

重み付き PLSA と DPC データを用いた患者と診療行為等の同時クラスタリングにおける重み値によるクラスタリングの特徴の違い

Clustering Difference of the features of the by Weight Values in the Weighted PLSA using DPC data

山下 和也^{*1}

Kazuya Yamashita

村田 知佐恵^{*1}

Chisae Murata

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

The accumulation of DPC data has progressed, and the importance of these medical big data is recognized. At present, there are few reports such as application for selection and comparison of a new treatment strategy using DPC. We used DPC data and PLSA to clarify simultaneous clustering of patients requiring intensive care and clinical practice and to extract clusters time transition patterns of patients and to evaluate feasibility of doctor's treatment strategy decision support algorithm we have verified. Furthermore, by using weighted PLSA, clustering can be performed which clearly shows the difference of attention variables by giving weight to 'attention' medical treatment, objective variables such as mortality rate, number of hospital days and medical expenses. It is confirmed that it is. In this research, we treat all DPC items including variables that we have not used so far as variables, and investigate the differences in clustering features by changing the weight of each variable. As a result, it was shown that it is possible to generate clusters according to the purpose by adjusting the weight of each variable as a parameter according to the variable to be noticed.

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ー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ー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ーを遷移する確率からその後の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ー分類されることを明らかにした [山下 2019]。ただし 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指してその為に、重み付き PLSA の重み値を 0 及び正の数とすることで、これまでの解析と DPC 全体を用いた新たな解析を重み値の違いとして統一的に議論し、重み値を変化させることによるクラスタリングの特徴の違いとして整理し、重み付き PLSA の医用応用における有用性を調べた。

2. 重み付き PLSA

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。

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)。

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 (weighted PLSA)」の重みを y の変数別に各行に対して、0, 1, 20 等に変えて、変数 x 側の患者のクラスター遷移パターンや y 側のクラスタリングの特徴が重み値を変えることによりどう変わるかを調べた。

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)

本研究は、研究施設の倫理審査委員会で承認を取得した。また、PLSAの実行は、産業技術総合研究所の知財ソフトウェア「PLASMA」を使用した。

際にはその診療行為に高い重みを与える(山下,2019)他、見たい現状に応じて各変数の重みをパラメータとして調整することで目的に応じたクラスタを生成出来ることが可能であることが示された。

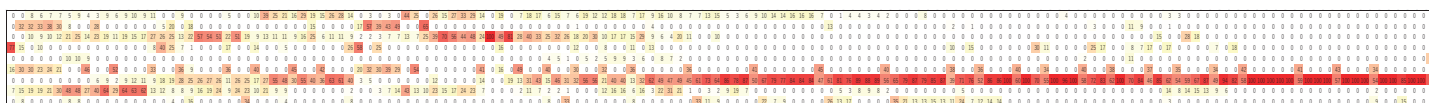


図1 診療行為名称のみに重み1他は0の生存患者のクラスタ遷移



図2 DPC各項目に重み1、死亡の有無のみ20倍の重みの生存患者のクラスタ遷移

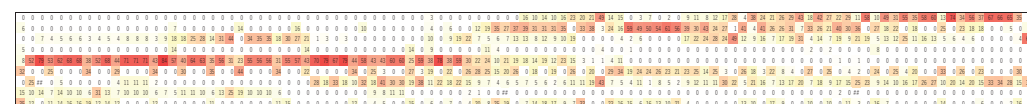


図3 診療行為名称のみ重み1他は0の死亡患者のクラスタ遷移

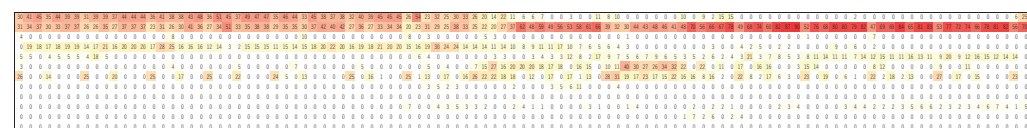


図4 DPC各項目に重み1、死亡の有無のみ20倍の重みの死亡患者のクラスタ遷移

3. 重み付き PLSA の重み値の違いによるクラスタリングの特徴

PLSA の y の変数として診療行為名称のみに重み1を与え、他の変数は0とした場合の生存したある患者のクラスタ遷移を図1に示す。図の横方向は時間軸(日)であり、左端が入院初日で右端が退院日である。縦方向は AIC 基準で最適とされた9つのクラスタをそれぞれのクラスタに属する患者の死亡率の低い順に下から並べたものである。各マス目の色(数字)は患者のその日の各クラスターへの所属確率(%)を示していて赤が濃い程確率が高い。入院初期は様々なクラスタに所属しているが、後期は下から3つ目のクラスタに集中している。このクラスタは投薬や診察、検査項目が中心のクラスタである。

次に y の変数として診療行為名称以外にも表1の各項目に重み1を与え、死亡の有無のみ重み20を与えた場合の12クラスタ(AIC最小)に対して上記と同じ生存患者の遷移を図2に示す。入院初期は下から2番目のクラスタの所属確率が高くまたこのクラスタへの所属は退院まで一貫して続いている。後半には下から3番目のクラスタへの所属確率も高まっている。この3番目のクラスタは投薬、診察、検査項目が中心のクラスタで、診療行為名称にのみ重み1の場合の入院後期に所属した項目と共通の診療行為名称が多い。

死亡した患者の事例についても上記と同様に診療行為名称のみに重み1とDPC各項目も入れた場合の結果を図3、図4にそれぞれ示す。診療行為名称に重み1倍の方では日々ある程度複数のクラスタへ所属している一方、各DPC項目を入れた方では上から2番目への所属確率が高くそのクラスタへの所属は入院期間にわたって一貫している。

4. 考察

各DPC項目を入れたPLSAでは入院全期にわたって所属しているクラスタが見られ、そのようなクラスタは患者の属性が反映されていると思われる。日々の診療行為の違いにあまり左右されない総合的な患者の状態をみられるクラスタが生成されていると捉える事が出来る。特定の診療行為の効果に注目したい

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.
- [本村 2018]本村陽一;ビッグデータを活用する確率モデリング技術～社会実装の取り組みと課題～,統計数理 第66巻 第2号,(2018)
- [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.
- [本村 2019]本村陽一,山下和也,阪本雄一郎:DPC診療群分類包括評価データに対する確率的潜在意味構造モデリング,BIO Clinica 34(3),2019