

医療・介護レセプト情報に基づく 将来の要介護状態の推定方法の検討と実験に基づく考察

佐藤 淳平[†] 合田 和生[†] 喜連川 優[†] 満武 巨裕[‡]

[†] 東京大学 生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

[‡] 医療経済研究機構 〒105-0003 東京都港区西新橋 1-5-11 第 11 東洋海事ビル 2F

E-mail: [†] {jsato, kgoda, kitsure}@tkl.iis.u-tokyo.ac.jp, [‡] mitsutake@ihp.jp

あらまし 国家や企業に於いて、蓄積されたデータの解析結果に基づく組織の運営方針の決定が行われつつある。医療の分野では、公的保険制度のもと、我が国で提供される医療・介護サービスの殆どが電子レセプト情報と呼ばれる請求データとして蓄積されている。労働人口の減少が見込まれる我が国に於いて、現行の医療・介護サービス提供体制を維持するには、当該データを活用した医療・介護サービスの需要の推定と推定結果に基づく医療政策の立案・実行が求められている。特に、今後 20 年間に渡り需要増加が見込まれる介護サービス需要を精緻に推定し、医療・介護資源を適切に配分することは医療政策に於いて重要である。本論文では、将来の介護サービス需要の推定を可能とすること目的とし、医療・介護レセプト情報に基づく将来の要介護状態の推定手法を提案し、その推定精度の検証結果を示す。

キーワード 電子レセプト情報, 要介護状態, 推定

1. はじめに

近年、国家や企業に於いて、ログデータ等の様々なデータを取集・蓄積し、蓄積されたデータの解析結果に基づく組織の運営方針の決定が行われ始めている。例えば、医療分野に於いては、公的医療保険制度および公的介護保険制度のもと、我が国に於いて提供される医療・介護サービスの殆どが電子レセプト情報と呼ばれる医療費・介護費の請求データとして継続的に収集されている[1-3]。2018 年度末時点に於いて 9 年分の医療レセプト情報（148.1 億件）と 6 年分の介護レセプト情報（8.6 億件）が医療・介護データベースとして蓄積されている[3]。少子高齢化に伴う労働人口の減少が見込まれる我が国に於いて現行の医療・介護サービス提供体制を維持するためには、これらの蓄積されたデータを活用した将来の医療・介護サービスの需要の推定を行い、推定結果に基づいた医療政策の立案・実行が求められている。特に、今後 20 年間に渡り増加が見込まれる要介護状態の患者に対する介護サービスの需要を精緻に推定し、医療・介護資源を適切に配分することは、我が国の医療政策に於いて重要である[4]。

一般的に、患者が必要とする介護サービスの種類やその量は、患者の要介護度（厚生労働省が定める介護サービスの必要度）ごとに異なる。そのため、将来の介護サービスの需要を精緻に推定するためには、患者の将来の要介護状態を正確に把握する必要がある。これまでに、蓄積された介護情報を用いた将来の要介護状態の推定方式に関する先行研究が報告されている[5-6]。武田は、要介護状態の患者の属性情報（性別や年齢階級、初回要介護度）などを特徴量とし、多変量

ロジスティック回帰を用いた将来の要介護状態の推定方式について報告している[5]。また福岡らは、現在の要介護度と認知自立度を説明変数とし、二項ロジスティック回帰による要介護状態の変化の推定方式について報告している[6]。将来の介護サービスの需要を推定するため、Lee らはステップワイズ法と回帰分析を用いて介護サービスの利用量と関連の高い因子を探索する方式について報告している[7]。先行研究では、介護サービスに関連する情報（以下、介護情報）のみを用いた将来の要介護状態の推定を行っており、医療レセプトなどの診療に関連する情報（以下、診療情報）は使用されていない。しかしながら、診療情報には将来の要介護状態の推定において重要な情報が含まれていると考えられる。例えば、転倒による大腿骨近位部骨折などの骨折やの脳卒中、認知症は生活の自立度（ALD スコア）の低下に繋がり、また要介護度状態となる主要な要因の一つとして知られている[8-12]。そのため、診療情報に骨折や脳卒中、認知症の診断名やその治療に関連する診療行為等の情報を有している患者は、将来的に要介護状態となる可能性が高いと考えられる。先行研究で使用されている介護情報に加え、診療情報に含まれる傷病名や患者に提供された診療行為・医薬品などの情報を用いることにより、患者の将来の要介護状態を高い精度で推定可能になるものと期待される。

