一般口演 | 医療アセスメント

一般口演2
医療アセスメント

2019年11月22日(金) 09:00 ～ 11:00 E会場 (国際会議場 3階中会議室301)

[2-E-1-05] トピックモデルを用いた病院機能の分析手法の提案

○田中 昌昭[1] （1. 川崎医療福祉大学）

キーワード：Hospital Function, Topic Model, Latent Dirichlet Allocation, Natural Language Processing, Healthcare Professions

医療機関が有する医療機能は，医療政策や病院経営が動機付けとなり，往々にして保険点数に誘導されて形成されていく。しかしながら，病院機能には必ずしも明確な定義はなく，あるべき機能が実現されているか否かを客観的に評価する手法は確立されていない。本研究は，入手可能な公表データを用いて病院機能を分析する手法を提案し，実態の把握と可視化を目的とする。
テキストマイニング分野で利用されるトピックモデルにおいて，文書を病院に，文書中に現れる単語とその頻度を病院で働く職員とその人数に置き換えて，医療機関に潜む隠れたトピックを推定した。分析対象データとしてH29年度病床機能報告データを用いた。解析に当たってトピック数は20とし，各トピックに含まれる職員の割合と病院ごとのトピック含有率を求めた。さらに得られたトピック含有率を用いて k平均法によって病院を20のクラスターに分類した。
抽出されたトピックを見ると，いずれのトピックもそのトピックを特徴づけるいくつかの職種によって構成されていた。また，病院のトピック含有率を見ると，クラスターごとに特徴的に表れるトピックが見られた。たとえば常勤の病棟医師が支配的なトピックを多く含むクラスターには特定機能病院の承認を受けた病院が他に比べて多く現われ，平均病床数も600床以上と大規模病院が多かった。
ドナベディアンは構造，過程，結果の3つの側面から医療の質を評価できるとした。職種ごとの医療従事者の数を構造の指標の一つと考えるなら，本手法で得られたトピックは医療機関が提供する医療の質，ひいては病院機能を表していると見ることができる。本研究は，テキストマイニングで培われた手法によって病院機能を明らかにできることを示唆するものである。

©一般社団法人 日本医療情報学会

トピックモデルを用いた病院機能の分析手法の提案

田中昌昭[*1]
*1 川崎医療福祉大学医療福祉マネジメント学部医療情報学科

Proposal of Analysis Method for Hospital Functions Using Topic Models

Masaaki Tanaka[*1]
*1 Department of Health Informatics, Faculty of Health and Welfare Services Administration
Kawasaki University of Medical Welfare

The medical functions possessed by medical institutions are formed by medical policies and hospital management, and are often guided by the medical systems such as health insurance and reimbursement systems. However, there is not always a precise definition of hospital functions, and no method has been established for objectively evaluating whether or not the desired function has been realized. So, this study proposes a method to analyze hospital functions using available published data, and aims to grasp and visualize the actual situation. By using the topic model cultivated in the field of natural language processing, the functions hidden in the medical institution were estimated by replacing the documents with the hospitals, the words appearing in the document and their frequencies with the healthcare professions and their number. The 2017 fiscal year clinical bed function report data was used as analysis target data. In the analysis, the number of topics was set to 40, and the distribution of healthcare professions making up each topic and the distribution of topic for each hospital were obtained. Furthermore, hospitals were classified into 20 clusters by k-means using their topic distribution as features. Looking at the identified topics, each topic was composed of several healthcare professions that characterized the topic. In addition, looking at the topic distribution of hospitals, there were topics that appeared characteristically for each cluster. For example, in a cluster that includes many topics that are dominated by full-time doctors, there are more hospitals approved as advanced treatment hospitals than others, and there are many large hospitals with an average number of beds of 500 or more. Donabedian argued that medical quality can be assessed from three aspects: structure, process, and outcome. If the number of healthcare professions for each occupation is considered as one of the structural indices, the topic obtained by this method can be viewed as representing the quality of medical care, and thus the hospital function provided by medical institutions. This study suggests that the hospital function can be clarified by the method cultivated by natural language processing.

Keywords: Hospital Function, Topic Model, Latent Dirichlet Allocation, Natural Language Processing, Healthcare Professions

1. 緒言

限られた医療資源を有効に活用するためには医療機関が機能分化と連携を進め，地域で完結する医療提供体制を構築することが望まれる。機能分化の促進には，現状において各医療機関が担っている役割を明確化し，地域における需給体制の把握，医療資源の過不足の定量化・可視化が不可欠である。このような背景から平成 26 年の医療法改正において「病床機能報告制度」が導入され，医療機関が有する病床の医療機能の現状と今後の方向性についての報告が義務付けられた [1)]。

