

クラスタリングによる看護クリニカルパス生成

Clinical Pathway Generation based on Cluster Analysis

津本周作 *¹
Shusaku Tsumoto

木村知広 *³
Tomohirno Kimura

平野章二 *¹
Shoji Hirano

岩田春子 *²
Haruko Iwata

*¹ 島根大学医学部医療情報学

Department of Medical Informatics, Faculty of Medicine, Shimane University

*² 島根大学医学部附属病院入退院管理センター

Center for Bed Control, Shimane University Hospital

*³ 島根大学医学部医療サービス課

Medical Services Division, Faculty of Medicine, Shimane University

This paper proposes the following two-fold clinical care generation method. First, the system proposes how clinical cases with the same DPC code are characterized by mixture model clustering, and construct classification model by the analysis of discharge summaries. Then, cases are classified by using the classification model and a clinical pathway is generated for each new class. The proposed method was evaluated on the datasets extracted hospital information system, whose results show that plausible clinical pathways were obtained, compared with previously introduced methods.

1. はじめに

著者らは、これまで看護オーダーの実施歴から、クラスタリングを用いて看護クリニカルパスを生成する仕組みについて提案してきた [岩田 11, 岩田 12a, 岩田 12b, 岩田 15, Iwata 15, Tsumoto 15]。この中で、同一 DPC コード内の症例を用いてパスの評価を行ったが、眼科領域等外科領域ではこれまで看護師が作成したパスに類似のものが生成できたのに対し、脳梗塞・肺がんといった症例では複雑なパスになった [岩田 12b]。これは DPC コード内に複数の病態が含まれ、それぞれの病態によって、入院日数が異なってくる場合があり、コーディングと臨床の病態の間との乖離がそのままパスに投影された形となってしまうことによる。

今回、この問題を克服するべく、同一 DPC コードの症例に対して、入院日数を指標として、混合分布モデル (いわゆる EM クラスタリング) [McLachlan 00] を適用し、DPC コード内の症例の分類を行った上で、退院時要約を用いて特徴付けし、分類モデルを生成、生成したモデルを用いて、改めて当該 DPC コード内の症例をサブクラスに分類、分類後に、看護クリニカルパスの生成法を適用する形での拡張を試みた。具体的には、まず症例の選別を自動化するために、同一 DPC コード内での疾患の類別を入院日数を指標として、混合分布モデルを用いて行い、それぞれのクラスへの帰属確率を推定する。次に、帰属確率を用いて、DPC コード内の症例をサブクラスに分類し、それぞれのサブクラスの看護実行歴のデータに分割、それぞれに対して、これまで提案してきたクラスタリングに基づく看護クリニカルパスの生成法を適用した。本研究では、本プロセスの性能について 2015 年度の退院時要約のうち、症例数の多い 10 コードについて評価を試み、それぞれのサブクラスで生成されたパスは、従来法で得られたパスよりも簡易なものが得られた。

2. これまでの方法

著者らは電子的に蓄積された看護オーダー実行歴を頻度情報として用いて、頻度情報の距離の差に基づくクラスタ分析を適用し、データ解析でパスの作成支援を提唱してきた (図 1)。こ

