

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

Temporal Data Mining on Treatment Processes and Stay Time in Outpatient Clinic

平野章二 *1 津本周作 *1

Shoji Hirano Shusaku Tsumoto

*1 島根大学医学部医療情報学講座

Department of Medical Informatics, Shimane University, School of Medicine

This paper presents a method for finding minimal patterns of treatment processes that characterize the stay time of patients in an outpatient clinic based on the temporal pattern mining algorithm proposed by Batal et al. We could observe MPTPs (Minimal Predictive Temporal Patterns) of treatment processes containing 'co-occur' relations as well as injections in the classes where the patients spent long time in receiving outpatient services.

1. はじめに

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を二次活用することで新たな診療知識の獲得や病院サービスの向上につながる事が期待されている [1]。

本研究では、外来診療における待ち時間短縮に資することを目的として、受付から会計に至るまでの一連の診療プロセスと診療所要時間との関係を時系列マイニングにより分析する。診療プロセスは質的な行為系列であるが、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全体の所要時間に違いが生じるため、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考慮して行為間の時間関係を記述する必要がある。本稿では、区間の取り扱いが可能である Batal ら [2] の時系列マイニング法を応用し、所要時間を特徴付ける診療パターンの導出を試みた結果について述べる。

2. 方法

2.1 Batal らの頻出パターン抽出法

Batal らの時系列マイニング法は、抽象化と Temporal Logic[3] を採り入れることで、多次元の量的時系列や質的な変数を含むデータセットから、イベントの時間関係 (前に生じる / 同時に起こるなど) を含む頻出パターンの生成を可能としている。

量的な時系列に対してはトレンド抽象化及び値抽象化を適用し、観測値を Decreasing(D), Steady(S) など記号による状態表現へ変換する。このときに、抽象化されたトレンドや値を表す記号 E と、対応する時間区間 (始点 b , 終点 e) の情報を組み合わせ、状態を $S = (E, b, e)$ の形式で表現する。例えば、ある患者の血小板の値が時間 $t_1 \sim t_5$ の区間において減少している場合は $S = (D, t_1, t_5)$ となる。これにより、任意の時系列を、 (E, b, c) の 3 つ組からなる状態 S の系列へと変換することができる。

パターンはこれらの状態を組み合わせることで表現できるが、各々が区間を伴うため、区間の重なりなど関係を定義しな

ければならない。Batal らは、Allen の論文 [3] で定義される 13 種類の関係の中から、“ E_i before E_j ” (ある状態 E_i が終わってから別の状態 E_j が始まる) と “ E_i co-occurs with E_j ” (ある状態 E_i が始まり、終わる前に別の状態 E_j が始まる) の 2 種類に絞って利用している (その他の関係は時間の同一性を評価するもので、時間の精度が悪い場合は問題が生じるため利用していない)。これにより、長さ k のパターン P_k を、 k 個の状態とその関係行列 R を用いて $P_k = (\langle S_1, S_2, \dots, S_k \rangle, R)$ と表現できる。以降、各分類クラスごとに頻出パターンマイニングのアルゴリズムを適用し、各クラスごとの最小サポート基準を満たす頻出パターンを抽出する。

Batal らのアルゴリズムでは、自身のサブパターンよりも有意に予測能力が高い MPTP (Minimal Predictive Temporal Pattern) を選ぶことで、spurious なパターンが選ばれることを抑制していることも大きな特徴である。MPTP の評価は、サブパターンで獲得された最良の confidence 値を母数とする片側二項検定により行われる。

2.2 診療プロセスの状態系列表現

まず、病院情報システムから受付・診察等の診療イベントを抽出し、イベントの時系列を生成する。本研究では、島根大学病院の病院情報管理システムに蓄積された外来診療データから次の項目を抽出して使用した。

- [受付] 窓口もしくは再来受付機での受付時刻、受診科及び初診再診区分。同日に複数科を受診するケースおよび救命救急を受診するケースは対象から除外した。
- [診察] 外来診察室の電子カルテにおいて診察開始及び診察終了の操作がなされた時刻。中座を伴う場合、あるいは開始終了操作を複数回行っている場合は最初の開始時刻と最後の終了時刻を使用した。
- [処置/自科検査] 診察室での処置 (創傷処置など) や診療科内で行われる自科検査 (屈折検査など) の実施時刻。
- [注射] 注射薬の投与開始時刻と投与終了時刻。
- [検体検査] 中央検査部において施行される検体検査 (血液検査、尿検査など) の実施時刻。
- [放射線] 中央放射線部において施行される画像検査 (レントゲン、CT など) の実施時刻。