一方，一言で医療機能と言っても明確な定義があるわけではなく，視点によって様々な定義が可能であり，目標に掲げた機能分化がどの程度達成されたかを定量的に示すのは容易ではない。また，医療機能は施設基準と密接な関係があるが，施設基準は医療機関による届出制となっている。たとえば在宅療養支援病院の施設基準の届出が承認されたら在宅患者診療・指導料や在宅療養指導管理料などの各種加算が算定できる。これは病院機能が機能分化の必要性からではなく診療報酬によって誘導される可能性を示唆している。しかしながら，多かれ少なかれ病院機能の形成にはそのような側面が関与しており，地域医療構想調整会議などの場において医療機関の思惑がせめぎあうことは否定できない。

とはいえ，近年では DPC 公開データ，病床機能報告公表データ，NDB オープンデータなど，病院が有する医療機能を客観的に裏付けるデータが利用できるようになった。そのため，地域において各医療機関が担う役割をデータに基づいて評価することが可能になった。これは医療機関の機能分化と連携の進捗を定量化・可視化する上で，また，医療機関自らが病床機能報告や施設基準等で主張する医療機能と実態との齟齬を明らかにする上で画期的な前進である。

本研究では，医療資源の中でも医療従事者など人的資源に焦点をあてて病院機能を分析する手法を提案する。具体的には，自然言語処理分野で培われたトピックモデル [2)]を用いて病床機能報告公表データの病院ごとの職員数から医療従事者の組み合わせパターンを特定し，そのパターンの構成比率によって病院機能を定義した。本研究の目的は，提案手法によって病院が担う機能をどの程度識別できるか検証することにある。使用するデータが職員数と限定的ではあるが，公表されている客観的なデータから医療提供体制の現状を把握する分析手法を開発することは，今後の地域医療構想の方向性を考える上でも意義深い。

2. トピックモデル

2.1 トピックモデルの概要

トピックモデルは，文書に隠されたトピックを推定する教師なし機械学習法である。トピックモデルでは，文書はいくつかのトピックから構成され，その各々のトピックからある一定の確率で単語が生成されると考える。たとえば「AI による画像診断」について書かれた文書は，技術と医学のトピックをある一定の割合で含み，「技術」というトピックからある確率で「機械学習」や「ディープラーニング」といった単語が出現し，同様に「医学」というトピックからある確率で「CT」や「腫瘤」といった単語が出現すると考える。このような考え方を用いて大量の文書から各文書のトピック分布と各トピックの単語分布を学習するのがトピックモデルである。なかでも，トピック分布の事前分布にディリクレ分布を仮定して事後分布をベイズ推定する手法を潜在ディリクレ配分モデル（LDA, Latent Dirichlet allocation）という[2)]。本研究では，この LDA を用いて，文書を病院に，文書中に現れる単語とその出現頻度を病院で働く医療従事者とその人数に置き換えて病院に潜むトピック（医療従事者の組み合わせパターン）を特定し，そのトピック分布を用いて病院の機能を推定する。

2.2 先行研究

トピックモデルを自然言語処理以外の目的で使用する研究は数多く報告されているが，ここでは医療分野に限定して紹介する。Hatakeyama 等は電子請求データ（DPC データ）を文書，診療行為を単語，疾患をトピックとみなして LDA の拡張版である Labeled LDA を用いて最も医療資源を投入した疾患を推定した[3)]。得られた結果を従来の方法（サポートベクタマシン，ナイーブベイズ）と比較したところ，教師データを使っていないにも拘わらず提案手法の推定結果は従来手法より正確であったと報告している。Bhattacharya 等は関連する症状のパターンを明らかにすることを目的として電子カルテに記録された患者のデータにトピックモデルを適用した[4)]。彼らは患者ファイルを文書，患者ファイルに記載されている SNOMED-CT でコーディングされた症状を単語とみなして LDA モデルを構築し，ある特定のトピックに高い確率で現れる症状を関連する症状と仮定した。得られた結果は医学関連文献によって裏付けることができたと報告している。そのほか，トピックモデルを使って入院患者のオーダーパターンを予測し，通常のオーダーセットと比較した研究[5)]や大量の保険請求データに拡張版トピックモデルを適用し，疾患と処方薬の関係から処方パターンを特定した研究[6)]などがある。いずれの先行研究も患者のデータを文書とみなし，そのデータに含まれる診療行為や症状などを単語とみなし，トピックモデルを適用して隠れたトピック（疾患やオーダーパターン，処方パターン）を特定するという点において共通している。これに対して本研究では患者ではなく病院を文書とみなし，病院の医療従事者およびその人数を単語およびその出現頻度とみなしてトピックモデルを構築している点が異なる。

