

一般口演 | 医療データ解析

一般口演16

医療データ解析

2019年11月23日(土) 14:40 ～ 16:40 B会場 (国際会議場 2階コンベンションホールB)

[3-B-2-02] 時系列クラスタリングを用いたクリニカルパスの自動作成

○五十嵐 吉輝¹、井上 悠¹、井川 澄人¹、片山 洋子²、砂野 由紀²、羽藤 慎二²、河村 進²（1. 株式会社ソフトウェア・サービス, 2. 国立病院機構四国がんセンター クリニカルパス管理・推進委員会）

キーワード：Clinical Pathway, Data Mining, Clustering, Machine Learning

【はじめに】クリニカルパス（パス）システムの新規導入時等において、パスマスタ作成が業務逼迫の一因となっている。そこで、パス自動作成システムによって業務負担軽減を目指し、株式会社ソフトウェア・サービスと国立病院機構四国がんセンターの共同で研究開発に取り組んでいる。本研究においては、対象症例の実施タスクデータに対して時系列クラスタリングを行い、パスの雛形となり得る標準タスクを選抜した。

【方法】① DPCコード14桁中の傷病名(1-6桁)と手術(9・10桁)を指定し、A)症例、B)実施タスク（注射、検査、看護指示）を取得。

② Aの在院日数に対して、四分位範囲法により外れ値を除外。

③②に関連する Bに対して、One-Hotエンコーディングでベクトル化し、主成分分析により次元圧縮。

④③で得たベクトルに対して、動的時間伸縮法による距離に基づく時系列クラスタリングを実施。

⑤クラスタの中心近傍付近における上位症例で実施タスクの集計を取り、過半数を標準タスクとして選抜。

【結果】既にパスが作成・運用されている胃切除の症例に対して本手法を用いたところ、各症例に適用されているパス改訂前後の2つの版に対応する2つのクラスタに完全分離できた。選抜した標準タスクとパスのタスクとの一致度の指標である Jaccard係数は0.809であった。

パス未作成の甲状腺全摘においては、DPCの甲状腺切除に該当する症例に対して本手法を用いたところ、甲状腺全摘のクラスタが概ね分離でき、その標準タスクを選抜した。

同じくパス未作成の大腸ポリペクトミーでは、主に「病理組織標本作製」の1臓器と2臓器に対応する2つのクラスタに概ね分離できた。

【おわりに】時系列クラスタリングによって、時系列タスクに内在するクラスタ構造を見つけ出すことが可能であった。標準タスクの選抜方法は課題である。システム化する際には利用者が閾値を選択できるようにしたい。

時系列クラスタリングを用いたクリニカルパスの自動作成

五十嵐 吉輝^{*1}, 井上 悠^{*1}, 井川 澄人^{*1},
片山 洋子^{*2}, 砂野 由紀^{*2}, 羽藤 慎二^{*2}, 河村 進^{*2}
^{*1} 株式会社ソフトウェア・サービス,

^{*2} 国立病院機構四国がんセンター クリニカルパス管理・推進委員会

Automatic creation of clinical pathway using time series clustering

Yoshiki Igarashi^{*1}, Yu Inoue^{*1}, Sumito Igawa^{*1},
Yoko Katayama^{*2}, Yuki Sunano^{*2}, Shinji Hato^{*2}, Susumu Kawamura^{*2}
^{*1} Software Service, Inc,

^{*2} NHO Shikoku Cancer Center, Management and Promotion Committee of Clinical Pathway

We have been aiming to develop an automatic creation system of clinical pathway which is the standard clinical practice plan, research and development is being conducted jointly by Software Service Inc. and the National Hospital Organization Shikoku Cancer Center. We think to support by the system to solve a problem that has become a heavy workload to create manually clinical pathways when a new medical information system is introduced. In this paper, the standard tasks of the cases, which was narrowed down using the DPC code, as the template of the clinical pathways were selected using the time series clustering based on the distance by Dynamic Time Warping. We selected gastrectomy, thyroidectomy and colorectal polypectomy as target diseases and operations. The Jaccard similarity coefficient with the tasks of the already created clinical pathway was 0.802 for the gastrectomy. This method can be applied to any facility as long as DPC is introduced and the treated tasks is stored in the database.

