情報処理学会論文誌 Vol.59 No.10 1837–1852 (Oct. 2018)

介護施設紹介コールセンタ記録のアンサンブル学習による将来予測と傾向分析

松木 萌[1,a)] 井上 創造[1,b)] 清田 陽司[2,c)]

受付日 2017年12月25日, 採録日 2018年7月10日

概要：介護施設紹介コールセンタの記録データをアンサンブル学習を用いて傾向分析や将来予測を行い，「コールセンタ」「介護施設」「相談者と入居者」の三者にとって有用な知見を取得する．介護分野における問題に対して，現場のデータを分析し問題解決に有用な知見を得る研究は多くあるが，それらは入居後の問題解決に向けたデータ分析がほとんどである．本研究では，介護施設入居前の分析も重要と考え，入居前の相談を受付けるコールセンタの記録データを分析する．分析方法は，アンサンブル学習を用いて，説明変数の属性値を 2 値化し，属性の重要度と目的変数に対する正負の相関を同時に可視化した．記録データは，相談者 1 名に対し 1 行，「性別」や「介護度」のような属性を列とした，7,685 行 106 列の行列データである．分析の結果，記録に欠損が多い人は成約に至らないことや，交通手段が見学に行くかどうかの重要な要因であること，「見学するかどうか」は機械学習による精度が 96.8%であること，入居決定の決め手は「介護サービス」ではなく，「きれいさ」や「印象」，「距離」であることなど，10 件ほどの知見を取得した．

キーワード：データマイニング，将来予測，XGBoost，介護，コールセンタ

Future Prediction and Trend Analysis by Ensemble Learning of Call Center Log Introducing Caregiving Facilities

Moe Matsuki[1,a)] Sozo Inoue[1,b)] Yoji Kiyota[2,c)]

Received: December 25, 2017, Accepted: July 10, 2018

Abstract: In this paper, we analyze the trend and predict the future with the recorded data at a call center which introduces caregiving facilities to residents who will be living in. We exploit ensemble machine learning in order to get useful knowledge for three parties, who are “the call center”, “caregiving facilities” and “prospective residents”. Although there are many data analysis research to solve the problems in the caregiving field, they focus on when living in the facilities. In this paper, we analyze the recorded data at a call center to introduce caregiving facilities. The recording data are matrix data of 7,685 rows and 106 columns, with one row for one client, and each column represents the attribute such as “gender” and “degree of care”. In the analysis, first, we changed the explanatory variables to binary. Second, we visualized the importance of attributes and the correlation between positive or negative values for objective variables at the same time by using ensemble learning. As a result, we obtained about 10 findings such that the customers with missing recorded data tends not to contract with a facility, the accesses to facilities are important factor for the decision, the accuracy predicting “whether to observe” achieved 96.8%, and the factor of the decision is not “caregiving service” but “cleaning”, “impression”, “distance”.

Keywords: data mining, future prediction, XGBoost, caregiving, call center

[1] 九州工業大学
Kyushu Institute of Technology, Kitakyushu, Fukuoka 804–8580, Japan
[2] 株式会社 LIFULL
LIFULL Co., Ltd., Chiyoda, Tokyo 102–0083, Japan
[a)] matsuki.sousisu@gmail.com.jp
[b)] sozo@mns.kyutech.ac.jp
[c)] KiyotaYoji@lifull.com

1. はじめに

近年介護の需要は急激に増加しているため，介護職員の不足や介護難民という問題が深刻なものとなっている [1]．同時に，施設での事故の発生，要介護者の虐待など多くの問題が存在する [2]．

© 2018 Information Processing Society of Japan

これらの介護分野における問題解決のために，介護データを分析し，事故防止や業務効率化に対する解決策の提案を行う研究が行われている [3], [4], [5], [6], [7]．これらのデータ分析による介護問題への取り組みは施設に入居した後に対するものがほとんどである．介護業界の全体の問題を解決するためには，入居前のデータを分析し現場以外の問題にも取り組む必要がある．たとえば，介護施設入居後のトラブルに関しての事例には，「入居前に想定していた月額費用に比べて高い」，「突然施設側から退去を求められる」のようなトラブルがある．これらの問題を軽減し，満足度の高い介護サービスと高齢者社会を実現するには，介護施設入居前の状況把握をする必要がある．

このような問題意識に基づき，本稿では，介護施設に入居する前の人々に対する相談サービスを行うコールセンタの記録データを分析することで，「コールセンタ」「介護施設」「相談者と入居者」の三者にとって有用な知見を取得する．分析に用いたデータは，相談者 1 名に対し 1 行，「性別」や「介護度」のような属性を列とした，7,685 行 106 列のデータである．分析方法として，結果に対して重要な要因を算出できるアンサンブル学習を用いて，傾向分析と将来予測を行う．このとき，分析目的の設定として，データに関係する「コールセンタ」「介護施設」「相談者と入居者」の三者に有力な知見を探索する．そうすることで，

