

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

Treatment Strategy of Septicemia Using DCP Data with pLSA which especially can Model Clients' Change along with Time

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

*¹ 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

*² 佐賀大学
Saga University

This paper show our research about treatment strategy of septicemia using DCP data with pLSA, probabilistic latent semantic analysis, which especially can model clients' change along with time. We used a variable which combine clients' ID and the days from hospitalization when we made some medical treatment and what we do as treatment to cluster our clients with pLSA, then summarized death rate and medical cost, days clients stay in this hospital of each cluster. Through researching clients' change along with time, We found some pattern how a client in some cluster move to another cluster those in this tend to die more, although there are some other clients who stay same cluster enough days after some treatments. Our research and method are accepted as some possibility of DCP data to assist treatment strategy in medical field.

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大半であり、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. 入院日数」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」、「短い入院日数」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[Hoffman 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として治療戦略に有益である可能性を検証した。

pLSA を用いた解析の実社会への応用としては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, 2017], [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。本研究は同様の手法を医療分野に適応した初の例である。

2. 時間変化も考慮した pLSA によるモデリング手法

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。pLSA では、観測される変数の分布に対する事前の仮定が不要であり、異なる分布や形式の変数を同時に厳しい制約なく扱えるため、本研究のような医療に関する変数を扱う際に適している。pLSA の実行には産業技術総合研究所の知財ソフトウェア「PLASMA」を用いた。

本研究では、上記の x, y のうち x を患者 ID と薬剤注射実施日を組み合わせたとした。実施日は「入院初日から3日まで」、「4~30日まで」、「31日目以上」と3期間に離散化して、「患者 ID_実施日の変数を x とした。 y を注射薬剤(分類名が注射の診療行為名称)とした。上記の変数 x, y を入力として pLSA を行い、同時に使用されやすい薬剤と、患者及び実施日を同時に分類した。使用した患者 ID 数は458であった。得られた潜在変数 z を薬剤クラスタと呼ぶこととし、このクラスタ毎に、死亡率(死亡の有無の平均)、出来高金額合計(医療費)、在院日数を

連絡先: 山下和也, 国立研究開発法人 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

集計した。更に同一患者の時間経過によるクラスター遷移を調べた。

3. 敗血症治療戦略への時間変化も考慮した pLSA の適応結果

3. 1. 薬剤クラスター毎の死亡率・医療費・在院日数の集計

2章で示した方法で pLSA を行ったところ薬剤クラスター数は 22 で AIC 最小となった。その各 22 クラスターに所属する ID について、死亡率・医療費・在院日数の平均を求めた結果を表 1 に示す。死亡率の低い順に c01, c02, .. とクラスターに名前を付けその順に示している。

表 1. 薬剤クラスター毎の死亡率・医療費・在院日数
死亡率の昇順 患者 ID 数: 458
患者 ID_実施日 データ数: 1040

クラスター	死亡率	医療費(万円)	在院日数(日)	データ数
c01	0.08	792	83	13
c02	0.08	395	58	61
c03	0.09	345	56	11
c04	0.12	186	28	99
c05	0.14	315	50	95
c06	0.14	223	37	70
c07	0.17	255	40	84
c08	0.17	241	34	122
c09	0.18	332	48	28
c10	0.20	428	75	75
c11	0.20	449	75	5
c12	0.22	661	116	18
c13	0.23	259	40	39
c14	0.31	358	40	55
c15	0.35	502	46	97
c16	0.38	702	75	34
c17	0.43	560	78	14
c18	0.48	366	57	27
c19	0.51	235	38	45
c20	0.57	798	104	14
c21	0.60	1035	86	10
c22	0.68	916	67	22

上記のクラスター c01~c22 を、死亡率-医療費の平面にマッピングした結果を図 1 に示す。図 1 をみると概ね死亡率が高いクラスター程、医療費も高くなっている傾向が見られるが、同様の死亡率の他のクラスターと比べて c18, c19 のように医療費が安いクラスターや、反対に c01 や c12 のように医療費が高いクラスターもある。

次に死亡率-在院日数の平面にマッピングした結果を図 2 に示す。c20 や c21 のように死亡率が高くかつ在院日数が長いクラスターがある一方、c04, c06, c08 など死亡率が低く在院日数が短いクラスターがある。c01 や c12 は同じ死亡率の他のクラスターと比べて在院日数が長い。

