

重み付き PLSA を用いた敗血症患者の DPC データ分析結果における 各クラスタの特徴およびクラスタ遷移パターンの検討

Cluster Features and Transition Patterns in the Analysis Results of Sepsis-related DPC Data with Weighted PLSA Methods

村田 知佐恵^{*1}
Chisae Murata

山下 和也^{*1}
Kazuya Yamashita

阪本 雄一郎^{*2}
Yuichiro Sakamoto

櫻井 瑛一^{*1*2}
Eiichi Sakurai

本村 陽一^{*1*2}
Yoichi Motomura

^{*1} 産業技術総合研究所人工知能研究センター
Artificial Intelligence Research Center, AIST

^{*2} 佐賀大学
Saga University

This study aimed to investigate factors related to mortality in patients with sepsis by analyzing all items of their diagnosis procedure combination (DPC) data with weighted probabilistic latent semantic analysis (PLSA) methods and to clarify features and transition patterns of the clusters extracted from the analysis. The results showed that 12 clusters were optimal and among those the one with high severity and the one receiving palliative care tended to have a high mortality rate. In addition, focusing on the main cluster for each patient, five cluster transition patterns were revealed.

1. はじめに

現在、著者らの研究チームは、医療ビッグデータの利活用の一環として、日常的に収集蓄積が進む Diagnosis Procedure Combination (診断群包括分類、以下 DPC) データに着目し、治療効果の検証のみでなく、治療戦略や医師の意思決定支援への応用の可能性を検討している。

これまで、診療科を問わず発症の可能性や対応の必要性がある敗血症に焦点を当て、DPC データに基づき、Probabilistic Latent Semantic Analysis (確率的潜在意味解析、以下 PLSA) 手法 [Hofmann 1999] を用いた患者と診療行為の同時クラスタリング、さらにそこで得られたクラスタを状態空間とみなして同一患者が入院中にたどるクラスタ間の時間遷移パターンを解析し、治療戦略決定支援アルゴリズムの実現可能性を検証してきた [山下 2017, 山下 2018a, 山下 2018b]。その結果、DPC データを用いた診療行為のクラスタリングが可能であることのほか、患者が入院中の治療経過に伴いクラスタを遷移する確率から、状態遷移パターンに違いがあり、ある状態で特定の治療を行った場合の在院日数や死亡率、その後どのような経過をたどる可能性があるかなどについて、予測可能なことが明らかになった。さらに、注目している変数に重みを与える「重み付き PLSA (weighted PLSA)」を初めて適用したクラスタリングも実施し、血液浄化の種類によりクラスタ分類されることを明らかにした [山下 2019]。ただし、これらは DPC データのうち「診療行為」に特化してクラスタリングしたものであり、年齢や金額など DPC データのすべての項目を同時に変数とした解析は行っていなかった。

そこで本研究では、DPC データ全体から死亡率と関連する要因を探索的に見つけることを目的とした解析を行い、抽出されたクラスタの特徴やクラスタ遷移パターンについて検討する。

2. 方法

研究施設において平成 24 年度～平成 26 年度の期間に入院していた敗血症患者 458 名の DPC データを用いて、PLSA によるクラスタリングを実施した。

PLSA は、2 つの変数 x, y の背後に共通の特性となる潜在変数 z があると仮定し、それらの関連性について、期待尤度最大化法によって x および y が潜在変数 z に所属する確率値を推定し、所属確率が最大となる z にクラスタリングする手法である。なお、本研究では、PLSA 解析において、共起行列の特定の行の頻度に重みを付ける「重み付き PLSA」を適用した。

分類された各クラスタについては、変数 y 側からクラスタの特徴、および、変数 x 側からクラスタ遷移パターンをそれぞれ分析し、考察において臨床の観点から統合した。

