

# 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 Treatment strategy of Septicemia based on DPC data analysis by using pLSA method

山下 和也 \*<sup>1</sup> 阪本 雄一郎 \*<sup>2</sup> 櫻井 瑛一 \*<sup>1</sup>\*<sup>2</sup> 本村 陽一 \*<sup>1</sup>\*<sup>2</sup>  
Kazuya Yamashita Yuichiro Sakamoto Eiichi Sakurai Yoichi Motomura

\*<sup>1</sup> 産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology

\*<sup>2</sup> 佐賀大学  
Saga University

This paper shows our study about treatment strategy of septicemia based on DPC data by using pLSA, probabilistic latent semantic analysis, which find hidden segments in data. Our aim of using this method is modeling patients' change along with time. In this study, we estimate segments from the data that consists of times when each patient had some medical treatment and name of treatments. We compared each segment from the points of death rate, medical cost and patients' hospitalization. We analyze patients' movement pattern between segments. We also pay attention to that these movements are related to medical decision.

## 1. はじめに

医療の質を向上させながら増え続ける医療費の圧縮が求められているなか、DPC (Diagnosis Procedure Combination=診断群包括分類) データの蓄積が進み、これら医療ビッグデータの重要性が益々認識されその利活用に大きな期待が寄せられている。DPC データを用いた集中治療関連の報告は散見されるが、治療法の効果などを検証した結果が多く、治療戦略への応用などの報告は認められないのが現状である。本研究では患者にとって重要な「1. 予後 (特に死亡率)」、医療制度を支える国にとって重要な「2. 医療費」、安定的に患者を受け入れる為に病院にとって大事になる「3. 在院日数」の3つの変数に注目し、他の変数2つが同じならもう1つの変数はより「生存率が高く」、「安価で」、「短い在院日数」となるような治療戦略を持つことを目的とした上で、多彩な治療薬が存在する敗血症に対してDPCデータで得られた治療薬をpLSA (Probabilistic Latent Semantic Analysis) 手法 [Hofmann, 1999] によりクラスタ分類するとともに同一患者におけるクラスタの時間遷移を解析する事で、DPC 入力データが敗血症治療ナビゲーション (医療判断支援システム) として治療戦略に有益である可能性を検証している。

pLSA を用いた解析の実社会への応用としては、[井手, 2016] が高齢者の大規模アンケートを用いた高齢者生活行動モデルを構築し、[井手, 2017], [Ide, 2017] では個人の代わりに空間 (地域) につい

ての潜在クラスを抽出する確率的潜在空間意味解析を用いて地域の自治体等の政策決定の提言に活用出来る事を示している。またスーパーの購買履歴であるID-POSデータを用いた顧客理解・来店人数予測 [山下, 2017a], [石垣, 2011]、需要予測 [本村, 2011] などマーケティング分野への例がある。更に顧客IDと月を組み合わせた変数により購買行動について時間軸上での変化のモデリングを行った研究 [原田, 2016] がある。同様の手法を医療分野に適応した初の例として [山下, 2017b] は敗血症患者と注射薬剤を同時に分類し、これまでの年齢や重症度による分類とは異なる、診療行為 (この場合は注射薬剤) によって情報量最大となる患者分類を行っている。[山下, 2017b] では「入院初日から3日まで」、「4~30日まで」、「31日目以上」と3期間における患者の時間変化も考慮した解析であり、[山下, 2018] では上記のように固定した期間に離散化する代わりに1日毎の時間変化を解析し、患者の変化の変わり目をデータから検出出来るようにしている (動的離散化)。また診療行為についても注射だけでなく、投薬、処置、手術等も含めた診療行為としている。[山下, 2018] では同一患者の時間変化に伴うクラスタ遷移パターンについていくつかの代表例が示されたが、本研究では全体像の解析を行った。

## 2. 時間変化も考慮した pLSA によるモデリング手法

連絡先: 山下和也, 産業技術総合研究所・人工知能研究センター, 〒135-0064 東京都江東区青海 2-4-7, yamashita-kazuya@aist.go.jp