図 3 に在院日数-医療費の平面へのマッピングも示す。入院日数が長くなると医療費が増える基本傾向 d であるが c21 や c22 は同じ入院日数の他クラスターと比べて医療費が高く、c12 は入院日数が最長だがその割に医療費はやや抑えられている。

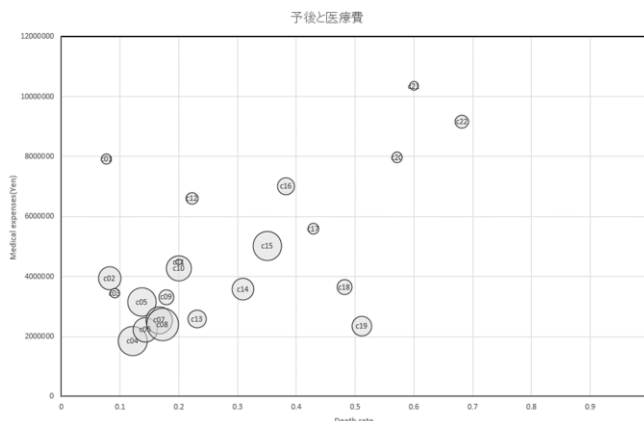


図 1. [死亡率-医療費]平面への薬剤クラスターマッピング
クラスターの円の大きさはデータ数を示す

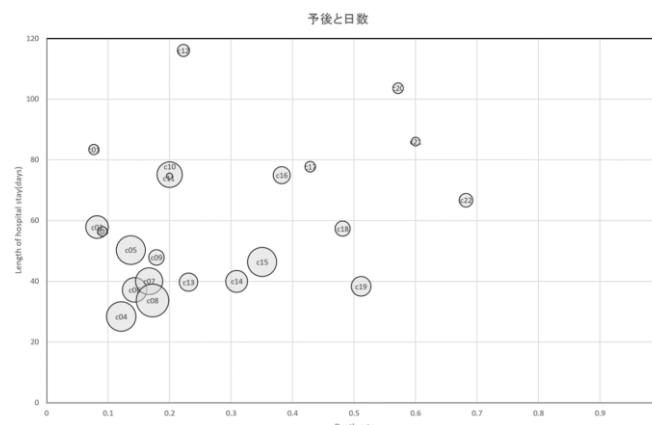


図 2. [死亡率-在院日数]平面への薬剤クラスターマッピング
クラスターの円の大きさはデータ数を示す

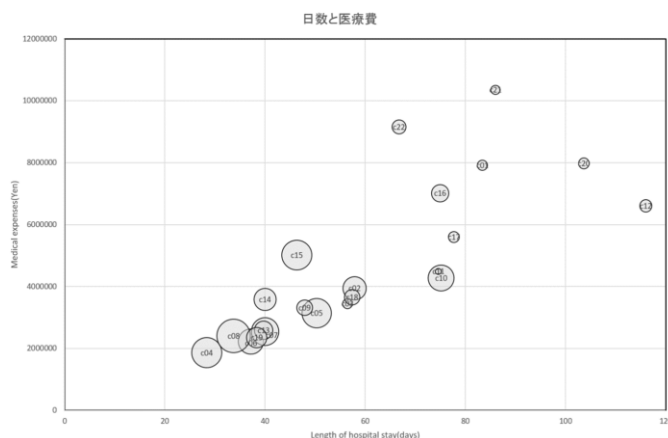


図 3. [在院日数-医療費]平面への薬剤クラスターマッピング
クラスターの円の大きさはデータ数を示す

3. 2. 薬剤クラスタの注射薬剤

3. 1. 節では変数 x の ID 部の集計を行った。本節では変数 y の注射薬剤の分類結果を示す。前節で見出した典型的及び特徴的なクラスタを以下に示す。

表 2 には、死亡率が低く・医療費が安く・在院日数が短いクラスタの薬剤を示した。P(z) は薬剤 y のクラスタ z への所属確率でありその降順で示した。

表 3 には、同様の死亡率の他クラスタと比べて医療費が安いクラスタを示した。

表 2. 死亡率が低く、医療費が安く、在院日数が短い薬剤クラスタ (a) c04 (b) c06 (c) c08

(a) c04 クラスタ

診療行為名称	P(Z)
スペインルディスボ関節注25mg1%2. 5mL	1.00
ノボラピッド注フレックスタッチ300単位	1.00
リユープリン注射用1. 88	1.00
中心静脈用カテーテル(酸素飽和度測定機能付き) ¥	1.00
栄養カテーテル(経鼻用・経腸栄養用) ¥	1.00
気管切開後留置用チューブ(輪状甲状腺切開チューブ) ¥	1.00
献血グロベニン-I 静注用2500mg2. 5g50mL溶解液付	1.00
セフトリアキソンNa 静注用1g「サンド」	0.91
ダウノマイシン静注用20mg	0.79
ソルデム3A輸液500mL	0.71
レベミル注フレックスペン300単位	0.62
ジゴシン注0. 25mg0. 025%1mL	0.58
ランタス注ソロスター300	0.55
アトニン-Q注5単位	0.55
ノボラピッド注フレックスペン300単位	0.43
ノボリンR注フレックスペン	0.30
大塚生食注2ポート100mL	0.24

(b) c06 クラスタ

診療行為名称	P(Z)
ウテメリン注50mg1%5mL	1.00
エリル点滴静注液30mg30. 8mg2mL	1.00
キサンボン注射用40mg	1.00
ドブトレックスキット点滴静注用600mg0. 3%200mL	1.00
マグセント注100mL	1.00
ユナシン-S 静注用1. 5g	1.00
