

テキストマイニングを使った病院の評判分析

—KHcoder を利用して—

Hospital reputation analysis using text mining
— Using KHcoder —

保健医療経営大学 川島 秀樹 ・ 後藤 浩士

本稿では福岡県八女・筑後医療圏における公立病院について、テキストマイニング手法の一つである KHcoder を用いて評判分析を行った。評判分析は、経済学の不確実性理論の一部であり、情報学のテキストマイニングを用いることで可能になる。考察結果として、患者側の意見は分析者の恣意的な判断に影響を受けず、客観性または信頼性を維持し、バランスよく評価していることを確認できた。評判分析は、顧客の意思行動を分析するのに有用なツールであることを確認することができた。

【キーワード】評判分析、KHcoder、公立病院、統合問題、テキストマイニング

はじめに

厚生労働省が地域医療構想¹で設定している福岡県八女・筑後医療圏（圏内人口約13万人）において、八女市の Y 病院は毎年約7億円の赤字となっており、移動距離8 km に位置する筑後市の C 病院との統合が検討されている。八女市は市町村合併により広域化しており、Y 病院まで自家用車で1時間以上かかる地域もある。特に山間部は過疎化し、バスなどの公共交通手段もない。しかし、Y 病院の利用者（患者及び家族）が C 病院に通院しようとするれば手段・時間などにおいて困難を伴う。このような状況の中で、八女市は救急医療体制や周産期医療体制等を維持していくために必要な医療資源を確保しなければならないという深刻な問題を抱えている。

まず、現状を正確に把握するために、Y・C 病院関係者、八女・筑後市役所担当者、八女市議会議員、福岡県議会議員などを対象に2018年4月から12月までの期間にインタビューを行うとともに、市の広報誌や市議会の議事録などの収集、分析を行った。次に、2018年10月1日から11月8日までアンケート（郵送依頼400通、回答数137通、回収率34.3%）を実施し、自治体（公立）病院の果たす役割について考察²した。しかしながら、病院は企業と異なり経営に関する情報を公にしていないため、情報の非対称性が生じている。そのため、本稿では、八女・筑後医療圏の自治体（公立）病院をケーススタディとして、アンケートを基にした地域におけるテキストデータを使って病院統合・再編に関する評判を、テキストマイニング手法を使って明らかにしたい。テキストマイニングとは、牛澤賢二（2018）によればアンケート調査における自由回答のような文書形式のデータを品詞単語に分解し、頻度を数え、統計手法などの分析手法を駆使して、文書自体を理解する方法であると定義される。本研究において使用したツールは、樋口耕一（2018）が開発した KHcoder である。

1. 評判とは

私たちの意思決定は評判に大きく影響を受けている。例えば、あの病院の誰々医師は名医で信用できる。多くの病院に通ったが、あの病院で治ったという評判は患者の行動に多大な影響をもたらす。最近では、新型コロナウイルスの院内感染拡大によって評判を下げた病院も少なくない。

神戸伸輔（2004）は、ゲーム理論の解釈において、相互依存関係の中で、評判を用いて相手方の行動に関して予測を行っていることを指摘した。たとえば、満足できる治療をしてくれるなら、A 病院に行くが、そうでなければ A 病院は行かずに、他に行くというものである。評判には、良い評判と悪い評判（評判を失う）があり、評判が存在するためには、繰り返しの取引が前提となる。

まず、評判を達成するための手段として、コミットメント（誓約）があり、自分の利得をあげるために、「その行動しかとれないようにする実効性のある仕組みをつくること」が挙げられ、ゲーム理論では、支配戦略と呼ばれている。たとえば、医療サービスでは、良い評判のところが多くの顧客を集める。特に、利用する前にサービスの質がわからない状況での意思決定に有効である。競争上で優位に立つことを目的として、相手の利得を下げ自分の利益を上げるために、評判を使うこともある。

次に、評判を達成するための手段として考えられるのは、協力である。評判の良いプレイヤーに対しては協力をを行い、評判の悪いプレイヤーに対しては、協力しない（非協力に行動する）という行動である。この場合、相手の評判を知る必要があり、相手の評判が良いのに協力しないと、こちらの評判を失うこともある。

評判の起こる仕組みは、ゲーム理論のトリガー戦略と密接な関係をもっている。トリガー戦略とは、1度きりでなく繰り返し行われるゲームにおいて、最初は協調し、相手が協調する限りこちらも協調するが、いったん相手が裏切れば、そ