3. 材料と方法

3.1 材料

平成 29 年度病床機能報告公表データ[1)]の施設票から「退院調整部門に勤務する職員数」と「職員数」を取り出して分析対象データとした。「退院調整部門に勤務する職員数」は，医師，看護職員，MSW，事務員，その他の 5 職種が各々専従と専任に分けて集計されている。また，「MSW」については「MSW のうち，社会福祉士の資格を有する者」が再掲として集計されているので「MSW」から再掲分を差し引いて社会福祉士保有者とそうでない者に分けた。「職員数」は，「施設全体の職員数」，「手術室の職員数」，「外来部門の職員数」，「その他の部門の職員数」の 4 つのカテゴリに分けて集計されているので「施設全体の職員数」から「手術室の職員数」，「外来部門の職員数」，「その他の部門の職員数」を差し引いて「病棟部門の職員数」を求めた。ここで，「その他の部門」は，透析室，外来化学療法室，放射線照射外来室，退院調整部門，薬剤部門，リハビリ部門，訪問看護部門，医事部門，管理部門，健診部門，一般病床・療養病床以外の病床等が該当する。こうして病棟，手術室，外来，その他の 4 部門別に医師，歯科医師，看護師，准看護師，看護補助者，助産師，理学療法士，作業療法士，言語聴覚士，薬剤師，診療放射線技師，臨床検査技師，臨床工学技士，管理栄養士の 14 職種について各々常勤と非常勤の職員数を求めた。ただし，4 つの部門すべてにこれら 14 職種が集計されているわけではない。医師，歯科医師，診療放射線技師，臨床検査技師は「施設全体の職員数」にのみ集計されている。したがって集計単位は全部で 100 種類になる。以後，便宜上，この集計単位を医療従事者と呼ぶことにする。これら医療従事者をトピックモデルにおける単語とみなし，その人数を単語の出現頻度とみなした。なお，1 つでも未報告のデータがある場合は，その病院は分析の対象から除外した。

得られたトピックモデルを評価するために，各病院の属性として施設票から「DPC 群の種類」，「特定機能病院の承認の有無」，「地域医療支援病院の承認の有無」，「総合入院体制加算の届出の有無」，「在宅療養支援病院の届出の有無」，「在宅療養後方支援病院の届出の有無」，「三次救急医療施設の認定の有無」を，病棟票から「稼動ベッド数」，「医療機能」を取得した。なお，「医療機能」は施設ごとではなく病棟ごとに高度急性期，急性期，慢性期，回復期のいずれかが設定されているが，今回の解析は病棟単位ではなく病院単位で行うため病院ごとに「医療機能」を設定する必要がある。そこで，病院内で医療機能ごとに病棟数を集計し，最も病棟数が多かった医療機能をその病院の医療機能とした。

3.2 分析方法

前節で述べた医療従事者を列，病院を行，行列の要素に医療従事者の数を設定した「病院×医療従事者行列」を入力として R（バージョン 3.5.1）の topicmodels パッケージ（バージョン 0.2-7）を用いて LDA モデルを構築した。次いで，構築したモデルからトピック分布 $\boldsymbol{\theta}_d = (\theta_{d1}, \cdots, \theta_{dK})$ と医療従事者分布 $\boldsymbol{\phi}_k = (\phi_{k1}, \cdots, \phi_{kv})$ を求めた。ここで，θ_{dk} は病院 d にトピック k が割り当てられる確率で，$\theta_{dk} \geq 0$，$\sum_{k=1}^{K} \theta_{dk} = 1$ を満たす。K はトピック数で，R の ldatuning パッケージ（バージョン 1.0.0）[7)]を用いて決定した。また，ϕ_{kv} はトピック k に医療従事者 v が割り当てられる確率で，$\phi_{kv} \geq 0$，$\sum_{v=1}^{V} \phi_{kv} = 1$ を満たす。V は医療従事者の種類の数である。

3.2.1 JS ダイバージェンス

得られた分布の類似度を次式で与えられる JS（ジェンセン・シャノン）ダイバージェンスで評価した。

$$D_{JS}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{2} D_{KL}\left(\boldsymbol{p}, \frac{\boldsymbol{p}+\boldsymbol{q}}{2}\right) + \frac{1}{2} D_{KL}\left(\boldsymbol{q}, \frac{\boldsymbol{p}+\boldsymbol{q}}{2}\right)$$

ここで，$D_{KL}(p, q)$ は次式で定義される KL（カルバック・ライブラー）ダイバージェンスである。

$$D_{KL}(\boldsymbol{p}, \boldsymbol{q}) = \sum_i p_i \log_2 \frac{p_i}{q_i}$$

上式において，p_i，q_i はそれぞれ分布 $\boldsymbol{p}$，$\boldsymbol{q}$ の要素（成分）である。KL ダイバージェンスは確率分布の隔たりを表す情報量で，常に 0 以上の値をとる。この値が 0 になるのは $\boldsymbol{p}=\boldsymbol{q}$ のときのみで，2 つの分布の差異が大きいほどより大きな値になる。しかしながら，KL ダイバージェンスは $\boldsymbol{p}$ と $\boldsymbol{q}$ について非対称なので，分布の類似度として扱うことができない。そこで，2 つの分布について対称になるように拡張したものが JS ダイバージェンスである。JS ダイバージェンスは 0～1 の値をとり，2 つの分布が似ているほど値は小さくなる。これを使って 2 つの病院 d と d' の類似度 $D_{JS}(\boldsymbol{\theta}_d,\boldsymbol{\theta}_{d'})$ を求め，病院情報局のランキングと比較した。また，2 つのトピック k と k' の類似度 $D_{JS}(\boldsymbol{\phi}_k,\boldsymbol{\phi}_{k'})$ を求め，そのヒートマップを描いてトピックが医療従事者の組み合わせパターンを特定しているかどうかを確認した。これらは，得られたトピックモデルの妥当性を検証するために行った。

