

平成 30 年 6 月 21 日現在

機関番号：15201

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26330253

研究課題名(和文) 位置情報を付与した「質的プロセス - 動態データマイニング」に関する研究

研究課題名(英文) A research on process-dynamics mining from spatiotemporal process sequences

研究代表者

平野 章二 (Hirano, Shoji)

島根大学・医学部・准教授

研究者番号：60333506

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究では、質的時系列に位置情報を加えた「プロセス - 動態データマイニング法」の開発に取り組んだ。主な成果は以下のとおりである。(1)外来診療イベントの発生源情報を元に位置情報を質的診療プロセス系列へ組み込む方法を考案し、受付から会計に至るまでの各部門の患者数動態を可視化した。(2)外来診療プロセスから診療所要時間を予測する診療プロセスマイニング手法を考案し、所要時間が長いケースでは処置や検査等が診察の間に実施される、あるいは注射が含まれる等の知見を得た。(3)期間付き状態関係に基づく新たな診療プロセスマイニング法を考案、イベントの継続期間や間隔の長短を含んだ頻出パターンの獲得を可能とした。

研究成果の概要(英文)：This research investigated process-dynamics mining, which facilitates knowledge discovery from spatiotemporal process sequences. Followings are main contributions. (1) Proposed a method to visualize the dynamics of outpatients in a hospital based on the clinical action records and input device locations. (2) Proposed a method to analyze the relationships between the stay time of outpatients and treatment processes they received during their hospital visits. The results suggested that a visit needs long time if treatments or examinations should be executed during the consultation, or if injections such as intravenous drips were involved in the process. (3) Proposed a temporal pattern mining method based on ranged relations. Experimental results demonstrated that the proposed method could generate frequent patterns with abstracted time ranges embedded in their temporal relations.

研究分野：データマイニング

キーワード：データマイニング 診療プロセス 位置情報

1. 研究開始当初の背景

医療機関においては診療業務の実施に伴い様々な部署で多種多様な情報が生成される。それらは患者さん自身の治療のために一次利用されるものであるが、電子化され蓄積された情報は、専門家である医師らにより組み立てられた診療計画や実際の治療経過を収めた大規模な時系列診療データベースを構成する。その有効な二次利用に向けて、慢性疾患における臨床検査データの時系列パターンマイニング、疾患共通のケアプロセスを獲得するプロセスマイニングなど、様々な医療データマイニング法が提案されてきた。一方、従来の手法は静的な情報あるいはそれに時間軸を加えたものが主体であり、位置情報を取り入れたアプローチは極めて少ない。転倒や徘徊の検知、あるいは救急などの限定された区域における位置追跡に RFID や各種センサ等を活用する方法は多く研究されているが、外来を含む一般患者さんに位置検出用デバイスを携帯して頂くことは現実的に難しく、医療機関における位置情報の活用は進んでいない現状にある。

2. 研究の目的

本研究では、質的な診療プロセスに位置情報という新たな軸を加えたプロセス-動態データマイニング法の開発に取り組む。病院窓口等における受付履歴、検査室や処置室における検査/処置等の実施履歴から患者さんごとに質的な診療プロセス時系列を生成し、これに発生源(入力端末)の位置情報を付与することで、位置の移動を伴う診療プロセス時系列のデータベースを構築する。そのマイニングを通じて、従来の時間軸のみによる分析では困難であった、プロセスと人の動きを関連づける知識の獲得を目指す。

3. 研究の方法

(1) 診療プロセスに関する基礎的分析と位置動態の可視化

外来診療に係るオーダ情報、実施情報、発生源(検査機器や診療端末等)の識別情報及び設置場所情報を病院情報システムから抽出、位置情報を伴う質的な診療プロセスの時系列データを構築し頻度情報もとに分析した。抽出対象としたイベントは、受付、診察、処置自科検査、各種検査(検体、生理機能、放射線画像、超音波・内視鏡)、注射、会計である。データの抽出後、前処理として、イベントの発生時刻を30分単位の粒度へ、また、診察室の位置を階単位の粒度へ抽象化した。続いて、来院から会計へ至るまでに記録されたイベントを事例ごとに時系列に並べ、診療プロセス系列を構成した。抽出期間は2013年度においてランダムに選択した1週間であり、含まれる事例数は延べ3,443例、イベント数は延べ21,548件である。