- オペレータの業務の効率化（例：相談の進行方法，アドバイス方法）
- これから入居を考えている「相談者と入居者」にとっての参考（例：準備時期，内容，選定方法の参考）
- 「介護施設」にとってビジネス上の参考（例：見学して成約しない理由や，施設選びの決め手）

のためになる知見を得る．さらにこれらの知見から，改善策や解決策を考察する．その結果，(1) 欠損値があると進捗が進まない傾向にあること，(2)「続柄」や「介護度」属性が成約に対して重要属性であること，(3) 入居決定の決め手は「介護サービス」ではなく，「きれいさ」や「印象」，「距離」であることが分かった．そして，(4) 相談者の情報から「見学するかどうか」を機械学習を用いて推定すると，精度が 96.8%であることが分かった．

これらの結果は，ユビキタスコンピューティングの研究においても有用な知見になりうる．たとえば，自動運転やドライブ共有サービスとの連携による送迎の効率化が見学率を向上させる可能性や，自宅における見守りシステムなどのセンサ出力から，介護施設に入居する必要があるかどうかを予知できる可能性，自宅や，コールセンタのサービスに AI チャットボットシステムを組み込むことによるデータ収集および業務効率化の可能性が結果から示唆される．

本稿の貢献をまとめると，以下のようになる．

- 介護施設入居前の状況把握やサービス最適化を目的とし，7,685 行 106 列の介護施設紹介コールセンタのデータを機械学習により傾向分析および将来予測を試みた．
- アンサンブル学習を用いて，重要項目と目的変数に対する相関を同時に可視化し，説明変数の 2 値化を行うことで，分析結果の理解を容易にした．
- 分析の結果，記録に欠損が多い人は成約に至らないことや，交通手段が見学に行くかどうかの重要な要因であること，「見学するかどうか」は機械学習による精度が 96.8%であること，入居決定の決め手は「介護サービス」ではなく，「きれいさ」や「印象」，「距離」であることなど，10 件ほどの知見を取得した．

2. 背景と関連研究

2.1 背景

近年，高齢化にともない介護サービスの需要は年々増加している．日本創世会議が 2015 年に発表した「東京圏高齢化危機回避戦略」では，2025 年までに 75 歳以上の高齢者の増加にともない，介護需要が大幅に増加し，全国で平均 32%の増加が見込まれると予測している [1]．さらに，その増加は急速であるため介護職員が不足し，同時に，適切な介護サービスを受けられない介護難民と呼ばれる高齢者は全国で約 43 万人になるとも予測している [1]．厚生労働省による需供推計 [2] では，2025 年には 37.7 万人の介護職が不足する見込みであるとされている．このように，介護分野では高齢者の急速な増加，介護職員の不足，介護難民などの問題が存在する．

これらの介護分野における問題解決をするために，介護データを分析し，事故防止や業務効率化に対する解決策の提案を行う研究はさかんに行われている．しかし，データ分析による介護問題への取り組みは，施設に入居した後に対するものがほとんどであり，入居前のデータについて分析をする必要があると考えられる．たとえば，介護問題には，「事故」や「職員の過労働」以外にも，「突然の退去命令」[8] や「職員による虐待」[9] などの問題がある．これらを事前に防ぐためには，適切な介護施設を選択できているのか，職員や施設の相性は良さそうかということを考える必要があると考えられる．そこで，本稿では，介護施設に入居する前の人々に対する相談サービスを行うコールセンタの記録データを分析する．

分析には，アンサンブル学習を用いる．アンサンブル学習は以下のような特徴がある，

(1) 自動的に過学習を抑え，汎用能力の高いモデルを生成する．
(2) 説明変数間の交互作用を中和したうえでの説明変数の目的変数への重要度が分かる．
(3) 弱学習器の決定木を抽出すれば，枝の分岐条件を直感的に理解しやすい．

つまり，(1) により説明変数から「重症度」などの目的

© 2018 Information Processing Society of Japan

変数を精度良く推定しながら，(2) によりその際の説明変数がどの程度有効だったかを効率良く推し量ることができる．また (3) により可視化することで，どのような場合の組合せにおいて目的変数の結果に影響したかを理解しやすい．

(2) については，XGBoost をはじめとするアンサンブル学習法においては，各弱学習器の学習にで使わなかったサンプルのある説明変数の値をランダムに入れ替えて精度評価を行い，その精度が全弱学習器で平均してどの程度下がったかがその説明変数の**重要度**として定義される [10]．

