

オーダー実行歴からのクリニカルパス作成の試み

Mining Clinical Pathway in Service Logs of Hospital information System

津本周作 *1 平野章二 *1 岩田春子 *2
Shusaku Tsumoto Shoji Hirano Haruko Iwata

*1 島根大学医学部医療情報学

Department of Medical Informatics, Faculty of Medicine, Shimane University

*2 島根大学医学部附属病院入退院管理センター

Center for Bed Control, Shimane University Hospital

This paper proposes a data mining based method which constructs clinical process conducted by nurses. The methods consist of three process. First, data on counting sum of executed orders are extracted from hospital information system with a given temporal granularity. Then, similarity-based methods, such as clustering and multidimensional scaling (MDS) are applied to the data and the labels for grouping are obtained. By using the labels, rule induction is applied, and classification power of each attribute is estimated. The attributes are sorted by an index of classification power, the original dataset is decomposed into subtables. Clustering, rule induction and table decomposition methods are applied to the subtables in a recursive way. The method was applied to datasets stored in hospital information system stored in 10 years. The results show that the reuse of stored data will give a powerful tool for construction of clinical process, which can be viewed as data-oriented management of nursing schedule.

1. はじめに

病院情報システム (Hospital Information System: HIS) は 1980 年代後半より大学病院をはじめとした大規模病院で運用が開始された。導入以前, 病院では, 医師が指示箋を書き, それを患者あるいはメッセンジャーが各部門に送ることで, 検査・処方等が行われてきた。指示伝票を送付する仕組みを電子計算機で肩代わりさせ, 時間的負担を軽減することが目的で, オーダーエントリーシステムが導入された [松本 06]。その後, 計算資源の著しい進展を背景に, オーダーエントリーではなく, 検査や処方等の実施歴, 医師・看護師他医療スタッフの記載をも電子計算機で蓄積・管理することが提唱された。1990 年代後半から, これらの仕組みを統合的に実現した病院情報システムいわゆる電子カルテ [電子 13][篠田 08] が登場, この導入は徐々に進み, 2012 年現在では, ほぼ 100% の大学病院, 70% 以上の大規模病院が電子カルテシステムを導入するに至っている。

HIS は, 病院における様々な診療情報を集約し, それをもって, 診療における意思決定支援, 病院における診療の品質管理を目的として構築されてきた。ここに蓄積される情報は大きく診療行為の指示, 指示の実行結果に大別できるが, システムの性質上, 指示を効率よく入力させる, 実行結果を可視化・出力させる, 指示・実行結果をリアルタイムに蓄積・検索する方法が必須とされ, ここに AI で研究されている方法が革新的な技術として貢献できる可能性がある [津本 08, 津本 11, Shusaku Tsumoto 14]。これまでは紙ベースのシステムを電子化するということを目標とされてきたが, ほぼその開発過程は終了し, 病院情報システムの知能化に目を向けられつつあり, 現在研究・開発中の技術がここに投入されることで, 病院情報システムが大きく変わっていく可能性がある。

本抄録では, クラスタリング等の類似度解析を用いて, HIS

に蓄積されたオーダー履歴から自動的にクリニカルパスを作成するシステムについて提案する。

2. クリニカルパス

クリニカルパスとは, 入院時の診断・治療を効率的に行うためのスケジュール表と定義できる。表 1 は白内障手術の看護クリニカルパスの例である。治療過程は 3 つのステージ: 術前, 術日, 術後に分かれるが, 入院日を 0 とし, それぞれの日にちに行われる看護行為のリストが与えられている。

3. 方法

前節のパスを自動的に生成するにはどのような手法が考えられるか。病院情報システムには, 各入院患者の看護行為が記録されており, これらを取り出すことによって, 実施された看護行為が上記のパスに沿ったものかどうかの類似度を確認できる。この向きを逆にし, ある疾患の入院患者のデータを集め, そこから実施された看護行為の時間的推移を観察, 類似した看護行為をグループ化することによって, 上記のスケジュールと同様の看護行為のリストを推定できることが予想される。

3.1 データの準備

HIS のオーダー歴は業務系システムの履歴として蓄積されている。この蓄積されたデータから, まず各入院患者の看護履歴を抽出し, 表 1 に近い形式の表を作成する。図 1 は, 病院情報システムに蓄積された診療行為の履歴を時間経緯に従ったプロットとして表示した。このようなデータをアルゴリズム 1 を用いて表 2 のような形式として作成する。

