

自己組織化マップを用いた診断群分類番号データの分析と 医療情報システムへの応用

仲濱 正大^{†1} 納富 一宏^{†1} 斎藤 恵一^{†2} 外山 比南子^{†3}

2003年より始まった急性期入院医療を対象とした診療報酬の包括評価制度（DPC/PDPS：Diagnosis Procedure Combination / Per-Diem Payment System）により、患者の疾病別医療行為等の診療情報が標準化され、分析可能な全国統一形式の患者臨床情報である DPC データが収集可能となった。我々は過去の DPC データを対象に、自己組織化マップ（SOM：Self-Organizing Maps）を用いた新しい患者の在院日数予測を試みている。本稿では DPC コードと合併症数を属性ベクトルに変換し、SOM を用いて分析を行った。その結果、DPC コード別に在院日数が分かれていることが確認できた。また、SOM による分析結果を利用した医療情報システムへの応用についても検討した。

Analysis of Diagnosis Procedure Combination Data by Self-Organizing Maps and its Application to Medical Information System

Masahiro Nakahama,^{†1} Kazuhiro Notomi^{†1} Keiichi Saito^{†2} Hinako Toyama^{†3}

1. はじめに

2003年より始まった急性期入院医療を対象とした診療報酬の包括評価制度（DPC/PDPS：Diagnosis Procedure Combination / Per-Diem Payment System）は、従来の出来高払い制度に代わるものとして特定機能病院等で導入された。診療報酬の包括評価制度の対象病院は段階的に拡大され、2013年4月1日見込みで1,496病院・約47.5万床となり、全一般病床の約52.8%を占めるに至った。DPC/PDPSにより、患者の疾病別医療行為等の診療情報が標準化されたことで他施設との比較が可能となり、在院日数の適正化や診療プロセスの見直しなど医療の質の向上に貢献している¹⁾。

我々は、複数の病院から収集された分析可能な全国統一形式の患者臨床情報である DPC データ²⁾を対象に、自己組織化マップを用いた新しい患者の在院日数予測を試みている。入院早期における在院日数予測が可能となれば、病院は病床管理の補助に、患者は在院日数・入院費用の概算の参考にすることができ、医療従事者・患者双方のメリットとなることが期待される。

これまでの研究から、自己組織化マップに学習させる属性ベクトルに診断群分類番号（DPCコード）と合併症数を含めることにより、在院日数予測精度が向上することが予想された³⁾。

本稿では、過去に収集された DPC データに含まれる DPC コードと合併症数を自己組織化マップ学習用の属性ベクトルに変換し、在院日数ごとの分類・情報視覚化を試みた。また、自己組織化マップを用いた DPC データの視覚化による医療情報システムへの応用を考察する。

2. 関連技術

2.1 診断群分類

診断群分類（DPC：Diagnosis Procedure Combination）とは、入院期間中に医療資源を最も投入した傷病名と、入院期間中に提供される手術、処置、化学療法などの診療行為の組み合わせにより、設定される分類である。

DPC においては、診断群分類として 14 桁の DPC コードが使用されている。その 3 桁目から 6 桁目が「臓器、病理コード」（傷病名）となる。

DPC コードは以下のように決定される。

- ① 医療資源を最も投入した傷病名を決定する。
- ② 医療資源を最も投入した傷病名に対応する ICD コードを決定する。
- ③ 決定した ICD コードと診療行為等に基づき、厚生労働大臣の告示により、ICD コードの属する DPC 傷病名と診断群分類番号を決定する。

DPC コードの各桁の意味付けを表 1 に示す。

ICD コードとは、異なる時点で集計された死亡や疾病のデータの体系的な記録、分析、解釈及び比較を行うため、世界保健機関憲章に基づき、世界保健機関が作成した分類である。22 の章から構成されており、それぞれの章は、3 桁分類（アルファベット 1 文字+数字 2 文字）と、より詳

^{†1} 神奈川工科大学大学院工学研究科
Graduate School of Engineering, Kan-agawa Institute of Technology

^{†2} 国際医療福祉大学情報教育室
Education Center of Medical Informatics, International University of Health and Welfare

^{†3} 国際医療福祉大学大学院
International University of Health and Welfare Graduate School

細な分類である 4 桁分類（アルファベット 1 文字+数字 3 文字）からなっている。ICD-10 コードによる分類符号の例を図 1 に示す。

表 1 DPC コードの各桁の意味付け

桁番号	意味
1 桁～2 桁	主要診断群
3 桁～6 桁	分類コード
7 桁	入院種別
8 桁	年齢・体重・JCS 条件
9 桁～10 桁	手術等サブ分類
11 桁	手術・処置等 1
12 桁	手術・処置等 2
13 桁	副傷病名
14 桁	重症度