通常の線形回帰やロジスティック回帰のような手法では，学習された結果の各説明変数の係数の値は互いに依存することが**交互作用**として知られているため，純粋にどの説明変数がどの程度有効だったかを知ることは難しいが，アンサンブル学習においては変数もランダムにサンプリングするため，変数が目的変数の推定にどの程度寄与したかを知ることができる．

2.2 関連研究

介護分野におけるデータを分析し，傾向を把握し，解決策を述べている研究例をあげる．文献 [3] は，介護の現場における事故防止に向けたデータ収集を行い，収集したデータを分析し，介護施設における事故の傾向についてまとめている．また，介護施設の記録は電子化されていることが少なく，テキスト（自然言語）による記録が多いため，分析するのが難しいとされている．文献 [11] は，介護施設の申し送りデータを電子化し収集することを可能にし，アンケート結果の分析を行っている．しかし，日常の介護業務内では，項目を選択するような形式ではなく，テキスト文や自然言語の文章での記録業務が多い．自然言語は表現の揺らぎ（たとえば，「朝ご飯」と「朝食」が混在するなど）や文章構造の統一性がないなど，分析するにおいて難しい点が多くある．しかし，テキスト文は自由な表現ができるため，大切な情報が潜在しており，分析することで，有用な情報を抽出することができる．文献 [12] は，コールセンタのやりとりの記録をテキストマイニング技術を用いて分析を行う研究を行った．その結果，顧客満足度を向上につなげることができた．文献 [13] は，介護施設における事故事例のテキスト文（自然言語）から，テキストマイニング技術を用いて分析を行い，介護施設の職員に有用な情報，改善策の提案を行った．このように，データを分析することで，データに関係する人に有用な知見を抽出しフィードバックすることで，業務，ビジネス，問題点の解決に貢献できる．さらに，データ化，分析方法の技術も発展しており，項目の統計分析だけでなく，テキスト（自然文）による分析も可能である．しかし，これらの分析は現場の解決策を提案したものが多く，本稿で扱う入居前の状況を分析し，介護分野における問題に取り組む研究はない．

3. 分析データと手法

介護施設紹介コールセンタの記録データから分析を行う．まず，用いるデータについて 3.1 節で説明する．そして，3.2 節で分析の目的を述べ，3.3 節で分析概要を説明する．次に 3.5 節と 3.4 節で分析に用いる変数の設定方法や生成方法について説明する．

3.1 データ

本研究では，介護施設探しに関する相談を受け付け，施設紹介コールセンタの 2016 年中の記録データから，無作為に抜粋したものを分析対象とした．本データの内容はオペレータが相談者から聞き取りを行った各種情報や，相談者とのやりとりをオペレータ自身が記録したものであり，(1) 相談者 1 名につき 1 行で表され，(2) 106 個の項目を持ち，(3) 7,685 行 106 列のデータである．オペレータによる記録は，主に次回相談の際の申し送りを目的に行われている．項目には，数値型や因子型（離散的な値の変数），日付型，テキスト型がある．個人の特定につながる情報は匿名化されている．**表 1** に主な項目を示す．ADL は，Activities of Daily Living（日常生活動作）の略称である．

3.2 分析の目的

分析の目的は，**表 2** に示したような，介護施設，将来の相談者や入居者，コールセンタに有用な情報を抽出することである．施設探しの方法は段階的なものがある．まず，施設や介護施設紹介コールセンタなどへの相談や資料請求がある．そして，実際に施設への見学をし，気に入れば成約をする．つまり，(1) 相談から見学へ進捗を進める段階，(2) 見学から成約に進捗を進める段階があると考える．本稿では，この 2 つの段階まで進捗を進める人がどのくらい存在し，また，どのような人が進めるのかを，相談者の情報から分析を行う．この分析結果，コールセンタの人にとって，効率的な相談進行のヒントになる．たとえば，進捗が進まない原因が分かれば，相談方法の改善に役立てることができる．相談者にとっては，いつごろから準備を行い，どのような準備を行うかの有用な知見となる．介護施設にとっては，見学から，契約しなかった人の傾向を知ることで，サービスの質の改善につながる．また，(1) 見学施設の種別，(2) 成約施設の種別の 2 つのことを相談者の情報から推定できるのか，相関はどのようにあるのかを分析する．分析の結果は，コールセンタの人や相談者にとって，入居する際の参考，または，サービスの参考になる知見である．介護施設にとっては，入居者とのマッチング方法の知見になると考えられる．そうすることで，適切なサービスを入居者に提供でき，職員のストレスの負担を減らすことにつながる．相談者にとっては，施設選びの参考になる．これらの分析内容と登場人物（相談者，コールセンタ，

© 2018 Information Processing Society of Japan

表 1 分析対象データの主な項目（カッコ内は因子の例を表す．表中の ADL は，Activities of Daily Living（日常生活動作）の略称．CM=ケアマネージャ，MSW=メディカルソーシャルワーカ）

Table 1 Major categories, in the analysis data (Examples of factors shown in the parentheses), ADL means Activities of Daily Living (CM = a care manager, MSW = a medical social worker).

