

介護記録自動生成のための記録内容の推定の試み

金子 晴^{†1,a)} 井上 創造^{†1,b)}

概要：近年日本では高齢化が進んでいる。それに伴い、介護サービスの需要も増している中、介護人材の不足が社会問題になっている。介護人材の不足を解消するために、介護職員の負担軽減が重要である。我々は、介護記録入力効率化のため、介護記録アプリを開発した [1][2]。本研究では、介護記録入力効率化のため、機械学習技術を応用し「対象の被介護者」と「介護行動の内容」の自動入力機能を目指した。介護記録アプリケーションにおける、「対象の被介護者」と「介護行動の内容」の推定機能の開発を行った。また、介護記録アプリから収集された介護記録データを用い、推定機能を評価した。結果として、「対象の被介護者」の推定において平均 F1 値 74%、「介護行動の内容」の推定において平均 F1 値 58%を得た。

キーワード：機械学習、介護記録

Estimation of Record Contents for Automatic Generation of Care Record

1. はじめに

近年日本では高齢化が深刻化している。内閣府によると [3]、現在の日本の高齢化率は 28.1%であり今後高齢化率は増していくとされている。それに伴い、介護サービスの需要も増している中、介護人材の不足が深刻になってきている。厚生労働省の推定 [4]によると、2025 年に必要とされる介護人材は 253 万人とされている。それに対し人材供給は 215 万人と推定されており、38 万人の介護人材が不足する見込みだ。人材不足の原因として、介護業界の離職率があげられている。離職率改善のため、介護職員の負担軽減は重要である。特に、介護記録作業は時間がかかる場合が多い。三輪らの研究 [5]によると、介護職員は記録の作成・共有・確認に勤務時間の 25%を割いているという結果が出ている。我々は、この記録作業にかかる時間を短縮する事が、介護現場の負担軽減に繋がると考えた。

我々は介護職員の負担軽減のため、Android 端末を用いた介護記録アプリ Gotolog を開発した [1][2]。介護職員は、スマートフォンを携帯し介護行動を行う際、Gotolog で介護行動の種類やその詳細を記録する。

本研究では、介護記録入力効率化のため、「対象の被介護者」と「介護行動の内容」の自動入力を目指す。機械学習を用い、これら二つの推定機能を開発する。また、介護記録アプリ Gotolog で収集された 2 ヶ月間のデータを用いて、「対象の被介護者」と「介護行動の内容」の 2 つの推定機能の評価を行う。その結果、被介護者の推定について平均 F1 値 74%、介護行動の内容の推定について平均 F1 値 58%を得た。

本稿の構成について述べる。まず、2 章にて関連研究を示す。次に、3 章で評価に用いるデータの概要やデータ数を示す。そして、4 章で被介護者の推定・介護記録の内容の推定の方法についてそれぞれ述べ、5 章にてその手法をそれぞれ評価する。最後に 6 章にて考察を述べ、7 章にてまとめや今後の展望について述べる。

2. 関連研究

介護職員の負担軽減や業務の効率化に向けた研究はいくつかある。例えば内平ら [6] は、看護・介護職員が携帯端末とヘッドセットを携帯し、それを用いて共有すべき事項や気づいた事を音声入力するシステムを提案している。共有された情報は自動的に分類され、共有すべき相手に送信される形になっている。この研究は、介護記録の分類を自動で行っているが、本研究のように介護記録の自動生成は行っ

^{†1} 現在、九州工業大学
Presently with Kyushu Institute of Technology

a) kaneko@sozolah.jp

b) sozo@sozolah.jp



図 1 介護記録アプリ上での推定システムの利用イメージ図

ていない。

野亦ら [7] は、介護記録の自動生成に向け、介護分野オントロジーの生成を目指している。介護学会で発表された論文をテキストマイニングする事で、日常生活行動についてのトピックモデルの構築に成功している。この研究では、構築されたオントロジーを元に、介護職員がキーワードを入力すると自動で申し送りの文章を作成するシステムや、センサなどを用いて誰が何をしているという情報を取得し、その情報を用いて自動で介護記録を文章を作成する事などを応用例として上げている。本研究では、介護記録の内容が入力される前に記録を自動生成することを目指しており、この点で野亦らの研究と異なる。