本研究は、研究施設の倫理審査委員会にて承認を取得した。また、PLSA による解析は、産業技術総合研究所の知財ソフトウェア「PLASMA」を使用した。

2.1 使用した変数

本研究では、変数 x を患者 ID と診療行為実施日 (入院初日から 1 日単位、在院日数と同じ) を組み合わせた「患者 ID_実施日」とし、その他の DPC データ項目を変数 y とした (表 1)。

さらに、死亡率と関連する要因を探索するため、「死亡の有無」を 20 倍、その他の変数項目はすべて 1 倍で重みを付けた。

表 1 変数 y に使用した DPC データの項目

年齢	出来高金額合計 (円)
性別	薬: 出来高 (円)
在院日数	材料: 出来高 (円)
ICU 日数	10 診療料合計 (円)
診療行為名称	20 投薬料合計 (円)
MDC6 名称	30 注射料合計 (円)
診療科	40 処置合計 (円)
医師コード	50 手術合計 (円)
死亡の有無	60 検査合計 (円)
	70 画像合計 (円)
	90 入院料合計 (円)
	97 食事合計 (円)

2.2 クラスタ遷移の加工

変数 x 側からクラスタ遷移を確認するため、各患者の実施日ごとに、所属確率が 15%以上のクラスタに焦点を当て、特に 60%以上を「強」として区別し、それらのクラスタの組み合わせの特徴が、目視でも明瞭なほど大きく変化する日を遷移日とした(図 1)。

実施日	遷移日	クラスタ	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
1	1	C6,C7(強),C9	12%	0%	0%	0%	0%	0%	63%	0%	25%	0%	0%	0%
2			0%	0%	0%	0%	0%	31%	38%	0%	31%	0%	0%	0%
3			4%	0%	0%	0%	0%	26%	47%	0%	23%	0%	0%	0%
4			0%	0%	0%	0%	0%	0%	62%	0%	38%	0%	0%	0%
5			0%	0%	0%	0%	0%	0%	62%	0%	38%	0%	0%	0%
6			0%	0%	2%	0%	0%	20%	38%	0%	41%	0%	0%	0%
7			2%	0%	0%	0%	0%	0%	59%	0%	38%	0%	0%	0%
8			0%	0%	0%	0%	0%	23%	41%	0%	36%	0%	0%	0%
9			0%	0%	0%	0%	0%	0%	55%	0%	45%	0%	0%	0%
10	10	C1,C6,C7,C9	18%	0%	0%	0%	2%	18%	26%	0%	36%	0%	0%	0%
11			26%	0%	0%	0%	2%	0%	38%	0%	36%	0%	0%	0%
12			28%	0%	0%	0%	0%	0%	36%	0%	36%	0%	0%	0%
13			21%	0%	0%	0%	1%	18%	27%	0%	33%	0%	0%	0%
14			38%	0%	0%	0%	0%	0%	27%	0%	35%	0%	0%	0%

図 1 クラスタ遷移加工のプロセス

3. 結果

変数 x 「患者 ID_実施日」の総数は 19658、変数 y のうち「診療行為」の総数は 4406 となった。

重み付き PLSA によるクラスタリングの結果、AIC 最小は 12 となった。12 の各クラスタは、分析結果として出力された順に、C1, C2, C3, ..., C12 と付番した。変数 y 側から分析したクラスタの特徴、および、変数 x 側から分析したクラスタ遷移パターンを以下に示す。

3.1 各クラスタの特徴

重み付き PLSA により得られた 12 のクラスタのうち、C1 および C7 は死亡率が高く、C8、C10、および C11 は死亡率が低かった。また、所属していた変数 y の内容を分析すると、重症度、手術実施や集中治療室(Intensive Care Unit、以下 ICU)への入院の有無、年齢期などでの特徴が得られた。ただし、C6 については、目立った特徴は見当たらなかった。詳細を表 2 に示す。