Keywords: Clinical Pathway, Data Mining, Clustering, Machine Learning

1. はじめに

診療行為の標準診療計画であるクリニカルパス(以下パス)の自動作成システムの開発を目指し、株式会社ソフトウェア・サービスと国立病院機構四国がんセンターの共同で研究開発を行っている。背景として、医療情報システムの新規導入時等に、パスを手動で作成しなければならず、業務逼迫の一因となっており、パス自動作成システムによって業務支援を行うことが狙いである。

パスの自動作成における先行研究では、空のパスを適用し、パスシステムのバリエーション統計機能を用いてバリエーションタスクを抽出し、パスの雛形とする方法[1]や、診療データベースに格納された看護指示履歴をクラスタリングし、パスの雛形として用いる方法[2]が提案されている。

本稿においては、時系列の診療行為の実施データに対して、動的時間伸縮法(Dynamic Time Warping; DTW)[3]による時系列クラスタリングを行い、パスの雛形となり得る標準タスクの選抜を試みた。標準タスクの選抜対象の病名・術式には、胃切除、甲状腺切除、大腸ポリペクトミーの3種を選出した。

2. 方法

次の5つの手順により、パスの雛形となり得る標準タスクの選抜を試みる。

2.1 DPCコードの指定とデータ取得

作成対象のパスの傷病名と術式の指定には、DPCコードを用いる。DPCコード14桁における上位6桁の傷病名コードと9・10桁目の手術コード、及び取得期間を指定し、それらの条件に該当する症例と実施タスクである注射・検査・看護指示データを取得する。

2.2 外れ値の除外

取得した症例の在院期間における入院日から術日までと術日から退院日までの日数に対して、四分位範囲(IQR)法を用いて外れ値を判定し、標準タスクの選抜対象症例から除外する。

2.3 エンコードと次元圧縮

実施タスクである注射・検査・看護指示を One-Hot エンコーディングによりベクトル化する。1症例の入院期間がM日で、実施タスクがN項目存在すると、1症例をM×N行列で表現する。

エンコードした実施タスクは高次元になるため、クラスタリングは困難である。そこで、エンコードした全対象症例のタスクに対して、主成分分析(Principal Component Analysis: PCA)を行い、累積寄与率が8割を超える主成分ベクトルまでを用いて、実施タスクを表現する。

2.4 クラスタリング

各症例の時系列データ間の距離をDTWによって定義し、クラスタリングを行う。

DTWは、位相や周期の異なる時系列データ間において類似度を算出するアルゴリズムである。この手法を用いることで、症例間の時系列上で実施タスクのずれが生じていても、症例間の類似度を計算できるようにする狙いがある。

クラスタリングのアルゴリズムには、k-meansを用いる。

DTWとk-meansによるクラスタリングには、Pythonの時系列解析ライブラリであるtslearn[4]を使用した。

2.5 標準タスクの選抜

クラスタリング結果からパスの雛形となり得る標準タスクを選抜する。クラスタリング結果のうち、クラスタの中心最近傍の症例がクラスタの代表症例となるが、その 1 症例のタスクを標準タスクとみなした場合、そこに患者固有のタスクが含まれている可能性がある。

そこで、クラスタの中心近傍付近にある上位症例を取得し、それらの症例のタスクの集計を取り、過半数を超えるタスクを標準タスクとして採用する。

3. 胃切除に対する実験結果

時系列クラスタリングの性能評価を行うために、既にパスが作成され運用されている傷病名・手術である胃切除に対して本手法を適用した。

国立病院機構四国がんセンターにおける胃切除パスの、連続した 2 つの版である Ver.23(以下版 A)と Ver.24(以下版 B)の症例を対象に、実施タスクに対して時系列クラスタリングを行い、パスの雛形となり得る標準タスクの選抜実験を行った。