分析では、1. 多数例が迎る外来診療プロセ

スや、検査/診察など診療行為の順序分布に関する統計解析、2. イベント発生源の位置=患者位置とみなしたときの位置動態(受付、検査部、2階診察室など各区域に位置する患者数の時間推移)の可視化を試みた。

(2) 外来診療プロセスと診療所要時間に関する時系列マイニング

診療プロセスを条件として位置情報に関するマイニングを行う場合、例えば検査や処置を診察前に実施できる場合と診察途中に実施して結果を評価する必要がある場合では患者さんの動きや診療所要時間に違いを生じることが想定される。そのため、ある行為の途中に、あるいは並行して、別の行為が行われることを前提に行為間の時間関係を記述する必要がある。本研究では、区間の取り扱いが可能である Batal らの時系列マイニング法(ACM Trans Intell Syst Technol. 4:4 (2013))を応用し、以下の方法により診療所要時間を特徴付ける頻出診療パターンの導出を試みた。

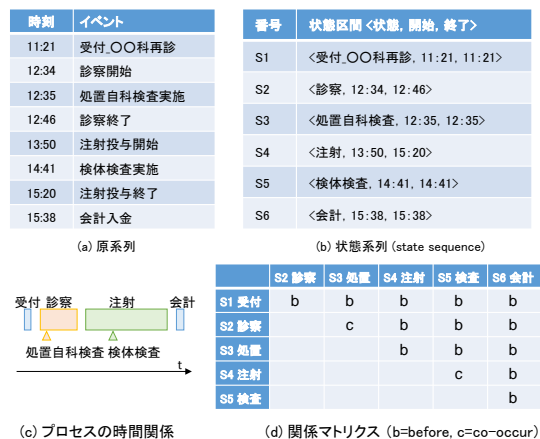


図1 診療プロセスの状態区間表現

まず、(1)と同様の診療イベントを、1事例1日ごとにまとめて図1(a)に示すようなイベント時系列を構成する。続いて、各イベントをその開始時刻及び終了時刻と組み合わせ、同図(b)に示すように状態と時間区間により記述する。例えば診察であれば、<診察, 12:34, 12:46>となる。それらを開始時刻に関して昇順に並べ、S1, S2等の番号を付したものを状態系列とする。次に、各状態の時間関係を表す関係マトリクスを生成する。この例の場合、各イベントは同図(c)に示す時間関係で発生しており、その関係は同図(d)に示すマトリクスで表現できる。診察中に行われる処置自科検査などは co-occur の関係で表される。以上で生成した状態系列と関係マトリクスをあわせて診療プロセスのパターンとする。

クラス属性となる診療所要時間は受付から会計までの経過時間とし、それを次の4段階にカテゴリ化して短い順にクラスAからDとした。

A: 1時間未満(2,822例)

- B: 1 時間以上 3 時間未満 (8, 219 例)
- C: 3 時間以上 5 時間未満 (1, 593 例)
- D: 5 時間以上 (423 例)

マイニングは、対象となる時間クラスを 1 つ選択し、当該クラスを正例、それ以外の 3 クラスをまとめて負例として Batal らのアルゴリズムを適用し、当該時間クラスの予測に寄与する診療プロセスの Minimal Predictive Temporal Pattern (MPTP) を生成する方式で実施した。対象データは、ランダムに選択した 1 ヶ月分の外来診療に係るイベントデータであり、計 13, 057 例が含まれる。