表 2 AIC 最小となった 12 クラスタの特徴

クラスタ名	死亡率	特徴
C1	高	・重症度が高い ・在院日数が長い
C2		・緊急入院が多い ・在院日数が短く、入院費が低い
C3		・手術が多い ・人工心臓/人工血管/人工弁/ステント関係が多い
C4		・リハビリ加算が多い
C5		・緊急入院が多い ・血液浄化や緊急手術が多い
C6		(目立った特徴は見当たらなかった)
C7	高	・緩和ケアを受けている患者が多い
C8	低	・31~45歳、61~75歳が多い ・薬剤や検査費が高い
C9		・褥瘡管理加算が高い
C10	低	・46~60歳が多い ・ICUに入院している患者が多く、入院期間は短い
C11	低	・76歳以上が多い ・ICUに入院している患者が多く、入院費が高い
C12		・小児患者が多い

3.2 クラスタ遷移パターン

患者ごとのクラスタ遷移パターンを確認すると、実施日ごとの所属クラスタにおいて、クラスタの強さの観点から 5 つの遷移パターンが見られた。また、患者の特徴に応じた遷移パターンも見られた。

(1) クラスタの強さに関する 5 つの遷移パターン

各患者の実施日ごとのクラスタリング結果において、所属が 60%以上と示されたクラスタを「強」と設定し、「主クラスタ」と定義すると、以下のような 5 つの遷移パターンが見られた。

- ・最初から最後まで 1 つのクラスタが継続して強い。
(主クラスタが変化しない)
- ・最初から最後まで複数のクラスタが現れ、強いクラスタが見られない。
(主クラスタが存在しない)
- ・最初から最後まで 1 つのクラスタが間欠的に強く現れる。
(主クラスタがときどき現れる)
- ・初期に現れた複数のクラスタのうち、1 つのクラスタが途中から強くなる(表 3)。
(主クラスタが途中から現れる)
- ・初期に強いクラスタがあるが、途中で複数クラスタに分散する(表 4)。
(主クラスタが途中で消える)

なお、初期に強いクラスタが現れ、途中で別のクラスタが強くなるパターンは 3 例のみであった。

表 3 主クラスタが途中から現れる遷移例

ID	遷移日	クラスタ	遷移日	クラスタ	遷移日	クラスタ	遷移日	クラスタ
***50	1	3,6,9,11	6	5,9,11	21	9,11(強)	82	6,11(強)
***03	1	2,3,5,6	2	2,5,6	7	2(強),10		
***98	1	5,9,11	15	9,10,11	21	4,8,9,10	28	4,8(強)

表 4 主クラスタが途中で消える遷移例

ID	遷移日	クラスタ	遷移日	クラスタ	遷移日	クラスタ	遷移日	クラスタ
***28	1	6,10(強),11	12	4,6,10,11				
***93	1	2(強),6,10	10	2,6,10	19	2,6,8,10		
***94	1	6,10(強)	8	6,10,11	17	5,6,10,11	29	6,10,11

(2) 患者の特徴に応じた遷移パターン

死亡の有無や年齢期など、患者の特徴に注目すると、以下の 7 つの遷移パターンが見られた。

- ・死亡に至った患者の場合、最終クラスタに C7 または C1 が含まれていた。
- ・死亡の有無に関わらず、初期クラスタに C2 が含まれる場合が多かった。
- ・死亡の有無に関わらず、C3 が強い実施日の場合、翌日に C5 が現れる場合が多かった(図 2)。
- ・死亡の有無に関わらず、C6 が間欠的に現れる患者が多かった。
- ・0~15 歳の小児患者は、死亡 1 例を除き、すべて最終クラスタが C12 であった。
- ・回復した 16 歳以上の患者は、最終クラスタに C8、C10、C11、C12 のいずれかを含んでいた。特に C10 および C11 の場合が多かった。