3.2.2 ハーフィンダール・ハーシュマン指数

シェア s_i $\left(s_i \geq 0, \sum_{i=1}^{N} s_i = 1\right)$ が与えられたとき，

$$HHI = \sum_{i=1}^{N} s_i^2$$

をハーフィンダール・ハーシュマン指数といい，経済学分野で市場の競争状態を測る指標として用いられている。ハーフィンダール・ハーシュマン指数は，市場が独占状態（ある要素iについてのみ $s_i = 1$ で，他の要素はすべて 0）の場合は 1 に，シェアが均等（$s_1 = \cdots = s_N = 1/N$）である場合は$1/N$になり，それ以外では寡占状態に応じて 0～1 の間の値をとる。したがって，ハーフィンダール・ハーシュマン指数の逆数は市場を支配している寡占者のおおよその数とみなすことができる。そこで，$N_d = 1/\sum_k^K \theta_{dk}^2$を計算して病院 d が持つおおよそのトピック数を求め，病院規模との相関を調べた。また，$M_k = 1/\sum_v^V \phi_{kv}^2$を計算してトピック k に含まれるおおよその医療従事者の種類の数を求め，トピックを特徴づける量とした。

3.2.3 k-means 法によるクラスター分析

最後に，トピック分布 $\boldsymbol{\theta}_d = (\theta_{d1}, \cdots, \theta_{dK})$ を病院 d の特徴量とみなして k-means 法を使ってクラスター分析を行った。その際，クラスター数は20とした。得られたクラスターを使ってクラスターごとに「DPC 群の種類」，「特定機能病院の承認の有無」，「地域医療支援病院の承認の有無」などの病院属性を集計してクラスターと病院属性の相関を調べた。これは，施設基準等によって恣意的に決められた病院機能と客観的なデータから得られた病院機能がどの程度関連しているかを見るためである。

4. 結果

4.1 データの抽出結果

平成29年度病床機能報告公表データの施設票には7,244病院のデータがある。そのうち職員数を報告していない病院が485件，病床機能を報告していない病院が10件あった。それら495病院を除く6,749病院を処理対象として抽出した。なお，施設票には「手術室の職員数」，「外来部門の職員数」，「その他の部門の職員数」を加えた人数が「施設全体の職員数」を超えるものがあったが，不適切データとして除外した。また，非常勤の職員数を常勤換算することにより小数点以下に端数が付く場合は四捨五入して整数値にした。医療従事者数は，トピックモデルにおいて単語の出現頻度に対応するので整数でなければならないからである。

医療従事者の種類は全部で100種類あるが，このうち手術室の言語聴覚士（常勤），言語聴覚士（非常勤），作業療法士（非常勤），理学療法士（非常勤）には該当者がいなかった。そのため，これら4種類を除く96種類の医療従事者を処理対象として抽出した。抽出した医療従事者の総計は 1,588,170 人で，このうち病棟看護師（常勤）が最も多く，516,164 人で全体の 32.5%を占めていた。次いで，医師（常勤）が 170,305 人で 10.7%，病棟看護補助者（常勤）が 126,813 人で 8.0%と続いた。96 種類の医療従事者のうち上位 12 種類が全体の約 8 割を占めており，下位 66 種類は 1%未満であった。

4.2 最適なトピック数

図 1 は ldatuning パッケージを使って 4 つの指標 Griffiths2004，CaoJuan2009，Arun2010，Deveaud2014 のトピック数依存性を描いた図である。Griffiths2004 はモデルの当てはまりの良さを表す対数尤度が最大になるトピック数を求める手法で，それ以外はトピック間の医療従事者事後分布の類似性に着目して最適なトピック数を求める手法である。

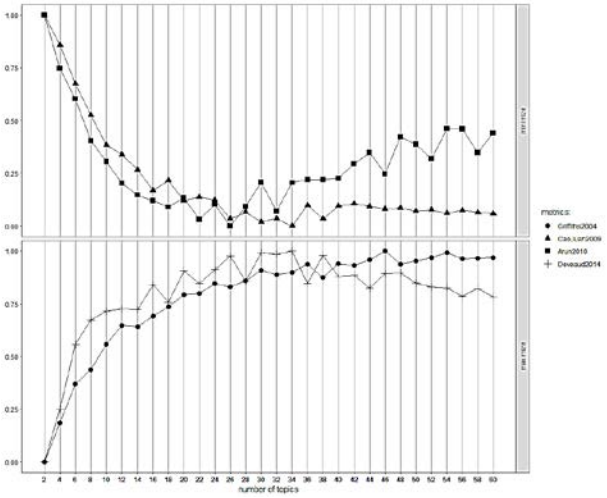

図 1. パッケージ ldatuning を用いたトピック数の決定

図 1 上の CaoJuan2009（上の▲印）と Arun2010（上の■印）は最小値を与えるトピック数が最適なトピック数で，Griffiths2004（下の●印）と Deveaud2014（下の＋印）は最大値を与えるトピック数が最適なトピック数である。明確な最大値や最小値は見られないがトピック数が 30～50 のあたりで各指標に極値あるいは飽和が見られるので，本研究ではトピック数として 40 を用いた。