福原ら [8] は、情報推薦を用いた介護施設向けの申し送り業務支援システムを開発している。iPhone や iPad 上で動作するアプリを開発しており、介護職員はそれを用いて記録を行う。音声や写真での記録が可能のほか、システムから申し送り文の推薦を受ける事ができる。介護職員は推薦された申し送り文を元に一旦記録を保存し、時間が開いた際に修正などを行うことができる。このシステムを実際に介護職員が使ったところ記録の作成・確認作業にかかる時間が削減できたという結果を得ている。

我々は、介護記録アプリ GtoLog を開発し、現場での導入実験を行っている [1] [2]。介護施設において介護職員は、勤務中にスマートフォンを携帯する。そして、Android アプリである Gtolog を用いて介護記録を逐次入力する。入力された介護記録データはサーバに送られ、WEB ページから見る事が出来るほか、従来の紙での介護記録のような形式での印刷もできる。加えて、介護行動中の加速度をスマートフォンで収集し、介護行動の自動認識の研究を行っている。

また我々は、介護記録の自動生成の試みも行っている [9]。介護記録アプリ Gtolog で収集した介護記録データを用いて機械学習を行い、介護記録の記録内容を推定した。結果として、79%の平均正答率を得た。しかし、F1 値が低い物も多く改善の余地がある。本研究では、この結果を踏まえ特徴量の追加や被介護者個人ごとでの学習・推定を行い精度の向上をめざした。また、まだ行っていない介護行動の対象となる被介護者の推定を行った。

3. 介護アプリとデータについて

本章では、評価に用いるデータの概要について述べる。評価には、先述した実験 [1] において、23 名の介護職員に Gotolog を使用してもらい集取された介護記録データを用いる。使用するデータは 2018 年 5 月 1 日 2018 年 7 月 1 日の期間に収集されたものである。以下では、介護記録アプリ Gotolog と評価に使用するデータについて述べる。

3.1 介護記録アプリ Gotolog

まず、介護記録の目的について説明する。介護記録とは、入居者の方の食事摂取量やバイタル、日々の様子や入浴の有無などを記録したものである。これらの記録は、健康管理やご家族への報告、サービス向上のために用いられている。介護記録には、介護職員が行った介護行動の内容とその対象となる被介護者の記録が必要である。その記録作業の効率化のため、被介護者と介護職員の行動を自動推定する機能をつけた介護記録アプリ Gotolog の開発を進める。

図 1 に推定機能を介護記録アプリ上で利用するイメージ図を示す。本研究では、介護職員がアプリから介護記録を作成する際、事前に推定された値を初期値として入力しておく事を想定している。まず介護職員がこれから行う行動を入力する。次にアプリ内部にて、入力された行動の種類や時刻などの情報から行動の対象となる被介護者を推定し、予測された対象者として選択肢の上位に表示する。介護職員は対象者となる被介護者を入力し、行動を開始する。行動を行っている間に、アプリ内部で介護行動の詳細を推定し、推定結果を初期値とした詳細入力画面の準備を行う。行動が終わった後、介護職員は介護行動の詳細を入力する。そうする事で介護職員は、推定が間違えている部分を修正するのみでよくなる。このような推定機能を用いたシステムを介護記録アプリに導入する事で、記録作業の効率化に繋がると考える。

3.2 データ

次に、介護記録データの構造について述べる。今回用いるデータには、11 種類の行動がある (以下行動クラスと呼ぶ)。例えば、「食事・服薬」や「排泄」などである。また、