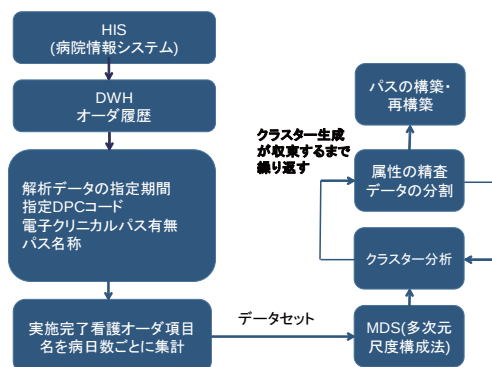


図 1: クラスタリング+決定木

の方法では、まず病院情報システムから DPC を指定し、与えられた DPC をもった症例について、入院日を基準として、各症例の看護オーダー数を算出し、行を看護オーダー、列を入院日数とした表形式データを入力データとして出力する。この入力データを元に、各オーダー間の距離を算出し、距離行列を計算する。この距離行列を用いて、多次元尺度構成法、階層クラスタリング [Everitt 96] を適用し、看護オーダーの類型化を試みる。この段階では、病日数を属性として用いるが、これらの属性のクラスで与えられたクラス分類の寄与度を考えるために、決定木で使用する Information Gain を用いて、各属性の寄与度を求める。これによって、日数が Information Gain の大きさによってグループ分けが可能になる。この寄与度によるグループ分けを用いて、元のデータを分割する。この分割した各データについて、距離行列の計算、クラスタリングを行う。クラスタ分析が収束するまで、この手続きを繰り返し、最終的

連絡先: 津本周作, 島根大学医学部医療情報学講座,
出雲市塩冶町 89-1, 0853-20-2171, 0853-20-2170,
tsumoto@med.shimane-u.ac.jp

に、分割した入院日数での類型化した看護オーダーを統して、パスを構築する。しかし、この手法では、指標の値によって、属性のグループ分けは複雑になる可能性があるため、これを避けるため、属性の選択をクラスタリングによる属性の類型化による方法、Dual-Clusteringを開発した。(図 2)

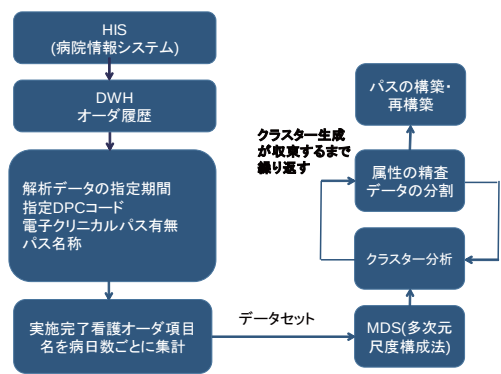


図 2: Dual Clustering

3. 混合正規分布モデルによる類型化

3.1 混合正規分布

DPC は診療に医療資源を一番投入した疾患によって、コーディングされる。したがって、例えば、治療薬は同じであれば、違う病態を持ちうる疾患群でも、同じコードに属する可能性がある。したがって、DPC をキーとして検索し、それによって得られたデータについての解析を行う場合、複数の病態をできるだけ分離することが必要であると考えられる。同一コード内に複数の病態が含まれる時、各病態が独自の時間発展を遂げると考えると、コードでの入院日数の分布は複数の病態を反映した混合分布となると考えられる。もし、入院日数の分布が正規分布で表現できるとすれば、これらを分離する方法として、混合分布モデルが適用できる。もともと本手法は判別分析における方法として提案されたが、現在、EM アルゴリズムを用いたモデルベースのクラスタリングの一つと位置づけられている。

混合正規分布では、パラメータが平均と共分散行列で x が出現確率は:

$$p(x, \theta) = \sum_{m=1}^M \pi_m \phi(x, \mu_m, \Sigma_m)$$

で与えられる。ここで、 M はクラスタ数、 μ_m, Σ_m はそれぞれクラスタ m の平均と共分散行列を示し、正規分布の線型結合で表現される。

クラスタ m が現れる確率 (事後確率) が

$$p(m|x, \theta) = \frac{p(x, m; \theta)}{p(x; \theta)} = \frac{\pi_m \phi(x; \mu_m, \Sigma_m)}{\sum_{m=1}^M \pi_m \phi(x; \mu_m, \Sigma_m)}$$

