大規模医療データの分散表現に基づいた医療費の地域差要因の可視化

吉本 廣雅 [1]　満武 巨裕 [2]　合田 和生 [3]

日本は高齢化に伴い死亡者が増加する高齢多死社会を迎える．介護や終末期医療といった高齢者への医療サービスの安定供給と，その格差の是正が重要な社会問題となっている．本稿は社会問題の実例として医療費の地域差を取り上げ，自治体が保有する膨大なデータから地域差の要因を定量的に抽出し可視化する手法を提案する．この手法は，まず地区の住民に関する医療・介護サービスの利用履歴からなるデータを分散表現ベクトルへと変換するモデルを作成し，クラスタリングにより地域差の要因を分析する方法である．本稿は，複数のモデルの性能評価実験と分析実験の結果について報告し，提案法の有効性について考察をまとめる．

1　はじめに

日本の高齢化率は世界的にも高水準である．団塊の世代が平均寿命に達する今後は高齢多死社会を迎えると予測されている．厚生労働省の人口動態統計によれば，2020 年に約 140 万人だった年間死亡数は，2040 年には約 170 万人にまで増加すると推計されている [12]．これは高齢化の進行に伴い，医療・介護サービスの需要がさらに拡大することを意味する．一方で，地方の過疎地域では高齢化に伴う多死社会，すなわち人口減少がすでに始まっている．その地域では医療需要が減少し，病院の経営難・廃院が社会問題となっている．このように世界に先駆けて高齢多死社会を迎える日本は，医療・介護サービスの需要と供給バランスが劇的に変動するという前例のない事態に直面しつつある．その対処策の検討は重要な社会的課題の一つであると言える．

戦後日本は国民皆保険，つまり国民全員が医療機関を自由に選択でき，高度な医療サービスを安価に利用できる体制を整備してきた．医療・介護サービスへの需給バランスを国は監視し，格差が小さくなるように国と各地の医療機関が連携している．たとえば厚生労働省は地域差指数と呼ばれる指数を算出している．地域差指数は，地域の住人 1 人当たりの医療費の平均値を表しており，これにより西日本の医療費は全国平均より高く，東北，関東，中部地方では低いという“西高東低”の傾向があることが示されている [11]．

厚生労働省や各自治体には医療関連データとそれを集計するツールが存在しているものの，それら情報を横断的かつ体系的に分析する手法やツールは十分に整備されていない．例えば前述の地域差指数を計算することはできても，その差の要因を分析し明らかにすることはまだ困難である．そこで厚生労働省は，国や自治体が保有する全国民のデータを研究者や事業者が活用し社会実装できる体制づくり（レセプト情報・特定健診等情報データベース，NDB）などを整備し，公益目的でのヘルスケアデータの分析や利活用を促進している [15]．地方自治体でも同様な取り組みがなされており，住民のヘルスケアデータを活用した様々な研究成果が発表されている [5, 6, 8, 13, 14] ．

本研究は，社会問題の実例として医療費の地域差を取り上げ，自治体が保有する膨大なデータから地域差の要因を定量的に抽出し可視化する手法を提案する．医療費の地域差やその要因を可視化・数値化できれば，それは各地域の医療政策立案へのエビデンスとしての活用が期待できる．以下，本稿は，提案の概要と予備実験の結果を示し，提案法の特性・有効性についての考察をまとめる．

2　大規模医療データの分散表現

本稿は，自治体が保有する大規模データとして，1）被保険者台帳，2）医療・介護レセプト，の 2 つを処理対象とする．

被保険者台帳は，健康保険に加入しているすべての住民に関する，ID，性別，年齢，加入期間（加入資格取得日と加入資格喪失日の対）の一覧である．後述するが，本実験では岐阜県全域について，国民健康保険と後期高齢者保険の被保険者台帳を用いており，後期高齢者保険の加入者（75 歳以上のほぼすべての住民），国民健康保険の加入者（75 歳未満の住民の約 20%）が含まれたデータとなっている．

医療・介護レセプトは，病院などの医療機関で医療サービスを受けた際の明細書である．各患者 ID に対する治療行為，レントゲンなどの医療器材の利用，処方した薬の製品名や数量などと，その請求額を記録したものである．