それぞれの行動クラスには、それぞれに対応した行動の詳細が複数記録されている(以下詳細レコードと呼ぶ)。そのため、一つの介護行動を行うと、複数の詳細レコードが保存される。表1に「食事・服薬」という行動クラスを例に行動クラスと詳細レコードの対応を示す。詳細レコードには、予め値の選択肢が決められており、介護職員はアプリから介護記録を作成する際、選択形式で値を入力する。

次にデータ数について述べる。総データ数は行動クラスが約1万8千個、詳細レコードが3万個ある。図2に各行動クラスのデータ数とその内訳を示す。この図から、行動クラスによってデータ数に偏りがあることが分かる。図3に各詳細レコードのデータ数とその内訳を示す。この図から、詳細レコードにもデータ数に偏りがあることが分かる。また、詳細レコードの値についても偏りがあることが分かる。

これは実際に、介護施設において収集されたデータである。つまり、実際の介護施設で入力される介護記録データは、行動クラスや詳細レコード、詳細レコードの値に偏りが大きいデータであることが予想される。今回は、データ数の偏りに注意して評価を行う。

表1 「食事・服薬」クラスにおける詳細レコードとその値の対応(参考文献[9]表2より転記)

行動クラス	詳細レコード	詳細レコードの値
食事・服薬	食事介助	全介助 一部介助 セッティングのみ 自立
	食事量(主食)	0, 1, 2, ..., 10
	食事量(副食)	0, 1, 2, ..., 10
	水分量	0, 50, 100, ..., 500
服薬	介助	
	自立 服薬なし	

4. 分析方法

本章では、過去の介護記録を用いて詳細レコードを推定する手法を述べる。本研究では、機械学習アルゴリズムの一つである Random Forest[10]を用いる。Random Forestは決定木を弱学習機とするアンサンブル学習アルゴリズムの一種である。

4.1 被介護者の推定

ここでは、介護記録からその介護行動の対象となる被介護者の推定を試みる。

行動クラスのうち、「食事・服薬」や「レクリエーション」などの行動クラスは1人の介護職員が同時に複数人に対して記録を行うことが多い。そのため、アプリ上で介護記録を作成する際も、一つの行動クラスに対し、複数人の被介護者を選択する形で記録する。なので本研究では、介護

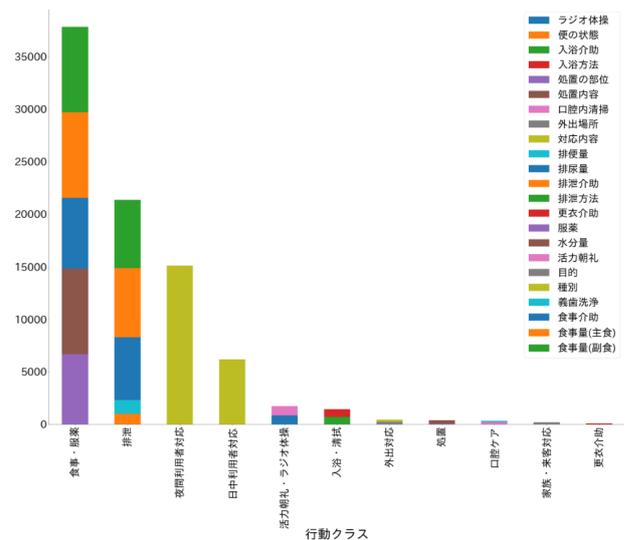


図2 各行動クラスのデータ数とその内訳(参考文献[9]図1より転記)