胃体部癌の場合

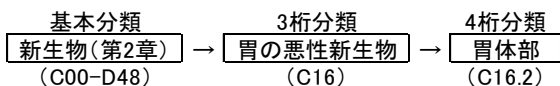


図 1 ICD コードによる符号化の例

2.2 自己組織化マップ

自己組織化マップ (SOM : Self-Organizing Maps) とは、1982 年に T.Kohonen により発表された競合学習を基礎とした人工ニューラルネットワークの一種である。入力層と競合層の 2 階層で階層型ネットワークを構成しており、教師信号を必要としない教師なし学習を行う。SOM の特徴は高次元 データをそのデータ空間での位相を保存した状態で低次元 (主に 2 次元) に写像することである。

本稿では、トーラス型 SOM を使用する。トーラス型 SOM は出力層の 2 次元座標空間が、トーラス状に結合したものである。トーラス型 SOM は、マップの端で学習が起きても偏りが生じないことと、2 次元マップの視認性を両立できる点で 2 次元 SOM より優れている。

3. 自己組織化マップによる DPC データの分析

3.1 分析対象となる DPC データ

分析対象となる DPC データは、2010 年 4 月から 2012 年 3 月までに複数の病院から収集された 22,001 ケースのうち、主症病名および医療資源を最も投入した疾病名の ICD コードが C162 (胃体部癌) であるものとした。データの総数は 119 件となった。

DPC データは 118 の項目から構成されている。DPC データの構成要素の一部を表 2 に示す。

表 2 DPC データの構成要素

データ名	意味
IN_DAYS	在院日数 (患者が病院で過ごす日数)
M_DISEASE_CODE	主症病名の ICD コード
DPC_BUNNRUI_NO	DPC 分類番号
compli_post1_code_max	合併症 1 の ICD コード
N_BREAKpost	合併症数 (合併症を発症した数. 最大 5)

3.2 属性ベクトルのエンコーディング

SOM への入力として、DPC データに含まれる手術や処置情報を示す DPC コードの一部と合併症数を用いて、SOM 学習用の属性ベクトルを作成した。DPC コードは、9, 10 桁目を 5 種類、12 桁目を 3 種類、13 桁目を 3 種類、合計 11 種類のダミー変数 $D_i (0 < i \leq 11)$ で表現した。合併症数は 0 から 0.133 刻みで 0.665 まで 6 段階の値 C で表した。このことにより、属性ベクトルの要素数は 12 となる。これを図 2 に示す。

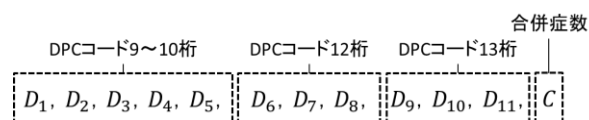


図 2 属性ベクトルの構成

合併症数を属性ベクトルに加えた理由は、合併症数と在院日数の間に正の相関があり、合併症数別の在院日数の分布に差が生じると考えたためである。

3.2.1 DPC コードのエンコーディング方法

手術等サブ分類を表す 9, 10 桁目を 5 種類のダミー変数にし、属性ベクトルにエンコードした。DPC コード 9, 10 桁目は診断・診療データに含まれる手術内容を表す術式 K コードによって決定される。本分析の対象となる胃体部癌 (DPC コードの 1~6 桁が 060020) における 9, 10 桁目の DPC コーディングの条件を表 3 に示す。

表 3 DPC コード 9, 10 桁目のコーディング条件

術式 K コード	DPC コード
K655-42 or K6552 or K657-22 or K6572	01
K636 or K636-3 or K636-4 K662 or K662-2 or K664	03
K6531 or K6532 or K6533 or K6534	04
K654 or その他の K コード	97
手術なし	99

DPC コードの 9, 10 桁目を p とした場合の属性ベクトル $D_1 \sim D_5$ のエンコーディング方法を表 4 に示す。

表 4 DPC コード 9, 10 桁目のエンコード条件

属性値	条件	値
D_1	$p = 01$	1
	$p \neq 01$	0
D_2	$p = 03$	1
	$p \neq 03$	0
D_3	$p = 04$	1
	$p \neq 04$	0
D_4	$p = 97$	1
	$p \neq 97$	0
D_5	$p = 99$	1
	$p \neq 99$	0

