

平成28年度厚生労働科学研究費補助金（地域医療基盤開発推進研究事業）
地域医療構想・地域医療計画を効果的に実装するためのデータ解析・活用方法の開発
報告書（第二部）

都道府県で策定された医療計画の 全国参照データベース構築に向けて —言語処理による統計処理の活用—

研究代表者：

今中 雄一（京都大学大学院医学研究科医療経済学分野 教授）

研究協力者：

國澤 進（京都大学大学院医学研究科 医療経済学分野 講師）

佐々木 典子（京都大学大学院医学研究科 医療経済学分野 講師）

山下 和人（京都大学大学院医学研究科 医療経済学分野 研究員）

目的：地域医療構想の根幹をなす地域医療計画は、各都道府県で策定される。医療計画自体は公開されているものの、これらが全国でどのような共通部分があり、取り組みの差が生じているのかは明らかになっていない。本研究は、統合型検索エンジン機能と高度なテキストアナリティクス機能を有する IBM Watson Explorer（以下「WEX」）を用いた自然言語解析を通じて、全国の医療計画の地域間比較・参照を容易にし、医療計画内容向上に資することを目的とした。

方法：各都道府県の地域医療計画から分析、定義した17因子に対して、同文書内に出現するフレーズや単語の頻度や、相対的な傾向分析を行うことで、同義語となるものを探索・抽出させ、相関分析を行った。具体的には、一つの章を一つ（あるいはある特定）の事項の説明単位と考えることで、同一章内に出現する頻度と、文章全体で出現する頻度とを比較することで、類似した概念を表すフレーズや単語の可能性を相関分析により数値的に示した。

結果：PDF形式からテキスト形式に変換可能だった44都道府県の地域医療計画を解析対象とした。本研究では「一次予防、健診・早期発見、体制」についての検討を行うことを想定し、初期の因子を設定した。まずこれらの一般的同義語を一般的辞書より抽出し、機械の初期学習に与えた。同時に、文章のクリーニングとしてWEXにより、文書の形態素解析を行ってフレーズや単語に分解し、初期因子の存在をフラグした。続いて、これら初期の因子と相関が高いフレーズや単語を抽出した。

考察：本研究では、自然言語解析を行うための準備としての、辞書の拡充を行うための手段について、機械的な処理の応用を試みた。結果的に様々な類似語を同定することができた。この中にはいわゆる正しくない結果も含まれているため、これらの中からさらに人の目によるなど何らかの方法での選出が必要になるものの、研究者が当初予測していないようなフレーズも抽出されてくるため、辞書の機械的な拡充を行うための非常に有力な手段

の一つとして活用できた。今後この作業をシステマティックに繰り返していくことで、さらに特徴的な表現を拡充あるいは洗練することが可能になると考えられ、ここで構築された「辞書」を用いることで、医療計画だけではなく、それに関連する文書の言語解析を行うための有用な基盤になることが期待される。今回見えてきたデータ処理上の課題から、データ提供元としての行政に対し、フォーマットの統一や、少なくともデータとして活用しやすいフォーマット等の提言を行っていく。

結論：本研究では、地域医療計画の文書解析を行うための「辞書」を機械的に拡充した。これら自然言語解析をより発展させること通じて、全国の地域医療計画の地域間比較・参照を容易にし、計画内容の向上に資することが期待される。

A. 研究背景と目的

各都道府県で策定し、5年に1度見直しがなされる医療計画によって、効率的かつ質の高い医療提供体制を構築し、地域における医療及び介護の総合的な確保を推進することが期待されている。この計画は、各都道府県が地域の実情を勘案し、独自に策定することが求められる。このため、5疾病5事業を始めとして重点的な課題や基本方針が定められているものの、全国画一の計画にはならない。医療計画自体は公開されているものの、これらが全国でどのように共通部分があり、取り組みの差が生じているのかは明らかになっていない。この計画は計画書として文書で記されたものであり、計画自体の定量的な評価や全国での比較は困難である。

最近では、機械学習など、ソフトウェアにより様々なデータの処理の可能性が開かれてきた。その中の一つとして統合型検索エンジン機能と高度なテキストアナリティクス機能を有するIBM Watson Explorer (以下「WEX」)がある。IBMによれば、WEXは、自然言語解析を核としたテクノロジーに基づく対話型ソリューションを有するWatsonファミリー群におけるプラットフォームとして位置づけられている。