3.1 DPC コードとデータの取得

表 1 の条件に従い、胃切除の症例を取得した。症例数は 141 症例であった。

表 1 胃切除の症例の取得条件

DPC の傷病名コード	060020(胃の悪性腫瘍)
DPC の手術コード	01(胃全摘術)、02(胃切除術)
対象期間	2017 年 3 月 1 日～2018 年 2 月 28 日

対象期間においては、胃切除パスの版 A と版 B が運用されている。対象症例のパス適用率は 100%であり、それぞれの症例に対して版 A または版 B のいずれかが適用されている。

表 2 に全 141 症例及び版 A と版 B の統計値を示す。そして図 1 に全症例の在院日数、図 2 に版 A と版 B ごとの在院日数のグラフをそれぞれ示す。

表 2 症例の統計値

	症例数 (件)	最頻値 (日)	平均 (日)	標準 偏差
版 A	46	18	25.8	18.8
版 B	95	18	22.4	8.9
全症例	141	18	23.5	13.0

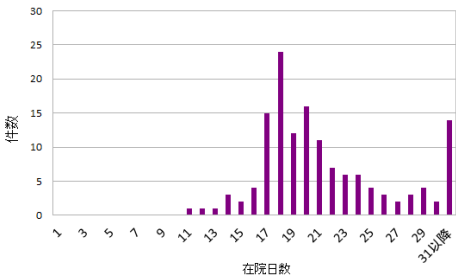


図 1 在院日数のヒストグラム

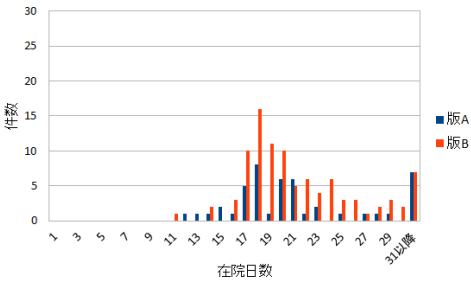


図 2 在院日数ヒストグラム(版 A と版 B)

3.2 外れ値の除外

全症例の在院日数を術前と術後に分け、IQR 法によって外れ値を除外した。

図 3 は術前・術後の箱ひげ図であり、○の症例が外れ値である。

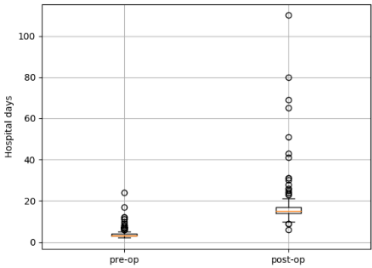


図 3 在院日数の箱ひげ図

外れ値除外後の統計値を表 3 に示す。表 2 と表 3 を比較すると、平均在院日数と標準偏差の減少が見られる。そして外れ値除外後の全症例の在院日数のグラフが図 4 であり、版 A と版 B の在院日数のグラフが図 5 である。IQR 法により、極端に長い或いは短い症例を除外できた。

表 3 外れ値除外後の統計値

	件数	最頻値 (日)	平均 (日)	標準 偏差
版 A	35	18	18.7	2.9
版 B	66	18	19.2	2.6
全症例	101	18	19.1	2.7

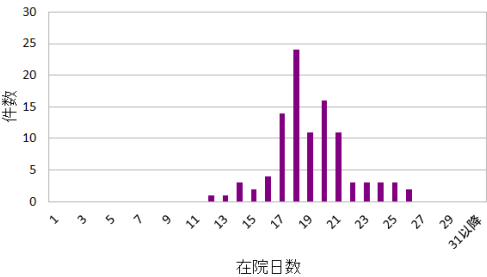


図 4 外れ値除外後の在院日数ヒストグラム

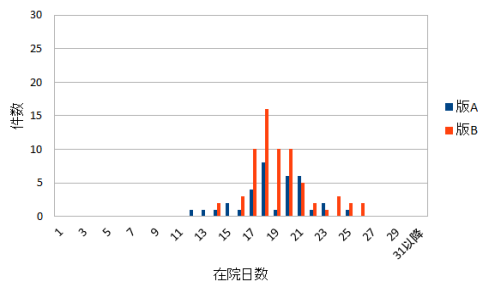


図 5 外れ値除外後の在院日数ヒストグラム(版 A と版 B)