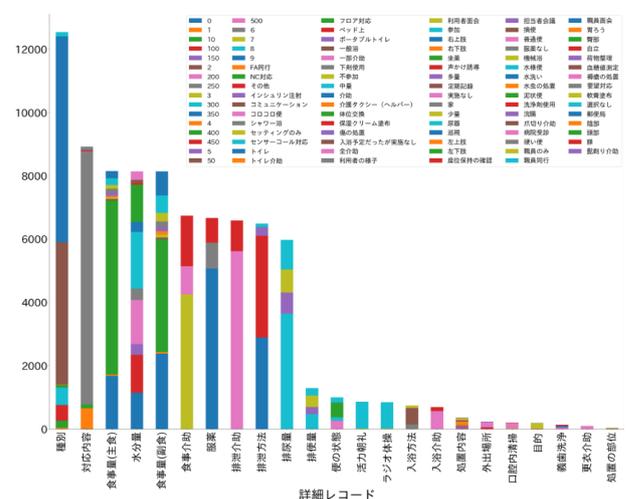


図3 各詳細レコードのデータ数とその内訳(参考文献[9]図1より転記)

行動の対象となる被介護者の推定という問題を、同時に記録された記録の中に、任意の被検者が含まれるかどうかという二値分類問題であると考えた。目的変数は「含まれる」または「含まれない」となる。

次に、説明変数について述べる。本研究では介護行動を行う際、アプリ上で対象者を入力される前に介護行動の対象となる被介護者を推定する事を目指している。そのため、介護記録のうち説明変数に用いることができる特徴量がかぎられる。今回は、「時刻」「介護職員」「行動クラス」を説明変数として用いる。

4.2 介護行動の内容の推定

ここでは、過去の介護記録や行動クラスから介護行動の内容、つまり詳細レコードの値の推定を試みる。

詳細レコードの値の推定は、介護職員がアプリから詳細レコードの値を入力する際、事前に初期値として推定され

た値を入力しておく事を想定している。

説明変数には、「行動クラス」「詳細レコード」「被介護者」「介護職員」「時刻」「過去の値」「前回の“食事・服”クラスの“主食量”コードの値」「前回の“食事・服薬”クラスの“副食量”レコードの値」「前回の“食事・服薬”クラスの“水分量”レコードの値」「前回の“排泄”クラスの“排泄量”レコードの値」を用いる。目的変数は詳細レコードの値となる。

5. 評価

本章では、4節にて述べた推定手法の評価を行う。それぞれの推定手法について評価結果を示す。今回評価には、3章で述べたように、実際の介護施設において収集されたデータを用いる。このデータは、データ数に偏りのあるデータであるため、評価には主に正答率とF1値を用いる[11]。F1値は偏りのあるデータを用いた機械学習の評価に有用である。

5.1 被介護者の推定

ここでは、4.1節にて述べた介護記録の推定の評価結果を示す。まず、図4～図7に被介護者の推定を行った際の正答率・F1値・適合率・再現率をバイオリンプロットで示す。図5から分かるように、推定の対象となる被介護者によってF1値の差が大きいことが分かる。全体の平均として、平均正答率92%・平均F1値74%・平均適合率75%・平均再現率74%を得た。行動クラスごとに見ると、「更衣介助」クラスのF1値が最も高く94%であった。F1値が最も低いのは「排泄」の59%であった。