pLSA は、2つの変数  $x, y$  の背後に共通の特性、潜在変数  $z \in Z = \{z_1, \dots, z_k\}$  があると仮定し、 $x, y$  が各  $z$  に関係する確率をデータから期待尤度最大化法によって推定し、その確率値が最大となる潜在変数  $z$  に  $x$  と  $y$  を分類する。本研究では、DPC データを用い上記の  $x, y$  のうち  $x$  を患者 ID と診療行為実施日（入院初日から 1 日毎）を組み合わせた変数“患者 ID\_実施日”とした。 $y$  を注射、投薬、処置、手術等の各診療行為名称（注射や投薬の場合は各薬剤名、処置の場合は各処置名等）とした。上記の変数  $x, y$  を入力として産業技術総合研究所の知財ソフトウェア「PLASMA」を用いて pLSA を行い、同じような診療行為を受けている患者及び実施日と、同時に行われやすい診療行為を同時に分類した。使用した患者 ID 数は 458 (患者 ID\_実施日の総数は 19658)、診療行為数は 4406 である。得られた潜在変数  $z$  を診療クラスタと呼ぶこととし、クラスタ数 9 が AIC 最小となった。このクラスタ毎に、各クラスタに所属する患者 ID について死亡率（死亡の有無の平均）を集計して死亡率の低い順に  $aC1, aC2, \dots, aC9$  とクラスタに名前を付けた。各クラスタの詳細は[山下, 2018]に示されている。本研究では前日から当日へのクラスタ遷移の事例数や遷移時期を、全体、生存・死亡の各ケース、在院日数の長短別、入院初期かそれ以降かによる別に分けて集計を行う事でクラスタ遷移パターンを調べた。

### 3. 同一患者の時間変化に伴うクラスタ遷移パターン結果

#### 3. 1. 前日から当日へのクラスタ遷移数

2章で示した方法で pLSA を行い死亡率の順に番号を付けた診療クラスタ  $aC1, aC2, \dots, aC9$  について、前日から当日へのクラスタ遷移の事例数を図 1 に示す。全体の傾向として遷移なし（同一クラスタへの遷移）が多い事が分かる。クラスタ  $aC3$ （以後単にクラスタ 3 と示す）に留まっている事例が最も多い事が分かる。また入院初日はクラスタ 6 で以後はクラスタ 6 への遷移は少なく、クラスタ 6 は入院最初期の特徴を持っている事が分かった。

図 2 に、結果的に(左)生存したケース、(右)死亡したケースそれぞれに分けたクラスタ遷移パターンを示す。生存した患者は特にクラスタ 3 に留まっている事が分かる。全体として死亡率が低いクラスタを遷移している。クラスタ 4 とクラスタ 2 の間の両方向の遷移が多い特徴が見てとれる。一方死亡ケースでは、死亡率の高いクラスタに留まっている特徴が見られる。クラスタ 8 にある日数が最も多い。ここまですべてをまとめると、全体として死亡率の高いクラスタ間の遷移は少なく、それに対して、死亡率の低い 4 や 2 等のクラスタ間では遷移が多いことが分かる。

図 3 では生存した患者について、在院日数が短かった患者（在院 28 日以下）と長かった患者（在院 29 日以上）に分けてそれぞれの遷移を示した。なお 28 日は在院日数の中央値である。長かった患者については、入院初期（入院初日から 28 日以前）とそれ以降（入院初日から 29 日以降）に分けて示している。

在院日数が短い患者はクラスタ 2 の日が多い。またクラスタ 2 と 4 の双方向の遷移も見られる。一方在院日数が長い患者の入院初期でも上記と同様の傾向が見られるがそれに加えてクラスタ 3 状態の日も多く、クラスタ 4 と 3 の遷移も見られる。またクラスタ 8 の日も比較的多い。入院 29 日目以降ではクラスタ 3 に留まる事がより多くなっている。以上まとめると生存した患者は入院初期にクラスタ 2 と 4 を遷移し、そのまま早期に退院する患者と、その後、クラスタ 3 に移行し滞留した後に退院する患者がいる。

図 4 では死亡した患者について、図 3 と同様の図を示した。早期に死亡する患者は死亡までクラスタ 8 の状態が多い。長期入院後死亡する場合、入院初期にはクラスタ 3 や 5 の状態が続いたり他のクラスタの合間にクラスタ 4 にいる状態を経過した後、クラスタ 7 か 9 に移行して亡くなる様子が見られる。

#### 3. 2. クラスタ遷移日

図 5 に各クラスタ遷移した日（入院初日からの日数）の平均を示した。生存・死亡別の図をみると、死亡では入院後期に死亡率の高いクラスタにあることが分かる。

図 6 は前日からクラスタ遷移したか否かを示した図である。遷移した割合を示して、全体として入院後期程、遷移はせず同じクラスタに留まる傾向が分かる。長期入院してかつ死亡したケースでは入院初期からクラスタ遷移が比較的少ない事が分かる。