本人属性	年齢	数値型（例：85）
	性別	因子型（男性，女性）
	介護度	因子型（例：自立，要支援 1，要介護 3，申請中，未申請）
	現在地	因子型（例：在宅，入院中，施設入居中）
本人の状況	認知症有無	因子型（あり，なし）
	認知症備考	テキスト型（例：会話の受け答えは可能だが，物忘れが激しい）
	ADL 食事	因子型（例：自立，一部介助，胃ろう）
	ADL 歩行	因子型（例：自立，一部介助，杖歩行，車イス）
	ADL 入浴	因子型（例：自立，一部介助，機械浴）
	ADL 排泄	因子型（例：自立，一部介助，オムツ）
	たん吸引	テキスト型（例：1 日 3 回，希望があるとき，夜間もあり）
	インスリン	テキスト型（例：1 日 2 回）
	透析	テキスト型（例：週 3 回）
	主な病歴	テキスト型（例：高血圧，白内障，脳梗塞）
相談者について	続柄	因子型（例：父，母，配偶者，兄弟，友人知人，おじ，おば，CM，MSW）
	相談経緯	テキスト型（例：叔母様の入居相談で姪御様からお電話．いま入居中の施設から移りたいとご本人がおっしゃっている）
入居に関する希望	入居希望時期	因子型（例：2 カ月以内，1 年以内，できるだけ早く，未定）
	希望エリア	テキスト型（例：東京 23 区内）
	居室要望	テキスト型（例：居室では静かに過ごしたい，2 人部屋希望，夫婦部屋希望）
相談状況	進捗	因子型（例：相談，資料，見学，体験入居，契約）
	相談備考	テキスト型（主に申し送り事項がフリーテキストで記録されている）
日時情報	受付日時	日付型
	最終更新日時	日付型

表 2 介護施設紹介コールセンタに関する登場人物と分析における利点（行は登場人物を表し，列は分析内容を表す）

Table 2 The benefits of analysis to people related with call center introducing the care-giving facility (rows show characters, columns show analysis content).

登場人物 / 分析内容	進捗過程内に見学，成約が含まれるかの有無	見学施設や成約施設の種別の推定
コールセンタ	効率的な相談進行の改善に有用な知見	相談者へのアドバイスの参考
介護施設	見学をして成約しなかった理由	施設と適切な相談者とのマッチング
将来的な相談者・入居者	準備の時期や内容に対する知見	施設選びの際に参考になる知見

介護施設）の利点を表 2 にまとめる．つまり，本稿では介護施設紹介コールセンタの記録データを分析することで，介護施設，将来的な相談者，入居者，コールセンタに有用な知見を取得する．さらに，扱うデータは介護施設に入居する前のデータであるため，現場のデータ分析とは違う観点から介護の分野で起きている問題解決に有用な知見を取得し，改善策の提案を試みる．本稿では，「見学したかどうか」「成約したかどうか」「見学施設は介護付きか住宅型か」「成約施設は介護付きか住宅型か」の 4 つを分析目的を設定した．詳しい分析方法は次の節で述べる．まず，「見学したかどうか」「成約したかどうか」を記録データから傾向分析をする．たとえば，見学した人は，部屋がきれいかどうかを気にする人が多い傾向にある，などの知見を習得する．そうすることで，介護施設にとっては，入居してもらうために部屋のきれいさを意識するなどの改善案をすることができる．相談者には介護度が要介護 3 になったら入居を考える時期である点などの各登場人物にそれぞれの有益な情報の提示を行うことができる．また，「見学施設は介護付きか住宅型か」「成約施設は介護付きか住宅型か」を記録データから傾向分析する．介護施設には，ほかにも「サービス付き」や「グループホーム」といった種別があるが，2 つにした理由は 3.5.2 項で述べる．介護付きと住宅型の入居傾向は，介護サービスの質が違うため，介護度による選択をする．しかし，「長期的なことを考えて」や「予