れ以降はこちらも裏切り続ける戦略である。評判は、裏切ったら無くなってしまう。したがって、あの病院、あの医師は「信頼できる」と評判を立てられることが重要である。

病院の評判において、コミットメントを達成するような評判は、そのプレイヤーの利得をあげることに貢献する。顧客の信頼を得ることによって、商品やサービスの価値は上がり、その評価が集計されて評判となる。評判の担い手は、個人を含む組織体であり、その病院の存続性の指標になっていると考えられる。

2. テキストマイニング (KHcoder) とは

テキストマイニングとは、アンケートなど、自由記述などの文章形式のデータを、定量的な方法で分析することを行う。定量的に分析することで、数値データと同じように、結果を視覚化することができる。

本研究における分析手法としては、樋口耕一氏が開発したテキストマイニング用のフリーソフト (KHcoder) を用いた。いかに適切に、分析対象となる言葉を抽出できるかに着目して、検証を行った。

2.1. KHcoder 概要

KHcoder は、テキスト型データの計量的分析を行うものである。この分析は、テキストマイニングと呼ばれる比較的新しい技術であり、Web の記事内容や SNS での書き込み、アンケートにおける自由回答のテキストデータを、個人の恣意的な判断によらず、客観的に分析しようとするものである。樋口耕一 (2018) によると、このようなテキストマイニング分析を行う場合、人間の判断や独創性が不可欠であり、人間の独創性を活かしつつ、どのようにして客観性ないし信頼性を維持するかというバランスが問題になるという。樋口は、大まかに 2 つの段階からなる分析手順³ を想定して、KHcoder を製作している。

段階 1 では、データ中から語を自動的に取り出して、その結果を集計・解析する。これによって、分析者の予断をなるべく交えずに、データの特徴をさぐったり、データを要約したりする。

段階 2 では、分析者が「こういう表現があれば、コンセプト A が出現していたと見なす」といった指定 (コーディングルール作成) を積極的かつ明示的に行い、データ中からコンセプトを取り出している。その結果を集計・解析することで、分析を深めている。

2.2. テキストマイニングの手順

テキストマイニングの手順として、まず、文書を言葉にばらす系形態素解析を行う。すなわちテキスト形式のデータを定量的に分析できる形に変換する。日本語は英語と違い、単語の間に空白が入らない「べた書き」であるため、辞書を用いて文を単語単位にばらす必要があり、本研究ではフリーの研究ソフトを用い、それを使う。

次に、言葉相互の関係を分析する。共起ネットワークでは、抽出した言葉の関連性を分析することができる。関連性 (共起性) は線のつながりとして、視覚的に確認できる。

2.3. Y・C 病院に関する評判分析の結果について

テキストマイニングで使うデータは、アンケートで集めた自由意見欄における記述および Web 上で集めた Y 病院や C 病院に関する 407 件の書き込み事項であり、抽出した語数は 9,221 語であった。Web 上の書き込みは、予想よりも少なかったため、分析に用いたデータ大半はアンケートによるものである。(図 1 参照)

Database Stats	
総抽出語数 (使用)	9,221 (3,737)
異なり語数 (使用)	1,571 (1,235)
集計単位 ケース数	
文	407
段落	178

図 1 プロジェクトメニュー

2.3.1. 品詞別抽出語リスト

品詞別の抽出語リストを、出現回数順に 150 位まで、図 2 に記した。キーワードとしては医療、医師、救急、安心、運営、改善、信頼、存続、痛み、連携、合併などがよく見られる。

次に、抽出語を品詞別に並べたところ (図 3 参照)、組織名では筑後圏の大学病院である久留米大学や Y 病院に並ぶほどに医療環境が充実した川崎病院の語が見られた。また、形容動詞では、必要、丁寧、健康、可能、不安、急、緊急、困難、重要、大切などが見られた。