リン酸二カリウム補正液1mEq/mL0. 5モル20mL	1.00
ヴィーン3G注500mL	1.00
ヴィーンD注(500mL)	1.00
ヴィーンF注(500mL)	1.00
低分子デキストラン注(500mL)	1.00
注射用水PL「フソー」100mL	1.00
血管造影用ガイドワイヤー(一般用) ¥	1.00
サンディミューン点滴静注用250mg5%5mL	0.94
セフメタゾン静注用1g	0.94
献血ベニロン-I 静注用2500mg2. 5g50mL(溶解液付)	0.91
セフィローム静注用1g	0.85
モダシン静注用1g	0.83
ノボリンN注フレックスペン	0.83
点滴静注用バンコマイシン0. 5「MEEK」	0.80
ネオフィリン注250mg2. 5%10mL	0.79
カタボンHi注600mg0. 3%200mL	0.63
大塚生食注(500mL)	0.56
セファメジンa 注射用1g	0.56
プロジフ静注液2008%2. 5mL	0.54
フレスミンS注射液1000µg1mg	0.50

(c) c08 クラスタ

診療行為名称	P(Z)
アミカシン硫酸塩注射液200mg「サワイ」	1.00
セビドナリン静注用1g	1.00
リユープリン注射用キット3. 753. 75mg	1.00
注射用ベニシリンGカリウム100万単位	1.00
クリンダマイシンリン酸エステル注射液600mg「サワイ」	0.96
ヴィーンF輸液500mL	0.95
ラセナゾリン注射用1g	0.89
ネオラミン・スリーピー液(静注用)10mL	0.80
ヴィーンD輸液500mL	0.77
ヴィーン3G輸液500mL	0.77
セフォチアム塩酸塩静注用1g「日医工」	0.75
イリノテカン塩酸塩点滴静注液100mg「NK」5mL	0.61
ケタラール静注用200mg20mL	0.59
ゾメタ点滴静注4mg/5mL	0.50

表 3. 同様の死亡率の他クラスタと比べて医療費が安いクラスタ

(a) c18 クラスタ

診療行為名称	P(Z)
アザクタム注射用1g	1.00
オキサロール注5µg1mL	1.00
グリセリン点滴静注200mL	1.00
ソリター-T2号輸液500mL	1.00
ダントリウム静注用20mg	1.00
ノーベルパール静注用250mg	1.00
ビペラシリンナトリウム注射用1g「マルコ」	1.00
ブリディオ静注200mg2mL	1.00
マンニトール注射液「YD」[20%]300mL	1.00
リプル注10µg2mL	1.00
新鮮凍結血漿-LR「日赤」血液400mL相当に由来する血漿	1.00
ゾシン静注用2. 252. 25g	0.93
ザイボックス注射液600mg300mL	0.90
トブラシン注60mg	0.84
アンソロピンP500注射用500単位(溶解液付)	0.72
プラスチックカニューレ型静脈留置内針(標準型) ¥	0.44
ラシックス注20mg	0.37
中心静脈用カテーテル(標準・シングルルーメン・セルジンガー型)	0.29

