

DPC データを用いた在院日数予測を行う Web アプリケーションの開発

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

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 DPC データの活用の一環として在院日数予測を試みている。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。本稿では、22,001 件の DPC データのうち、ICD-10 コードが C162 (胃体部癌) である 116 件を対象に、「在院日数」、「DPC コード」、「合併症数」、「病院コード」、「入院区分」を自己組織化マップ学習用の属性ベクトルに変換し、在院日数予測を行った。また、在院日数予測を行う Web アプリケーション開発について検討した。

Development of Web Application using DPC Data: Predicting Length of Stay in Hospital

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

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、2014年4月1日見込みで1,585病院・約49万床となり、全一般病床の約55%を占めるに至っている¹⁾。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。

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⁾ 応用の一環として、新規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。DPC データは統一形式で記録されているため、その応用は、DPC/PDPS に参加する全ての病院で利用可能である。そのため、DPC データの応用手法の検討は重要であると考えられる。

本稿では、DPC データに含まれる、「在院日数」、「DPC コード」、「合併症数」、「病院コード」、「入院区分」を自己組織化マップ (SOM : Self-Organizing Maps) 学習用の属性

ベクトルとした場合における在院日数予測成功率の評価・考察、および在院日数予測を行う Web アプリケーション開発について述べる。

2. 自己組織化マップによる在院日数予測実験

2.1 実験対象となる DPC データ

在院日数予測実験の対象となる 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件のうち、医療資源を最も投入した疾病名の ICD-10 コード (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 (WHO : World Health Organization) が作成した分類) が C162 (胃体部癌) であるものとした。123 件の DPC データがこの条件に合致した。胃体部癌を実験対象としたのは、22,001 件ある DPC データの在院日数の歪度が 2.90、尖度が 16.0 と分布の偏りが大きく、多くの患者が1週間以内に退院していることがわかったため、在院日数の歪度が 1.77、尖度が 4.64 である胃体部癌を対象にした。

また、123 件中その他の手術に該当する DPC データを在院日数予測実験の対象から除外した。これは、その他の手術に該当する DPC データの件数が7件、在院日数の標準偏差が 16.1 となり、予測が困難であると判断したためである。この結果、在院日数予測実験の対象となる DPC データは 116 件となった。

2.1.1 DPC データとは

DPC データとは、DPC/PDPS 導入影響評価のための調査 (退院患者調査) による調査データを指す²⁾。DPC データは5つのファイルから構成されている。今回分析に使用し

^{†1} 神奈川工科大学大学院工学研究科
Graduate School of Engineering, Kanagawa Institute of Technology
^{†2} 国際医療福祉大学大学院
International University of Health and Welfare Graduate School

たデータは、簡易診療録情報からなる様式1および医科点数表に基づく出来高点数情報からなる EF 統合ファイルから抽出した 118 項目で構成されている。

2.2 属性ベクトルの構成

SOM への入力として、DPC データに含まれる「在院日数」、「DPC コード」、「病院コード」、「入院区分」、「合併症数」を用いて、SOM 学習用の属性ベクトルを作成した。属性ベクトルの構成を図 1 に示す。

2.2.1 在院日数および合併症数から属性ベクトルへの変換

在院日数 (v_1) および合併症数 (v_{18}) は次式を用いて、0~1 の数値に正規化した。

$$v_i = \frac{x_j - \min(x)}{\max(x) - \min(x)} \dots \dots \dots (1)$$

x : v の元となる数値(在院日数または合併症数)

2.2.2 DPC コードから属性ベクトルへの変換

DPC コードとは入院期間中に医療資源を最も投入した傷病名と、診療行為の組み合わせによって決定される 14 桁のコードである。DPC コードの各桁の意味付けを図 2 に示す。

DPC コードは、手術等サブ分類を表す 9, 10 桁目を 4 種類、手術・処置等 2 を表す 12 桁目を 3 種類、副傷病名を表す 13 桁目を 3 種類、合計 11 種類で表現した。DPC コードの 9, 10 桁目を p , 12 桁目を q , 13 桁目を r とした場合の DPC コードから属性ベクトルへの変換テーブルを表 1 に示す。

2.2.3 病院コードから属性ベクトルへの変換

分析対象となる DPC データは複数の病院から収集されたため、提供元の病院がコードとして保存されている。病院の特性によって在院日数が変化すると考え、病院コードを属性ベクトルに変換した。病院コードを s とした場合の属性ベクトルへの変換テーブルを表 2 に示す。