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
病院	136	長い	9	職員	5
思う	64	良い	9	信頼	5
医療	46	機関	8	診断	5
自治体	34	市立	8	待つ	5
医師	33	受ける	8	退院	5
地域	32	少ない	8	担当	5
行く	30	場合	8	地区	5
公立	29	診療	8	痛み	5
八女	28	特に	8	年	5
患者	27	利用	8	不安	5
多い	26	連携	8	不明	5
対応	23	運営	7	本当に	5
看護	21	受付	7	麻酔	5
経営	21	出す	7	もう少し	4
必要	20	紹介	7	レントゲン	4
言う	19	答	7	移動	4
検査	18	評価	7	維持	4
総合	18	問	7	印象	4
入院	18	業	7	環境	4
説明	17	お世話	6	機能	4
先生	17	スタッフ	6	結果	4
手術	16	改善	6	現在	4
民間	16	顔	6	個人	4
医者	14	近く	6	向上	4
待ち時間	14	健康	6	広い	4
筑後	14	見る	6	考え	4
通院	14	今	6	自分	4
原因	13	住民	6	収益	4
診察	13	出る	6	助かる	4
救急	12	存続	6	少々	4
安心	11	痛い	6	伸びる	4
今後	11	電話	6	親切	4
市立病院	11	努力	6	針	4
治療	11	統合	6	昔	4
充実	11	分かる	6	絶対	4
人	11	聞く	6	専門医	4
赤字	11	母	6	存在	4
筑後	11	問題	6	体制	4
悪い	10	来る	6	待たず	4
確保	10	意識	5	大きい	4
感じる	10	一番	5	大学	4
考える	10	可能	5	駐車	4
高い	10	外来	5	提供	4
受診	10	含める	5	内科	4
前	10	行う	5	難しい	4
丁寧	10	高齢	5	入れる	4
家族	9	受信	5	年度	4
久留米	9	重要	5	不足	4
時間	9	初めて	5	法人	4
専門	9	状況	5	満足	4

図 2 品詞別抽出語リスト

2.3.3. KWIC コンコーダンス2

KWIC コンコーダンスを用いることで、このポジションによって抽出語（この例では自治体）と文脈との関連性をみる事ができた。また、2つの言葉の関連性を測る尺度（共起性の尺度、通常のデータ分析の相関係数に該当）に基づいて関連する語を検索することができる。抽出語をあげると、病院、経営、赤字、運営、民間、ダメ、維持、協力などに関するキーワードを確認することができた。（図5参照）

■ コロケーション統計

Node Word		抽出語: 自治体 品詞: 活用形: ヒット数: 34														
Result																
N	抽出語	品詞	合計	左合計	右合計	左5	左4	左3	左2	左1	右1	右2	右3	右4	右5	スコア
1	病院	名詞	37	2	35	0	0	2	0	0	32	2	0	1	0	33.917
2	思う	動詞	3	3	0	1	0	1	1	0	0	0	0	0	0	1.033
3	地域	名詞	4	1	3	0	0	1	0	0	0	0	0	2	1	1.033
4	経営	サ変名詞	2	1	1	0	0	0	1	0	0	1	0	0	0	1.000
5	高い	形容詞	1	1	0	0	0	0	0	1	0	0	0	0	0	1.000
6	必ずしも	副詞	1	1	0	0	0	0	0	1	0	0	0	0	0	1.000
7	ない	否定動詞	3	3	0	1	1	0	1	0	0	0	0	0	0	0.950
8	ほしい	形容詞(非自立)	3	2	1	0	1	0	1	0	0	0	0	0	1	0.950
9	赤字	名詞	3	1	2	0	0	1	0	0	0	0	1	1	0	0.917
10	考える	動詞	2	2	0	0	0	1	1	0	0	0	0	0	0	0.833
11	必要	形容動詞	3	0	3	0	0	0	0	0	0	1	1	1	1	0.783
12	運営	サ変名詞	2	0	2	0	0	0	0	0	0	0	2	0	0	0.667
13	民間	名詞	2	0	2	0	0	0	0	0	0	0	1	1	0	0.583
14	紹介	サ変名詞	2	0	2	0	0	0	0	0	0	1	0	1	1	0.533
15	環境	名詞	1	1	0	0	0	0	1	0	0	0	0	0	0	0.500
16	現在	副詞可能	1	1	0	0	0	0	1	0	0	0	0	0	0	0.500
17	姿勢	名詞	1	1	0	0	0	0	1	0	0	0	0	0	0	0.500
18	他	名詞C	1	1	0	0	0	0	1	0	0	0	0	0	0	0.500
19	都合	名詞	1	1	0	0	0	0	1	0	0	0	0	0	0	0.500
20	入院	サ変名詞	1	1	0	0	0	0	1	0	0	0	0	0	0	0.500
21	八女	地名	1	1	0	0	0	0	1	0	0	0	0	0	0	0.500
22	医療	名詞	2	1	1	1	0	0	0	0	0	0	0	0	1	0.400
23	アップ	サ変名詞	1	1	0	0	0	1	0	0	0	0	0	0	0	0.333
24	スタッフ	未知語	1	0	1	0	0	0	0	0	0	0	1	0	0	0.333
25	ダメ	形容動詞	1	1	0	0	0	1	0	0	0	0	0	0	0	0.333
26	維持	サ変名詞	1	0	1	0	0	0	0	0	0	0	1	0	0	0.333
27	開業医	名詞	1	0	1	0	0	0	0	0	0	0	1	0	0	0.333
28	患者	名詞	1	0	1	0	0	0	0	0	0	0	0	1	0	0.333
29	求める	動詞	1	0	1	0	0	0	0	0	0	0	1	0	0	0.333
30	協力	サ変名詞	1	0	1	0	0	0	0	0	0	0	1	0	0	0.333