手術・処置等 2 を表す 12 桁目を 3 種類のダミー変数にし、属性ベクトルにエンコードした。12 桁目は DPC データに含まれる処置内容によって決まる。12 桁目のコーディングの条件を表 5 に示す。

表 5 DPC コード 12 桁目のコーディング条件

処置内容	DPC コード
なし	0
人工呼吸, 中心静脈注射	1
化学療法ありかつ放射線療法なし	3

DPC コードの 12 桁目を q とした場合の属性ベクトル $D_6 \sim D_8$ のエンコーディング条件を表 6 に示す。

表 6 DPC コード 12 桁目のエンコード条件

属性値	条件	値
D_6	$q = 0$	0.5
	$q \neq 0$	0
D_7	$q = 1$	0.5
	$q \neq 1$	0
D_8	$q = 3$	0.5
	$q \neq 3$	0

副傷病名を表す 13 桁目を 3 種類のダミー変数にし、属性ベクトルにエンコードした。13 桁目は DPC データに含まれる副傷病名によって決まる。13 桁目の DPC コーディングは手術等サブ分類と手術・処置等 2 および副傷病名によって決まる。13 桁目の DPC コーディングの条件の一部を表 7 に示す。

表 7 DPC コード 13 桁目のコーディング条件

術式 K コード	12 桁目の値	副傷病	DPC コード
なし	0	なし	0
		あり	1
	1 or 2	なし	0
		あり	1
3 or 4	なし	0	
	あり	1	
その他の手術			x
K653\$			x
K636 等			x
K655-22	0		x
	1	なし	0
		あり	1
	2 or 3 or 4		x
K6572 等			x

DPC コードの 13 桁目を r とした場合の属性ベクトル $D_9 \sim D_{11}$ のエンコーディング条件を表 8 に示す。

表 8 DPC コード 13 桁目のエンコード条件

属性値	条件	値
D_9	$r = 0$	0.5
	$r \neq 0$	0
D_{10}	$r = 1$	0.5
	$r \neq 1$	0
D_{11}	$r = x$	0.5
	$r \neq x$	0

$D_6 \sim D_8$ および $D_9 \sim D_{11}$ が 0 または 0.5 の値をとる理由は、SOM の学習において手術等サブ分類によってクラスタリングを行いたいためである。DPC コードを基にした属性ベクトルの全てが 0 または 1 の値になった場合、表 1 における「手術・処置等 2」や「副傷病名」によってクラスタが生成される可能性があるからである。

3.3 学習と考察

総ユニット数 2,500 個 (マップサイズ 50×50)、学習回数は 100,000 回、属性ベクトル数 118、属性ベクトル次元数 12 の SOM 出力層を図 3 に示す。各領域は、DPC データから属性ベクトルにエンコードされた際に与えられた DPC コードに基づく識別子別にプログラムによって塗り分けられている。図中の矢印と長方形の中の文字は、塗り分けられた領域の DPC コードを表している。図中のラベルは次のように意味付けされている。