2.2.4 入院目的から属性ベクトルへの変換

DPC データには患者が予定された入院であったか、緊急入院であったかがコードとして記録されている。入院目的を t とした場合の入院目的から属性ベクトルへの変換テーブルを表 3 に示す。

表 1 DPC コードから属性ベクトルへの変換テーブル

属性	条件	値	説明
v_2	$p = 01$	1.0	腹腔鏡下胃切除術
	$p \neq 01$	0.0	
v_3	$p = 01$	1.0	試験開腹術等
	$p \neq 03$	0.0	
v_4	$p = 04$	1.0	粘膜切除術
	$p \neq 04$	0.0	
v_5	$p = 99$	1.0	手術なし
	$p \neq 99$	0.0	
v_6	$q = 0$	1.0	処置なし
	$q \neq 0$	0.0	
v_7	$q = 1$	1.0	人工呼吸, 中心静脈注射
	$q \neq 1$	0.0	
v_8	$q = 3$	1.0	化学療法ありかつ 放射線治療なし
	$q \neq 3$	0.0	
v_9	$r = 0$	1.0	副傷病なし
	$r \neq 0$	0.0	
v_{10}	$r = 1$	1.0	副傷病あり
	$r \neq 1$	0.0	
v_{11}	$r = x$	1.0	副傷病名の項目なし
	$r \neq x$	0.0	

表 2 病院コードから属性ベクトルへの変換テーブル

属性	条件	値
v_{12}	$s = AU$	1.0
	$s \neq AU$	0.0
v_{13}	$s = J$	1.0
	$s \neq J$	0.0
v_{14}	$s = R$	1.0
	$s \neq R$	0.0
v_{15}	$s = S$	1.0
	$s \neq S$	0.0

表 3 入院目的から属性ベクトルへの変換テーブル

属性	条件	値	説明
v_{16}	$t = 1$	1.0	予定入院
	$t \neq 1$	0.0	
v_{17}	$t = 2$	1.0	救急医療入院
	$t \neq 2$	0.0	

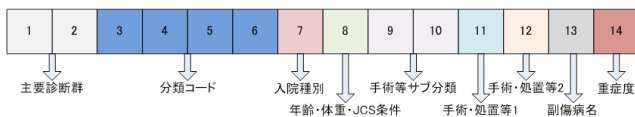


図 1 DPC データ各桁の意味

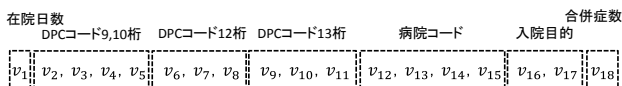


図 2 属性ベクトルの構成

2.3 自己組織化マップによる学習および在院日数予測

2.3.1 自己組織化マップによる DPC データの学習

実験対象となる DPC データ 116 件のうち、半分を SOM 学習用として無作為に非復元抽出した。その後、DPC データを属性ベクトルに変換し、総ユニット数 4,900 個（マップサイズ 70×70）、学習回数は 100,000 回、属性ベクトル数 58、属性ベクトル次元数 18 で SOM 学習を行った。

SOM は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。入力層と出力層の 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。

2.3.2 自己組織化マップによる在院日数予測

DPC データを学習した SOM の出力層に対して、学習に使用しなかった DPC データを属性ベクトルに変換した後投入し、在院日数予測を行った。

実際に予測を行う場合、在院日数の値は未定となる。そのため、 m_i を出力層のノードの重み、 v を投入データの重みとした場合における 1 次元目（在院日数）を抜いたユークリッド距離 $|m_i - v|$ を計算し、 $|m_i - v|$ が最短となる SOM 出力層上のノード m_i に投入データを配置する。この計算の概念を図 3 に示す。

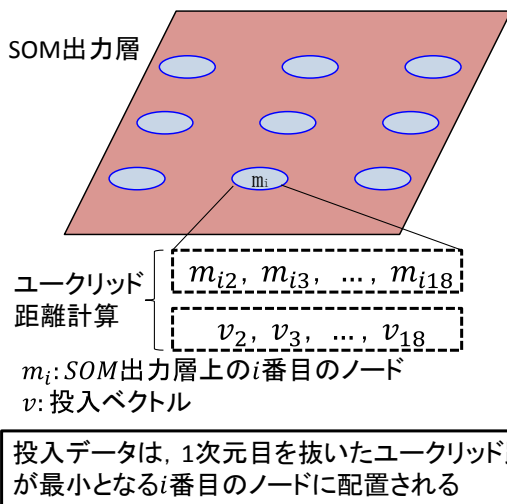


図 3 学習後の SOM マップへのデータ投入

在院日数予測値は、投入データの SOM 出力層上での配置先の重みを利用して計算する。 m_{i1} を投入データとのユークリッド距離が最小となるノードの 1 次元目の重みとし、この場合における在院日数予測値 a を次式で定義した。この式は、(1)の逆関数である。