図5 「自治体」のコロケーション統計

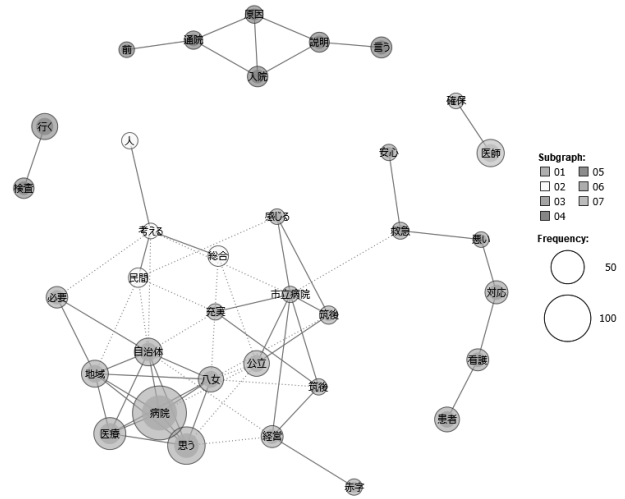


図6 共起ネットワーク1 (口コミなどのWeb情報を入れた場合)

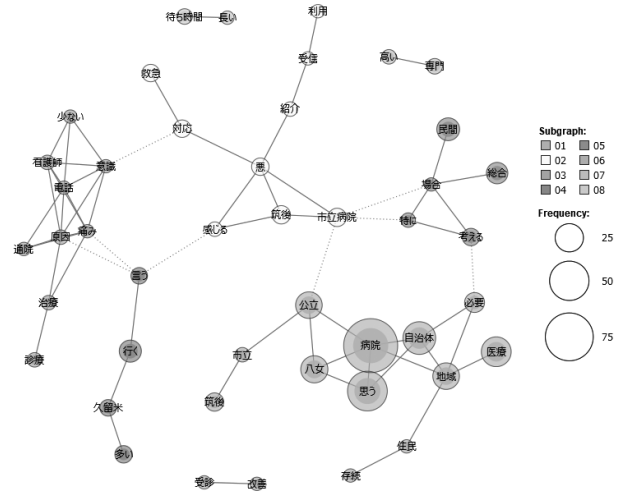


図7 共起ネットワーク2 (アンケート情報のみの場合)

2.3.4. 共起ネットワーク

共起ネットワークによって、抽出語間の共起性と抽出語と外部変数の間の共起性を分析することができる。出現回数は図形に比例し、共起性、関連性の強さは図形の位置や大きさではなく、線での接続および、その太さで示される。

図6を見ると、「原因、説明、通院、入院」「八女の病院」、「筑後市の病院」、「行く、検査」、「統合、民間、考える、人」「確保、医師」、「救急、安心、悪い、対応、看護、患者」の7グループに分かれており、充実と経営赤字に反応している。図7は、アンケートの自由記述の内容のみを用い作成している。対応というキーワードに関連して、詳細に確認してみると意識、看護師、電話、痛み、原因、治療、診療という用語があった。

2.3.5. 自己組織化マップ

自己組織化マップでは、語と語の関連性を探索することができる。ここでは、赤字経営の中で、診療、長い待ち時間、紹介、悪、救急、改善が見受けられた。(図8参照)