### (3) 期間付き状態関係に基づく頻出パターンマイニング

前出(2)の方法で生成されるパターンは、診察や検査などの行為の前後関係や共起関係を反映する一方、それらの持続時間や間隔の長短に関する情報を含まない。すなわち、処置や注射の実施に要する時間、他のプロセスとの間隔の長短などをパターンとして記述できない。そこで、本研究では Batal らの時系列マイニング法を拡張し、状態の継続期間や間隔を明示的に組み込んだ新たな診療プロセスマイニング法を考案した。提案法は、before, co-occur with の 2 種類の状態関係を、分、時間、日、週など抽象化された期間を伴う期間付き状態関係へと拡張することで、状態の持続時間や間隔の長短を含む進展パターンの記述を可能とするものである。

Batal らの方法における頻出パターンの抽出は、候補の生成と支持度に基づく候補選別の 2 段階で行われる。候補の生成では、長さ  $k$  の頻出パターンの先頭に頻出 1-pattern を 1 つ挿入することで、新たな候補  $(k+1)$ -pattern を生成する。例として、 $S_1, S_2, S_3$  という 3 つの頻出 1-pattern があり、 $S_2 \langle \text{before} \rangle S_3$  という頻出 2-pattern から候補 3-pattern を生成するケースを考える。簡単のため、以降は関係 before と co-occur をそれぞれ  $\langle b \rangle, \langle c \rangle$  と略記する。まず、この 2-pattern の先頭に  $S_1$  を挿入する。新たに挿入した  $S_1$  と既にある 2 つの状態との関係は未定であるため、論理的に矛盾のない関係の組み合わせ ( $\langle b \rangle, \langle b \rangle$ ), ( $\langle c \rangle, \langle b \rangle$ ), ( $\langle c \rangle, \langle c \rangle$ ) の 3 種類を列挙して関係マトリクスの先頭行を埋めることで、候補 3-pattern が 3 つ生成される。それらの内、支持度が閾値を上回るものが頻出 3-pattern となる。

提案法では、新たに「期間付き状態関係」を考案し、上記のプロセスに組み込む。まず、開始時刻で昇順ソートされた状態系列  $S$  において、2 つの状態  $(S_i, S_j)$ :  $i < j$  の時間差  $d_{ij}$  を次式により定義する。

$$d_{ij} = b_j - e_i$$

ここで、 $b_j$  は状態  $S_j$  の開始日時、 $e_i$  は状態  $S_i$  の終了日時である。 $S_j$  は  $S_i$  より後にあるため、 $d_{ij} > 0$  の場合は before 関係、 $d_{ij} \leq 0$  の場合は co-occur 関係となる。次に、 $d_{ij}$  のレンジに応じて、表 1 に示す 6 種類の期間付き状態関

表 1 期間付き状態関係

Relation	Meaning	Range $d_{ij}$ [days]
$\langle cd \rangle$	co-occur several days with	[-7, 0]
$\langle cw \rangle$	co-occur several weeks with	[-31, -7]
$\langle cm \rangle$	co-occur several months with	(inf, -31)
$\langle bd \rangle$	several days before	(0, 7]
$\langle bw \rangle$	several weeks before	(7, 31]
$\langle bm \rangle$	several month before	(31, inf)

係を定義する。状態の数及び期間の定義は任意であるが、ここでは日、週、月の単位で 3 段階に抽象化した。これらの関係においても同様に時間制約があり、関係マトリクスにおいて、いずれかの before 関係が先行する場合、それ以降、同じ行では before 関係のみが許容される。また、 $bd < bw < bm$  の関係性から、後ろに来る before 関係は前の before 関係と同等以上の長さである場合のみ許容される。同様に、 $cd < cw < cm$  の関係性から、後ろに来る co-occur 関係は、前の co-occur 関係と同等以下の長さである場合のみ許容される。これらの制約に基づき、不要な候補パターンを除外する。その他の手続きは基本的に元のアルゴリズムと同じである。