本論文では、将来の介護サービスの需要の推定を可能とすることを目的とし、患者の将来の要介護状態を推定する手法を提案し、我が国の特定地域より提供を受けた医療レセプト情報と介護レセプト情報を用いた推定精度の検証結果を示した後、将来に向けた課題と

展望を纏める。本論文の構成は以下の通りである。2.では、医療レセプト情報・介護レセプト情報の概要と要介護度と要介護状態の定義について述べる。3.ではレセプトの診療情報・介護情報に基づく将来の要介護状態の推定方法の提案し、4.にて提案手法の検証結果を示す。5.に於いて将来に向けた課題と展望を示した後、6.に於いて本論文を纏める。

2. 医療レセプト情報・介護レセプト情報と要介護度および要介護状態の定義

2.1. 医療レセプト情報・介護レセプト情報の概要

我が国に於ける公的医療保険・公的介護保険が適用される保険診療・介護サービスに於いて、保険医療機関（病院、診療所等、歯科医院等）及び保険薬局、または介護サービス事業者は、毎月、保険者に対する医療費・介護費の請求書を決められた様式で作成し、請求内容の適切性を判断する審査支払機関へ送付する。その後、審査支払機関での審査を行った上で、保険者への請求を行い、保険者は医療費のうち保険者負担分を医療機関に支払う。この際の請求書は、レセプト（診療報酬明細書；medical insurance claim;（独）Rezept）と呼ばれ、請求元である医療機関・介護サービス事業者の情報に加えて、被保険者毎に、患者の属性情報、傷病名、提供した医療行為や医薬品、介護サービス、診療報酬等が記載される[13]。従来は、紙媒体を用いた請求書処理が行われていたが、平成11年に磁気媒体等による電子的な手段による提出が認められ、審査支払機関でのレセプト電算処理システムの構築に伴い、電子化したレセプト情報を電子媒体で提出することが可能となった。公的医療保険のもと医療機関に於いて作成されたレセプト情報は医療レセプト情報、公的介護保険のもと介護サービス事業者に於いて作成されたレセプト情報は介護レセプト情報と呼ばれる。

医療レセプト情報は、医療機関の種別等に応じて記載様式が定められており、当該様式は主に以下の4つに分類される。

- 内科レセプト：病院や診療所等の医療機関に於いて患者が外来診療もしくは入院診療を受けた際に発行される。
- DPCレセプト：一部の急性期病院に於いて患者がDPC/PDPS制度が適用された入院診療を受けた際に発行される。
- 歯科レセプト：病院や診療所等の医療機関に於いて患者が歯科診療を受けた際に発行される。
- 調剤レセプト：調剤薬局に於いて患者が調剤を受けた際に発行される。

我が国に於ける公的介護サービスは大きく3つに分類されるが、介護サービス種別間で共通化された記載様式が定められている。3つの介護サービスの種別は

以下の通りである。

- 居宅サービス：サービス享受者が、現在の居宅に住んだまま提供を受けられるサービスであり、訪問介護や訪問入浴介護等が当該サービスに該当する。
- 施設サービス：介護老人福祉施設等の特定の介護施設に入所した要介護状態のサービス享受者に対して提供されるサービスであり、介護老人福祉施設入居者生活介護等が当該サービスに該当する。
- 地域密着型サービス：市町村が指定・監督を行うサービスであり、定期巡回・随時対応型訪問介護看護等が当該サービスに該当する。

レセプトは医療費・介護費の請求を元来の目的としているため、臨床検査の値や診断画像等の臨床情報は含んでいない。しかしながら、我が国は公的医療保険制度および公的介護保険制度を採用しているため、労災や自費治療等の例外的なケースを除き、国内で提供される医療行為および介護サービスの殆どは公的保険の対象となっている。そのため、医療レセプト情報と介護レセプトを用いることにより、我が国において提供されている診療行為と介護サービスを横断的に把握することが可能だと考えられる。

2.2. 要介護度と要介護状態の定義

要介護度は、介護サービスの必要度を表す指標であり、厚生労働省が定める要介護認定の審査判定に基づき決定される[14]。要介護状態とは、いずれかの要介護度であると認定された状態である。要介護度と各要介護度に於ける患者の平均的な状態を要介護度が高い順に示す。