$$a = m_{i1} \{ \max(x) - \min(x) \} + \min(x) \dots \dots \dots (2)$$

学習用・投入用データの抽出、SOM 学習、在院日数予測値の計算までの流れを 100 回行った。その結果、在院日数予測値と投入用データが本来持つ在院日数との差の絶対値の平均とその標準偏差は 5.14 ± 6.02 となった。また、予測値に対してマージンを設定し、予測値±マージンの範囲内に投入用データが本来持つ値が含まれた際に予測が成功と判断した場合のマージンと予測成功率のグラフを図 4 に示す。また、過去の研究³⁾との在院日数予測成功率の比較を表 4 に示す。

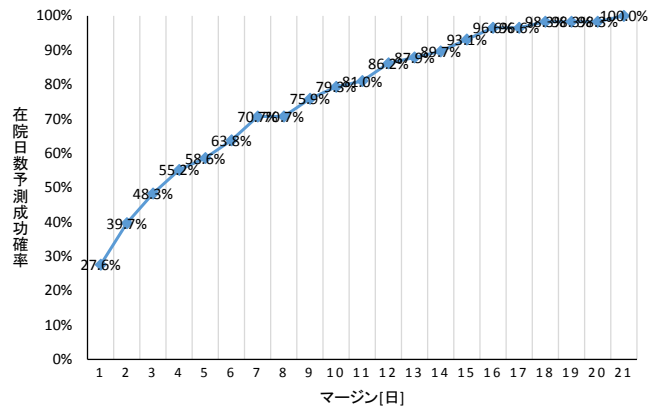


図 4 在院日数予測成功率

表 4 在院日数予測成功率の比較

		予測成功率	
		今回の手法	過去の手法
マージン[日]	3	48.8%	51.3%
	7	70.6%	69.4%
	14	89.7%	81.7%

2.4 考察

在院日数予測が低下させる要因を調べるため、在院日数予測実験で用いた属性ベクトルから在院日数属性を抜き、総ユニット数 4,900 個（マップサイズ 70×70）、学習回数は 100,000 回、属性ベクトル数 58、属性ベクトル次元数 18 で SOM 学習を行った。学習結果の SOM 出力層を図 5 に示す。図中のラベルは以下の意味を持つ。

在院日数、DPC コード、病院コード、入院目的、合併症数、データ番号

また、出力層は手術の種類別に着色した。

SOM 学習では、属性ベクトルがクラスタリングされ、類似度の高いベクトル同士が出力層上で近くに配置される。出力層上で同じ位置に配置されているにも関わらず、在院日数の値が大きく異なる属性ベクトルに注目し、その属性ベクトルの元となった DPC データを分析した。

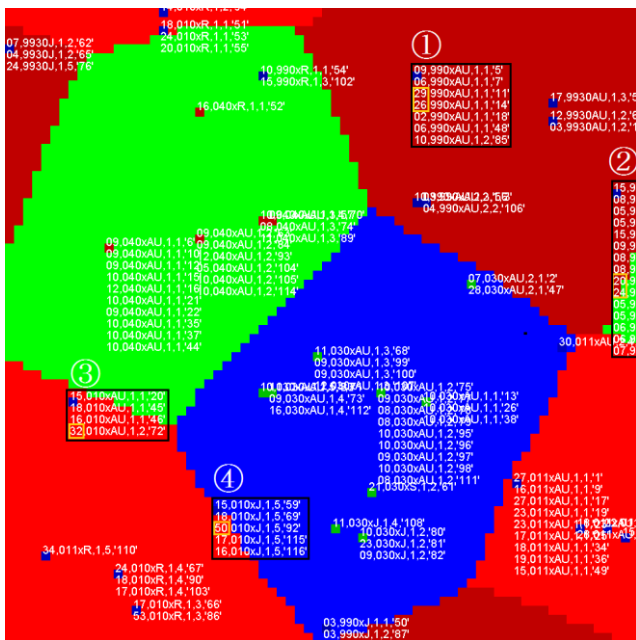


図 5 DPC データを学習した SOM 出力層

4 つ以上の属性ベクトルが集まり、かつ在院日数の値が大きく異なる属性ベクトルを含むノードを黒い四角で囲み、在院日数の値が大きいデータを黄色い四角で囲んだ。また、それぞれのノードに①～④の番号を割り振った。以下にそれぞれのノード上の在院日数が長いデータの特徴を示す。