### 4. まとめ

診療クラスタの時間遷移には予後や在院日数毎にそれぞれのパターンがある事が分かった。これらクラスタ遷移は医師の何らかの医学的判断と結びついていて、その判断により診療行為を変えた事を意味する。また同じクラスタが継続されているのも医学的判断の結果である。医学的判断には「その際の患者の状態における標準的な診療を行っている」や「集学的治療を行った」、「ある種の挑戦的な診療を行った」、または「with draw」や「with hold」などがあり、今後これら医学的判断とクラスタ遷移、また検査値を紐づけ、さらにそれらの結果として、目的変数としての予後、在院日数、医療費の症例データを蓄積する事で、医療現場の医師の判断も含めた現象を、ベイジアンネットワーク等により計算空間にモデル化する事が出来る。そのモデルは診療判断システムとして実際の集中治療室で活用する事が出来、さらにその活用結果は新たな事例の蓄積として医療現場にフィードバックが可能である。このように医療現場と計算空間でのループを回していく事で、経験を積んだ医師の知見の蓄積と伝承が可能となり、全体の医師の知見とモデルの両方が共に成長していく事が可能となる。本研究ではそのようなシステム実現に向けて診療クラスタ遷移の全体像を把握した。

t-1 \ t	1	2	3	4	5	6	7	8	9	total
1	598	30	76	237	26	8	15	10	15	1011
2	25	1271	87	812	77	86	49	30	36	2473
3	63	56	2904	698	94	27	40	19	105	4074
4	254	863	731	467	375	78	288	168	149	3373
5	24	28	91	313	1820	26	27	26	11	2366
6	16	103	19	182	11	78	21	181	14	625
7	12	41	70	244	45	18	1718	62	36	2246
8	8	52	19	214	38	32	83	1271	96	1813
9	16	49	105	150	18	12	47	61	781	1219
Enter	14	47	4	105	1	268		12	9	458
total	1028	2540	4198	3410	2505	629	2298	1840	1232	19658

図1. クラスタ遷移数 (全ケース)

縦軸が前日のクラスタ番号 (Enter は入院)

横軸が当日のクラスタ番号

t-1 \ t	1	2	3	4	5	6	7	8	9	average
1	72	13	58	62	32	13	35	17	45	64
2	15	13	23	16	14	10	19	12	16	15
3	65	25	60	47	63	47	56	57	59	58
4	60	15	46	29	38	23	48	33	63	36
5	35	23	64	40	52	28	50	48	41	50
6	5	7	34	14	28	23	34	15	14	16
7	34	17	52	51	43	37	89	57	56	64
8	10	8	98	30	41	42	44	26	50	29
9	41	13	59	67	44	39	51	38	72	65
Enter	1	1	1	1	1	1	1	1	1	1
average	63	14	57	35	48	15	63	28	64	44

t-1 \ t	1	2	3	4	5	6	7	8	9	average
1	72.8	14	61.3	64.2	42.4	12.7	35.3	5.67	29.8	66
2	17.7	13.3	23	16.5	15.2	10.2	19.4	13.1	16.6	14.9
3	68.8	24.4	64	48.9	66.9	45.8	59.6	111	43.6	60.5
4	62.4	15.2	47.7	29.7	39.3	22.6	44.6	23.7	28.4	34.4
5	47	24.8	68.6	42.2	55.6	24.1	51.2	31.3	18.8	53.3
6	5.6	7.51	38.6	14.5	30.2	25.2	31.4	12.4	13.7	15.7
7	34.2	17.5	52	47.9	42.8	24.7	59.6	56.5	29	55.9
8	10.3	7.71	65	18.8	27.4	47.3	44.8	19.3	26.1	22.6
9	33.8	12.8	43.2	30.4	27.7	29.5	39.3	29.5	33.7	32.8
Enter	1	1	1	1	1	1	1	1	1	1
average	65	13.9	59.6	33.4	51.3	14.7	55	21.6	31.8	42.14

図5. クラスタ遷移した日 (入院初日からの日数)の平均 (上: 全体、下左: 生存、下右: 死亡)

t-1 \ t	1	2	3	4	5	6	7	8	9	total
1	588	24	67	226	17	6	15	6	12	961
2	19	1148	75	726	65	69	47	24	30	2203
3	54	48	2515	622	86	24	36	13	52	3450
4	239	775	666	324	307	60	226	79	59	2734
5	15	24	81	247	1443	18	26	12	5	1871
6	15	82	16	143	6	70	19	129	11	481
7	12	34	64	199	42	16	1278	39	26	1711
8	8	41	17	127	21	16	57	578	43	908
9	14	41	54	61	11	8	30	21	282	522
Enter	13	38	1	75	1	208		7	5	948
total	977	2255	3556	2750	1999	493	1735	908	524	15197

図2. クラスタ遷移数 (左: 生存、右: 死亡)

t-1 \ t	1	2	3	4	5	6	7	8	9	total
1	47	5	22	2	1	4				81
2	4	309	21	232	35	14	18	8	1	942
3	11	70	27	10	4	3	1	1	1	127
4	21	271	45	48	72	10	24	31	6	528
5	5	3	49	222	1	1	1	1	1	282
6	6	45	1	56	2	15	3	41		169
7	2	14	4	21	5	2	90	3	1	132
8	6	19	1	48	4	1	9	122	3	213
9	3	2	6	1					20	32
Enter	3	11	30	121		3				168
total	88	883	147	538	351	170	138	213	33	2574