- 要介護5：ほぼ寝たきりであり、意思の伝達が困難かつ自力での食事が困難な状態。
- 要介護4：移動には車いす等が必要であり、日常生活に於いて常時介護を要するものの、医師の伝達が可能な状態。
- 要介護3：移動には杖や歩行器等が必要であり、食事や排せつ、入浴の際に介助が必要な状態。
- 要介護2：歩行が不安定であり、買い物や家事、移動などの手段的日常生活動作に介助が必要な状態、または認知症の症状がある状態。
- 要介護1：歩行が不安定であり、手段的日常生活動作の一部に介助が必要な状態。
- 要支援2：歩行が不安定であり、将来的に手段的日常生活動作介助を要する可能性がある状態。
- 要支援1：将来的に手段的日常生活動作介助を要する可能性がある状態。

我が国の介護保険制度では、要介護度別に給付される介護費の上限額が異なる[15]。そのため、要介護状態

である患者は、介護費の上限額を参考として、各々の患者の要介護度に併せた介護サービスの提供を受けることが多い。そのため、将来の介護サービスの需要を精緻に推定するためには、患者の将来の要介護状態を正確に推定することを可能とする手法が必要である。

3. レセプトの診療情報・介護情報に基づく将来の要介護状態の推定方法の提案

本論文では、レセプトの診療情報・介護情報に基づく将来の要介護状態の推定手法を提案する。

厚生労働省が定める要介護認定の審査判定の基準の一つとして、認知症の症状の有無が含まれている[14]。また、転倒による大腿骨近位部骨折などの骨折や脳卒中・認知症の発症はADLスコアの低下に繋がり、また要介護状態となる主要な要因の一つとして知られている[8-12]。即ち、診療情報に認知症や骨折、脳卒中の診断名を有する患者は、将来的に要介護状態となる可能性が高いと考えられる。また、例えば、要介護状態の改善に寄与する介護サービスを受けている患者は、将来的に要介護状態が改善する可能性を有していると考えられる。そのため、先行研究で使用されている介護レセプト情報に含まれる介護サービス種別などの介護情報に加え、医療レセプト情報に含まれる傷病名や患者に提供された診療行為・医薬品などの医療関連情報を用いることにより、患者の将来の要介護状態を高い精度で推定可能となり、将来の介護サービスの需要を推定できると考えられる。

そこで、本手法では、医療・介護レセプト情報の両方を用いて、各患者に紐づく両レセプトに記載された診療情報および介護情報に基づき特徴量を作成する。当該手法で作成する特徴量は以下の通りである。

- 個人の属性情報（性別、年齢）
- 現在の要介護度
- {医科入院、医科外来、DPC、歯科、調剤、介護}レセプトの件数
- {医科入院、医科外来、DPC、歯科、調剤}レセプトに記載された医療費の合計
- 介護レセプトに記載された介護費の合計
- {医科入院、医科外来、DPC、歯科、調剤}レセプトに記載された{傷病名、診療行為、医薬品、特定器材}コードの出現頻度
- 介護レセプトに記載された介護サービス項目コードの出現頻度

当該特徴量から構成される特徴量ベクトルを用いて機械学習に基づく分類器の作成を行うことにより、将来の要介護状態の推定を可能とする。

4. 評価実験

提案方式の有効性を示すため、岐阜県下の地域保険者より提供を受けた医療レセプト情報および介護レセ

プト情報を用いた評価を行う。表1に提供を受けた医療レセプト情報および介護レセプト情報の概要を示す。これまでに著者らは、当該保険者より2014年度から2018年年度の4年分の医療レセプト情報および介護レセプト情報の提供を受けている。本論文では、2017年度の介護レセプト情報を用いた正解ラベルの作成、2016年度の医療レセプト情報（医科入院、医科外来、DPC、歯科、調剤）および介護レセプト情報を用いた特徴量ベクトルの作成を行うことにより、2016年度に於ける患者の状態から2017年度に於ける要介護状態を推定し、提案方式の評価を行う。