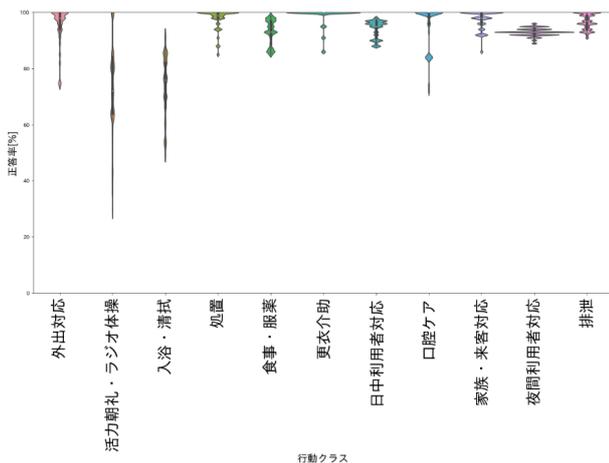


図4 被介護者の推定を行った際の正答率。縦軸に正答率 [%]，横軸は行動クラス

5.2 詳細レコードの値の推定結果

ここでは、4.2節にて述べた介護記録の推定の評価結果を示す。表2にバリテーション・学習手法ごとの平均正答率

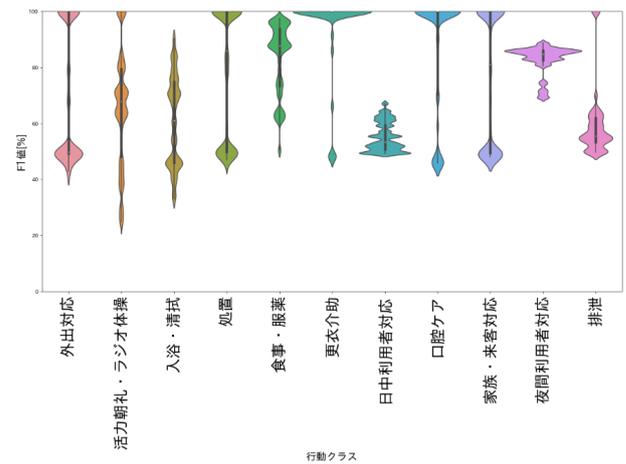


図5 被介護者の推定を行った際のF1値，縦軸にF1値 [%]，横軸は行動クラス

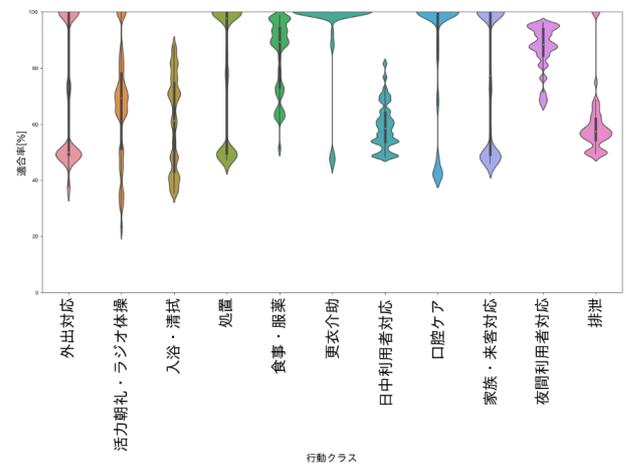


図6 被介護者の推定を行った際の適合率，縦軸に適合率 [%]，横軸は行動クラス

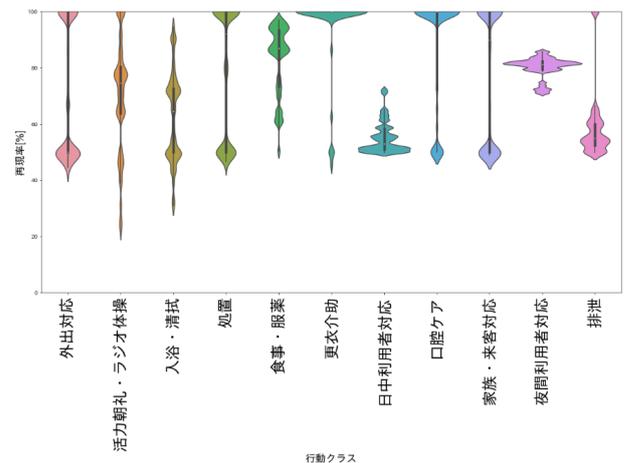


図7 被介護者の推定を行った際の再現率，縦軸に再現率 [%]，横軸は行動クラス

と平均F1値を示す。今回三種類の学習・テストを行った。一つ目は、期間でバリテーションを行った場合である。これは、データを前半45日と後半15日に分けて行った。二つ目は、被介護者でバリテーションを行った場合である。40