© 2018 Information Processing Society of Japan

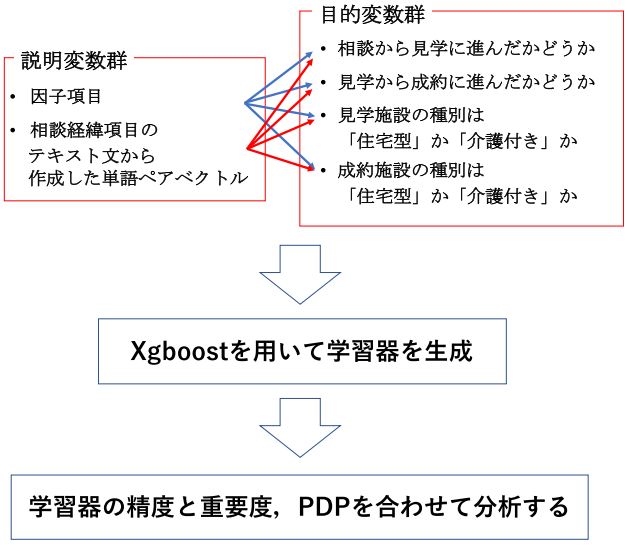


図 1 分析の流れ
Fig. 1 Overview of the analysis.

算の計画設計」の難しさが複雑に混在している．したがって，機械学習による分析によって，コールセンタのサービスのアドバイスのしやすさの助けになると考えられる．もちろんこの知見は，相談者の施設決定に役立つ知見にもなりうる．以下の節では，分析の詳細について説明する．

3.3 分析の概要

介護施設紹介コールセンタの記録データから，(a)「見学したかどうか」，(b)「成約したかどうか」，(c)「見学施設は介護付きか住宅型か」，(d)「成約施設は介護付きか住宅型か」の傾向分析を行うために，アンサンブル学習法を用いて，重要因子と相関を同時に可視化する．分析環境には，R 言語を用いて，XGBoost（eXtreme Gradient Boosting）ライブラリ [14] を用いて分析を行う．XGBoost とは，アンサンブル学習の 1 つで，決定木を弱学習器としたブースティングベースのアルゴリズムである．XGBoost はデータマイニングにおいてデータサイエンティストをつなぐプラットフォーム Kaggle で[*1]注目を浴びている．さらに，データ分析の国際大会 KDD Cup 2015[*2]で入賞した上位 10 チームすべてが XGBoost を用いた．

分析の概要を図 1 に示す．

全体的な分析方法の流れは以下のようである．

(1) 説明変数集合を生成する（3.4.1 項 3.4.2 項で説明）．
(2) 目的変数を生成する（3.5.1 項 3.5.2 項で説明）．
(3) 説明変数と目的変数の組合せ（図 1）を行い，それぞれでアンサンブル学習法を用いる．
(4) 学習によって生成された関数モデル F を用いて変数重要度と相関関係を可視化する．

本章の以下の節では，各過程について詳しく説明する．

[*1] Solutions come from of top-3 teams of each competitions.
[*2] https://biendata.com/competition/kddcup2015/

3.4 説明変数の生成方法

説明変数集合は 2 つ生成した．3.1 節のデータの中から，「属性項目」の 35 種類のベクトル集合 X_{atr} と，「相談経緯」項目からテキスト分析し TF-IDF を用いて生成した単語ベクトル集合 X_{tfidf} の 2 つである．これらを説明変数に設定することで，目的変数にする重要な属性が分析できる．以下では，各説明変数の生成方法について述べる．

3.4.1 「属性項目」の 35 種類のベクトル集合 X_{atr} の生成方法

3.1 節から，今回用いるコールセンタデータは因子型，数値型，テキスト型と，主に 3 つの形式が存在することが分かる．この中で，因子型を説明変数に設定した．なぜなら，因子型は，テキストから生成される単語ベクトルと違い，言葉の揺らぎがなく，また離散的な情報であるため，決定木を用いた分析において表現の理解が容易であるからである．以降，因子型のデータで生成したベクトル集合を「属性項目群」と呼び，各項目のことを属性と呼ぶ．

属性項目群は，以下の 35 属性のベクトル集合の集まりである．「続柄 1」「性別 1」「介護度 1」「ADL. 食事 1」「ADL. 歩行 1」「ADL. 入浴 1」「ADL. 排泄 1」「続柄 2」「性別 2」「介護度 2」「ADL. 食事 2」「ADL. 歩行 2」「ADL. 入浴 2」「ADL. 排泄 2」「利用期間」「入居希望時期」「移動手段」「相談経路」「現在地 1」「現在地 2」「申請. 区分変更 1」「申請. 区分変更 2」「見込介護度 1」「見込介護度 2」「医療行為有無 1」「医療行為有無 2」「認知症有無 1」「認知症有無 2」「AP ステータス」「AP 担当エリア」「問合せ施設以外の資料送付」「運営会社からの資料送付のアナウンス」「送付先」「連絡がいくことを」「資料が届くことを」

3.4.2 相談経緯テキストから生成した単語ベクトル集合 X_{tfidf} の生成方法

「相談経緯」項目のテキスト部分には，相談者の相談目的や背景など，コールセンタとのやりとりのうえで重要な項目が含まれると考えられる．しかし，テキスト型は，自然文で書かれているため，明確な目的や背景を抽出することは困難である．