- ① 診療科が外科、癌が再発、退院時紹介が有であった。
- ② 女性かつ退院時紹介が無であった。
- ③ 他のデータが腹腔鏡下胃切除術を行っているのに対して、腹腔鏡下胃全摘術を行っている。
- ④ 転科している。

①～④に挙げた要素を SOM 学習の属性ベクトルに含めることによって在院日数予測成功率が向上すると考えられる。

3. 在院日数予測システムへの応用

今回の実験結果を用いた応用システムとして、傷病名および処置情報から在院日数予測を行う Web アプリケーションの開発を進めている。DPC データを用いた在院日数予測は DPC/PDPS に参加する全ての病院で利用可能である。また、在院日数予測は病院運営、入院費用の算出に利用できるため、DPC データを用いた在院日数予測システムの有用性は高いと考える。システム構成を図 6 に、システムの動作例を図 7 に示す。

4. おわりに

本校では、22,001 件の DPC データのうち、ICD-10 コードが C162 (胃体部癌) である 116 件を対象に、「在院日数」、

「DPC コード」、「合併症数」、「病院コード」、「入院区分」を自己組織化マップ学習用の属性ベクトルに変換し、在院日数予測を行った。その結果、在院日数予測値と投入用データが本来持つ在院日数との差の絶対値の平均とその標準偏差は 5.14 ± 6.02 となった。また、在院日数予測を行う Web アプリケーション開発について検討した。

今後の課題は、在院日数に影響を与える変数の特定と、様々な属性ベクトルの構成で在院日数予測を行うことである。また、在院日数予測を行う Web アプリケーションに様々な機能を追加していきたい。

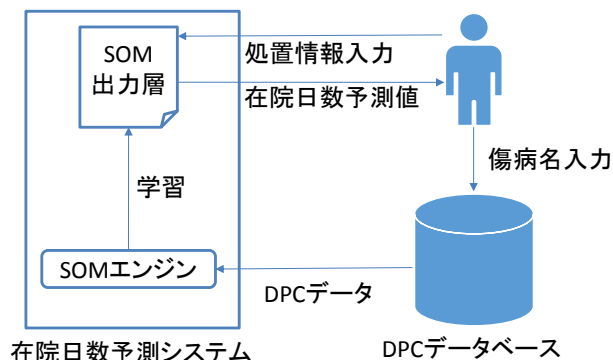


図 6 在院日数予測を行う Web アプリケーション

病名, ICDコードを入力:

胃体部癌(スキルス胃癌)の検索結果

m_disease	胃進行癌,胃体部癌,胃体部癌(スキルス胃癌),胃体部癌の疑い,胃体部癌の
length	116
lndays_mean	13.7759820689655
lndays_max	53
lndays_min	2
lndays_skewness	1.77921332771162
lndays_kurtosis	4.64238385075033

手術
 手術なし 試験開腹術 胆嚢摘出術 胆嚢嚢性腫瘍手術 膵頭部腫瘍切除術

手術・処置等2
 なし 人工呼吸, 中心静脈注射 化学療法ありかつ放射線療法なし

副傷病名 あり なし

合併症数

病院コード
 AU J R S

入院目的
 予定された入院 救急入院

予測在院日数: 8日

図 7 在院日数予測 Web アプリケーション動作例

5. 参考文献

- 厚生労働省保険局医療課：平成 26 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/file/06-Seisakujouhou-12400000-Hokenkyoku/0000039616.pdf>> (参照 2014-05-13)
- 厚生労働省保険局医療課：DPC データの提供について, 厚生労働省 (オンライン), 入手先 <<http://www.mhlw.go.jp/stf/shingi/2r98520000033u5s-at/2r98520000033ubu.pdf>> (参照 2014-05-13)
- 仲濱正大, 納富一宏, 斎藤恵一, 外山比南子：自己組織化マップを用いた在院日数予測と評価, 第 26 回バイオメディカル・フュージ・システム学会年次大会 講演論文集, pp85-88, (2013)