データ構造としては，医療・介護レセプトは，独自のコード体系をもっており，すべての医療サービスには一意のコードを割り当ててあり，それらコードの可変長の系列データで医療・介護サービスの利用履歴を記録している．医療レセプトのコード体系は複雑であるが，今回の実験では以下のコードを処理対象とした．

- SY コード：糖尿病や胃がんのような傷病名に対応する
- IY コード：インスリンのような医薬品の製品名に対応する
- TO コード：ガーゼのような機材に対応する
- SI コード：縫合や注射のような診療行為に対応する
- SS コード：抜歯のような歯科診療に対応する

各コードは最大 11 桁の整数値で表現され，定義済みコードは約 6 万種類ある．ただし，例えば IY コードは，医薬品の製品名毎に定義されており，廃止された医薬品はそのままコードも廃番となるなど，廃番となったコードも多数ある．またコロナのような新しい病名，新しい治療薬が随時新規コードとして登録されるためコード数はデータの収集期間により変動する．今回の実験で用いたデータセットに含まれるコードは 35,207 種類であった．

可変長のコード列という意味で，医療・介護レセプトは自然言

[1] 正会員　医療経済研究機構
yoshimoto@tkl.iis.u-tokyo.ac.jp
[2] 非会員　医療経済研究機構
mitsutake@ihep.jp
[3] 正会員　東京大学生産技術研究所
kgoda@tkl.iis.u-tokyo.ac.jp

語処理と似た構造，似た課題を持つ．たとえば高血圧の診断に対応するコードが履歴内にある場合でも，その症状の度合いは患者によって異なり，一つのコードだけでは患者の健康状態は特定できない．しかし高血圧に対応するコードの後ろに高血圧の薬の処方に対応するコードが複数あれば，それらコードの共起関係（薬の組み合わせ）から高血圧症の病状や健康状態を類推できるようになる．つまり自然言語処理における分布仮説がそのまま適用できる．

本研究は分布仮説に従った方法として Word2Vec [7] を用いた分散表現モデルを用いる．分散表現とは単語をベクトルで表現する手法である．ベクトルの各基底は似た使われ方をする単語は似た方向のベクトルになるよう調整されており，これにより似た意味をもつ単語は似たベクトルに変換される性質を持つ．つまりベクトル間のコサイン類似度を求めることで，単語間の意味的な類似度が計算できるようになる．本研究は，ある期間内のレセプトを Word2Vec における一つの文章と捉え，そのレセプトに含まえるコードを Word2Vec における単語と捉え，コードをベクトルに変換する分散表現モデルを得る．

分散表現モデルが得られれば，たとえば，類似した医療・介護レセプトや類似した患者，などのクラスタを抽出できる．このクラスタを基準に，医療レセプトに含まれるコード以外の情報，医療費（点数）や，住民が加入している地域の保険行政の区分コードを分析することで，地域間の医療費の差やその要因の分析が可能になると期待できる．

3 実験１：分散表現モデルの生成と検証

3.1 データセットと実験環境

実験は岐阜県から受領したデータセットを用いた．データは 2014 年 4 月から 2023 年 3 月までの期間において，国民健康保険または後期高齢者に加入した住民についての被保険者台帳と，医療・介護レセプトである．住民数は，のべ 3,136,613 名，データサイズは約 550GB であった．データは匿名化加工がなされており，氏名や住所が削除され，個人が特定出来ない状態になっている．そのうえで研究倫理規定に従って実験を実施した．

実験にあたり先ず，無作為に $\frac{1}{8}$ の住民を抽出した small データセットと，すべての住民を抽出した large データセットの２つを用意した．

実験で用いた計算機の構成は以下の通りである．CPU は Intel Xeon Gold 6430（2 コア，128 スレッド），メモリは 512GB，OS は Linux，プログラムの実装には C++ 及び python を用いた．

3.2 分散表現モデルの生成

手順は以下の通りである．まず下記の手順で患者ごとのコード列を得る．

- 台帳情報の抽出：各患者について，保険への加入資格取得日と加入資格喪失日を抽出した．保険の資格取得は，出生時に新規に加入するケース，転入により加入するケースがある．保険の資格喪失は転出，死亡のケースがある．そこで，各患者について（出生，転入，転出，死亡）のコードを付与し，コード列を作成する．
- 患者 ID 単位での医療介護レセプトの抽出：各患者について，医療介護レセプトを集計し，医療・介護サービスのコードをコード列に追加する．
- 上記のコード全てを時系列順にソートし，患者のコード列とする．