3.3 エンコードと次元圧縮

外れ値除外後の症例の実施タスクである注射・検査・看護指示の 470 項目を、One-Hot エンコーディングにより 470 次元にベクトル化し、PCA を行った。

PCA 実行結果の累積寄与率の図 6 より、第 14 主成分ベクトルで累積寄与率が 8 割を超えていたため、第 14 主成分ベクトルまでを用いた。

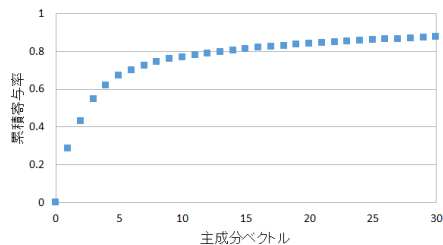


図 6 累積寄与率

3.4 クラスタリング

第 14 主成分ベクトルまでに射影した実施タスクに対して、DTW 距離に基づく k-means クラスタリングを行った。

k-means におけるクラスタ数は 2 を指定した。本実験においては、版 A と版 B の 2 種類の症例が含まれていることが事前に判明していたため、クラスタ数は 2 を用いた。

クラスタリングを行った結果、症例の 2 クラスタにおいて版 A と版 B は完全分離できた。

各症例の術日のタスクを第 8-9-10 主成分ベクトルへ射影した結果が図 7 である。版 A と版 B のクラスタが目視でも確認できた。

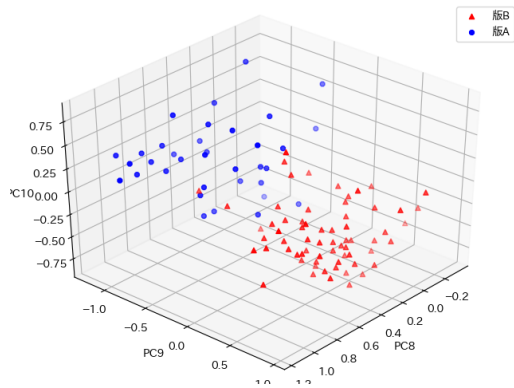


図 7 胃切除の術日のタスクを第 8-9-10 主成分ベクトルに射影した結果

3.5 標準タスクの選抜

クラスタリングによってクラスタ構造が得られたと仮定すると、クラスタの中心最近傍の症例のタスクが標準的なタスクと解釈できる。しかし、中心最近傍の 1 症例のみでは、患者の背景固有のタスクが含まれている可能性が高い。

そこで、本実験においては、クラスタの中心近傍の上位 2 割を取得し、日ごとにタスクの集計を取り、過半数を超えるタスクを選抜した。

結果の確認のために、版 B クラスタの「中心最近傍 1 症例のタスク」と「中心近傍上位 2 割 13 症例の過半数のタスク」の 2 つに対して、胃切除パス版 B のタスクとの一致度の指標である Jaccard 係数を算出した。表 4 はその結果である。

	Jaccard 係数	
	版 A	版 B
中心最近傍	0.633	0.707
中心近傍上位 2 割	0.730	0.809

3.6 考察

クラスタリングによって版 A と版 B が完全分離できたことから、本手法によって症例の時系列タスクにおいてクラスタ構造を見つけ出すことが可能であると言える。

標準タスクの選抜については、版 A・B 共に中心最近傍のタスクのみを用いるよりも、中心近傍上位 2 割のタスクを用いた方が、Jaccard 係数が高かった。このことから、標準タスクの選抜にはクラスタの中心近傍付近の複数症例のタスクの集計を取ることが有効であると考えられる。