図3. 生存したケースのクラスタ遷移数 (左: 在院短期、中: 在院長期の入院初期、右: 在院長期の入院 29 日目以降)

t-1 \ t	1	2	3	4	5	6	7	8	9	total
1	3	1	1	3	1					9
2	47	1	29	2	8	1	3	2		93
3	2	5	6	4	1		1	1	1	20
4	4	29	7	45	8	4	4	22	8	131
5	1	4	7	14					1	27
6	1	9	1	21	3		24	1		60
7	5	2			14	2	3			26
8	8	1	13	1	2	8	105	7		223
9	2	1	8		1	5	5	27	47	82
Enter	3	14	42		2	1				62
total	10	105	21	146	30	61	32	244	51	700

図4. 死亡したケースのクラスタ遷移数 (左: 在院短期、中: 在院長期の入院初期、右: 在院長期の入院 29 日目以降)

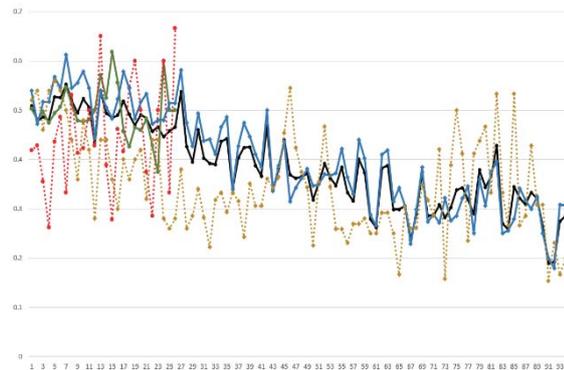


図6. クラスタ遷移の有無 横軸は入院初日からの日数、縦軸は遷移率の平均 (29 日以降も伸びている線の実線&丸点: 全体、実線&◆点: 在院長期かつ生存、点線&◆点: 在院長期かつ死亡、28 日までしかない線の実線&丸点: 在院短期かつ生存、点線&丸点: 在院短期かつ死亡)

## 謝辞

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。

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,(2018)

## 参考文献

[Hofmann,1999] Hofmann, Thomas. "Probabilistic latent semantic analysis." Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., (1999)

[Ide,2017]Ayae Ide,Kazuya Yamashita,Yoichi Motomura and Takao Terano,"Analyzing Regional Characteristics of LivingActivities of Elderly People from Large Survey Data with Probabilistic Latent Spatial Semantic Structure Modeling", IEEE BigData2017 Workshop,(2017)

[石垣,2010]石垣司, 竹中毅, 本村陽一: 百貨店 ID 付き POS データからのカテゴリ別状況依存的変数間関係の自動抽出法, オペレーションズ・リサーチ, 56, 2, (2010)

[石垣,2011] 石垣司, 竹中毅, 本村陽一: 潜在クラスモデルによる流通量販店舗の来店人数予測の精度改善の評価, 人工知能学会全国大会, (2011)

[井手,2016] 井手絢絵, 本村陽一:高齢者の生活行動推論のための大規模アンケートデータを用いた確率的潜在意味構造モデル構築, 人工知能学会「社会における AI」研究会第 27 回研究会,(2016)

[井手,2017] 井手 絢絵,川本 達郎,山下 和也,本村 陽一:マイクロアグリゲーションを用いた匿名化による確率的潜在空間意味解析, 人工知能学会全国大会,(2017)

[近藤,2017] 近藤那央,竹内理人,山下和也,櫻井瑛一,本村陽一,大規模イベントにおける人の行動履歴情報収集とそれを利用したインタラクティブシステムの開発に向けて,人工知能学会全国大会,(2017)

[阪本,2018] 阪本雄一郎,山下和也,櫻井瑛一,本村陽一:統計技術による DPC データの敗血症治療戦略への活用, 第 45 回日本集中治療医学会学術集会,(2018)

[原田, 2016] 原田奈弥, 山下和也, 本村陽一: ID 付 POS データによる購買行動の季節変化の分析と視覚化, 人工知能学会社会における AI 研究会 27 回研究会,(2016)

[本村,2011] 本村陽一, 竹中毅, 石垣司: 条件付層別差分モデルによる需要予測の高精度化, 人工知能学会全国大会, (2011)

[山下, 2017a] 山下和也,原田奈弥,黄冬陽,吉開朋弘,本村陽一: ID-POS データによる来店行動・購買行動の潜在的時空間意味構造分析, 人工知能学会全国大会,(2017)

[山下, 2017b] 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, 人工知能学会 社会における AI 研究会 27 回研究会,(2017)

[山下,2018] 山下和也,阪本雄一郎,櫻井瑛一,本村陽一:診療クラスター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