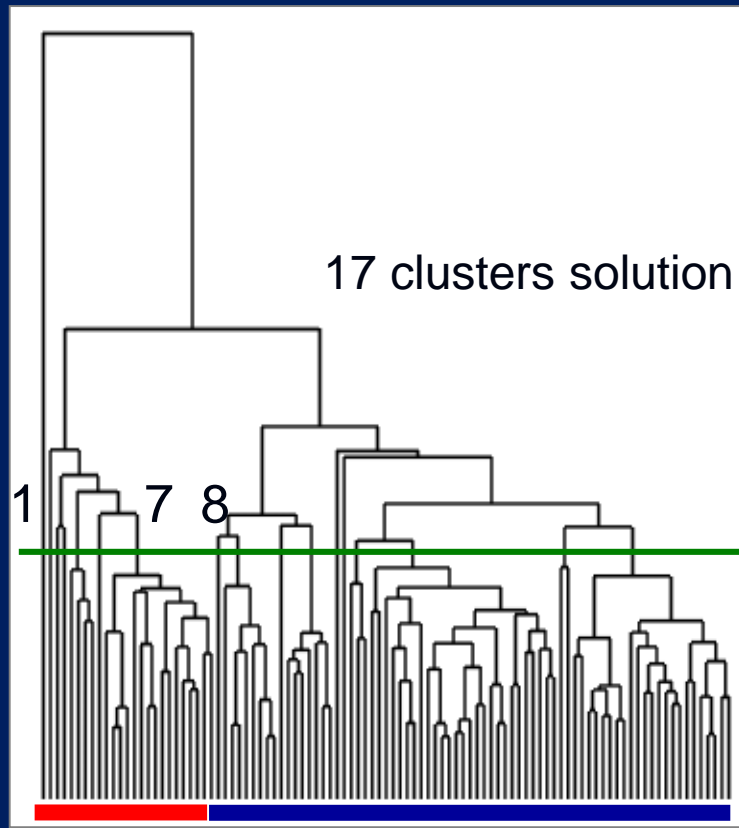
- **軌跡**: 2変量以上の時間的发展



Tsumoto S and Hirano S
(2010)
J Intell Inf Syst .

クラスター (アルブミン-血小板数)

- デンドログラム



進行例

非進行例

- 肝生検による繊維化度との関連

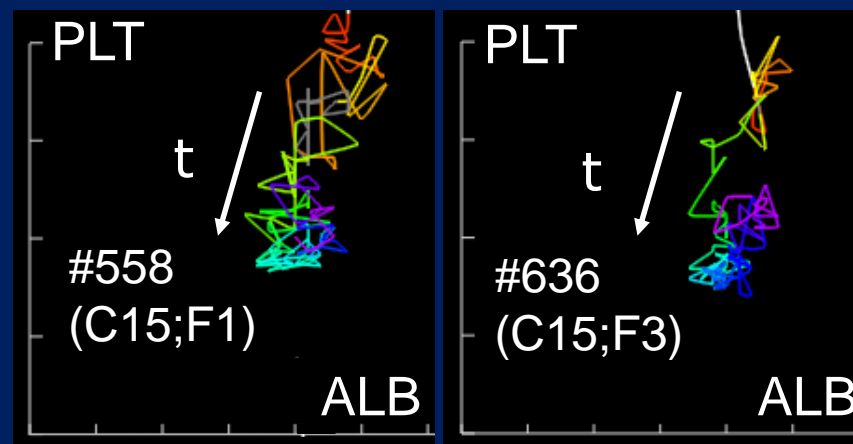
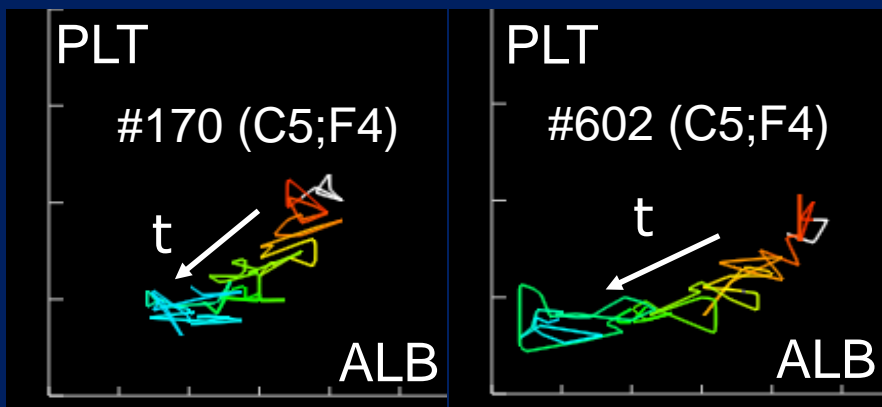
Cluster	Number of Cases/Fibrotic Stage				Total
	F0,F1	F2	F3	F4	
5	0	1	0	3	4
7	3	2	2	9	16
9	6	2	0	0	8
11	7	0	0	0	7
14	2	1	0	0	3
15	17	2	7	1	27
16	1	0	1	0	2
17	20	2	1	0	23