## 4. 研究成果

### (1) 診療プロセスに関する基礎的分析と位置動態の可視化

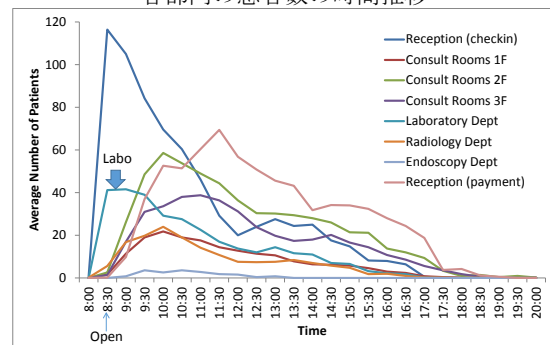
表 2 に、頻度の高い外来診療プロセス系列の上位 5 件を示す。受付 (CHK)-診察 (CSL)-会計 (PAY) が約 22.9% と最も多く、続いて診察の間に処置や自科検査を含むプロセスが約 13% を占めた。第 3 位と第 4 位のプロセスは診察前に臨床検査 (LAB) あるいは放射線画像検査 (RAD) などを行うもので、フォローアップなどのケースが想定される。同様に診察前検査を伴うプロセスが全体の約 17% であった。

図 2 に、イベント発生源の位置を患者位置とみなした場合の、各部門での患者数の時間推移 (5 日平均) を示す。8:30 の開院直後に

表 2 頻度の多い診療プロセス系列

Rank	Course	# of Cases	Ratio [%]
1	CHK-CSLs-CSLe-PAY	788	22.9
2	CHK-CSLs-TRE-CSLe-PAY	448	13.0
3	CHK-LAB-CSLs-TRE-CSLe-PAY	398	11.6
4	CHK-RADs-RADe-CSLs-CSLe-PAY	102	3.0
5	CHK-LAB-CSLs-CSLe-PAY	95	2.8

図 2 発生源位置から推定した各部門の患者数の時間推移



まず受付と臨床検査部門が混雑のピークを迎え、しばらくして診療部門へピークが移行する。午前中の終わり頃には会計窓口がピークを迎え、以降、夕方まで時間をかけて緩やかに減少する動態であった。これらの結果について、文献⑦及び⑥にて発表した。

(2) 外来診療プロセスと診療所要時間に関する時系列マイニング

表3 クラスBを正例としたMPTP

No	Pattern	Supp. (%)	Conf.
1	[処置自科検査]	0.58	0.69
2	[検査]	0.25	0.69
3	[検査] before [処置自科検査]	0.17	0.73
4	[検査] before [診察]	0.14	0.72
5	[受付_眼科再診]	0.11	0.75

表4 クラスCを正例としたMPTP

No	Pattern	Supp. (%)	Conf.
1	[処置自科検査]	0.61	0.14
2	[検査]	0.39	0.21
3	[診察] co-occurs with [処置自科検査]	0.37	0.15
4	[放射線]	0.30	0.24
5	[注射]	0.16	0.29

表5 クラスDを正例としたMPTP

No	Pattern	Supp. (%)	Conf.
1	[検査]	0.39	0.05
2	[放射線]	0.31	0.07
3	[注射]	0.24	0.11
4	[処置自科検査] before [注射]	0.16	0.17
5	[検査] before [注射]	0.14	0.20

クラスB (所要時間が1時間以上3時間未満) を正例として抽出したMPTPを表3に示す(\*クラス内支持度順上位5件;以下同様)。このクラスは例数が多く、全体の約63%が該当する。最初の2つは診療科において実施される処置や自科検査、あるいは検査部において実施される血液や尿などの検体検査が含まれることを条件とする1-patternである。次の2つは関係を含む2-patternであり、処置や診察の前に検体検査を受けるパターンがMPTPとして抽出されている。