人の被介護者を学習データ、残り 17 人のデータをテストデータとして用いた。三つ目が、被介護者個人ごとに学習・テストを行った場合である。

次に 3 種類それぞれの学習結果を詳しく示す。はじめに、図 8 に前半 45 日を学習データ、後半 15 日をテストデータとして詳細レコードの内容を推定した際の正答率と F1 値を詳細レコードごとに示す。平均正答率 70%、平均 F1 値 42%を得た。「活力朝礼」レコードは、正答率・F1 値がともに 100%となっているが、これは活力朝礼の値(参加・不参加)のうち「不参加」の値をとったデータがテストデータの一つもなかったためである。活力朝礼を除き F1 値が最も高いのは、排泄方法の 86%である。逆に最も F1 値が低いのは、処置の部位の 8%であった。

次に、図 9 に被介護者 40 人の前半 45 日分を学習データ、残りの 17 人の後半 15 日分をテストデータとして詳細レコードの内容を推定した際の正答率と F1 値を詳細レコードごとに示す。平均正答率 62%、平均 F1 値 22%を得た。F1 値が最も高いのは、ラジオ体操の 49%であった。F1 値が最も低いのは処置の部位と義歯洗浄とともに 0%であるがこれは、学習データに多くあった「実施なし」という項目がテストデータに一切含まれていなかったからである。これら 2 つの詳細レコードを除くと、処置の内容の 6%が最も低い F1 値となる。

最後に、図 10 に前半 45 日を学習データ、後半 15 日をテストデータとして、被介護者ごとに詳細レコードの内容を学習・推定した際の正答率と F1 値の平均を詳細レコードごとに示す。平均正答率 62%、平均 F1 値 58%を得た。被介護者個人ごとに学習した場合、複数人の被介護者で学習した場合と比べ、正答率・F1 値が高くなった。最も F1 値が高いのは、活力朝礼の 94%であり、ラジオ体操の 89%が次に続く。最も F1 値が低いのは処置の部位で 7%である。

表 2 バリテーション、学習手法ごとの平均正答率 [%] と平均 F1 値 [%].

学習方法	バリテーション方法	正答率	F1 値
複数人で学習	期間でバリテーション	79	42
	被介護者+期間でバリテーション	62	22
個人ごとに学習	期間でバリテーション	75	58

6. 考察

まず、被介護者の推定について述べる。外出対応・更衣介助・口腔ケアなどは F1 値が高い場合と低い場合の二極化している。しかしこれらは、正答率・F1 値・適合率・再現率の全てが 100%の物を多く含んでいる。原因として、テストデータに更衣介助や口腔ケアなどを行ったデータが含ま

れていなかった被介護者がおり、推定結果が全て「含まれない」という値になったためであると考えられる。その場合を除くと平均 F1 値は 64%となる。加えて、F1 値が被介護者によって大きく異なる事が分かった。上記と同様に推定結果が全て「含まれない」という値になった被介護者が居ることが理由であると考えられる。平均として、平均 F1 値 74%・平均再現率 74%を得ている。「被介護者」の推定の場合、再現率とは、実際に被介護者が「含まれる」という場合のうち、推定結果において「含まれる」と推定された割合である。つまり、どの程度見逃さなかったかという割合である。記録アプリにおいて、被介護者の推定結果は入力選択肢の上位に表示される。もし見逃しが多い場合、介護職員は見逃された被介護者を探す手間が増えるため、見逃さない事が重要である。今回は、再現率が平均 74%であるため、この推定機能は従来の介護記録アプリに比べ記録入力の効率化になると考える。