4.1. 正解ラベルおよび特徴量ベクトルの作成

表2に2016年度および2017年度の医療レセプト情報・介護レセプト情報に含まれる患者数を示す。本論文では、2016年度に於ける患者の状態から2017年度に於ける要介護状態の推定を行うため、2016年度と2017年度の両年度に存在する287,680名の患者を対象として正解ラベルおよび特徴量ベクトルの作成を行う。

介護レセプト情報には、各月に於ける患者の要介護度が格納されている。介護レセプトに格納されている患者の要介護度は、要介護認定を受ける毎に値が変化する可能性があり、2017年度内に複数の要介護度を有する患者が存在し得る。そこで、各々の患者の介護レセプト情報に格納された要介護度のうち、介護サービスの必要度が最も高い要介護度を、各々の患者の2017年度に於ける要介護度とし、各々の患者に対応する正解ラベルを作成した。この際、2017年度に介護レセプトを有さない患者の要介護度を、介護サービスが必要のない状態：自立として定義した。また、各々の患者の2016年度の医療レセプト情報・介護レセプト情報を使用することにより、各々の患者に対応する特徴量ベクトルを作成した。

表1 岐阜県下の地域保険者より提供を受けた医療レセプト情報および介護レセプト情報の概要

	医療レセプト 情報	介護レセプト 情報
データ期間	2014/04－2018/03	
レセプト数	65.989 Million	7.524 Million
レコード数	1.787 Billion	0.053 Billion

表2 2016年度および2017年度の医療レセプト・介護レセプトに含まれる患者数

	#
2016年度の医療レセプト・介護レセプトに含まれる患者数	310,986
2017年度の医療レセプト・介護レセプトに含まれる患者数	292,484
2016年度と2017年度の医療レセプト・介護レセプトの両方に含まれる患者数	287,680

4.2. 学習方法および評価方法

本論文では、将来の患者の要介護度を推定する分類器の学習モデルとして、Neural Network (NN)とGradient Boosting (GB)を使用する。287,680名の患者のうち無作為に抽出した半数の143,840名の患者に紐づく特徴量ベクトルを学習用データ、残りの143,840名の患者に紐づく特徴量ベクトルを評価用データとして使用する。

複数種類の正解ラベルを分類する分類器の評価指標として一般的に用いられるF値を学習の評価指標とし、学習データに対して3-fold Cross Validationを適用することにより分類器のパラメタの調整および学習を行う。学習の際には、過学習を抑制するため、線形回帰モデルに基づく変数選択により抽出された150個の特徴量を用いて学習を行う。評価用データに対して学習済みの分類器を適用することにより検出精度の評価を行う。評価の指標には、適合率：Precision、再現率：Recall、適合率と再現率の調和平均：F値を使用する。

本論文に於ける分類器は、Python 3.6.1上でNumpyパッケージ(1.12.1)とSklearnパッケージ(0.18.1)を用いることにより作成する。

4.3. 評価結果

各々の学習モデルに於ける各評価指標の値を表3に示

す。評価の結果、GBを学習モデルとして使用した分類器に於いて最も高いF値0.8572となり、NNを用いた学習モデルではF値0.8543となった。Precision, Recallの夫々の値に関してもGBを学習モデルとして使用した分類器に於いて最も高い値を示した。図1に各々の学習モデルに於ける学習曲線を示す。全ての学習モデルに於いて、学習データの増加に伴い、訓練データと交差検証データの精度が近接し、かつ、両者の精度が0.85近傍と高い値で飽和していることから、過学習を伴わず学習ができていると考えられる。

次に、表4に最も高いF値を示したGBを学習モデルとして使用した分類器に於ける、推定結果の混合行列(Confusion Matrix)を示す。表4より、正解ラベルが要介護度：要介護4から要支援1である患者のうち推定を誤った患者の多くは、介護サービス度が近い他の要介護度、もしくは要介護度：自立として推定される傾向を示した。学習および評価に使用したデータは、