次に患者のコード列を用いた学習データで分散表現モデルを得る．モデルは skip-gram を使い，パラメータとして，モデルの次元数 D，ウインドウサイズ ws の組み合わせを複数検証した．

表 1: モデルの内的評価

データセット名（保険加入者数）	モデルパラメータ D	ws	r_s
Small (392,072)	5	5	-0.85 *
	5	15	-0.85 *
	25	5	-0.85 *
	25	15	-0.82 *
Large (3,136,613)	25	5	-0.84 *
	25	15	-0.79 *
	25	25	-0.78 *
	25	50	-0.79 *
	50	5	-0.77 *
	50	15	-0.85 *
	50	25	-0.77 *
	50	50	-0.79 *

3.3 分散表現モデルの性能評価

Word2Vec などの分散表現モデルの評価手法は「内的評価」と「外的評価」に大別できる [10]．本実験は内的評価の指標として，２枚のレセプトの記載内容が類似した場合はレセプトで請求される医療費の金額が同額に近づくという関係を考え，この関係の有無を相関係数を用いて検証した．

医療費には外れ値や非線形な関係が含まれるため相関係数は Spearman の順位相関係数 r_s を採用した．つまり r_s が -1 に近づくほど「レセプト間でコサイン類似度が高いほど，医療費差は小さくなる」という対応関係が強いことを意味する．計算は患者 10,000 人をランダムに抽出し，患者間のレセプトについてコサイン類似度と医療費の差を求め，r_s を計算した．有意水準は .05 とした．検定の結果を表 1 に示す．

表 1 は，モデルのパラメータ（学習データ数，モデルの次元数，ウインドウサイズ (ws)）毎の相関係数を示している．例えば，$3,136,613$ 名分のレセプトから，次元数 50，ws15，で生成したモデルでは，コサイン類似度と医療費差の間には統計的に有意な負の相関（$r_s = -0.85$，$p < 0.05$）が認められた．これはレセプトの分散表現が類似しているほど，請求される医療費の差も小さい傾向にあることを示唆している．

本稿では以下，データセット Large から生成した次元数 50，ws 15 のモデルを用いて実験を行う．このモデルを選択した理由は，

表 2: 「“死亡” – “胃がん”」の結果．類似度が高いコード 13 個を示す．

コード	コードの説明
190024510	救命救急入院料１（３日以内）
190139470	充実段階Ａ加算
140000490	時間外特例加算（処置）
140010210	非開胸的心マッサージ
140000290	休日加算（処置）
190138110	救命救急入院料３（救命救急入院料）（３日以内）
140000390	深夜加算（処置）
140000190	時間外加算（処置）
8842543	来院時心肺停止
190217370	救急体制充実加算２（救命救急入院料）
140009010	救命のための気管内挿管
140040610	四肢ギプス包帯（半肢）（片）
5118002	血気胸

表 3: 「“死亡” – “胃がん” + “糖尿病”」 の結果．類似度が高いコード 13 個を示す．

コード	コードの説明
08833421	高血圧症
902724007	高脂血症
904279016	不整脈
904293004	心拡大
904289015	心不全
160167250	ＬＤＬ－コレステロール
902500015	２型糖尿病
902720004	高コレステロール血症
160010010	ＨｂＡ１ｃ
908844446	脂質異常症
612170709	ノルバスク錠２．５ｍｇ
160023410	ＨＤＬ－コレステロール
610432011	ブロプレス錠２　２ｍｇ

まず r_s が-1 に近いこと，また次元数が高いことの２点にある．次元数については，メモリ消費量・必要な演算量の削減という点では次元数は小さいほうが有利である．また一般論として過学習などの諸問題に対しても次元数は小さい方が好ましい．一方，コサイン類似度の一致不一致やその背後にあるであろうベクトルの「意味」をさまざまな次元・基底で精緻に分析できるという点では次元が高い方が有利とも言える．以上のトレードオフを念頭に，本稿では次元数 50 のモデルを用いて実験を行った．