3.2 マイニング

図 2. にマイニングの過程を図示した。まず, 得られたデータについて, クラスタリング [Everitt 11] を適用する。クラスタリングによって得られた看護オーダーの分割によって, ラベルをつけ, 病日を属性としてルール生成の方法を適用する。こ

連絡先: 津本周作, 島根大学医学部医療情報学講座,
出雲市塩冶町 89-1, 0853-20-2171, 0853-20-2170,
tsumoto@med.shimane-u.ac.jp

表 1: 白内障クリニカルパスの例

術前	術日			術後		
-1(日)	0	1	2	3	4	5
体温/脈拍 血圧						
	吐き気 嘔吐	吐き気 嘔吐	吐き気 嘔吐	吐き気 嘔吐	吐き気 嘔吐	吐き気 嘔吐
		清拭	清拭	清拭	清拭	清拭
	眼部疼痛	眼部疼痛	眼部疼痛	眼部疼痛	眼部疼痛	眼部疼痛
術前教育						

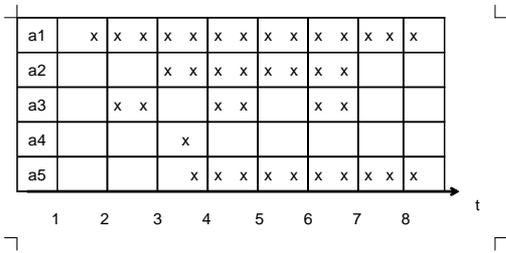


図 1: 実行歴の模式図

表 2: 図 1 における看護実施数

actions	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9
a ₁	1	2	2	2	2	2	2	1
a ₂	0	0	2	2	2	2	0	0
a ₃	0	2	0	2	0	2	0	0
a ₄	0	0	1	0	0	9	0	0
a ₅	0	0	1	2	2	2	0	0

のルールの情報から、病日の重要度を用いて、病日のグループ分けを行い、グループごとにデータを分割し、再度、上記プロセスを繰り返す。含まれる属性(病日)が一つになったところで、各子プロセスは停止する。これらの方法は、Rの上スクリプトとして記述した。

3.3 クリニカルパス生成

図 2 はマイニング結果を受けて、パス用のスケジュールをアウトプットするアルゴリズムを示した。

3.4 評価

専門家から獲得されたクリニカルパスが前もって与えられている時は、生成されたパスと比較することによって、生成法の妥当性を定量的に比較できる。一番簡単な方法は、それぞれの病日での看護リストを比較した分割表を作成し、各病日の類似度を計算し、全体の類似度を平均値として計算することである。以下の評価実験では、類似度として比較的性能が良く、

Algorithm 1 データ前処理

```

Input:  $L_p$  = List of Patients for a given Disease
Input:  $T_g$  = List of Temporal Granularity
Output: List of Counter
while  $L_p \neq \{\}$  do
   $Pt \leftarrow \text{car}(L_p)$ 
  Pick up the data for Pt
   $D_a \leftarrow$  date of admission
   $D_d \leftarrow$  date of discharge
  for  $i = 0$  to  $D_d - D_a + 1$  do
    for all  $j \in T_g$  do
       $List \leftarrow$  List of Nursing Orders for  $D_a + i$  and  $j$ 
      while  $List \neq \{\}$  do
         $Order \leftarrow \text{car}(List)$ 
         $Counter[i, j, Order] = +1$ 
         $List \leftarrow \text{cdr}(List)$ 
      end while
    end for
  end for
   $L_p \leftarrow \text{cdr}(L_p)$ 
end while
Return List of Counter
  
```

Algorithm 2 パス生成

```

procedure CONSTRUCTION_PROCESS( $Level_v, Level_h$ )
   $List \leftarrow \text{Labels}(Level_v, Level_h)$ 
  while  $List \neq \{\}$  do
     $Order \leftarrow \text{car}(List)$ 
    if  $Evaluation(Order) > Threshold$  then
      for all  $attr(Level_h)$  do
        Append Order into
         $List_{pathway}(Level_v, attr(Level_v, Level_h))$ 
         $\triangleright attr(Level_v, Level_h)$ : List of Dates
      end for
    end if
     $List \leftarrow \text{cdr}(List)$ 
  end while
end procedure
  