次に、詳細レコードの値の推定について述べる。詳細レコードの値の推定は、被介護者個人ごとで学習した場合、複数人の被介護者のデータを用いて学習する場合に比べ、F1 値が高くなる事が分かった。特にラジオ体操の F1 値が最も向上しており、約 49%から 88%になった。また、対応内容、食事量(主食)、食事量(副食)などの F1 値は 15%以上向上している。理由として、毎回同じ値しか取らない人が多いためであると考えられる。この詳細レコードの値の推定機能を導入する事で、従来の介護記録アプリに比べ記録入力の効率化になると考える。また、さらに F1 値を向上させるために、加速度情報を特徴量として用いると良いと考える。

7. まとめ

本研究では、介護記録の自動生成を目指し、介護行動の対象となる被介護者の推定と介護記録に記される記録内容の推定の 2 つの評価を行った。機械学習には Random Forest を用いて「被介護者」と「介護行動の内容」の二つの推定を行った。また、介護施設にて収集された介護記録データを用い推定結果を評価した。被介護者の推定において、平均 F1 値 74%を得た。また、平均再現率も 74%であるため従来の記録アプリに比べ記録入力の効率化に繋がる。

介護記録の記録内容の推定においては、複数人のデータを期間を用いバリテーションし学習を行った場合、平均正答率 79%・平均 F1 値 49%という結果を得た。また、被介護者個人ごとに学習を行った場合、複数人の被介護者のデータを用いて学習を行う場合に比べ F1 値が平均で 16%ほど高くなる事が分かった。

本研究の被介護者や詳細レコードの推定は介護アプリの機能として実装する事を前提に行った。今後は介護記録アプリ Gtolog 上での実装を行う予定である。現在は、サーバ上で学習器を作成した後にスマホに学習器を送信し、スマ

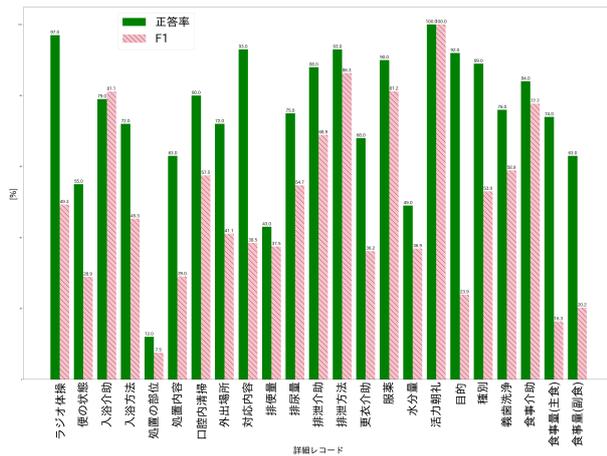


図 8 期間でバリテーション：前半 45 日を学習データ，後半 15 日をテストデータとして詳細レコードの内容を推定した際の正答率と F1 値を詳細レコードごとに示す。縦軸に正答率・F1 値 [%]，横軸に詳細レコード。

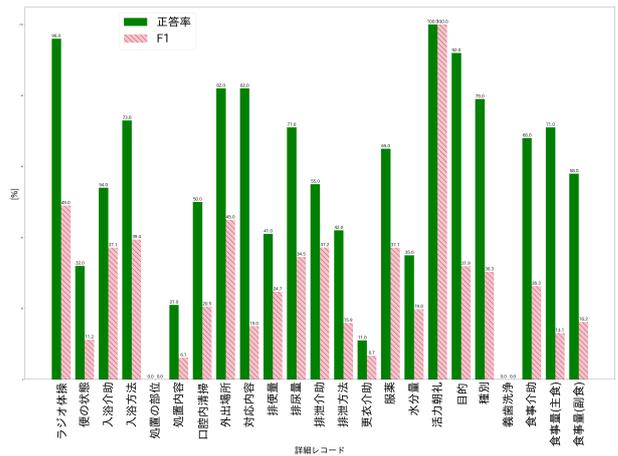


図 9 被介護者でバリテーション：被介護者 40 人の前半 45 日分を学習データ，残りの 17 人の後半 15 日分をテストデータとして詳細レコードの内容を推定した際の正答率と F1 値を詳細レコードごとに示す。縦軸に正答率・F1 値 [%]，横軸に詳細レコード。