4.3 構築したト LDA モデル

4.3.1 トピックの医療従事者分布

得られた LDA モデルからトピックごとの医療従事者の分布をヒートマップで描いたものを図2に示す。列（横軸）はトピック（左から順に 1～40）で，行（縦軸）は医療従事者である。明るいセルほど出現確率が高いことを表している。表 1 には，いくつかのトピックについて，出現確率の高い順に 5 つの医療従事者を示した。見出しの k はトピック番号，ϕ_k は出現確率，そして M_k は 3.2.2 で述べたハーフィンダール・ハーシュマン指数の逆数で，トピック k に含まれるおおよその医療従事者の種類の数を表している。

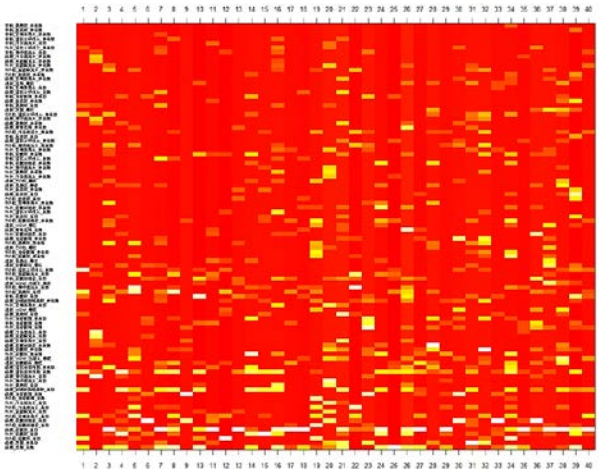

図 2.「医療従事者×トピック」のヒートマップ

表 1. トピックの医療従事者分布（抜粋）

k	医療従事者	ϕ_k	M_k
2	病棟理学療法士（常勤）	0.566	2.4
	病棟作業療法士（常勤）	0.294	
	病棟言語聴覚士（常勤）	0.109	
	退院ＭＳＷ社福士（専従）	0.015	
	病棟理学療法士（非常勤）	0.007	
6	病棟看護補助者（常勤）	0.999	1.0
8	医師（常勤）	0.358	3.9
	その他薬剤師（常勤）	0.308	
	病棟看護師（常勤）	0.125	
	臨床検査技師（常勤）	0.100	
	診療放射線技師（常勤）	0.082	
10	手術室看護師（常勤）	0.454	3.1
	医師（常勤）	0.315	
	臨床検査技師（常勤）	0.090	
	診療放射線技師（常勤）	0.087	
	病棟看護師（常勤）	0.042	
19	その他看護補助者（常勤）	0.391	3.8
	その他准看護師（常勤）	0.258	
	その他看護師（常勤）	0.188	
	その他作業療法士（常勤）	0.054	
	その他看護補助者（非常勤）	0.036	
25	医師（常勤）	0.987	1.0
	臨床検査技師（常勤）	0.012	
33	病棟看護補助者（常勤）	0.999	1.0
39	病棟助産師（常勤）	0.317	4.4
	医師（常勤）	0.308	
	手術室看護師（常勤）	0.129	
	病棟看護師（常勤）	0.105	
	臨床検査技師（常勤）	0.039	

4.3.2 病院のトピック分布

病院のトピック分布をヒートマップで描いたものを図 3，図 4 に示す。列（横軸）は病院で，行（縦軸）はトピック（下から順に 1～40）である。明るい線ほど出現確率が高いことを表している。病院は，基準となる左端の病院（基準病院）との JS ダイバージェンスが小さい順，すなわち基準病院に類似している順に並べている。図 3，図 4 における基準病院はそれぞれ独立行政法人国立がん研究センター中央病院および葛飾赤十字産院である。図 3 から，独立行政法人国立がん研究センター中央病院に類似した病院はトピック 8 と 10 を多く含んでいることがわかる。一方，図 4 から，葛飾赤十字産院に類似した病院はトピック 39 を多く含んでいることがわかる。