によって当たられる。この事後確率を最大化するパラメータ (μ_m, Σ_m) が EM アルゴリズムである [McLachlan 00]。

3.2 クラスタ数の決定

クラスタ数別のあてはめについては、AIC の値を用いて評価する。

一つのクラスタで 3 つのパラメータ (線形結合の係数、平均、標準偏差) が追加され、線形結合の係数の和が 1 であるこ

とから、線型結合の係数の自由度は 1 減ずる。したがって、この場合の AIC は:クラスタ数を k として、

$$AIC = -2 * (\text{対数尤度}) + 2 * (3k - 1)$$

で求められる。例えば、2015 年度に入院した脳梗塞患者について、クラスタ数を以下のように指定し、AIC を計算、表 1 のような結果を得た。、2 つの分布の混合とした場合のあてはめが最もよかった。(以下、入院日数の短いクラスをクラス 1、より長いものをクラス 2 と呼ぶことにする。)

表 1: クラスタ数と対数尤度、AIC

クラスタ数	対数尤度	AIC
2	-296.1122	600.2244
3	-293.6723	601.3446
4	-293.6574	607.3148
5	-291.5038	609.0076
6	-288.7785	609.5573
7	-287.4351	612.8702
8	-288.7281	621.4562
9	-286.2126	622.4252

3.3 各症例の分類

EM-clustering では、各例については、あるクラスタに属する確率が算出される。しかし、この場合、正確に症例を排反的に分離できていない。このため、もともとのデータをクラスタ数に分割するため、以下のように帰属群を判定する。

1. クラスタ数 k を与え、与えられた DPC に属する症例の入院期間のデータに対して、混合正規分布モデルによる EM クラスタリングを適用する。
2. 入院日数 (単位:日) に対する各症例のクラスタへの帰属確率を算出する。
3. 帰属確率の値が大きい 2 つのクラスタが算出した確率が等しくなる日数、なければ、前後の日数の平均を 2 つのクラスタの分界点と定義。
4. 分界点の前後によって、症例の分布クラスタを決定する。

4. パス生成プロセス

以上の方法と、Dual-clustering によるパス生成を組み合わせ、パス生成方法を図 3 のように拡張する。プロセスは R3.4.1 のスクリプトとして記述し、混合正規分布モデルには mixtools[Benaglia 09] を、階層クラスタリングには R 上の hclust を用い、距離行列はユークリッド距離を用いて計算した。

5. 結果

5.1 混合正規分布

島根大学医学部附属病院の 2015 年度に入院した患者について、DPC 使用頻度上位 20 位までの疾患 (表 2) を抽出し、入院日数について、EM-clustering のあてはめを試みた。