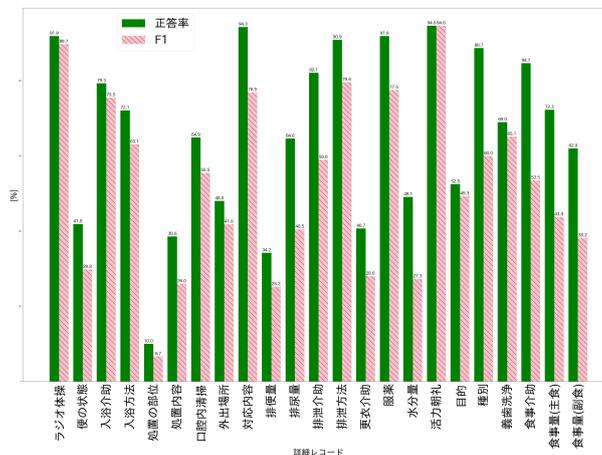


図 10 被介護者個人ごとに学習・推定：前半 45 日を学習データ，後半 15 日をテストデータとして，被介護者ごとに詳細レコードの内容を学習・推定した際の正答率と F1 値の平均値を詳細レコードごとに示す。縦軸に正答率・F1 値 [%]，横軸に詳細レコード。

ホ上で推定を行う事を考えている。そのため，今後はスマホのスペックなどが推定の実行に耐えうるのかの検証や，推定時間がどの程度かかるのかなどの検証が必要である。また，記録作業の簡略化という点で，適合率と再現率のどちらを向上させるべきなのか，検証が必要である。

参考文献

- [1] Inoue, S., Lago, P., Hossain, T., Mairittha, T. and Mairittha, N.: Integrating Activity Recognition and Nursing Care Records: The System, Deployment, and a Verification Study, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 3, No. 3 (2019).
- [2] 井上創造, 木村幸平, 内野百里, 大屋 誠: 介護施設における介護スタッフの行動センシング実験, 技術報告 (2017).
- [3] 内閣府: 令和元年版高齢社会白書, (オンライン), 入手先 (https://www8.cao.go.jp/kourei/whitepaper/w-2019/zenbun/01pdf_index.html) (2019).
- [4] 厚生労働省: 2025 年に向けた介護人材にかかる需要水系

- (確定値) について, (オンライン), 入手先 (<https://www.mhlw.go.jp/stf/houdou/0000088998.html>) (2015).
- [5] 三輪洋靖, 渡辺健太郎, 福原知宏, 中島正人, 西村拓一: 介護プロセスの計測と記述, *日本機械学会論文集*, Vol. 81, No. 822, pp. 14-00207-14-00207 (2015).
- [6] 内平直志: 音声つぶやきによる気づきの収集と活用で看護・介護サービスの質を向上する, *サービスロジー*, Vol. 1, No. 2.
- [7] 野亦 優, 沼尾雅之: 介護日誌自動生成のための介護分野オートロジー生成技術の研究, *マルチメディア, 分散協調とモバイルシンポジウム 2019 論文集*, Vol. 2019, pp. 1599-1603 (2019).
- [8] 福原知宏, 中島正人, 三輪洋靖, 濱崎雅弘 and 西村拓一: 情報推薦を用いた高齢者介護施設向け申し送り業務支援システム, *人工知能学会論文誌* (2013).
- [9] 金子 晴, 井上創造: 介護施設における介護記録自動生成の試み, 第 2 1 回日本知能情報ファジィ学会九州支部学術講演会.
- [10] Breiman, L.: *Random Forests*, p. 5-32 (2001).
- [11] 井上創造: ウェアラブルセンサを用いたヒューマンセンシング, *知能と情報*, Vol. 28, No. 6, pp. 170-186 (2016).