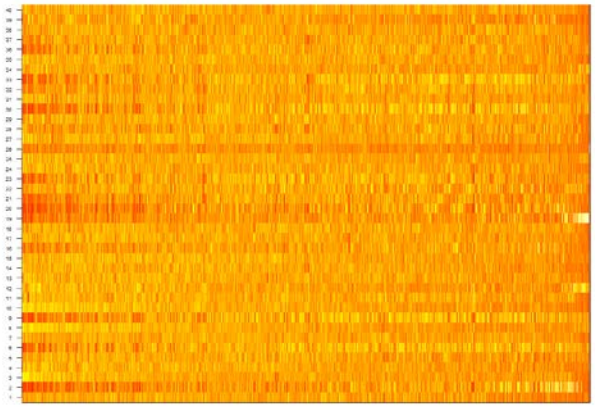

図 3. 独立行政法人国立がん研究センター中央病院に類似している順に病院を並べた「トピック×病院」のヒートマップ

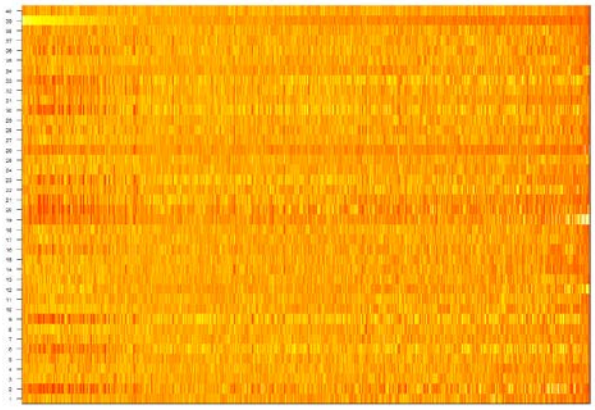

図 4. 葛飾赤十字産院に類似している順に病院を並べた「トピック×病院」のヒートマップ

4.4 病院情報局のランキングとの比較

病院情報局[8]の主な疾患別患者数ランキングにおいて，疾患名から「がん合計」，表示年度から「平成 29 年度」を選択して全国ランキングを取得し，その上位 20 病院に図 3 の独立行政法人国立がん研究センター中央病院に類似している病院の上位 100 病院がどのくらい入っているかを調べた。その結果 11 病院が全国ランキングの上位 20 病院に入っていた（$p < .001$）。

同様に病院情報局の産婦人科病院・診療所年間分娩件数ランキング（2017 年 12 月調査）の上位 20 病院に図 4 の葛飾赤十字産院に類似している病院の上位 100 病院がどのくらい入っているかを調べた。その結果 12 病院が全国ランキングの上位 20 病院に入っていた（$p < .001$）。

4.5 クラスター分析

クラスター分析によって得られたクラスターと病院属性の相関をまとめたものを表 2 に示す。医療機能として高度急性期

を掲げている病院の割合が最も高いのはクラスター12 で，340 病院の約半数にあたる 155 病院が 2017 年 7 月 1 日時点で高度急性期と報告していた。急性期も含めると 99%以上に達している。また，このクラスターに特徴的なことは，特定機能病院，地域医療支援病院，そして三次救急医療施設の占める割合が他のクラスターに比べて高いことである。急性期の割合が特に高いクラスターは 5, 8, 10, 11 などである。これらの病院に共通しているのは DPCⅢ群，地域医療支援病院，在宅療養後方支援病院の割合が比較的高い点である。慢性期の割合が高いクラスターは 3, 13, 18 などで，回復期の割合が高いクラスターは 2 である。

図 5 は，クラスターごとに平均病床数と平均トピック数を求めてバブルチャートを描いたものである。ここで，バブルの大きさは，そのクラスターに属する病院の数を表す。高度急性期の割合が最も高いクラスター12 は，平均病床数が 538 の大規模病院で，平均トピック数は 17 であった。急性期の割合が高いクラスター5, 8, 10, 11 は，平均病床数が 200～300 程度の中規模病院で，平均トピック数は 16～21 であった。慢性期の割合が高いクラスターのうち 3 と 18 は平均病床数が 130 くらいの小規模病院で，平均トピック数は 17 程度であった。同じく慢性期の割合が高いクラスターでも，クラスター13 は平均病床数が 59 でさらに小規模病院で，平均トピック数も 6 と少なかった。回復期の割合が高いクラスター2 は，平均病床数が 140 程度の小規模病院で，平均トピック数も 9 と少なかった。

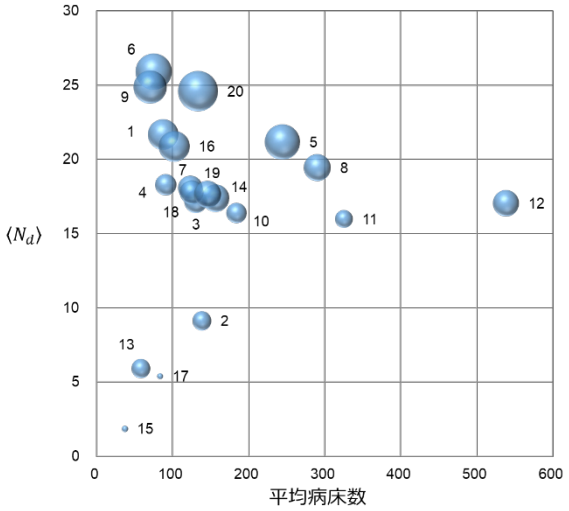

図 5. クラスター平均病床数と平均トピック数N_dの相関

クラスター別に病院のトピック分布をヒートマップで描いたのが図 6 である。行（縦軸）は病院で，上からクラスター，病床数の昇順に病院を並べている。列（横軸）はトピックで，左から順に 1～40 である。このヒートマップにおいて，明るい色ほどトピック確率が高いことを表している。