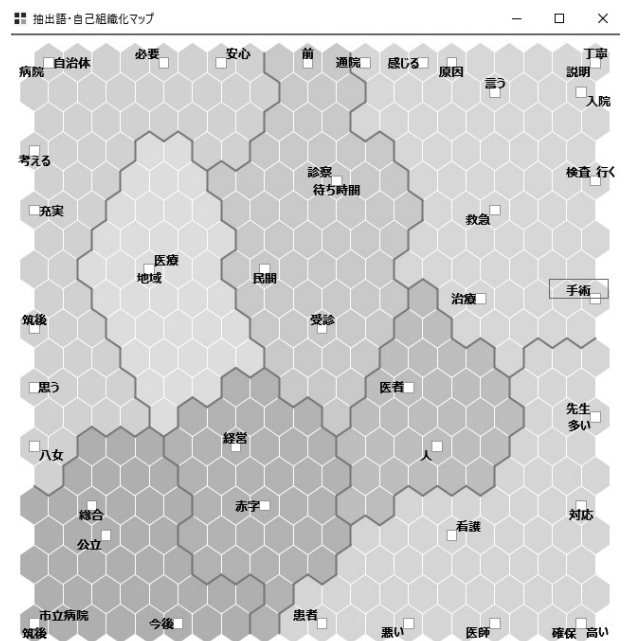


図8 自己組織化マップ

2.3.6. クラスタ分析

クラスタ分析とは、対象間の類似度または距離に基づいて、似ているもの同士をいくつかのグループに分類する手法である。(図9参照)

内容を見ると、住民、存続、赤字、統合、医師、患者、運営、民間、救急、対応、良、専門、高い、受診、改善、専門、高い、久留米、多いなどのキーワードが見られた。グラフの大きさを考慮すると、八女地区の公立病院の医療が必要であること、民間への移管も考えていることも見受けられる。それに、医師確保が必要であることや、赤字経営を気にしていることがわかる。

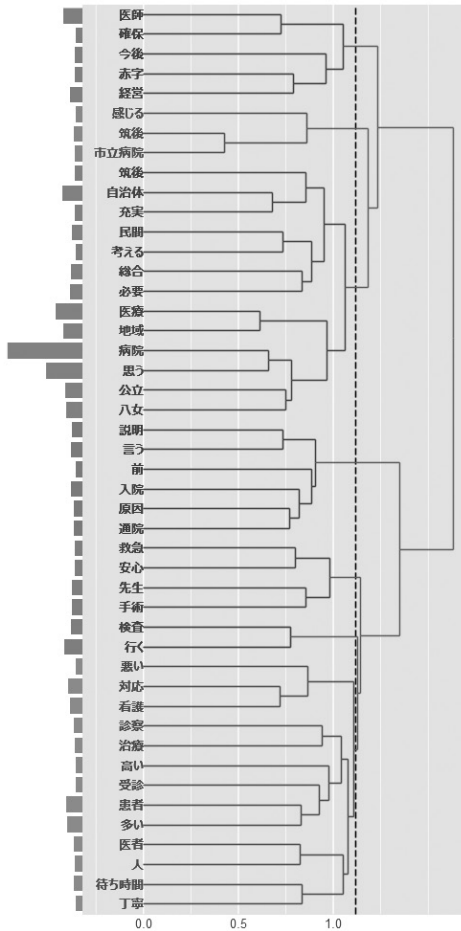


図9 クラスタ分析

2.3.7. 抽出語対応分析

抽出語の対応分析は、クロス集計を視覚化する方法である。クロス集計は、基本的に2つの項目あるいは変数を視覚化する方法である。設定はデフォルトのままである。この分析で、印象的だったのが、救急診療、赤字、検査、入院というキーワードである。(図10参照)

2.4. 評判分析の考察について

評判分析について、品詞別に抽出語を選ぶことができた。それを基に、日本語評価極性辞書を用いたテキストのネガティブ・ポジティブ分析について行った。極性辞書は、東北大学の乾・岡崎研究室⁴が構築し、評価極性情報を付与した

データ、評価極性タグである。ポジティブ(正の数)が最高値1からネガティブ(負の数)最低値-1で表現している。その辞書をAccessのT極性辞書に約5万5千件登録し、Q単語評判値クエリにおいて抽出した語(評判単語)をマッチングさせ、集計することにより評判値を出すことができた。(図11参照)