連絡先: 島根大学医学部医療情報学講座 平野章二

〒 693-8501 島根県出雲市塩冶町 89-1

Phone:(0853)20-2173, E-mail:hirano@ieee.org

時刻	イベント	番号	状態区間<状態、開始、終了>
11:21	受付,〇〇科再診	S1	<受付,〇〇科再診, 11:21, 11:21>
12:34	診察開始	S2	<診察, 12:34, 12:46>
12:35	処置,自科検査実施	S3	<処置,自科検査, 12:35, 12:35>
12:46	診察終了	S4	<注射, 13:50, 15:20>
13:50	注射投与開始	S5	<検体検査, 14:41, 14:41>
14:41	検体検査実施	S6	<会計, 15:38, 15:38>
15:20	注射投与終了		
15:38	会計入金		

(a) 原系列 (b) 状態系列 (state sequence)

図 1: 状態系列による診療プロセス表現.

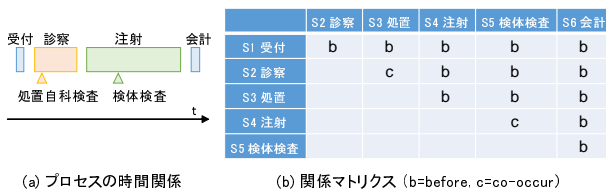


図 2: 状態間の関係マトリクス.

7. [会計] 窓口もしくは再来受付機での入金時刻。同日に時間を空けて複数回の入金操作が行われた場合は最初の入金時刻を使用した。支払いが受診当日でないケースは対象から除外した。

抽出したイベントを 1 患者 1 日ごとにまとめ、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。なお、開始と終了が明確に分かれておらず単一の実施時間のみ記録されている受付等のイベントにおいては、便宜上、実施時刻に 30 秒付加したものを終了時刻とした。それらを同図 (b) のように開始時刻に関して昇順に並べ、 s_1, s_2 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。図 1 の各イベントは図 2(a) に示す時間関係で発生しており、その関係は同図 (b) 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。

2.3 クラスの設定とマイニング

本研究では、診療プロセスから診療所要時間を予測するルールの生成を試みる。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:

- [1 時間未満 =A]
- [1 時間以上 3 時間未満 =B]
- [3 時間以上 5 時間未満 =C]
- [5 時間以上 =D]

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 Positive、それ以外の 3 クラスをまとめて Negative として Batal らのアルゴリズムを適用し、当該時間クラスの

表 1: 各クラスのデータ数と割合.

クラス名	所要時間	ケース数	割合
A	<1H	2,822	0.22
B	1H<, ≤3H	8,219	0.63
C	3H<, ≤5H	1,593	0.12
D	5H<	423	0.03
合計		13,057	1.00

予測に寄与する診療プロセスの MPTP を生成する方式で実施した。

3. 実験結果

3.1 データ及びパラメータ

直近 5 年からランダムに 1 ヶ月を選択し、2.2 節に示したデータを病院情報システムから抽出した。同節に挙げた条件の他に、欠損や不整合のあるデータを除外し、計 13,057 例からなるデータセットを構築した。なお、同一患者が月内に複数回受診している場合は各受診日につき 1 ケースのプロセス系列が作成されるので、ケース数と患者数の対応は必ずしも 1 対 1 ではない。所要時間クラスごとのデータ数を表 1 に示す。クラスの分布には偏りがみられ、クラス B が約 63% で最も多く、クラス A と併せて約 85% を占める。

マイニングには Batal らの論文における Algorithm 1 および Algorithm 2 を Python により独自に実装したプログラムを使用した。実験で用いたパラメータは、クラス内最小サポート値 $\sigma_y = 0.05$, MPTP を判別するための二項検定の片側有意水準は論文と同じく $\alpha = 0.01$ とした。データ読み込み後、MPTP 生成が完了するまでの時間は仮想マシン (Xeon 3.2GHz 1P/4C, 6GB Mem, SSD) 上で約 4 分であった。

3.2 頻出パターン

クラス A (所要時間が 1 時間未満) を Positive クラスとして抽出した MPTP を表 2 に示す。同表左から順にパターン番号、パターン、support, confidence, lift 値である。Batal らの論文では全データではなく Positive クラス内における support を評価基準としており、本稿の support 値もそれに倣って記載した。

このクラスで生成された MPTP は全て受付に関する 1-pattern であった。これは、受診科と初再診の情報だけで所要時間の短いケースを予測するものであり、逆に、それ以外の項目を条件に加えても予測が有意には良くならないことを示している。支持度は 4 割前後であるが、これらの診療科の再診においては比較的短時間で会計まで進む傾向が示されている。

次に、クラス B (所要時間が 1 時間以上 3 時間未満) を Positive クラスとして抽出した MPTP を表 3 に示す。このクラスは例数が多く、比較的 support 値の高い MPTP が生成されている一方、lift 値は低めのものが多い。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 Positive クラスとして抽出した 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 12% 程度が該当する。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

表 2: クラス A を Positive として抽出した MPTP.