そこで，先行研究 [15] で，我々は「相談経緯」項目のテキスト分析を行った．その結果，形容詞の表現に相談者の相談内容や，希望する施設の条件が表れやすいということが分かった．たとえば，「悪い」という形容詞の前には「足腰が」や「交通の便が」など相談者の悩みと考えられる情報が存在する．また，「良い」という形容詞の前には．「食事が」や「世話の」などの相談者が希望する施設の情報が存在する．この分析結果を参考に，本稿では「相談経緯」テキストから形容詞とそれに係る名詞単語のペアを作成し，そこから単語の特徴量をベクトルで表すことができる．TF-IDF 法を用いて単語ベクトルを作成した．この単語ベクトルを説明変数集合として用いる．

単語ベクトルを作成する概要は以下である．

© 2018 Information Processing Society of Japan

表 3 コールセンタとのやりとりが終了した時点の相談者の進捗状況と相談者の人数
Table 3 Customer number and progress when finishing the contact between the operator and the customer.

3 段階で分類した場合	第 1 段階		第 2 段階			第 3 段階	
属性の種類	相談	資料	見学	仮申込	体験入居	契約	成約
各属性の件数（件）	497	84	104	7	2	11	207

(1) 係り受け解析ツール Cabocha [16] を用いて，「相談経緯」テキストを係り受け解析する．
(2) 係り受け解析した結果から，形容詞単語 W_a の抽出．
(3) その形容詞単語 W_a に係る節から名詞単語 W_n を抽出．
(4) その形容詞単語 W_a と名詞単語 W_n のペア W_p を作成し，1 つの単語と見なす（以降ペア単語を呼ぶ）．
(5) 形容詞単語 W_a と名詞単語 W_n とペア単語 W_p を並べた単語列を，TF-IDF 法を用いて文章と単語の行列を生成する．
(6) 作成された行列からペア単語 W_p のベクトルを抽出した行列を X_{tfidf} とする．

TF-IDF 法 [17] とは，単語の特徴量をベクトルで表す単語ベクトル表現である．TF とは Term Frequency の省略語で，各の単語が文書内でどのくらい出現するかの出現頻度を表す．つまり，同じ文章内に多く出現する単語は重要であると仮定しているもので，式で表すと，

$$tf(t,d) = n_{t,d} / \sum_{s \in d} n_{s,d}$$

となる．$tf_{t,d}$ は文書 d 内のある単語 t の TF 値，$n_{t,d}$ は，ある単語 t の文書 d 内での出現回数 $\sum_s \in dn_s, d$ は文書 d 内のすべての単語の出現回数の和を表している．IDF（Inverse Document Frequency）は，各単語がどのくらいの文章で出現するのかを表しており，その出現頻度が高い単語は重要でないと仮定している．式で表すと，

$$idf(t) = \log \frac{N}{df(t)} + 1$$

となる．$idf(t)$ とは，ある単語 t の IDF 値，N は全文書数 $df(t)$ はある単語 t が出現する文書の数を表している．

アンサンブル学習を用いて重要因子属性を分析する際，重要因子の属性名が単語のみだと分析が難しいことがある．たとえば，「難しい」という形容詞単語が「見学するかどうか」を決定するのに重要な属性であるということが分かっても，なにが「難しい」のか分からないため，分析は難しい．しかし，「歩行–難しい」というペア単語 W_p が重要項目であるという結果であると，因子項目の意味が容易に理解でき，分析しやすくなる．

また，ペア単語 W_p だけでなく，W_a と W_n を含めて，1 つの単語列とした理由としては，IDF の効果を効かせ，疎な行列表現となるのを防ぐためである．ペア単語 W_p のみの単語列としてしまうと，TF-IDF で生成される行列は 1 か 0 の疎な行列表現となってしまう．なぜなら，同じ文章内でも，他の文章でも同じ単語ペア W_p は存在しにくいという問題がある．したがって，IDF の効果を反映させるため，形容詞単語 W_a と W_n と W_p を含ませた単語列とし，TF-IDF 法を用いて行列を作成した．

TF-IDF の処理を行った結果，2,793 単語，7,683 行の単語行列が作成された．そのうちの単語数の内訳は，W_p は 1,236 単語，W_a は 265 単語，W_n は 1,292 単語である．そこから，単語ペア W_p だけを抽出したため，1,236 個のベクトル集合 X_{tfidf} が生成された．