5.2 脳梗塞のパス生成

図 4 に脳梗塞のデータについて、クラスタ数を 2 とした場合の混合正規分布によるあてはめ結果を示した。

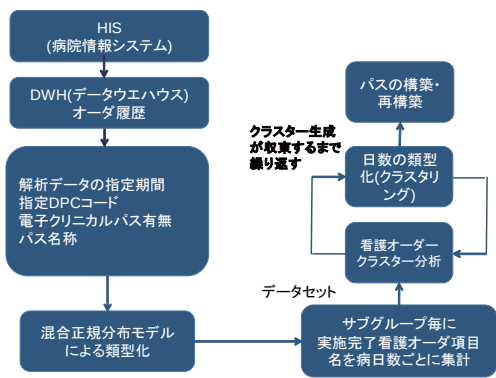


図 3: EM clustering + Dual Clustering

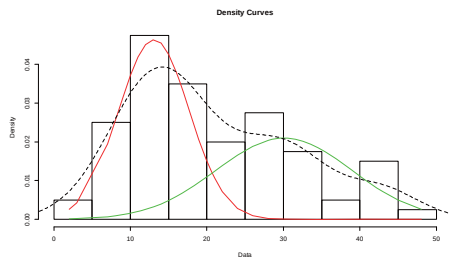


図 4: 混合正規分布モデルによるあてまめ

5.3 パスの生成

図 3 で示したパス生成プロセスを脳梗塞のデータに以下のよう適用した。

1. 病院情報システムから抽出した脳梗塞 (DPC コード:010060x099030x) の看護オーダー集計に Dual-clustering を適用し、看護オーダーは概ね 3~5 つに分割された。病日に関しても 3 つのクラスタが得られた。ここから得られるパスは図 5 の通りである。
2. 混合正規分布の推定 (EM clustering) によって、元の症例を与えられたクラスタ数での確率モデルを生成する。この場合、AIC の結果より、分布は二峰性 (クラスタ:2) が最もあてはめがよかった。Group1 は入院日数の短い症例、Group2 は長い症例からなる。
3. 分界点を求め、各症例を属すべきクラスタに配置する。
4. 配置した各クラスタを subgroup とし、subgroup に対して、Dual clustering でパス生成を行った。この Dual clustering は再帰的にデータを分割することが可能であるが、図 6 および 7 に、それぞれ 1 回再帰的に適用した際の 2 つのクラスタで生成されたパスを示した (パスに表示されたものは、最も頻度の高いグループのみを示した。)

3 図の比較から、より病態を反映したパスが生成できたと考えられる。

6. おわりに

本論文では、DPC コード内での病態の均一化を混合正規分布モデルにより、同一 DPC コードの症例を与えられた数のサ

表 2: データと混合分布モデルのあてはめの結果

DPC 名称	件数	正規性	クラスタ数
白内障、水晶体の疾患 手術手術あり 重症度等片眼	445		1
白内障、水晶体の疾患 手術手術あり 重症度等両眼	152		1
2 型糖尿病 (糖尿病性ケトアシドーシスを除く。)	145	○	3
肺の悪性腫瘍 手術手術あり 処置等 2 なし	131	○	1
子宮頸・体部の悪性腫瘍 手術手術なし 処置等 2 4 あり 副傷病名なし	121		1
肺の悪性腫瘍 手術手術なし 処置等 1 あり 処置等 2 なし 副傷病名なし	120		5
子宮の良性腫瘍 手術腹腔鏡下腔式子宮全摘術等	111		1
肺の悪性腫瘍 手術手術なし 処置等 1 なし 処置等 2 4 あり	110		1
妊娠期間短縮、低出産体重に関連する障害 (出生時体重 2500g 以上) 手術手術なし 処置等 2 なし 副傷病名なし	110		1
肘、膝の外傷 (スポーツ障害等を含む。) 手術縫合術等	99	○	2
全身性臓器障害を伴う自己免疫性疾患 手術手術なし 処置等 2 なし	96		
非ホジキンリンパ腫 手術手術なし 処置等 2 4 あり 副傷病名なし	94		5
肺炎、急性気管支炎、急性細気管支炎 (15 歳以上) 手術手術なし 処置等 2 なし	86		1
肺の悪性腫瘍 手術手術なし 処置等 1 なし 処置等 2 なし	85		4
慢性腎臓病・慢性間質性腎炎・慢性腎不全手術手術なし 処置等 1 なし 処置等 2 なし 副傷病名なし	83	○	3
肝・肝内胆管の悪性腫瘍 (続発性を含む。) 手術その他の手術あり 処置等 2 なし	82		1
胆管 (肝内外) 結石、胆管炎 手術限局性腹腔臓器手術等 処置等 2 なし 副傷病名なし	82		1
脳梗塞 (JCS10 未満) 手術手術なし 処置等 1 なし 処置等 2 3 あり 副傷病名なし	80	○	2
網膜剥離 手術手術あり 重症度等片眼	75		2
胎児及び胎児付属物の異常 手術子宮全摘術等	73		1