No	Pattern	Supp	Conf	Lift
1	[受付_精神科再診]	0.14	0.46	2.13
2	[受付_皮膚科再診]	0.11	0.41	1.90
3	[受付_小児科再診]	0.07	0.36	1.67
4	[受付_婦人科再診]	0.07	0.34	1.57
5	[受付_整形外科再診]	0.07	0.28	1.30

表 3: クラス B を Positive として抽出した MPTP.

No	Pattern	Supp	Conf	Lift
1	[処置自科検査]	0.58	0.69	1.10
2	[検査]	0.25	0.69	1.10
3	[検査] before [処置自科検査]	0.17	0.73	1.16
4	[検査] before [診察]	0.14	0.72	1.14
5	[受付_眼科再診]	0.11	0.75	1.19
6	[受付_泌尿科再診]	0.09	0.71	1.13
7	[放射線] before [診察]	0.09	0.70	1.11
8	[受付_泌尿科再診] before [処置自科検査]	0.06	0.76	1.21
9	[受付_内分泌再診]	0.05	0.76	1.21

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 として出現し、lift 値も上がる傾向にある。なお、10 番目のパターンは診察と会計が重なるものであるが、これはカルテを閉じる操作が後刻になされた等の状況が考えられる。

最後に、クラス D(所要時間が 5 時間以上) を Positive クラスとして抽出した MPTP を表 5 に示す。このクラスは全体の 3% 程度であり、得られた MPTP の lift 値も総じて高い。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。

4. おわりに

本稿では、Batal らの時系列マイニング法を用いて、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の一手法を提案した。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。一方、診察の開始終了時刻については、カルテのオープン/クローズ時刻を基準としており、後刻に記載を加えた場合などは患者さんの診察室からの退室時刻と大幅にずれる可能性がある。次患者の呼び込み時刻等を参照してより正確な終了時刻を得る方法について今後検討を進めていきたい。

謝辞

本研究の一部は JSPS 科研費 (基盤研究 (C) 26330253) の助成による。

参考文献

[1] Jensen PB, Jensen LJ, Brunak S.: Mining electronic health records: towards better research applica-

表 4: クラス C を Positive として抽出した MPTP*.

No	Pattern	Supp	Conf	Lift
1	[処置自科検査]	0.61	0.14	1.15
2	[検査]	0.39	0.21	1.72
3	[診察] co-occurs with [処置自科検査]	0.37	0.15	1.23
4	[放射線]	0.30	0.24	1.97
5	[注射]	0.16	0.29	2.38
6	[受付_泌尿科再診]	0.11	0.17	1.39
7	[診察] co-occurs with [放射線]	0.10	0.39	3.20
8	[検査] before [放射線]	0.09	0.33	2.70
9	[処置自科検査] before [注射]	0.09	0.37	3.03
10	[診察] co-occurs with [会計]	0.08	0.17	1.39
11	[診察] co-occurs with [検査]	0.07	0.37	3.03
12	[検査] before [注射]	0.07	0.40	3.28

*support 値降順に 12/17 パターンを掲載

表 5: クラス D を Positive として抽出した MPTP.

No	Pattern	Supp	Conf	Lift
1	[検査]	0.39	0.05	1.54
2	[放射線]	0.31	0.07	2.16
3	[注射]	0.24	0.11	3.40
4	[処置自科検査] before [注射]	0.16	0.17	5.25
5	[検査] before [注射]	0.14	0.2	6.17
6	[検査] before [放射線]	0.11	0.11	3.40
7	[受付_消化内再診]	0.11	0.07	2.16
8	[診察] co-occurs with [放射線]	0.09	0.10	3.09
9	[注射] co-occurs with [会計]	0.08	0.25	7.72
10	[受付_腫血科再診]	0.08	0.11	3.40
11	[受付_腫血科再診] before [注射]	0.07	0.34	10.49
12	[受付_消化外再診]	0.05	0.06	1.85

tions and clinical care. Nat Rev Genet. 13(6):395–405 (2012).

[2] Batal I, Valizadegan H, Cooper GF, Hauskrecht M.; A Temporal Pattern Mining Approach for Classifying Electronic Health Record Data. ACM Trans Intell Syst Technol. 4(4) (2013).

[3] Allen JF.: Maintaining knowledge about temporal intervals. In: Communications of the ACM. 26 (1983).