3.4 分散表現の演算例

Word2Vec の有用性を示す有名な例として「“king”– “man” + “woman” = “queen”」がある．つまり分散表現を用いることで意味の演算が可能になる．医療データについて同様の演算を行った例を２つ示す．

表 2 は「“死亡” – “胃がん”」の結果である．演算は，減算して得たベクトルについて，コサイン類似度が高いコードを検索する処理となる．表は，コサイン類似度が高いコードの上位 13 個を列挙している．各コードには，ラベルとして厚生労働省が作成したマスタデータから抽出した文字列を併記している．ラベルの文字列をみると「救急」や「時間外」「深夜」という単語が散見され，減算結果は想定外のタイミングで病院に担ぎ込まれた状況を想起させる結果となっている．

表 3 は「“死亡” – “胃がん” + “糖尿病”」の結果である．上位のコードは高血圧，高脂血症であり典型的な生活習慣病である．また不整脈，心拡大と続き循環器系が死因であると推察できる．これは自殺や事故死を除外すると，死因の上位は，がん及び生活習慣病になるという従来からある疫学的知見 [9] とも合致している．

重要な点は，モデル生成時にはこれら「コードの説明」に含まれる情報は一切付与していない点である．つまりこの２例の結果は，医療・介護レセプトのデータでも分布仮説が適用でき，相応の分散表現が利用できることを示唆している．

4 実験２：地域差要因の可視化

本実験では，分散表現を用いた患者のクラスタリングを行い，各クラスタにおける地域差を可視化した．

4.1 患者の分散表現の獲得

以下の手順で，各患者について 50 次元の分散表現ベクトルを取得した．

1. 各患者のレセプトの時系列から，タイムウインドウ内のレセプトを抽出する．実験では，タイムウインドウは 12 ヶ月とした．
2. 抽出したレセプトをそれぞれベクトルへ変換する．
3. Smooth inverse frequece(SIF) の方法 [1] を用いて，ベクトルの重み付け和を計算し，各患者のベクトルを得る．

4.2 クラスタリング

患者ごとのベクトルに対して，以下の手順でクラスタリングを行い，患者クラスタを得る．

1. 主成分分析により 50 次元のベクトルを次元削減する．
2. 次元削減したベクトル群に対してクラスタリングを行う．クラスタリングのアルゴリズムは mean shift clustering [3] を採用した

4.3 地域差の可視化

クラスタリング結果を用いて，各患者クラスタの地域差を可視化した．手順は以下の通りである．た．

- 台帳に含まれる地域コードにより，患者が加入している保険者番号（地域の市町村名に対応する）を得る
- 各クラスタに属する患者数を地域ごとに集計し，地図上に可視化する

可視化結果として，以下の図に主要な 3 つの患者クラスタの地

表 4: 各グループに対応するレセプトのコードと，コードから推察される患者像

グループ	コサイン類似度が高いレセプトのコード	推察される患者像
Group #0	外来管理加算, 特定疾患療養管理料（診療所）, 糖尿病，高血圧，高脂血症，腰痛症, 胃炎, 変形性腰椎症	生活習慣病（高血圧，高脂血症，糖尿病）の患者
Group #1	人工腎臓（慢性維持透析）, 透析液水質確保加算２, 人工腎臓（慢性維持透析），ダイアライザー，レグパラ錠２５ｍｇ	人工透析の患者
Group #2	看護補助加算２, 調剤料（入院）, 食堂加算（食事療養）, 入院時食事療養（１）, 療養環境加算, 摘便	医療的なケアが必要な長期入院患者

理的分布を示す．

- 図 2: 生活習慣病（高血圧、高脂血症、糖尿病）の患者分布
- 図 3: 人工透析患者の分布
- 図 4: 医療的ケアが必要な長期入院患者の分布

5 患者のクラスタリング

図 1 にクラスタリングの結果を示す．図の横軸，縦軸は主成分分析で得た第一主成分，第二主成分であり，図中の各点は患者のベクトルに対応する．

Mean shift クラスタリングの結果，ベクトルは 21 のクラスタに分類された．医療・介護サービスの観点で各クラスタの意味が解釈できるように，コサイン類似度を用いて各クラスタに属するベクトルに近いレセプトのコードを列挙した結果を表 4 に示す．表 4 に示すように，各グループには特徴のあるレセプトのコード，すなわち医療行為や薬剤の名称が含まれている．これら名称から，Group #0 は「生活習慣病（高血圧，高脂血症，糖尿病）の患者群」，Group #1 は「人工透析の患者群」，Group #2 は「医療的なケアが必要な長期入院患者群」と解釈した．