表3 各々の学習モデルに於ける Precision, Recall および F 値

	Precision	Recall	F 値
NN	0.8526	0.8604	0.8543
GB	0.8558	0.8628	0.8572

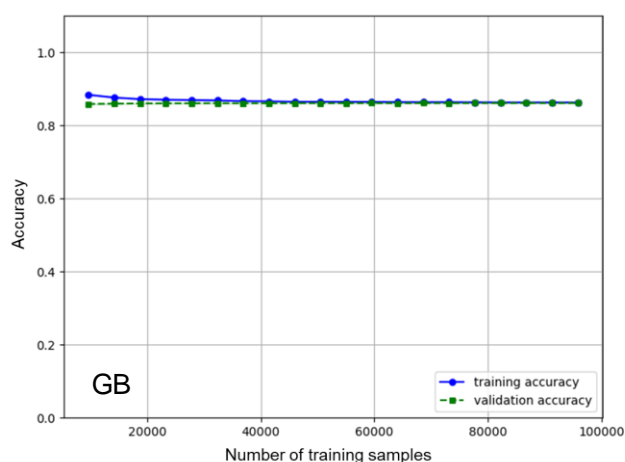
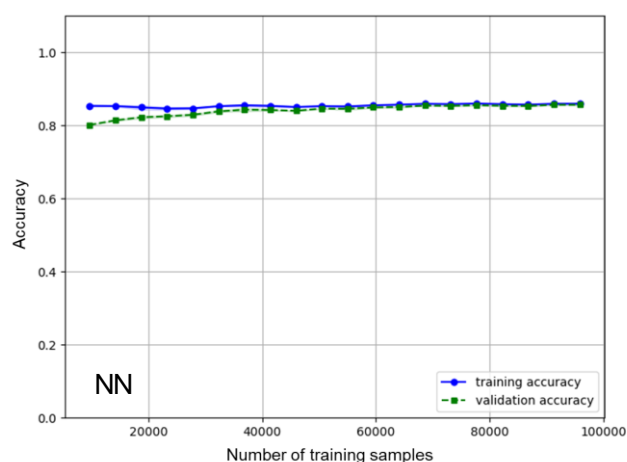


図1 各々の学習モデルに於ける学習曲線 (左：Neural Network, 右：Gradient Boosting)

表4 学習モデル：Gradient Boosting に於ける推定結果の混合行列

		推定							
		要介護5	要介護4	要介護3	要介護2	要介護1	要支援2	要支援1	自立
正解	要介護5	4,466	888	495	276	139	42	23	403
	要介護4	254	4,884	1,074	645	323	111	29	755
	要介護3	66	400	4,994	1,210	540	165	49	801
	要介護2	28	169	450	5,908	1,222	354	109	1,268
	要介護1	12	39	111	460	4,675	451	198	1,803
	要支援2	1	16	35	135	286	3,500	329	1,075
	要支援1	1	3	7	19	96	366	1,749	861
	自立	159	183	162	190	179	161	102	93,936

要介護度：自立の患者数が他の要介護度の患者数よりも多く存在する不均衡なデータであったため、要介護度：自立に対する学習の重みづけが高くなり、推定の精度が低下したと考えられる。今後、学習の重みづけの最適化やアンダーサンプリング・オーバーサンプリングなどの不均衡データに対する処理の追加が必要だと考えられる[16]。

以上の結果より、提案方式を用いることにより、F 値 0.85 以上の精度で将来の要介護状態を推定することが可能であることが示された。

5. 課題と今後の展望

本節では、今後の研究に向けた課題と展望を纏める。

a) 不均衡データに対する処理の検討

本論文では、岐阜県下の地域保険者の医療・介護レセプト情報を使用し、将来の要介護状態を推定する手法の検証を行った。提案手法を用いることにより、患者の将来の要介護状態を F 値 0.85 以上の精度で推定可能であることを示した。しかしながら、表 4 に示した通り、要介護度：要介護 4 から要支援 1 の患者に於いて、要介護度：自立として誤って推定される患者が多く存在した。評価実験に使用した学習データおよび評価データは、要介護度：自立の患者が他の要介護度の患者よりも多く存在する不均衡なデータであった。そのため、要介護度：自立に対する学習の重みづけが他の要介護度よりも高くなり、推定精度が低下したと考えられる。今後、不均衡データに対する学習に於ける重みづけの最適化処理や、学習に使用するデータ数を予め均等化するアンダーサンプリング・オーバーサンプリング処理の検討を行うことが課題である。

b) 汎化性能の検証

本論文における提案方式の評価は、岐阜県地域の医療・介護レセプト情報に含まれる限られた患者群のみを対象として実施した。介護サービスを必要とする患者数や要介護度別の構成比等は地域によって異なり[17]、また運動習慣や食生活等の患者の要介護状態に影響を与える背景因子も地域によって異なる可能性が高い。今後、異なる地域の患者群を対象とした交差差検証等の汎化性能の検証を行うことが課題である。