	1日目	2日目	3-7日目	8-16日目	9-16日目	17-29日目
(No)VS測定(※)項目	○	○	○			
(No)体温・脈拍測定	○	○	○			
(No)呼吸数	○	○	○			
ME機器作動状態の確認	○	○	○	○		
SPO2			○	○	○	
しびれ			○			
その他症状			○	○	○	
体位の調整			○			
体位変換			○	○	○	○
体温・脈拍測定			○	○	○	○
口腔清拭			○			○
呼吸			○	○	○	
嘔気・嘔吐	○	○	○	○	○	○
室内環境調整			○			
座位保持介助	○	○	○	○	○	○
嚥下			○	○	○	○
情報提供			○	○	○	○
排便チェック			○	○	○	○
排便回数			○	○	○	○
更衣			○			
歯磨き介助			○	○	○	○
洗面			○	○	○	○
消化・腹部症状			○	○	○	
留置経管管理	○	○	○	○	○	
疼痛			○	○	○	○
関節状況			○	○	○	○
褥瘡			○	○	○	
神経	○	○	○	○	○	○
移乗介助(車椅子)			○			
洗面洗浄			○			○
歯磨き指導			○			
自己排泄防止??			○			
血圧測定			○	○	○	○
起き上がり介助	○	○	○	○	○	○
転倒防止??			○	○	○	○
転落防止??			○	○	○	○
運動	○	○	○	○	○	○
陰部洗浄			○			

図 5: 元データから生成されたパス

ブクラスに類別化することで実現し、これによるデータの分割にクリニカルパス作成の方法を適用した。これらのクラスタ

	1日目	2日目	3-9日目	10-14日目	15日目	16日目	17日目
(Na)VS測定(※項目)	○	○	○				
(Na)体温・脈拍測定	○	○	○				
(Na)呼吸数	○	○	○				
ME機器作動状態の確認	○	○	○				
SPO2			○	○			
その他症状			○	○			
体温・脈拍測定			○	○	○	○	○
呼吸			○	○			
嘔気・嘔吐	○	○	○	○	○	○	○
座位保持介助	○	○	○	○			
循環			○	○			
排尿チェック			○	○			
排泄回数			○	○			
姿勢き介助			○	○			
洗面			○	○			
消化・腹部症状			○	○			
留置時管理	○	○	○				
疼痛			○	○			
躁動状態			○				
睡眠			○	○			
神経	○	○	○	○	○	○	○
血圧測定	○	○	○	○	○	○	○
起き上がり介助	○	○	○	○	○	○	○
転倒防止?	○	○	○	○	○	○	○
転落防止?	○	○	○	○	○	○	○
運動	○	○	○	○	○	○	○
食事摂取量	○	○	○	○	○	○	○

図 6: グループ 1 のデータから生成されたパス

	1日目	2日目	3-9日目	10-14日目	15-24日目	25-34日目	35-42日目	43-48日目
(Na)VS測定(※項目)	○	○	○	○				
(Na)体温・脈拍測定	○	○	○	○				
(Na)呼吸数	○	○	○	○				
ME機器作動状態の確認	○	○	○	○				
SPO2			○	○				
その他症状			○	○				
体位の変更			○	○	○	○		
体位変換			○	○	○	○		
体温・脈拍測定			○	○	○	○		
口腔清拭			○	○	○	○		
呼吸			○	○				
嘔気・嘔吐			○	○	○	○		
室内環境調整			○	○				
座位保持介助	○	○	○	○	○			
循環			○	○	○			
排尿チェック			○	○	○	○		
排泄回数			○	○	○	○		
日常生活習慣の獲得を促す??			○	○			○	
更衣			○	○				
座浴き介助			○	○	○	○		
洗面			○	○	○	○		
消化・腹部症状			○	○				
留置時管理			○	○				
疼痛			○	○	○	○		
躁動状態			○	○	○	○		
睡眠			○	○				
神経			○	○	○	○		
排尿介助(座椅子)			○	○	○	○		
経口吸引			○	○		○		
褥瘡洗浄			○	○	○	○		
褥瘡観察			○	○		○	○	
自己排泄防止??			○	○	○	○		
血圧測定			○	○	○	○		
起き上がり介助	○	○	○	○	○	○		
転倒防止??			○	○	○	○	○	
転落防止??			○	○	○	○	○	
運動			○	○	○	○		
排泄洗浄			○	○				
食事介助(ハイス)			○	○	○	○		
食事介助(ロイス)			○	○	○	○		
食事摂取量	○	○	○	○	○	○	○	
トイレ交換			○	○	○	○		
ハット交換			○	○	○	○		