T極性辞書とマッチングした件数は872件であり、その結果を集計すると、集計値は-376.99、平均値が-0.43となった。この結果を見ると八女筑後医療圏の公立病院はネガティブ・イメージが強い。

しかしながら、KHcoderのテキストマイニングの分析から、八女筑後医療圏において、広範囲な医療圏での救急医療の整備は不可欠であると考えられる。赤字であるものの、統合や民間化になったとしても維持してほしいという意思表示が確認された。

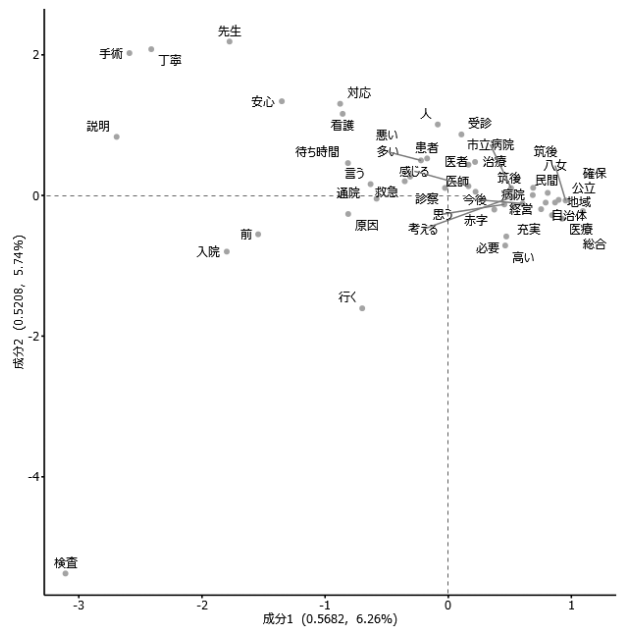


図10 抽出語対応分析

評判単語	出現回数	品詞	ポジティブ/ネガティブ値	評判値
優れる	2	動詞	1	2
良い	9	形容詞	0.998995	8.998995
嬉しい	1	形容詞	0.998971	0.998971
適切	2	名詞	0.998406	1.996812
人材	1	名詞	0.997049	0.997049
福祉	1	名詞	0.995217	0.995217
優秀	1	名詞	0.995178	0.995178
満足	4	名詞	0.994414	3.977656
美しい	1	形容詞	0.992888	0.992888
ズリット	1	名詞	0.99214	0.99214
評判	3	名詞	0.99087	2.97261
聖	1	名詞	0.990084	0.990084
支援	1	名詞	0.989732	0.989732
優しい	2	形容詞	0.987973	1.975946
広大	3	名詞	0.986627	2.959881
全力	1	名詞	0.986162	0.986162
上手	1	名詞	0.985536	0.985536
あっさり	1	副詞	0.983673	0.983673
利便	2	名詞	0.981406	1.962812
安心	11	名詞	0.981397	10.795367
信頼	5	名詞	0.981102	4.90551
儂かいしい	1	形容詞	0.980047	0.980047
きちんと	1	副詞	0.979967	0.979967
期待	1	名詞	0.978309	0.978309
癒す	1	動詞	0.977712	0.977712
望ましい	2	形容詞	0.975693	1.951386
助かる	4	動詞	0.970031	3.880124
向上	4	名詞	0.963925	3.8557
一番	5	名詞	0.963556	4.81778
好む	1	動詞	0.96078	0.96078
給与	1	名詞	0.958327	0.958327
明るい	1	形容詞	0.955372	0.955372
改善	6	名詞	0.954802	5.728812
安い	1	形容詞	0.952483	0.952483

図11 Q単語評判値クエリ

結語 まとめと今後の課題

福岡県八女・筑後医療圏における自治体病院の統合・再編に関する評判分析について述べた。

1章においては、ゲーム理論の解釈を用い、評判を達成するものとしてコミットメント（誓約）と協力の2つがあり、医療のような情報の非対称的な財では、評判を得ることが重要である。評判の担い手は、個人だけではなく組織でもあり、その病院が今後存続するかという指標になっていると思われる。

2章においては、樋口耕一氏が開発したテキストマイニング用のフリーソフト（KHcoder）を使って分析した。定量的に分析することで、結果を誰が見てもわかるように視覚化し、いかに分析対象となる言葉を抽出できるという点に注目した。