```

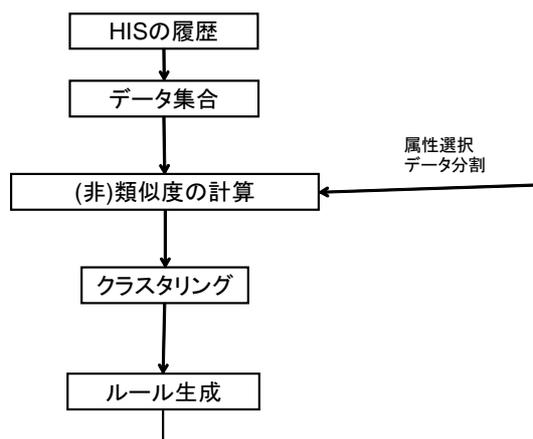


図 2: マイニングプロセス

表 3: 類似度の計算結果

	平均 Jaccard 係数
白内障 (両側)	0.75
白内障 (片側)	0.81
肺がん (手術あり)	0.70
膀胱がん	0.82

計算が簡単な Jaccard 係数 [Everitt 92, Cox 00] を用いた。

4. 実験結果

著者らが所属する施設の附属病院で 2013 年度に入院した全患者を対象に、DPC コード別に分類、上位 10 疾患についての患者の看護行為の履歴データを抽出した (表 4)。

4.1 基本的な手法の評価

まず、クラスタリングとして Ward 法、ルール生成として C5.0 の方法を用い、上記データからクリニカルパスを生成した。

表 5 に得られたクラスタの情報の概要: ステージ、看護オーダー数の多いクラスタ 3 つのステージ、それから看護オーダー数の推移から予測される最適な入院日数を示した。たとえば、白内障 (両側) では、主に 4 つのステージ: 0, {1,2,3}, 4, 5 以上に分かれ、上位 3 つのオーダー数は 9, 5, 3 となる。最後の 3 は、体温・脈拍、血圧、眼部疼痛のみであり、ほとんど外来での経過観察でも問題のないレベルといえるため、最適な日数は 4 と予想できる。

表 3 に、クリニカルパスが前もって与えられた場合の類似度の平均を示した。

4.2 クラスタリングとルール分割法の比較

上記評価では、Ward 法および C5.0 をマイニングの方法として用いた。これらの方法については、クラスタリングを別の方法、ルールマイニングに関しては、離散化の方法を変えて、その性能を比較することが可能である。そこで、クラスタリングとして、K means, Fuzzy c means, Complete Linkage, Ward 法、それぞれの非類似度としてユークリッド距離、カイ 2 乗統計量を使った得られたパスとの平均類似度の結果を表 6 に示した。

次に、離散化の方法として、Precision, Recall, F-値, Lift, Information Gain, χ^2 統計量, AUC について、生成されたパスと既存のパスとの平均類似度の結果を表 7 に示した。

これらの結果から、抽出したデータについては、クラスタリングとしては Ward 法、離散化の方法としては、AUC を使用するのが最も性能がよかった。

5. おわりに

本抄録では、クラスタリングとルール生成の方法とを組み合わせ、クリニカルパスを生成するシステムについて提案した。本稿では示さなかったが、クラスタリングでは、クラスタの分割数をどのように生成するかという問題について未解決である。これについても、さまざまな方法が提案されているので、本来は、これらの比較も必要であろう。今後、クラスタ分割数についての実験を行い、それらの評価も試みる予定である。

また、既存のクリニカルパスとの比較において、どのような類似性を定義するかも課題である。ここでは、きわめて単純に、各病日の類似度を平均化して考えたが、本来は得られた構造を比較すべきところであり、ここに構造物の類似性という困難な課題が存在する。この課題についても今後検討したい。