そこで、今回、このテクノロジーを利用することで、全国47都道府県の医療計画を分析し、比較を行うことを目標としつつ、まずその第1歩として医療計画の文章についての分析を試みた。

B. 研究方法

本研究では、医療計画の分析を行うための前段階として、文章の解析を行うための

第1歩目を行った。現在のテクノロジーでは、単純にデータを投げ込めば答えが返ってくるまでには成熟していない。文章であれば、大きく分けてまずデータの基礎的クリーニングを行い、意味づけを行い、分析を行うという過程が必要になる。本研究では、文章にデータとしての意味づけを行う過程の一つを行った。この方法にはさまざまなアプローチが考えられる。一般的に言語処理では、対象となる文書に応じた「辞書」を作成することが必須となるが、今回はこの辞書を、一定の統計処理により拡充する試みを行ったものである。各都道府県の地域保健医療計画から分析する17因子を研究者が任意に決定し、その同義語となるフレーズや単語を探索し、定義する処理を行った。この作業はWEXでは相関分析機能と呼ばれ、WEXの特徴の一つである。具体的には初期に定義した因子に対して、同文書内に出現するフレーズや単語の頻度や、相対的な傾向分析を行うことで、同義語となるものを探索・抽出させる過程である。ここでは「相関分析」と呼ばれる手法、具体的には一つの章を一つ（あるいはある特定）の事項の説明単位と考えることで、同一章内に出現する頻度と、文章全体で出現する頻度とを比較することで、類似した概念を表すフレーズや単語の可能性を数値的に示す手法を用いた。

C. 研究結果

47都道府県の地域医療計画をダウンロードし、解析の準備を行った。公開されている文書はすべてPDF形式であり、これらをテキスト形式に変換した。変換は汎用PDF読み取りあるいは編集ソフトウェア

を利用しつつ行ったが、3 都道府県のファイルにはテキスト情報が一切含まれておらず、変換を断念し、結果的に 44 都道府県のデータを解析対象とした。

本研究では「一次予防、健診・早期発見、体制」についての検討を行うことを想定し、初期の因子を設定した。まずこれらの一般的同義語を一般的辞書より抽出し、機械の初期学習に与える（表 1）。同時に、文章のクリーニングを行う。文章を章に分割し、さらに文単位に分割する。WEX により、文書を形態素解析し、フレーズや単語に分解した。ここでまず初期の因子の存在をフラグする（表 2）。続いて、これら初期の因子と、「相関」が高いフレーズや単語を抽出した。

具体的には（表 3）に示すように例えば「体制」を表すものとして設定した「維持期」を初期の因子とした場合、「回復期」では非常に高い相関を認め類義語として候補に挙げた。ここで示すように頻度だけではなく、文章全体での頻度との比をもとにした相関を係数としてあらわしている。例えば「患者」や「治療」なども頻度としては高いものに分類されるが、文章全体でも比較的頻出しているため、相関の係数としては低く、類似するものの候補からは外してもよい例となる。

D. 考察

本研究では、自然言語解析を行うための準備としての、辞書の拡充を行うための手段について、機械的な処理の応用を試みた。結果的に様々な類似語を同定することができた。この中にはいわゆる正しくない結果も含まれているため、これらの中からさら

に人の目によるなど何らかの方法での選出が必要になるものの、研究者が当初予測していないようなフレーズも抽出されてくるため、辞書の「機械的」な拡充を行うための非常に有力な手段の一つとして活用できたと考えられる。

この解析をさらに進めていくことで、この作業をシステムティックに繰り返していくことで、さらに特徴的な表現を拡充あるいは洗練することが可能になると考えられ、ここで構築された「辞書」を用いることで、「医療計画」だけではなく、それに関連する文書の言語解析を行うための有用な基盤になることが期待される。

一方で、今回の解析では、複数都道府県の文章データは、そのデータの性質上利用することができなかった。本解析の有用性を説明することで、行政の方々にもご理解いただき、計画書をはじめとするデータ提供のあり方についても、フォーマットの統一や、少なくともデータとしての活用しやすいフォーマットの提唱などを行っていけると考えている。具体的には、PDF だけではなく、テキストとして認識しやすい TEXT や Microsoft Word などの形式での並行した頒布がまずは期待されることのほか、少なくとも PDF にはテキストをフォントとともに埋め込むことなどのソフトウェア的な提唱なども可能と考えられた。