6 地域差の可視化

最後に各グループに属する患者群について，地域ごとに患者の割合を集計し，地図上に重畳表示する可視化を行った．結果を図 2，図 3，図 4 に示す．

図 2 は，生活習慣病（高血圧，高脂血症，糖尿病）の可視化結果である．これは，地域毎に，総患者数に占める生活習慣病の患者の割合を可視化したものである．図をみると岐阜県全域が同色（約 80%）で覆われている．この結果から，生活習慣病の患者の割合に地域差が無いことが確認できる．

図 3 は人工透析の患者の割合を可視化した図である．この図では，オレンジ色の部分が患者の割合が高い地域を表しており，人工透析の患者の割合には地域差があることを示している．オレンジ色の地域には大病院があり，人工透析が可能な病院がある地域では，人工透析の患者が多くなる傾向があることが一目で把握出来る結果となっている．また，黒色の地域は人工透析の患者がおらず，これら地域には人工透析が可能な大病院が存在しないという事実と対応する結果となっている．

図 4 は医療的なケアが必要な長期入院患者の割合を可視化している．こちらも図 4 と同様，長期入院が可能な医療機関が多い地域は，患者の割合が増加している．この結果は，医療資源と患者数に一定の相関があることを示唆するものである．

7 考察

7.1 レセプトデータと分散表現

実験１では，レセプトデータに特化した分散表現モデルの生成とその有効性を検証した．結果，レセプトデータにおいても分布仮説が適用可能であり，分散表現モデルの構築が可能であることを実証した．

7.2 分散表現に基づいた地域差要因の検出と可視化

本稿の提案法は，分散表現を用いて患者をクラスタリングし，その結果を地図上にマッピングして表示するシンプルな方式である．これは，教師なし機械学習の一種であり，学習データに事前の正解ラベルを与えない形で分析を行う方式である．

実験結果から，医療機関の有無といった医療資源に関する情報を明示的に与えなくとも，地域ごとの医療資源の分布が浮かび上がるような可視化結果が得られることが確認された．これは，提案手法がデータに含まれる知識を抽出し，定量化出来る可能性を示唆している．ただし，本稿の実験結果は新たな知見を示すものではなく，既存の知見をデータ駆動型の手法で確認した試みとして位置づけることができる．

7.3 先行研究との関係

医療情報を分散表現で扱う先行研究としては Med2Vec [2] がある．Med2Vec は｛患者の性別，年齢，傷名コード｝の系列を文章とみなし，傷名コードの部分を分散表現でエンコードする手法である．一方，本研究は，病名のコードに加え，医薬品，注射器のような医療器材，さらに初診料や深夜加算のような日本の医療サービス固有の情報も含めた多様なコードをエンコードしている．動作原理的には Word2Vec のサブセットに過ぎないが，実際の医療データに対する実用性を評価した点で，Med2Vec と本提案は相補的な意義があると言える．

次に，自然言語処理が対象とするテキストデータと，本研究が対象とする医療データの類似点と相違点について整理する．類似点は，両者は可変長のトークン（コード）の列であり，単一のトークンだけでは「意味」を一意に特定できない点が挙げられる．例えば，高血圧という診断名のコードが付与された患者でも，患者によって血圧や体重は異なり，健康状態は異なる．