また本稿においては、中心近傍上位の割合は固定で 2 割とし、標準タスクの選抜割合も固定で過半数と決めたが、それらの割合の最適な閾値の決定方法は今後の課題である。

4. 追加実験結果

パス未作成の疾患・術式である甲状腺切除と大腸ポリペクトミーを対象に、本手法を適用した。

甲状腺切除に関しては、甲状腺葉峡部切除のパスは既に作成されているが、甲状腺全摘のパスは未作成であるため、本手法により甲状腺全摘の標準タスクの選抜を試みた。

4.1 甲状腺切除

表 5 に従い、甲状腺切除の症例を取得した。症例数は 83 症例であった。

表 5 甲状腺の症例の取得条件	
DPC の傷病名コード	100020 (甲状腺の悪性腫瘍)
DPC の手術コード	01 (甲状腺悪性腫瘍手術(全摘及び亜全摘))
対象期間	2017 年 1 月 1 日～ 2018 年 12 月 31 日

本実験においては、得られた症例から、甲状腺全摘の標準タスクを選抜したい。83 症例中、甲状腺全摘は 27 症例あり、残り 56 症例は全摘以外である。

この得られた症例に対し、手法 3.1～3.5 を適用したところ、全摘と全摘以外のクラスタが得られた。

図 8 は術日のタスクを第 1-2-3 主成分に射影し、クラスタ毎にラベルを割り振った結果である。

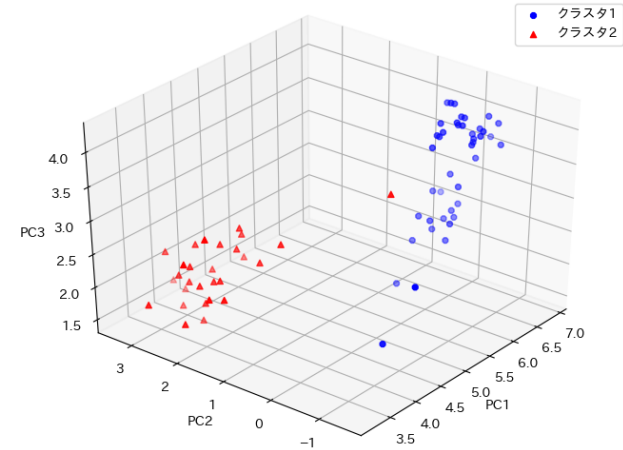


図 8 甲状腺の術日のタスクを 1-2-3 主成分に射影した結果

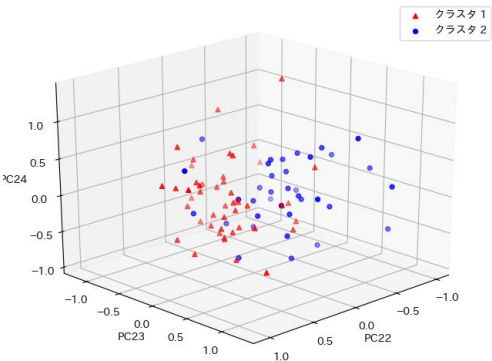
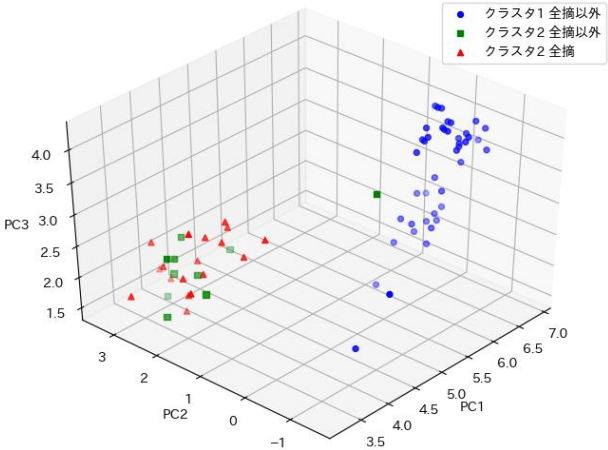


図 10 甲状腺の術日のタスクを 第 22-23-24 主成分に射影した結果

図 9 は、クラスタにおける全摘と全摘以外の症例にラベルを割り振った結果である。



4.2 大腸ポリペクトミー

表 6 に従い、大腸ポリペクトミーの症例を取得した。症例数は 98 症例であった。

表 6 大腸ポリペクトミーの症例の取得条件

DPC の傷病名コード	060100 (小腸大腸の良性疾患)
DPC の手術コード	01 (内視鏡的大腸ポリープ・粘膜切除術)
対象期間	2017 年 1 月 1 日～ 2018 年 12 月 31 日