図 7: グループ 2 のデータから生成されたパス

の意味については, [津本 16] で [Tsumoto 17] で提案した方法を用いて, 退院時要約から, サブクラスへの分類に関するパターンを抽出することを試み, Random Forest[Breiman 01], Deep Learning[Goodfellow 16] で分類器を構築できた., 今回は, 他の類別化の方法との比較検討ができていないが, 今後, 退院時要約内の情報を用いて, 入院日数という情報にたよることなく, 症例をサブグループに類別化する方法を検討していく予定である。

謝辞

本研究は日本学術振興会, 科学研究費基盤 (B) 15H2750 の補助を受けた。

参考文献

- [Benaglia 09] Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D.: mixtools: An R Package for Analyzing Finite Mixture Models, *Journal of Statistical Software*, Vol. 32, No. 6, pp. 1–29 (2009)
- [Breiman 01] Breiman, L.: Random Forests, *Machine Learning*, Vol. 45, No. 1, pp. 5–32 (2001)
- [Everitt 96] Everitt, B.: *Cluster Analysis*, John Wiley & Son, London, 3rd edition (1996)
- [Goodfellow 16] Goodfellow, I., Bengio, Y., and Courville, A.: *Deep Learning*, MIT Press (2016), <http://www.deeplearningbook.org>
- [Iwata 15] Iwata, H., Hirano, S., and Tsumoto, S.: Maintenance and Discovery of Domain Knowledge for Nursing Care using Data in Hospital Information System, *Fun-dam. Inform.*, Vol. 137, No. 2, pp. 237–252 (2015)
- [Kim 09] Kim, J.-H.: Estimating Classification Error Rate: Repeated Cross-validation, Repeated Hold-out and Bootstrap, *Comput. Stat. Data Anal.*, Vol. 53, No. 11, pp. 3735–3745 (2009)
- [McLachlan 00] McLachlan, G. J. and Peel, D.: *Finite Mixture Models*, Wiley, New York (2000)
- [Tsumoto 15] Tsumoto, Y., Iwata, H., Hirano, S., and Tsumoto, S.: Construction of Clinical Pathway Using Dual Clustering, *Neuroscience and Biomedical Engineering*, Vol. 3, (2015)
- [Tsumoto 17] Tsumoto, S., Kimura, T., Iwata, H., and Hirano, S.: Construction of Discharge Summaries Classifier, in *2017 IEEE International Conference on Healthcare Informatics, ICHI 2017, Park City, UT, USA, August 23–26, 2017*, pp. 74–82, IEEE (2017)
- [岩田 11] 岩田春子, 津本周作: 病院情報システムに蓄積された看護行為に関する時系列マイニング-看護パス作成支援に向けて-, 第 31 回医療情報学会連合大会 (2011)
- [岩田 12a] 岩田春子, 津本周作: 病院情報システムに蓄積された看護行為に関する時系列マイニング-看護パス作成支援に向けて-(その 2), 第 16 回日本医療情報学会春季学術大会 (2012)
- [岩田 12b] 岩田春子, 津本周作: 病院情報システムに蓄積された看護行為に関する時系列マイニング-看護パス作成支援に向けて-(その 3), 第 32 回医療情報学会連合大会 (2012)
- [岩田 15] 岩田春子, 津本周作: 病院情報システムに蓄積された看護行為に関する時系列マイニング-看護パス作成支援に向けて-(その 4), 第 35 回医療情報学会連合大会 (2015)
- [津本 16] 津本周作, 平野章二, 岩田春子, 木村知広: 同一 DPC コード内症例の類型化のための分類知識の獲得, 第 3 回医用人工知能研究会 (SIG-AIMED-003-06) (2016)