一方，相違点としては，近年の自然言語処理 [4] はトークン数が数十億ある問題を対象にしている一方で，本提案や Med2Vec

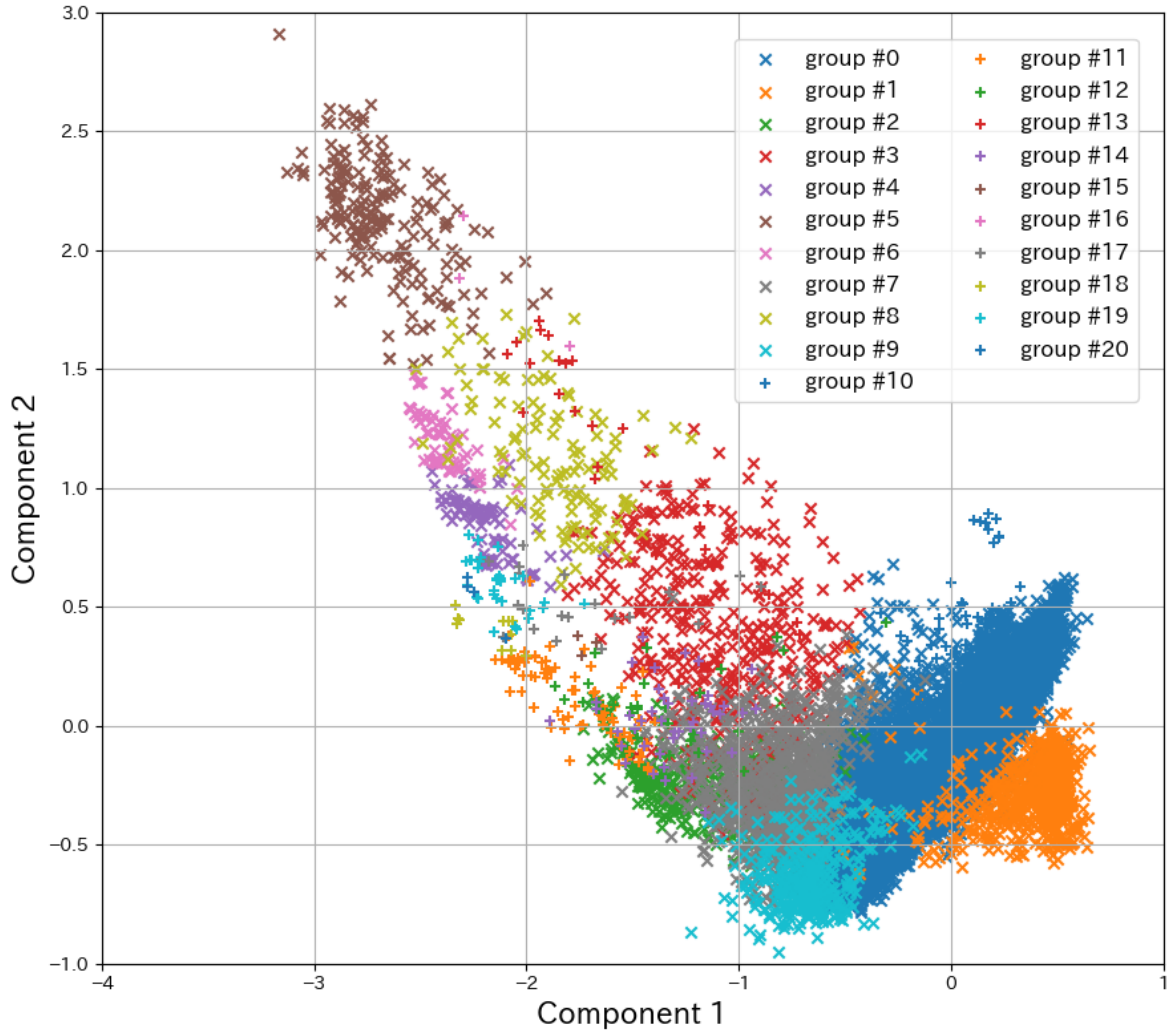

図 1: レセプトのクラスタリング結果

が扱うトークン（＝コード）の種類は高々数万種類であり，コードのバリエーションは圧倒的に少ない．このため医療データは自然言語と類似の課題を抱えるが，トークンの次元が圧倒的に小さい点が特徴的である．今後，この特徴を活かしつつ，機械学習や自然言語処理の知見を活用し，医療情報処理に特化した手法の開発が求められる．

8 まとめと今後の課題

本稿は社会問題の一例として医療費の地域差を取り上げ，自治体が保有する膨大なデータから地域差の要因を定量的に抽出し，可視化する手法を提案した．提案手法は，地区の住民に関する医療・介護サービスの利用履歴からなるデータを分散表現ベクトルへと変換し，ベクトルのクラスタリングにより地域差の要因を分析する処理を基本的枠組みとしている．