図 10 は術日のタスクを第 22-23-24 主成分に射影し、クラスタ毎にラベルを割り振った結果である。

各クラスタ内のタスクを確認したところ、「病理組織標本作製」タスクの臓器数でクラスタ間に違いが見られた。表 9 は、各クラスタにおける臓器数の内訳である。

表 7 大腸ポリペクトミー 各クラスタの病理組織標本作製タスクの種類の割合

	病理組織標本作製		
	1 臓器	2 臓器	3 臓器
クラスタ 1	0.729	0.042	0.229
クラスタ 2	0.250	0.667	0.083

その他に目立ったタスクの差異では、表 10 における「日常生活ケア」の 2 項目が挙げられる。「移動ケア/移乗介助(車椅子)/部分介助」と「整容・更衣ケア/更衣/部分介助」のタスクは、クラスタ内における割合がクラスタ 1 よりもクラスタ 2 で多く見られた。

表 8 各クラスタの二種のタスクの割合

	移動ケア/ 移乗介助(車椅子)/ 部分介助	整容・更衣ケア/ 更衣/ 部分介助
クラスタ 1	0.542	0.104
クラスタ 2	0.889	0.639

4.3 追加実験考察

甲状腺切除に関して、全摘の症例のタスクだけで完全なクラスタが構築されなかった原因は、全摘と全摘以外でタスクが似通っていたためと思われる。
大腸ポリペクトミーに関しては、「病理組織標本作製」や「移動ケア/移乗介助(車椅子)/部分介助」「整容・更衣ケア/更衣/部分介助」のタスク項目でクラスタが分かっていたことから、主にケアの必要度の高いクラスタで分離されたと思われる。

5. 結論

本研究では、対象症例の実施タスクの時系列データに対して時系列クラスタリングを行い、パスの雛形となり得る標準タスクの選抜を行った。
胃切除に対しては、時系列クラスタリングにより、胃切除パスの版 A と版 B でタスクのクラスタが分離できた。甲状腺切除に対しては、主に全摘出とそれ以外で、クラスタが分離できた。大腸ポリペクトミーに対しては、主に「病理組織標本作製」タス

クの1臓器、2臓器でクラスタが分離できた。

このことから、本手法の時系列クラスタリングによって、時系列タスクが内在するクラスタ構造を見つけ出すことが可能であると言える。

そしてクラスタの中心近傍における上位2割の症例のタスクに対して集計を取り、過半数のタスクを標準タスクとして選抜した。この標準タスクの妥当性については、専門医療従事者による確認が必要である。

本手法の利点は、施設に依存しないことが挙げられる。DPCが導入されていて、データベースに実施タスクが格納されていれば、施設を問わず適用可能である。

k-meansにおけるクラスタ数の指定と、得られたクラスタの解釈は今後の課題である。

また、標準タスクの選抜方法も課題であり、今後システム化する際は、一律に過半数で区切るのではなく、ユーザがパスのオーバービュー画面を見ながら閾値の数値を指定できる仕様を検討したい。

6. 倫理的配慮・謝辞

本研究は国立病院機構四国がんセンターの倫理審査委員会による承認を得て実施した。(承認番号:臨2019-38)

参 考 文 献

- [1] 井川澄人, 松本崇志, フレキシブルパス標準統計機能を活用した新規電子パス作成の検討, 日本クリニカルパス学会雑誌 Vol.18 No4, 445, 2016.
- [2] 岩田春子, 津本周作, 病院情報システムに蓄積された看護行為に関する時系列マイニング—看護パス作成支援に向けて(その4)—, 第35回医療情報学会連合大会(第16回日本医療情報学会学術大会), 646-649, 2015.
- [3] Sakoe, H. and Chiba, S.:Dynamic Programing Algorithm Optimization for Spoken World Recognition, IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, No.1, pp. 43-49 (1978)
- [4] tslearn's documentation,
[<https://tslearn.readthedocs.io/>]
(2019年6月4日アクセス)