* Nine clusters of N<2 omitted

得られたクラスタからの発見

ALB, PLT 共変量

ALB, PLT 共変性なし

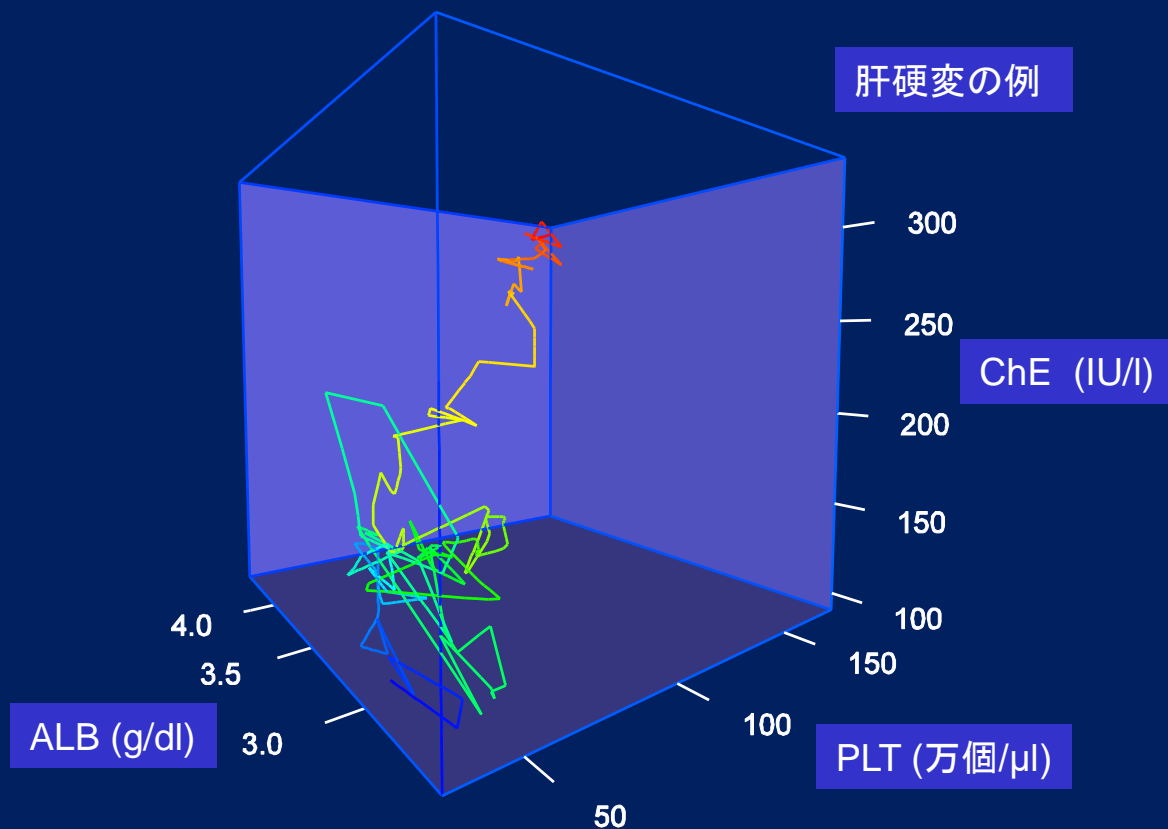


肝線維化の進展に関する2つのグループ

Group1: ALB, PLT: ともに減少 → 進行が早い

Group2: PLT: 減少, ALT: 一定

慢性C型肝炎の系列類型化 (3次元)



3次元軌跡のクラスタ分析結果

クラスタ8 (99例中15例)
血小板がまず減少,
その後ChEとALBが共起的に減少

世界ではじめて

3次元拡張にも成功:

Hirano S and Tsumoto S (2010) IJITD
9(6): 889-904

得られた結果からの反省

- アルブミン(ALB)と血小板数は肝障害のマーカー
- アルブミン: 体液中の蛋白, 血小板数: 出血傾向
- 経験的には, 血小板減少の方が早い。
- 検査マーカーに関する「時間に関する知識」がヒントに。

解析プロセス

- 解析目的の変数があるか？
 - あれば, ルール生成, 決定木, SVM, 深層学習
 - 選択された変量について「**解釈**」する。
 - なければ, クラスタリングで類型化
 - 類型化の「**解釈**」
 - 類型化の結果を使って, 教師あり学習
 - 類型化に関する変量について「**解釈**」する。
- 解釈しやすい変量とは？

まとめ

- ・ ここ20年の医療の変化
- ・ データマイニングの応用例
 - 髄膜脳炎データベース (1995)
 - ・ 選択された変量の解釈 → 診療支援
 - インシデントレポートの解析 (1) (2000)
 - ・ 選択された変量の解釈 → プロセスの改善
 - インシデントレポートの解析 (2) (2017)
 - ・ より大きなデータの解析 → 診療支援
 - 検査データの解析 (2000)
 - ・ 得られたクラスの解釈
- ・ 得られた結果の解釈とは？

まとめ (2)

- ・ 得られた結果の解釈を専門家に促す
 - 情報の粒度が重要
 - 粒度を変えることでわからなければ、より深く考える必要もある。
- 情報は何を「測定」しているのか？
- その測定結果がどのような管理に使えるのか？
 - ・ 専門家との議論で、それを引き出す。