3.5 目的変数の生成方法

目的変数は，(a)「見学したかどうか」，(b)「成約したかどうか」，(c)「見学施設は介護付きか住宅型か」，(d)「成約施設は介護付きか住宅型か」の 4 種類ある．以下では，これらの目的変数の生成方法 h について説明するが，(a) と (b)，(c) と (d) は生成方法はほとんど同じである．したがって，以下では (a) と (b) を進捗の経路 Y_{status} と呼び，(c) と (d) を施設の種別 Y_{type} と呼ぶ．これらの目的変数の設定理由は，3.2 節で述べたとおり，それぞれの目的変数の傾向分析を行うことで，コールセンタ，相談者と入居者，介護施設の 3 人の登場人物に有用な情報を提供する．

以下では，これらの目的変数の生成方法について説明する．

3.5.1 進捗の経路 Y_{status} の生成方法

進捗経路とは，相談者が相談から契約するまでに行う段取りのことである．段取りの種類は，「相談」，「資料請求」，「見学」，「仮申し込み」，「体験入居」，「契約」，「成約」がある．この段取りを 3 段階に分類すると，「相談」「資料請求」が第 1 段階，「見学」「仮申し込み」「体験入居」が第 2 段階，「契約」「成約」が第 3 段階である．この段階の定義はコールセンタのオペレータの方に調査を行った結果である．この場合，最終段階が契約であり，コールセンタの最終目的であるといえる．しかし，現実には，契約までいかず途中の段階でコールセンタとのやりとりを終了している状況が発生する．**表 3** はオペレータとのやりとりが終了した時点の相談者の進捗状況と相談者の人数を数えあげた結果を示している．

このように，第 1 段階で終了している相談者が多いことが分かる．また，第 2 段階で終了している相談者は全体の 12%いることが分かる．そこで，相談者の進捗状況の違いの背景にはそのような原因があるのか分析することで，

© 2018 Information Processing Society of Japan

コールセンタのサービスの質の向上や，介護施設の改善などに有用な情報が得ることができると考えられる．

そこで，3.1 節から，「見学日付」が欠損値の有無で，(a)「見学したかどうか」を 1，0 の 2 値で表し，「成約日付」が欠損値の有無で，(a)「成約したかどうか」を 1，0 の 2 値で表す．そうすることで，2 値分類の学習モデルを生成し，分析する．

3.5.2 相談者が見学，契約した施設の種別 $Y_{\mathbf{type}}$ の生成方法

相談者が成約した介護施設名，また，見学した施設名は 3.1 節のデータから抽出することができる．それとあわせて，各介護施設に介護施設種別をラベル付けされたデータが存在する．そのデータを統合させることで，各相談者の介護施設種別をラベル付けすることができる．その介護種別の種類として，「グループホーム」「サービス付き」「介護付き」「住宅型」が存在する．

見学施設の記入があったサンプル数は 7,683 件中 1,802 件あった．見学施設のラベル割合は，以下となっている．

- グループホーム：5 件
- サービス付き：1 件
- 介護付き：1,201 件
- 住宅型：408 件

その他 187 件は，見学施設は分かっているが，施設種別がラベル付けされていないものであった．

成約施設の記入があったサンプル数は 7,683 件中 238 件あった．成約施設のラベル割合は，以下となっている．

- グループホーム：0 件
- サービス付き：0 件
- 介護付き：152 件
- 住宅型：56 件

その他 30 件は，成約施設は分かっているが，施設種別がラベル付けされていないものであった．

このことから，「グループホーム」，「サービス付き」の件数はきわめて少ないことが分かる．そのため，「介護付き」「住宅型」を分類することに需要があると考えられる．介護付きと住宅型の違いは，主に介護サービスが施設内に付属しているかどうかである．そのため，介護度が重い高齢者は介護付きに入居するが，長期的な目でみたら，介護度が軽い高齢者も介護付きに入居したほうがいい場合もある．また，費用的な部分でも，介護度によって月々の支払い額は変わってくるため，絶対的な決定指標にはなりにくい．したがって，相談者の情報から「介護付き」か「住宅型」かを決定させる指標として，アンサンブル学習による傾向分析からの結果は有用となると考えられる．

目的変数は「介護付き」を 0，「住宅型」を 1 とし，目的変数 Y_{type} を生成した．

3.6 アンサンブル学習を用いた分析

アンサンブル学習とは，弱学習器を多く用いて強学習器にする機械学習の 1 つの方法である．アンサンブル学習を分析に用いる利点として，目的変数を推定する際に重要な説明変数が可視化できる点である．たとえば，文献 [18] は，介護施設の事故事例テキストデータから，事故の要因分析をアンサンブル学習を用いて分析を行い．事故発生の原因として重要な要因を可視化した．本稿では，アンサンブル学習の中でも，最近データマイニング分野で注目を浴びている XGBoost を用いる [14]．