今後の課題としては，考察で言及したように医療データに特化した分散表現の最適化が重要であると考えられる．具体的には，分散表現モデルのパラメータ（次元数やウインドウサイズ）などに関する医療情報処理固有の指標やその評価方法を明らかにすることなどが早急の課題であると考えられる．また医療・介護・医療行政に関する知見を有する専門家らとともに，結果の妥当性を様々な観点から検証する計画である．

謝辞

本研究の一部は，内閣府総合科学技術・イノベーション会議の戦略的イノベーション創造プログラム (SIP) 「統合型ヘルスケアシステムの構築」の補助を受けて行なった．

参考文献

[1] Sanjeev Arora, Yingyu Liang, and Tengyu Ma. A Simple but Tough-to-Beat Baseline for Sentence Embeddings. In *International Conference on Learning Representations*, February 2017.

[2] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. Multi-layer Representation Learning for Medical Concepts. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1495–1504, New York, NY, USA, August 2016. Association

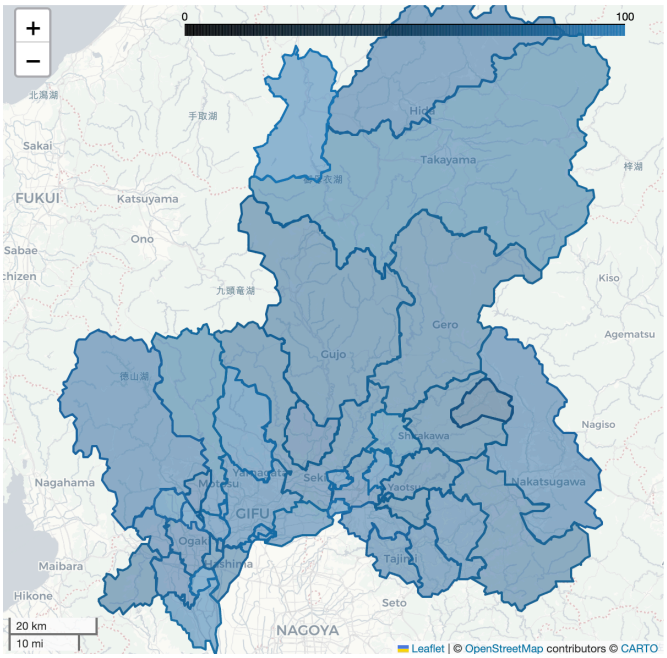

図 2: (Group #0) 生活習慣病（高血圧，高脂血症，糖尿病）の患者の分布

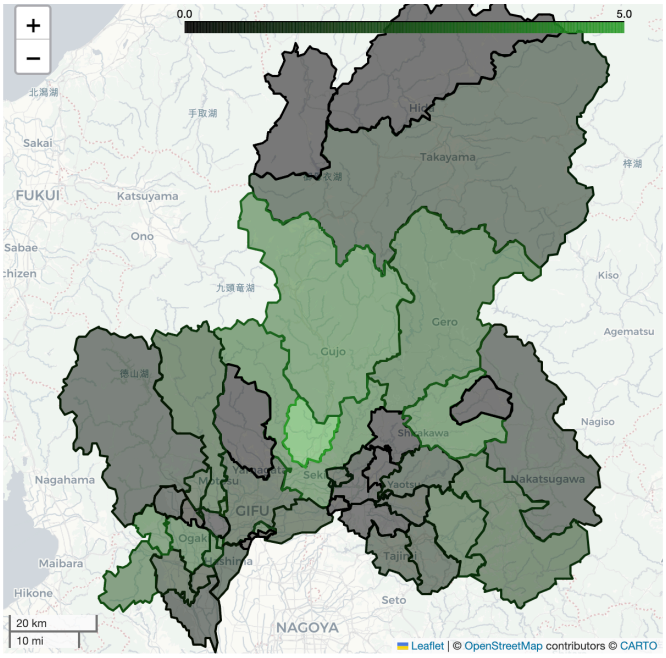

図 4: (Group #2) 医療的なケアが必要な長期入院患者の分布

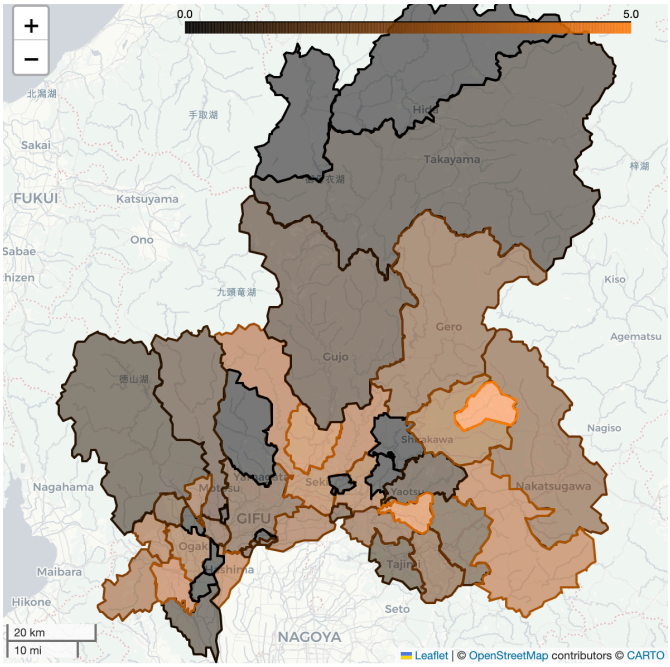

図 3: (Group #1) 人工透析の患者の分布

for Computing Machinery.
[3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.
[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
[5] Kaho Hirayama, Naoki Kanda, Hideki Hashimoto, Hiromasa Yoshimoto, Kazuo Goda, Naohiro Mitsutake, and Shuji Hatakeyama. The five-year trends in antibiotic prescription by dentists and antibiotic prophylaxis for tooth extraction: A region-wide claims study in Japan. *Journal of Infection and Chemotherapy: Official Journal of the Japan Society of Chemotherapy*, 29(10):965–970, October 2023.