6. おわりに

本論文では、将来の介護サービスの需要推定を可能とすることを目的とし、医療レセプト情報と介護レセプト情報に記述された診療情報と介護情報に基づく将来の要介護状態の推定方法を提案した。岐阜県下の地域保険者より提供頂いた医療・介護レセプト情報を用いた評価実験を行った。その結果、将来の要介護状態を F 値 0.85 以上と高い精度で検出可能であることを示し、提案手法の有効性を示した。今後の課題である不均衡データに対する処理の検討を行った後、異なる

地域のデータを使用した汎化性能の検証を進める。

参考文献

- [1] 藤森研司, レセプトデータベース (NDB) の現状とその活用に対する課題, 医療と社会, Vol. 26, No. 1, pp. 15-24, 2016.
- [2] 満武巨裕, 日本のレセプト情報・特定検診等データベース (NDB) の有効活用”, 情報処理, Vol. 56, No. 2, pp. 140-144, 2015.
- [3] 厚生労働省, 医療・介護データ等の解析基盤に関する有識者会議報告書について, <https://www.mhlw.go.jp/content/12601000/000439072.pdf>, 2020 年 1 月 9 日に参照.
- [4] 総務省 自治体行政局, 自治体戦略 2040 構想研究会 (第 3 回) 事務局提出資料. https://www.soumu.go.jp/main_content/000517853.pdf, 2020 年 1 月 9 日に参照.
- [5] 武田俊平, 介護保険における 65 歳以上要介護等認定者の 2 年後の生死と要介護度の変化, 日本公衛誌. Vol. 51, No. 3, pp. 157-167, 2014.
- [6] 福岡美紀, 塩飽邦憲, 介護保険制度発足後の居宅要介護者の要介護度変化, 日農医誌, Vol. 58, No. 5, pp. 516-525, 2010.
- [7] T Lee, M. E. Mill, Analysis of patient profile in predicting home care resource utilization and outcomes, J Nurs Adm Vol 30, No.2, pp 67-75, 2000.
- [8] 大高洋平, 高齢者の転倒予防の現状と課題, 日転倒予会誌, Vol. 1, No. 3, pp. 11-20, 2015.
- [9] 前野里恵, 井上早苗, 足立徹也, 転倒による高齢大腿骨頸部骨折者の退院後の日常生活状況と QOL, 理学療法学, Vol 31, No. 1, pp. 45-50, 2004.
- [10] I. Liepelt-Scarfone, M. Fruhmann Berger, D. Prakash, I. Csoti, S. Gräber, W. Maetzler, D Berg, Clinical characteristics with an impact on ADL functions of PD patients with cognitive impairment indicative of dementia, PLoS One, Vol 8, Issue. 12, e82902, 2013.
- [11] C. Hofgren, A. Björkdahl, E. Esbjörnsson, K. S. Sunnerhagen, Recovery after stroke: cognition, ADL function and return to work, Acta Neurol Scand, Vol 115, No. 2, pp. 73-80, 2007.
- [12] 厚生労働省, 平成 28 年国民生活基礎調査の概況, <https://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa16/dl/05.pdf>, 2020 年 1 月 9 日に参照.
- [13] 厚生労働省保険局, 診療報酬情報提供サービス, <http://www.iryohoken.go.jp/shinryohoshu/>, 2020 年 1 月 9 日に参照.
- [14] 厚生労働省, 要介護認定はどのように行われるか, https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi_kaigo/kaigo_koureisha/nintei/gaiyo2.html, 2020 年 1 月 9 日に参照.
- [15] 厚生労働省, 区分支給限度基準額について, https://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000049257.pdf, 2020 年 1 月 9 日に参照.
- [16] R. Longadge and S. Dongrem, Class imbalance problem in data mining review, arXiv preprint arXiv, 1305. 1707, 2013.
- [17] 内閣府政策統括官, 要介護 (要支援) 認定率の地域差要因に関する分析, <https://www5.cao.go.jp/keizai3/2018/09seisakukadai15-0.pdf>, 2020 年 1 月 9 日に参照.