- 初期にC2が強く、その他はC6のみの場合、全員回復し、入院期間が短く(最大で32日間)、クラスタの変化もほとんど見られなかった。

ID	実施日	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
***13	7	47%	0%	0%	4%	0%	0%	49%	0%	0%	0%	0%	0%
***13	8	20%	0%	64%	0%	3%	13%	0%	0%	0%	0%	0%	0%
***13	9	29%	0%	0%	3%	39%	21%	7%	0%	0%	0%	0%	0%
***13	10	29%	0%	0%	3%	39%	20%	9%	0%	0%	0%	0%	0%
***13	11	26%	0%	0%	0%	42%	19%	14%	0%	0%	0%	0%	0%
***13	12	27%	0%	0%	0%	43%	19%	11%	0%	0%	0%	0%	0%
***03	1	0%	7%	63%	0%	0%	16%	0%	3%	0%	11%	0%	0%
***03	2	0%	6%	2%	6%	36%	17%	0%	0%	0%	22%	10%	0%
***03	3	0%	8%	2%	8%	34%	8%	0%	0%	0%	24%	16%	0%
***03	4	0%	17%	0%	9%	27%	0%	0%	6%	0%	35%	7%	0%
***03	5	0%	10%	0%	15%	24%	20%	0%	0%	0%	24%	8%	0%
***03	6	0%	12%	3%	23%	30%	0%	0%	0%	0%	21%	11%	0%

図2 C3からC5の遷移例

4. 考察

4.1 DPC データに基づくPLSA 分析の有用性

分析結果から、クラスタの特徴と遷移パターンを統合すると、臨床における実際の状況が推察でき、各クラスタの臨床的な意味付けが可能であることが明らかになった。

例えば、C7 は、死亡率が高く、緩和ケアを受けている患者が多いクラスタであり、また遷移パターンにおいても最終クラスタとなることが多いという結果から、患者がターミナル(終末期)の状況にあると考えられる。同様に、死亡率が高く、重症度が高い状態の患者が多い C1も、遷移パターンで最終クラスタとなっており、最大限の治療を行い、他に治療の選択肢がない状況と考えられる。

敗血症以外の(入院する原因となった)疾患も含め、重症度が高く入院が長引いている状況や、緩和ケアを受ける状態にある患者において死亡率が高い傾向が見られることは、実際の臨床現場で想定される状況でもあるが、DPC データからの客観的な分析により、そのような状況が確率論として導けたことの意味は大きい。臨床における現象をデータに反映できることの証明であり、本研究により、DPC データを活用した治療選択ナビゲーションシステムの実現可能性が高まったと考える。

一方、上記と比較して、C10 や C11 など ICU に入院したケースでは死亡率が低い傾向を示したことも興味深い。集中治療を受けることができる、すなわち明確な治療方法がある状況下では、敗血症に罹患した場合でも、救命率が高いということが言えるかもしれない。この結果から、DPC データを活用した PLSA による分析は、システム構築のためのみでなく、臨床で起こる現象の検証方法として、新たな視点が提供できる可能性が示唆された。

さらに、クラスタ遷移パターンにおいて、一人の患者の入院期間中には主クラスタが一つのみ存在する、もしくは存在しないという今回の分析結果から、治療選択ナビゲーションシステムの開発において、主クラスタが注目すべきポイントの一つとなり得ることがわかった。主クラスタが変わることはほとんどないため、患者単位でのカテゴリー分けなどに主クラスタという考え方が役立つと思われる。また、経過とともに遷移しやすいのは、主クラスタ以外で所属するクラスタであることがわかったため、今後、その観点からより詳細な分析を行いたい。

4.2 臨床判断につながるクラスタ分類と遷移パターン

本研究における各クラスタの特徴やその遷移パターンから、臨床判断につながるいくつかの視点が得られた。

C6 は、主クラスタにはならない、付属的なクラスタであると考えられる。数日おきに現れる傾向があることから、定期的に行われる薬剤投与や処置などが含まれるクラスタではないかと推測される。また、入院期間が短い患者が多い C2 が強く、その他は C6 のみが所属している組み合わせが一定数見られたことから、薬剤投与などで順調に回復したケースが想定される。したがって、具体的な薬剤や処置については、今後さらなる分析・検討が必要であり、重要な視点となる可能性があることが示唆された。