参考文献

- [Cox 00] Cox, T. and Cox, M.: *Multidimensional Scaling*, Chapman & Hall/CRC, Boca Raton, 2nd edition (2000)
- [Everitt 92] Everitt, B.: *The Analysis of Contingency Tables*, Chapman & Hall/CRC, 2nd edition (1992)
- [Everitt 11] Everitt, B. S., Landau, S., Leese, M., and Stahl, D.: *Cluster Analysis*, Wiley, 5th edition (2011)
- [Shusaku Tsumoto 14] Shusaku Tsumoto, , Shoji Hirano, , Haruko Iwata, , and Yuko Tsumoto, : Similarity-based behavior and process mining of medical practices, *Future Generation Computing Systems*, Vol. 33, pp. 21–31 (2014)
- [篠田 08] 篠田英範: 病院情報システムの現況と今後の動向, 医療機器システム白書 2008-2009, 月刊新医療, 第 35 巻, pp. 261–265, エム・イー振興協会, 東京 (2008)
- [松本 06] 松本武浩, 宮下光世, 本多正幸: 病院情報システム導入効果の検討, 病院管理, Vol. 43, p. 129 (2006)
- [津本 08] 津本周作, 平野章二, 津本優子: 医療リスクの計量化, 計量生物学, Vol. 29, pp. 155–176 (2008)
- [津本 11] 津本周作, 平野章二, 阿部秀尚: 病院情報システムに基づく診療サービスの創出に向けて, 人工知能学会誌, Vol. 26, No. 2, pp. 139–146 (2011)
- [電子 13] 電子カルテ&PACS 白書 2013~2014 年版, 月刊新医療別冊, エム・イー振興協会, 東京 (2013)

表 4: データ集合の基本情報

	パスの有無	症例数	使用看護オーダー数	最小入院日数	中央値	最大
白内障 (両側)	有	1674	89	4	4	16
緑内障	無	1273	78	3	7	14
白内障 (片側)	有	1273	78	3	3	6
肺がん (手術あり)	有	1072	170	5	7	50
脳梗塞	無	1061	209	1	12	> 70
網膜剥離	無	887	125	5	14	21
膀胱がん	有	859	105	4	5	18
膝関節外傷	No	812	67	3	17	44
肺がん (手術なし)	No	776	135	1	6	34
胆管・胆嚢がん	No	683	113	1	6	55

表 5: 得られたクラスタの情報
オーダー数の多いステージ

	ステージ数	グループ			使用オーダー数			予想される 最適入院日数
		1	2	3	1	2	3	
白内障 (両側)	4	0	1 to 3	4	9	5	3	4
緑内障	3	-1, 3 to 5	0 to 2	others	7	7	14	3
白内障 (片側)	4	-1	0	1,2	3	2	9	3
肺がん (手術あり)	8	1	2, 3	4 to 7	7	13	10	7
脳梗塞	5	2, 10 to 15	3 to 9	17 to 24	34	38	16	15
網膜剥離	3	0 to 12	-2, -1, 13 to 21	others	7	10	19	12
膀胱がん	6	0	1	2	11	6	3	2
膝関節外傷	4	0 to 8	9 to 15	-1, 16 to 31	5	10	11	15
肺がん (手術なし)	5	2 to 4	3 to 9	1, 10 to 15	14	9	9	7
胆管・胆嚢がん	4	2 to 10	11 to 15	1, 16 to 31	10	18	10	15

表 6: クラスタリング法の比較

	K-means		Fuzzy C-means		Complete Linkage		Ward method	
	Euclidean	χ^2	Euclidean	χ^2	Euclidean	χ^2	Euclidean	χ^2
白内障 (両側)	0.67	0.58	0.74	0.67	0.81	0.79	0.84	0.79
白内障 (片側)	0.74	0.78	0.63	0.72	0.75	0.81	0.77	0.82
肺がん (手術あり)	0.84	0.68	0.75	0.78	0.81	0.74	0.85	0.73
膀胱がん	0.82	0.78	0.85	0.79	0.93	0.82	0.95	0.82

表 7: 離散化手法に関する比較

	Precision	Recall	F-value	Lift	Information Gain	χ^2 -square	AUC
白内障 (両側)	0.61	0.51	0.67	0.59	0.69	0.71	0.79
白内障 (片側)	0.65	0.41	0.65	0.71	0.78	0.79	0.85
肺がん (手術あり)	0.59	0.42	0.65	0.61	0.73	0.68	0.73
膀胱がん	0.77	0.64	0.74	0.71	0.83	0.91	0.91