[6] Naoki Kanda, Hideaki Hashimoto, Imai, Hiromasa Yoshimoto, Kazuo Goda, Naohiro Mitsutake, and S. Hatakeyama. Indirect impact of the COVID-19 pandemic on the incidence of Non-COVID-19 infectious diseases: A region-wide, patient-based database study in Japan. *Public Health*, 214:20–24, January 2023.
[7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*, September 2013.
[8] Jumpei Sato, Naohiro Mitsutake, Masaru Kitsuregawa, Tomoki Ishikawa, and Kazuo Goda. Predicting demand for long-term care using Japanese healthcare insurance claims data. *Environmental Health and Preventive Medicine*, 27:42–42, 2022.
[9] Shu-Yu Tai, Soyeon Cheon, Yui Yamaoka, Yu-Wen Chien, and Tsung-Hsueh Lu. Changes in the rankings of leading causes of death in Japan, Korea, and Taiwan from 1998 to 2018: A comparison of three ranking lists. *BMC Public Health*, 22(1):926, May 2022.
[10] Bin Wang, Angela Wang, Fenxiao Chen, Yuncheng Wang, and C.-C. Jay Kuo. Evaluating Word Embedding Models: Methods and Experimental Results. *APSIPA Transactions on Signal and Information Processing*, 8(1), 2019.
[11] 印南 一路. 医療費の決定構造と地域格差. **医療と社会**, 7(3):53–82, 1997.
[12] 厚生労働省. 人生の最終段階における医療・介護 参考資料. https://www.mhlw.go.jp/content/12404000/001104699.pdf.
[13] 吉本 廣雅, 満武 巨裕, and 合田 和生. 異種医療データの融合による医療需要予測手法の検討. In **第 *15* 回データ工学と情報マネジメントに関するフォーラム** *(DEIM 2023)*, March 2023.
[14] 吉本 廣雅, 満武 巨裕, and 合田 和生. GPU を用いた k –匿名化加工処理の高速化の検討. In **第 *16* 回データ工学と情報マネジメントに関するフォーラム** *(DEIM 2024)*, March 2024.
[15] 満武 巨裕. ウェルネスのための ICT：2. 日本のレセプト情報・特定健診等データベース（NDB）の有効活用. **情報処理**, 56(02):140–144, January 2015.