この図から，高度急性期の割合が高いクラスター12 に特徴的に現れるトピックは 25 と 39 であることがわかる。表 1 より，トピック 25 はほぼ常勤医師のみで構成されるパターンで，トピック 39 は常勤病棟助産師，常勤医師，常勤手術看護師などからなるパターンである。

急性期の割合が高いクラスター10 でひときわ目立つのがトピック 39 である。このクラスターに属する病院をトピック 39 の降順に整列すると，その上位 100 病院のうち 11 病院が病院情報局の産婦人科病院・診療所年間分娩件数ランキング（2017 年 12 月調査）の上位 20 病院の中に入っていた。

慢性期の割合が高いクラスター3, 13, 18 においてはそれぞれトピック 6, 19, 33 が際立っている。表 1 より，トピック 6 と 33 はほぼ常勤病棟看護補助者のみで構成されるパターンで，トピック 19 はその他の常勤看護補助者，常勤准看護師，常勤看護師，常勤作業療法士などからなるパターンである。

回復期の割合が高いクラスター2 に特徴的に現れているのがトピック 2 である。表 1 より，トピック 2 は病棟のリハビリテーション技師やメディカルソーシャルワーカーから構成されるパターンである。

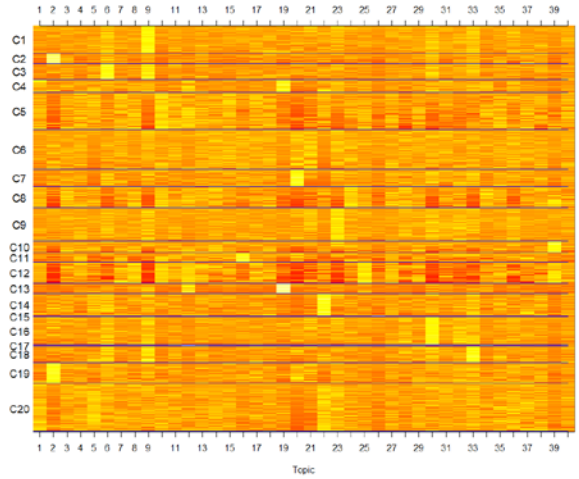

図 6. クラスター別トピック分布のヒートマップ

5. 考察

入力情報として各病院の医療従事者の人数しか使っていないにもかかわらず，提案手法は現実の病院機能をうまく捉えていた。具体的には，がんやお産など機能が似ている病院をトピック分布の類似性から収集して病院情報局のランキングと比較したら順序に有意な一致が見られた。また，トピック分布を用いてクラスタリングした結果，表 2 に示すように病床機能報告の医療機能や特定機能病院，地域医療支援病院といった病院機能と高い相関を示した。

トピックモデルはデータから隠れたトピックを推定し，そのトピック分布を分析対象の特徴量とみなして分類や検索に利用する。通常の機械学習であれば医師や看護師といった実体を伴う属性を特徴量に用いるが，トピックモデルではより抽象化されたトピックが特徴量として用いられる。トピックモデルではトピックの解釈は人間（研究者）が行う。本研究では医療従事者の組み合わせパターンがトピックになっている。「医療従事者の組み合わせ」から容易に想像できるのは多職種連携における医療チームである。そう考えると本研究におけるトピック分布は，その病院がどのような医療チームをどのような割合で有しているかを示すものと解釈できる。これはまさに病院機能そのものと言っても過言ではない。提案手法が病院機能をうまく捉えることができたのはこのためだと考える。

類似するトピック分布を持つ病院どうしの共通する特徴を探すことで病院機能を明らかにできるかもしれない。たとえば図 4 はトピックモデルによって産婦人科機能を有する病院が左端から順に集まるように並べた図であるが，共通した特徴はトピック 39 の割合が高いことである。表 1 からわかるようにトピック 39 は病棟助産師を筆頭に医師や手術室看護師などからなる産婦人科機能を持つ医療チームを表しているものと解釈できる。一方，図 3 はがん治療機能を持つ病院が左側に集まるように並べた図である。がん治療機能の医療チームとは

何であろう。図からトピック 8, 10 がこれらの病院に特徴的なトピックであることがわかる。したがって，がん治療機能を持つ病院に特徴的な医療チームは，これらのトピックを構成する医療従事者の組み合わせパターンであると推測できる。実際，表 1 からトピック 8 はその他の常勤薬剤師を多く含んでいることがわかるが，これはがん薬物療法が行われていることを示唆している。また，トピック 10 では手術室の常勤看護師が約半数を占めているが，代表的ながん治療に根治手術があることを考えると頷ける。このように本手法は病院機能を形成する医療チームは何であるかを特定する可能性を秘めている。