(b) c19 クラスタ

診療行為名称	P(Z)
▽注射用蒸留水20mL	1.00
アキネトン注射液5mg0. 5%1mL	1.00
アナフラニール点滴静注液25mg	1.00
アリクストラ皮下注2. 5mg0. 5mL	1.00
オキファスト注10mg1%1mL	1.00
カルボカインアンプル注1%10mL	1.00
クレキサン皮下注キット2000IU2千低ヘパ国際単位0. 2mL	1.00
ゲンタシン注10	1.00
ジアグノグリーン注射用25mg(溶解液付)	1.00
ソル・メドロール静注用1000mg1g(溶解液付)	1.00
ネオビタカイン注シリンジ5mL	1.00
ハイスコ皮下注0. 5mg0. 05%1mL	1.00
ビタメジン静注用	1.00
ビーエヌツイン-1号輸液	1.00
リンデロン注4mg(0. 4%)	1.00
動脈圧測定用カテーテル(末梢動脈圧測定用) ¥	1.00
塩酸モルヒネ注射液	1.00
ケイツーN静注10mg	0.89
スルバクタム・アンピシリン静注用1. 5g「サンド」	0.89
プロボフォール静注1%「KABI」500mg50mL	0.80
ソル・メドロール静注用40mg(溶解液付)	0.67
アレビアチン注250mg5%5mL	0.66
グルカゴン注射用1単位「イトウ」1U. S. P. 単位(溶解液付)	0.62
ベルジピン注射液2mg	0.56
レペタン注0. 2mg	0.55
ソル・メドロール静注用500mg(溶解液付)	0.54

3. 3. 薬剤クラスタの実施日

本節では変数 x の実施日部の集計を行った。図 4 にクラスタ毎の実施日のデータ数を示す。図 5 は各クラスタデータ数で正規化し実施日の割合を示した。

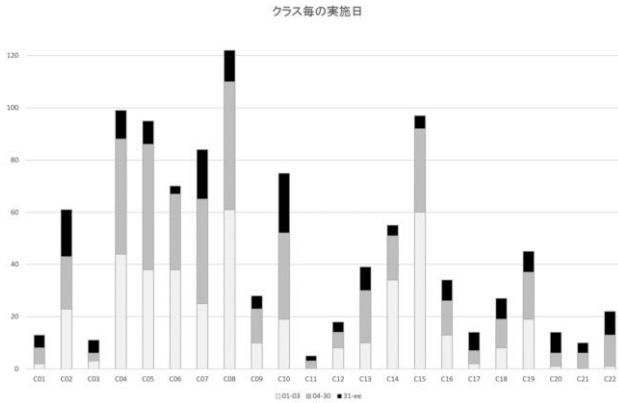


図 4. 各薬剤クラスタの実施日

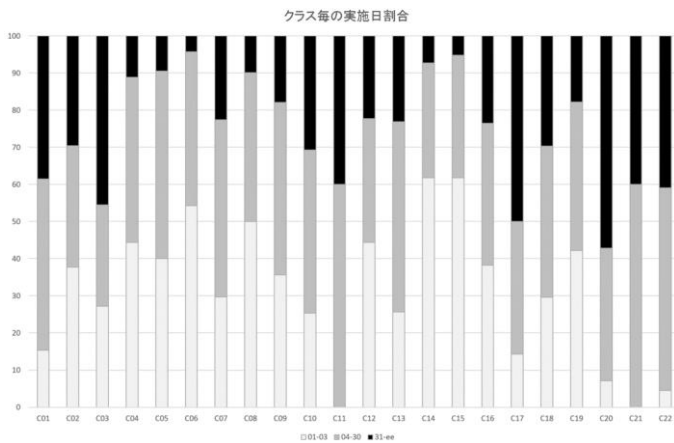


図 5. 各薬剤クラスタの実施日割合

死亡率が低く・医療費が安く・在院日数が短いクラスタ c04, c06, c08 は実施日の入院からの日数が浅いデータが多い。同様に入院からの日数が浅いデータが多いクラスタとしては c14 や c15 がある。C20～c21 等は入院日からの日数が長いデータの割合が大きいものに対して c18, c19 では実施日が浅いデータも含まれている。