E. 結論

本研究では、地域医療計画の文書解析を行うための「辞書」の機械的拡充を行った。

これら自然言語解析をより発展させること通じて、全国の医療計画の地域間比較・参照を容易にし、計画内容の向上に資する

ことが期待される。

F. 健康危険情報
特になし

G. 研究発表
特になし

(表1) 初期に任意に設定した因子(3因子)に対する、一般的同義語を設定

#	大分類	小分類	同義語			
1	一次予防	食生活	食事	粗飯	夕飯	昼御飯
1	一次予防	栄養	栄養	滋養分	養い	栄養メ
1	一次予防	運動	運動	スポーツ	エクササイズ	エキサ
1	一次予防	休養	休養	息抜	休止	ひと休
1	一次予防	睡眠	睡眠	ねんねこ	御寝	就眠
1	一次予防	飲酒	飲酒	呑む	飲む	一献
1	一次予防	喫煙	喫煙	スモッキング	一服	タバコ
1	一次予防	嗜好品	嗜好品	ぜいたく品	贅沢品	
2	健診、早期発見	特定健診	検診	健康診断	健診	人間ド
2	健診、早期発見	保健指導	保健指導			
3	体制	地域連携	連携	協同	協調	和協
3	体制	かかりつけ	かかりつけ	掛り付け		
3	体制	急性期	急性期			
3	体制	回復期	回復期			
3	体制	維持期	維持期			
3	体制	リハビリ	レハビリ	リハビリテーショ	社会復帰	リハビリ
3	体制	在宅医療	在宅医療			

(表2) 文章を章ごと、さらに文に分け、それぞれに初期因子の同義語の存在を検索する。

文章	大分類	小分類	単語
道においては、北海道医師会等の関係団体や体育大学、医療機関、市町村などで精	-	-	-
(1) 設定21圏域(現状維持) (2) 考え方 第一次医療圏のサービスの提供機	-	-	-
この圏域は、医療法第30条の4第2項第9号に規定する区域とし、医療資源の適	-	-	-
* 1平成24年3月30日付医政発0330第28号厚生労働省医政局長通知「医療計	-	-	-
こうした状況のもと、現在も各圏域において「自治体病院等広域化・連携、構想」や「地	体制	地域連携	連携
ただし、次期北海道医療計画(平成30年度?)の策定に向けて、北海道総合保健	-	-	-
【医療の需給状況の改善に向けた具体的な取り組み】 現行計画の「道民の医療に対	-	-	-
「地域医療再生計画」に基づき、各圏域における地域医療提供体制の確保を図る取	-	-	-
道の「自治体病院等広域化・連携構想」に基づく地域の議論を具体化するため「地域	体制	地域連携	連携
医療計画に基づき、第二次医療圏ごとに地域の実情を踏まえた「地域推進方針」を	体制	在宅医療	在宅医療
3第三次医療圏の設定とその考え方 (1) 設定6圏域 (2) 考え方 高度で専門的	体制	地域連携	連携
この圏域は、医療法第30条の4第2項第10号に規定する区域とし、厚生労働省令	-	-	-
-5【医療圏の区域】第三次第二次第一次南渡島函館市、北斗市、松前町、福島町、	-	-	-
第5節基準病床数 ○ 基準病床数は、医療法第30条の4第2項第11号に基づく	-	-	-
第5節基準病床数 基準病床数は、医療法第30条の4第2項第11号に基づく	-	-	-
なお、既存病床数が基準病床数を超える圏域においては、今後、新たな病院または有	-	-	-
1療養病床及び一般病床の基準病床数 療養病床*1及び一般病床*2は、第二	-	-	-

(表3) 「体制」を表す初期因子「維持期」と高い相関を示す単語の例

維持期	▼ 頻度 ▼	相関 ▼↓
回復期	495	42.12849
急性期	455	20.42662
リハビリテーション	379	17.60454
クリティカル	151	14.58668
共有する	131	13.89536
切れ目	129	13.88621
脳卒中	297	13.81873
パス	189	12.91645
地域連携	195	12.90417
介護サービス	132	12.37631
維持	171	11.87398