在院日数 (週), 合併症数

また、図 4 に SOM 出力層を在院日数別に色を塗り分けた結果を示す。

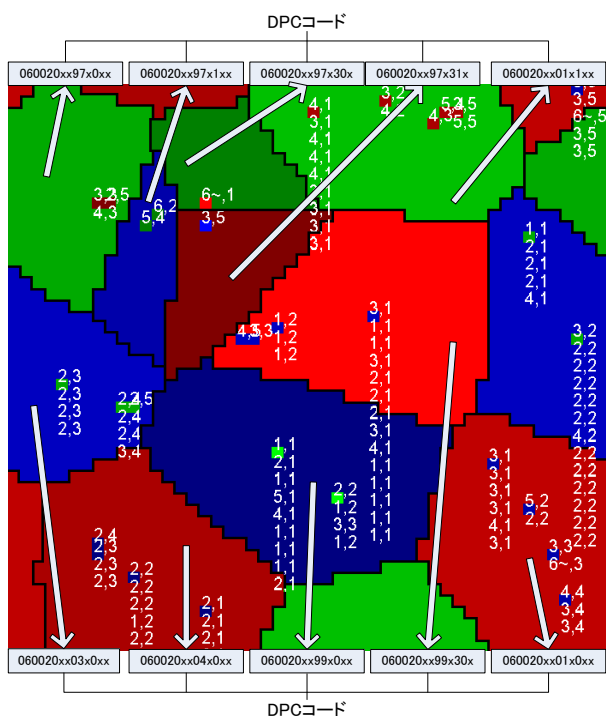


図 3 SOM の学習結果

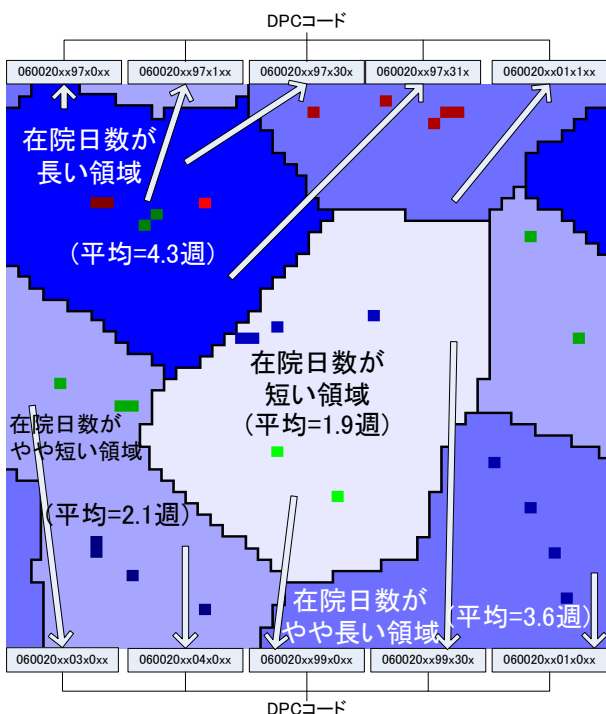


図 4 SOM の学習結果 (在院日数別に色分け)

SOM の学習によって生成された 7 つの領域において在院日数が近いデータが集まっていることが分かった。しかし、DPC コード 060020xx01x0xx, 060020xx99x0x は在院日数のばらつきが大きく、それぞれ平均と標準偏差は 1.92 ± 1.32 , および 3.61 ± 1.38 となり、双方に在院日数が長いデータが含まれる。また、DPC コード別の領域内に合併症数別で属性ベクトルが配置されており、それらは合併

症数が増えると在院日数の値が増大する傾向があることが分かった。これらの結果から DPC コーディングに採用される処置情報以外にも在院日数に影響を与える要素が存在していると考えられる。

4. 医療情報システムへの応用

これらの実験結果を利用した応用システムとして、医師と患者のコミュニケーションを支援する医療情報システムを検討している。医療行為を行う際は、インフォームド・コンセントを行う必要がある。インフォームド・コンセントとは、医師が患者に対して治療の必要性、治療期間、治療にかかる費用等の説明を行い、患者の合意を得ることである。しかし、治療に複数の選択肢が存在する場合、口頭の説明に加えて、視覚化された情報提示により、理解が容易になるものと期待される。

提案システムは、実験結果に示した SOM の視覚化を用いて、処置別の在院日数分布を把握することを容易とする。患者は医師が持つ経験的な情報と SOM マップが示す過去の DPC データの情報を得ることができ、システムの提供する情報が医師とのコミュニケーションの補助となることが期待される。図 5 に提案システムの利用場面を示す。



図 5 医療情報システムへの応用

5. おわりに

本稿では、過去に収集された DPC データの DPC コードと合併症数を属性ベクトルに変換し、SOM を用いて分析を行った。その結果、DPC コード別に領域が生成され、在院日数が近いデータが集まっていることが確認できた。また、同じ領域の中でも在院日数が長いデータが存在することから、DPC コード以外にも在院日数に影響を与える要素が存在していることが示唆された。また、分析結果の応用システムとして、SOM の情報視覚化を用いた患者・医師間のコミュニケーション補助について提案した。

参考文献

- 1) 厚生労働省保険局医療課：平成 24 年度診療報酬改定の概要 (DPC 制度関連部分), 厚生労働省 (オンライン), 入手先 <http://www.mhlw.go.jp/bunya/iryohoken/iryohoken15/dl/h24_01-05.pdf> (参照 2013-01-07)
- 2) 診療報酬調査専門組織各分科会：松田研究班からの報告 (D-1), 厚生労働省 (オンライン), 入手先 <<http://www.mhlw.go.jp/shingi/2008/10/dl/s1003-8a.pdf>> (参照 2013-05-13)
- 3) 仲濱正大, 納富一宏, 斎藤恵一, 黒田史博, 外山比南子：入院早期における在院日数予測手法の検討, 情報処理学会 第 75 回全国大会講演論文集 第 3 分冊 3ZG-3, pp.905-906, (2013)