3. 4. 同一患者の時間経過に伴うクラスタ遷移

同一患者の時間変化に伴うクラスタ遷移を無作為に選んだ 5 事例を以下に示す。

- 患者 A C04→C14→C18→死亡
- 患者 B C04→C08
- 患者 C C05→C05→C07
- 患者 D C09→C09→C09

患者 E C16→C16→C22→死亡

死亡率の高いクラスタへ推移していく患者や、日数経過後も同一クラスタに留まる患者などのパターンが見られる。

死亡率が高くまた比較的医療費が高く、在院日数も長い c20, c21, c22 クラスタに一度でも入る患者のクラスタ遷移の全ケースを表 4 に示す。

表 4. 死亡率が最も高く医療費が高く在院日数も長いクラスタ (c20, c21, c22) に一度でも入る患者の時間経過に伴うクラスタ遷移の全ケース 上の方は生存した患者、下部は死亡した患者でそれぞれに「初日～3 日」時点でのクラスタ番号 (クラスタの平均死亡率) が若い (低い) 順に示す

初日～3日	4～30日	31日～	死亡=1
2	20		
3	12	20	
4	22		
5	22	13	
6	15	21	
7	21		
8	20	20	
12	22	2	
15	21		
15	20	2	
15	22		
15	22	22	
20	7		
22	21	6	
22	13		
2	22	22	1
4	21	21	1
4	15	20	1
6	21	21	1
7	21	21	1
7	5	20	1
7	22	14	1
8	20	20	1
8	20		1
9	22	22	1
13	22		1
14	19	20	1
14	15	20	1
15	22	22	1
15	22	22	1
15	15	22	1
15	12	20	1
16	2	22	1
16	16	22	1

表4をみると死亡した患者が半数より若干多い。生存した患者は初期に死亡率が高いクラスタにいた場合やまたは期間途中で死亡率の高いクラスタに一度入った後でその後死亡率の低いクラスタに遷移している場合が見られる。一方最終クラスタが死亡率の高いクラスタでも生存している患者も見られる。死亡した患者は初期に比較的死亡率の低いクラスタにいたが期間途中で死亡率高いクラスタに遷移しその後もそのクラスタに留まるかまたはより死亡率高いクラスタに遷移して死亡するケースが多い。

次に死亡率が低く、医療費が安く、在院日数が短いc06クラスタに一度でも入る患者のクラスタ遷移の無作為サンプルを表5に示す。

表5. 死亡率が低く医療費が安く在院日数も短いクラスタ(c6)に一度でも入る患者の時間経過に伴うクラスタ遷移の無作為サンプル 上の方は生存した患者、下部は死亡した患者でそれぞれに初日～3日のクラスタ番号(クラスタの平均死亡率)が若い(低い)順に示す

初日～3日	4～30日	31日～	死亡=1
6	19	18	
6	4		
6	8	7	
6	4		
6	6		
6	6		
6	15	21	
6	8		
6	18		
6	10	8	
9	6		
15	6	6	
19	6		
6	21	21	1
6	6		1
6			1
14	6		1

表5をみると大方の患者は生存していて、死亡率の低いクラスタ間を遷移している場合が多い。一方死亡した患者は死亡率が高いクラスタに遷移した後に死亡した場合もあるものの、低いクラスタc6に留まったままやまたはc6に遷移した後に死亡したケースが見られる。

4. まとめと今後の展開

本研究により今後の症例の集積と医学的判断の関与によってDPCデータが治療戦略に関与し得る可能性が認められた。

ここでは実施日の離散化を固定した3期間を与えたが、もっと区切り幅を細かくし(たとえば3日毎)同

一患者の実施日のクラスタの分かれ目を見つけることで患者毎に動的に適切な離散化を行う事が出来ることも確かめている。

またここでは治療の種類として注射薬剤のみを変数yに加えたが、yに同時に注射以外の各診療行為(投薬、処置、手術等)も加える事も出来、同時に実施されやすい診療行為をクラスタリングする事が可能である。

本解析を進め上記のように手法を確立しシステム化する事で集中医療室に導入する検討を始めている。

本研究により時間変化もモデル化の対象としたpLSA手法の医療分野への応用の道を開いた。

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。原田奈弥氏には文章の校正をしてもらい、またアブストラクトの英文作成でお世話になり厚くお礼申し上げます。

参考文献

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

[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)

[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, 1B3-1, (2011)

[山下, 2017] 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