クラスC (所要時間が3時間以上5時間未満) を正例として抽出したMPTPを表4に示す。このクラスは会計まで比較的長い時間を要しており、全体の約12%が該当する。抽出されたMPTPのうち最初の2つの1-patternはクラスBと同様である。このクラスから関係co-occurを含む2-patternが出現し、診察の途中で処置自科検査あるいは放射線画像検査が行われるパターンが抽出されている。また、このクラスから注射を含むパターンがMPTPとして出現する。

最後に、クラスD (所要時間が5時間以上) を正例として抽出したMPTPを表5に示す。このクラスは全体の約3%である。クラスCと同様に注射及び放射線検査を含む2-patternが多く、注射を含む2-patternがより上位に出現している。診療科の傾向からも、外来化学療法点滴など長時間の注射を伴うケースが本クラスに多く含まれ、対応するMPTPが得られたと考えられる。

所要時間の長いクラスではMPTPの中に関係co-occurが比較的多く見られること、ま

た、注射を含むMPTPが多く見られることから、診察中に処置や検査を実施し評価する必要がある場合、あるいは点滴等の注射を含むプロセスである場合に所要時間が長くなる傾向が示唆された。これらの結果について、文献⑤及び④にて発表した。

(3) 期間付き状態関係に基づく頻出パターンマイニング

テストデータ及び慢性関節リウマチの検査・投薬データに対して提案法を適用し、期間付き状態関係に基づくMPTPの抽出実験及び所要時間の評価を行った。テストデータは、3つの状態からなる簡素な系列をベースとして、状態間の関係とデータ数を様々に変化させ人工的に生成したものである。慢性関節リウマチデータは、ある2年の観察期間を前半1年間と後半1年間に区分し、それぞれ活動性指標(Simplified Disease Activity Index値を4段階に区分したもの)の最大値を求め、その差から活動性減少群(n=102)、不変群(n=181)、増加群(n=64)へクラス分けしたものである。条件属性は上記2年間に実施された臨床検査の結果(HまたはLのみ)と、処方及び注射の実施歴であり、検査・投薬の時系列と疾患活動性の変化を関連づけるパターンをMPTPとして抽出する。

MPTP抽出の所要時間を計測した結果を表6及び表7に示す。本実験では、期間付き関係の増減による影響をあわせて評価するため、関係before, co-occurそれぞれが1種類(B1C1)から3種類(B3C3)である場合の3条件で計測を行った。同表の右から3列目は、データ読み込み後、期間付き関係に基づき全てのケースを状態系列へ変換する処理の所要時間である。この処理はnの増加に伴い著しく増加するが、関係数増加の影響は顕著には見られなかった。同表の右から2列目は、期間付き関係を用いて頻出1-patternから順次

表6 テストデータにおけるMPTP抽出所要時間

Type of relations	Number of cases	全ケースの状態系列作成		候補生成・選別・MPTP評価		Number of sequences compared
		1. Time for making state sequences [sec]*	2. Time for extracting MPTPs [sec]*	2-ptn: 1200	3-ptn: 100	
B1C1	100	0.02±0.00	0.01±0.00			1,300
Before 1種	1,000	0.64±0.00	0.04±0.00			13,000
Co-occur 1種	5,000	13.31±0.00	0.18±0.00			65,000
	10,000	67.34±0.27	0.37±0.00			130,000
B2C2	100	0.02±0.00	0.02±0.00			2,520
Before 2種	1,000	0.64±0.00	0.07±0.00			25,200
Co-occur 2種	5,000	13.33±0.02	0.29±0.00			126,000
	10,000	67.22±0.07	0.61±0.00			252,000
B3C3	100	0.02±0.00	0.03±0.00	2-ptn: 3600		3,870
Before 3種	1,000	0.64±0.00	0.10±0.00	3-ptn: 270		38,700
Co-occur 3種	5,000	13.35±0.01	0.42±0.00			193,500
	10,000	67.35±0.13	0.88±0.01			387,000