C12 は、小児患者が多いことから、小児特有の疾患が多いことや、薬剤の投与量が成人とは異なるなどの治療の違いが反映された結果のクラスタと考えられる。小児領域は、他領域と区別して検討する必要があることがデータから示されたと考えられる。

C3 は手術を受けた患者が多く、C5 は緊急入院した患者や、緊急手術、血液浄化が行われた患者が多いクラスタとして分類されていた。したがって、C3 が強く現れた翌日に C5 に遷移する傾向については、翌日に再手術や術後合併症への対応が行われた可能性もデータからは考えられた。今後、実際の診療記録との突合せを行い、クラスタ遷移に関連する現象を検証したい。

4.3 今後の課題

DPC データを治療戦略支援に活かす取り組みは、本研究も含め、これまでの研究によりある程度の実現可能性は示されている。しかし、臨床応用に向けては、複合疾患の状況や各疾患の治療期間など、DPC データのみでは得られない診療記録からの詳細情報も追加し、より大規模なデータベースを構築することや、多施設のデータによる分析を行い、結果の信頼性を高めていく必要があると考える。

また、今回は所属確率 60%以上のクラスタを主クラスタと定義したが、妥当性を確保するためには、他の変数項目での実施や条件を変えた上での解析を重ねながら検証する必要がある。

5. 結論

DPC データの全項目を用いて、「死亡の有無」のみに 20 倍の重みを付けた PLSA によるクラスタリングを行った結果、敗血症患者のうち、重症度が高く入院が長引いている患者や緩和ケアを受けている患者の場合は、死亡率が高いことが明らかになった。また、ICU に入院したケースでは死亡率が低い傾向が示された。

各患者の実施日ごとの所属クラスタのうち、所属確率が高い主クラスタに注目すると、5 つの遷移パターンが得られた。また、各クラスタの特徴に注目すると、小児領域は他領域とは区別して検討するほうがよいこと、薬剤や処置について詳細な分析が必要なこと、手術とその後の処置の関係性を診療記録から把握する必要があることなどの示唆を得た。

今後は、医師の意思決定を支援する治療選択ナビゲーションシステムの開発に向けた分析・予測の高度化と、現場ニーズに即した機能開発、試作、実証へと進めていく。

謝辞

本研究にご協力くださった関係者の皆様に心より感謝申し上げます。本研究は、NEDO 委託事業「人間と相互理解できる次世代人工知能技術の研究開発」の支援を受けて行いました。

参考文献

- [Hofmann 1999] Hofmann T: Probabilistic Latent Semantic Analysis, Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., 1999.
- [山下 2017] 山下和也, 阪本雄一郎, 櫻井瑛一, 本村陽一: 時間変化も考慮した pLSA 手法による敗血症治療戦略への応用, 第 27 回人工知能学会 社会における AI 研究会, 2017.
- [山下 2018a] 山下和也, 阪本雄一郎, 櫻井瑛一, 本村陽一: 医師の知見と経験を蓄積・伝承する判断支援システム実現に向けた pLSA による診療クラスタの時間遷移パターンの解析, 人工知能学会全国大会, 2018.
- [山下 2018b] 山下和也, 阪本雄一郎, 櫻井瑛一, 本村陽一: 診療クラスタ遷移と医学的判断の紐づけによる敗血症治療ナビゲーションシステム実現可能性の検証, 第 5 回「JAMI 医用知能情報学研究会-JSAI 医用人工知能研究会」合同研究会, 2015.
- [山下 2019] 山下和也, 村田知佐恵, 阪本雄一郎, 櫻井瑛一, 本村陽一: 敗血症患者への血液浄化を含む治療戦略決定支援に向けた重み付き PLSA による診療行為と患者の DPC データに基づく同時クラスタリング解析, 日本急性血液浄化学会雑誌, 2019, in print.