分析結果では、品詞別抽出語リスト、キーワードを中心としてどのような文脈で使われているかというKWICコンコーダンス、共起ネットワーク、自己組織化マップ、クラスター分析、抽出語対応分析を行った。抽出語は「医療、医師、救急、安心、運営、改善、信頼、存続、痛み、連携、合併」などがあげられる。また、KWICコンコーダンスでは自治体に対する期待があげられ、「病院、経営、赤字、運営、民間、ダメ、維持、協力」などのキーワードが見られた。共起ネットワークでは、対応というキーワードに関連して、「意識、看護師、電話、痛み、原因、治療、診療」という用語があった。さらに、自己組織化マップでは、語と語の関連を探索し、赤字経営の中で、「診療、長い待ち時間、紹介、悪、救急、改善」が見受けられた。クラスター分析や抽出語対応分析では、「救急」というキーワードが印象的であった。

八女筑後医療圏の公立病院の評判⁵については、ネガティブ・ポジティブ分析を行った。東北大学が構築した日本語評価極性辞書を使い、集計したところ、集計値は-102.26平均値が-0.46となりネガティブであった。この結果を見ると公立病院は、ネガティブ・イメージが強いことがわかる。しかしながら、公立病院を継続してほしいという意思表示がみられた。

今後の課題として、Webからのネット書き込みなどの情報を採集するツールを充実させ、テキスト情報を増やし、評判分析と自治体病院の経営の在り方について確実性を高める研究を進めていく。

付記

本研究は、JSPS 科研費 JP18K09960の助成を受けたものです。

注

- 1 厚生労働省，地域医療構想，URL<<https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000080850.html>> 参照日（2019-11-6）。
- 2 川島秀樹、後藤浩士（2019）「地域医療圏における自治体病院の役割一病院統合・再編に関するアンケート―」JAPA九州、第43号、pp.31-39。
川島秀樹（保健医療経営大学）・後藤浩士（保健医療経営大学）（2019）「公立病院の統合問題と産業連関表による経済効果」日本経済政策学会第76回全国大会予稿集。
- 3 詳細内容については、樋口耕一氏のホームページ「KH Coderの主な機能と分析手順」URL<<https://kncoder.net/diagram.html>>を参照されたい。
- 4 極性辞書 東北大学乾・岡崎研究室，URL<<http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources/Japanese%20Sentiment%20Polarity%20Dictionary>>、参照日（2020-07-12）。
- 5 公立病院の評判については、特定の病院を論文上で記述することはできなく、八女筑後医療圏という枠組みの中で行った。

参考文献

- [1] あずさ監査法人編（2016）『公立病院の経営改革』、同文館出版。
- [2] 牛澤賢二（2018）『やってみよう テキストマイニング 自由回答アンケート分析に挑戦！』、朝倉書店。
- [3] 樫田洋資，菊田遥平，谷田和章，森本哲也，（2017），『フリーライブラリーで学ぶ機械学習入門』，秀和システム。
- [4] 神取道宏（2014）『ミクロ経済学の力』、日本評論社。
- [5] 神戸伸輔（2004）『入門 ゲーム理論と情報の経済学』、日本評論社。
- [6] 厚生労働省（2017）「地域医療の確保と公立病院の推進に関する調査研究会報告書」。
- [7] 樋口耕一（2018）『社会調査のための計量テキスト分析』、ナカニシヤ出版。

Web 参照

- [8] 総務省、公立病院改革財務決算状況、URL<http://www.soumu.go.jp/main_sosiki/c-zaisei/hospital/kessan-bunseki/>、参照日（2018/11/22）。
- [9] 総務省、病院事業決算状況（独立行政法人）、URL<http://www.soumu.go.jp/main_content/000556627.pdf>、参照日（2018/11/23）。
- [10] 厚生労働省、八女・筑後地域医療再生計画、URL<<https://www.mhlw.go.jp/bunya/iryuu/saiseikikin/dl/fukuoka-keikaku.pdf>>、参照日（2018/11/18）。
- [11] 東北大学乾・岡崎研究室 極性辞書、URL<<http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources/Japanese%20Sentiment%20Polarity%20Dictionary>>、参照日（2020-07-12）。
- [12] 樋口耕一ホームページ、URL<<https://kncoder.net/diagram.html>>、参照日（2020/10/10）。