\*mean±SD on 5 trials

表7 活動性減少群(n=102)におけるMPTP抽出所要時間

Type of relations	1. Time for making state sequences [sec]*	2. Time for extracting MPTPs [sec]*	Number of sequences compared
B1C1	10.58±0.26	1317.76±9.64	2-ptn: 1,056,991
B2C2	12.41±0.23	1254.76±7.04	38.5万回
B3C3	14.74±0.31	1926.11±8.59	2-ptn: 1,517,132
			115万回

\*mean±SD on 3 trials

長い候補パターンを生成し、原系列と照合して全ての頻出パターン及び MPTP を抽出する処理の所要時間である。この処理では、関係の種類増加に伴い候補パターンの組み合わせが増大し、特に低次のパターンの照合回数が大きく増加する一方、高次のパターンではサブパターンの支持度が低下し頻出パターンの数が減少することなどから全体の照合回数の増加は比較的抑制されていることが明らかとなった。

慢性関節リウマチデータの活動性減少群において抽出された MPTP の例を表 8 に示す。[MMP3]=H→(数ヶ月)→[フォリアミン錠5mg]=ON→(数ヶ月)→[好中球数]=L など、期間に関する情報がパターンに組み込まれ、従来法と比べてより詳しい関係性を表現できることが示された。これらの結果について、文献③、②、①において発表した。

表 8 活動性減少群 (n=102) で  
関係 B3C3 により抽出された MPTP の例

No	Conf	Supp	Pattern
1	0.846	0.108	[MMP-3]=H [フォリアミン錠5mg]=ON [好中球数]=L <bm> <bm>
2	0.706	0.118	[赤沈1時間値]=H [リウマトリックスカアセル2mg]=ON [フォリアミン錠5mg]=ON <bm> <bm> <cd>
3	0.684	0.127	[血算-白血球数]=L <bm> MMP-3: マトリックスメタロプロテイナーゼ-3

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 0 件)

[学会発表] (計 6 件)

- ① Shoji Hirano: Towards Knowledge Discovery from Heterogeneous Time-series Medical Databases. The 6th Int'l Conf on Informatics, Electronics & Vision (ICIEV) & 7th Int'l Symp in Computational Medical and Health Technology (ISCMHT) (invited) (2017).
- ② Shoji Hirano, Shusaku Tsumoto: Frequent Temporal Pattern Mining for Medical Data based on Ranged Relations. 2017 IEEE ICDM Workshop on Data Mining for Services. (2017).
- ③ 平野章二, 津本周作: 期間付き状態関係に基づく時系列医療データからの頻出パターンマイニング. 2017 年度人工知能学会全国大会. (2017).
- ④ Shoji Hirano: Temporal Data Mining on the Stay Time of Outpatients and Treatment Processes. 2016 IEEE ICDM Workshop on Data Mining for Services. (2016).
- ⑤ 平野章二, 津本周作: 外来診療プロセスと診療所要時間に関する時系列マイニング. 2016 年度人工知能学会全国大会. (2016).
- ⑥ Shoji Hirano, Shusaku Tsumoto: Visualizing dynamics of patients in hos-

pitals using devise locations. 2014 IEEE International Conference on Systems, Man, and Cybernetics. (2014).

- ⑦ Shoji Hirano, Shusaku Tsumoto: Visualization of Patient Distributions in a Hospital Based on the Clinical Actions Stored in EHR. 2014 IEEE ICDM Workshop on Data Mining for Services. (2014).

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

○取得状況 (計 0 件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
取得年月日:  
国内外の別:

[その他]  
ホームページ等

## 6. 研究組織

(1) 研究代表者

平野 章二 (HIRANO, Shoji)  
島根大学・医学部・准教授  
研究者番号: 60333506

(2) 研究分担者

( )

研究者番号:

(3) 連携研究者

( )

研究者番号:

(4) 研究協力者

津本 周作 (TSUMOTO, Shusaku)