本研究の限界は，扱うデータが職員数だけという点にある。病院機能は人的資源だけで決まるものではない。医療機器などの設備や扱う症例，そして周辺地域の年齢構成や地理的条件など様々な要因が絡み合って形成される。したがって本研究で得られた結果は限定的なものだと言わざるを得ない。しかしながら，提案手法そのものは大きな可能性を秘めている。今回は人的資源に焦点をあててトピックモデルを構築したが，病床機能報告には CT，MRI 等の医療機器や扱う症例の情報が含まれている。これらを入力データに加えることで，より正確で詳細な病院機能の分析ができるかもしれない。実際，著者は公表 DPC データの症例実績を用いて本手法により DPC 病院の機能を分析したが，示唆に富む結果が得られている [9]。

6. 結語

本研究では，病床機能報告の職員数データにトピックモデルを適用して病院機能を分析する手法を提案した。提案手法を用いることにより，病院が病床機能報告の中で報告している自らの医療機能や施設基準等で標榜する病院機能を概ね識別することができた。しかしながら，今回の研究は人的資源にのみ焦点をあてているため，得られた結果は限定的なものと言わざるを得ない。より精度の高い分析を行うためには，医療機器などの物的資源や症例実績を加えてより高精度なモデルに改良する必要がある。今後は，今回の研究をさらに発展させて病院機能の高精度な分析モデルを構築し，恣意的に決められた現状の病院機能と比較して課題を明らかにしたい。

参考文献

1) 病床機能報告. 厚生労働省, [https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000055891.html (cited 2019-Aug-17)].
2) 佐藤一誠，奥村学監修. トピックモデルによる統計的潜在意味解析. コロナ社, 2015.
3) Hatakeyama Y, Ogawa T, Ikeda H and Haseyama M. A most resource-consuming disease estimation method from electronic claim data based on labeled LDA. IEICE TRANS INF SYST 2016 ; E99.D(3) : 763-768.
4) Bhattacharya M, Jurkovitz C, Shatkay H. Identifying Patterns of Associated-Conditions through Topic Models of Electronic Medical Records. BIBM 2016 : 466-469.
5) Chen JH, Goldstein MK, Asch SM, Mackey L, Altman RB. Predicting inpatient clinical order patterns with probabilistic topic models vs conventional order sets. JAMIA 2017 ; 24(3) : 472-480.
6) Park S, Choi D, Kim M, Cha W, Kim C, Moon IC. Identifying prescription patterns with a topic model of diseases and medications. Journal of Biomedical Informatics archive 2017 ; 75C : 35-47.
7) Nikita M. Select number of topics for LDA model, 2019. [https://cran.r-project.org/web/packages/ldatuning/vignettes/topics.html (cited 2019-Aug-17)].
8) 病院情報局. 株式会社ケアレビュー. [https://hospia.jp/ (cited 2019-Aug-17)].
9) 田中昌昭. トピックモデルによる DPC データの分析と病院機能についての考察. in press. 川崎医療福祉学会誌 2019 ; 29(1).

表 2. クラスターと病院属性の相関関係

Cluster	病院数	医療機能（2017/7/1時点）				特定機能病院	DPC病院群			総合入院体制加算			地域医療支援病院	在宅療養支援病院	在宅療養後方支援病院	三次救急医療施設
		高度急性期	急性期	慢性期	回復期		Ⅰ群	Ⅱ群	Ⅲ群	Ⅰ	Ⅱ	Ⅲ				
1	459	0	106	328	24	0	0	1	1	0	3	0	3	62	6	4
2	172	1	8	7	156	1	0	0	7	0	4	0	0	31	4	1
3	263	0	13	239	10	1	0	0	2	1	2	0	2	36	9	0
4	218	0	113	92	13	1	0	0	18	0	1	0	1	28	6	0
5	613	22	524	47	18	3	3	17	377	1	13	27	139	32	94	36
6	659	4	452	137	55	3	0	4	66	0	5	0	9	162	13	3
7	298	1	114	123	60	0	0	0	33	0	2	0	4	70	11	0
8	355	27	306	7	15	4	1	24	236	5	27	40	115	24	59	36
9	541	0	307	206	26	1	1	1	11	2	2	1	2	110	4	2
10	202	8	193	1	0	0	0	7	85	2	9	18	43	7	20	15
11	159	32	117	5	5	9	7	12	91	4	5	25	42	8	19	29
12	340	155	183	1	1	73	70	89	170	24	54	72	176	1	24	172
13	175	0	12	143	13	1	0	0	0	0	1	0	0	1	7	0
14	372	1	106	114	150	3	0	1	51	1	2	0	2	105	13	0
15	18	1	17	0	0	0	0	0	0	0	0	0	1	0	0	0
16	465	0	90	356	19	1	0	0	6	0	2	0	2	82	13	1
17	17	1	3	10	3	0	0	0	1	0	0	0	0	0	0	0
18	276	0	13	255	8	0	0	2	0	0	2	0	1	28	9	0
19	353	2	136	92	123	0	0	1	67	0	0	0	7	98	27	1
20	794	5	538	164	87	1	1	2	178	0	3	2	24	226	26	3
合計	6,749	260	3,351	2,327	786	102	83	161	1,400	40	137	185	573	1,111	364	303