しかし，重要因子は重要度によって分析できるものの，どのように重要かが分からない．たとえば，説明属性「共有スペース」が重要であることが分かったものの，目的変数の「事故が起きること」か「事故が起きない」のかどちらに推定しやすい傾向にあるのか分からない．そこで，本稿では，R 言語の Partial Dipendence Plot（PDP）ライブラリ [19] を用いる．PDP とは，説明変数属性が，目的変数を推定するとどのような相関で効いているのかを可視化する．この可視化されたグラフの傾きでどちらに相関が高いか判断できる．また，説明変数を 2 値にすることで，相関関係を簡単に表す．たとえば，説明変数の「続柄」属性の属性値が「母」「父」「祖父」「祖母」の 4 つの属性値が存在していたら，「続柄. 母」「続柄. 父」「続柄. 祖父」「続柄. 祖母」の 4 つの属性に増やし，属性値を 1 と 0 で表現する．そうすると，説明変数「続柄. 母」が 1 のとき，目的変数「見学をした」が 1 と推定されれば，PDP は右肩上がりの図を出力する．この右肩上がりの相関を「正の相関」と呼ぶ．反対に，目的変数「見学をした」が 0 と推定されれば，PDP は右肩下がりとなり，これを「負の相関」と呼ぶ．

本稿では，XGBoost を用いて，属性の重要度と正負の相関を同時に可視化し，分析する．その流れを以下に示す．

(1) 学習用データの説明変数 x をベクトルに変換（因子型の場合，因子項目値を分解しブール値に変換する）．
(2) 学習用データを XGBoost で学習し，関数モデル F を生成する．
(3) 関数モデル F から目的変数 Y に対する説明変数 X の重要度を出力する．
(4) また，PDP を用いて，説明変数 X が正の相関があるのか，負の相関があるのかを出力する．

本稿の分析の特徴は以下である．

- 目的変数 Y と説明変数 X の入力を 2 値に変換することで，相関分析を簡単に表現する．
- 説明変数を 2 値にしたことで，重要項目表現の制限を広げる．
- 重要度とどのように重要かを同じ図に可視化することで分析しやすくする．

© 2018 Information Processing Society of Japan

表 4 説明変数を属性群 X_{atr} にした場合のコンフュージョンマトリックス
Table 4 Confusion matrix when attribute count X_{atr} is the explaining variable.

目的変数	見学したかどうか		成約したかどうか		見学施設が介護付き（0）か住宅型（1）か		成約施設が介護付き（0）か住宅型（1）か	
予測値→	0	1	0	1	0	1	0	1
正解値 = 0	6,687	107	748	0	1,201	0	152	0
正解値 = 1	193	696	108	33	364	44	26	30

表 5 説明変数を単語ペア群 X_{tfidf} にした場合のコンフュージョンマトリックス
Table 5 Confusion matrix when word pairs group X_{tfidf} is the explaining variable.

目的変数	見学したかどうか		成約したかどうか		見学施設が介護付き（0）か住宅型（1）か		成約施設が介護付き（0）か住宅型（1）か	
予測値→	0	1	0	1	0	1	0	1
正解値 = 0	6,794	0	748	0	1,201	0	152	0
正解値 = 1	886	3	115	26	398	10	49	7

表 6 説明変数を結合させたもの X_{merge} にした場合のコンフュージョンマトリックス
Table 6 Confusion matrix when X_{merge} is the explaining variable.

目的変数	見学したかどうか		成約したかどうか		見学施設が介護付き（0）か住宅型（1）か		成約施設が介護付き（0）か住宅型（1）か	
予測値→	0	1	0	1	0	1	0	1
正解値 = 0	6,675	119	748	0	1,201	0	152	0
正解値 = 1	208	681	97	44	374	34	30	26

4. 分析結果

本章では，前章で説明した分析方法の結果について述べる．目的変数を，(a)「見学したかどうか」，(b)「成約したかどうか」，(c)「見学した施設はどのような種別か」，(d)「成約施設はどのような種別か」の 4 つ設定し，以下各節で分析結果を述べる．説明変数は，1. 相談経緯テキスト文から生成した形容詞特徴量ベクトル郡と，2. その他の属性郡，2 つの説明変数集合を用いた分析結果について述べる．

4.1 学習モデルの精度

評価指標は以下の 2 つの指標を用いる．

- 正解率 *accuracy*：推定された $\tilde{Y} = \{y_1, \ldots, y_n\}$ がどのくらい正解か．

$$w_0 + w_1 / w_0 + w_1 + z_1 + z_0$$

- BCR：目的変数値ごとに精度を計算でき，かつ不均一な正例と負例のサンプル数にも影響されない指標 BCR (Balanced Classification Rate)

$$((w_0/w_0 + z_0) + (w_1/w_1 + z_1))/2$$

上記で w は推定結果が正しかったデータ数で，z は間違えたデータ数を表している．そして，添字の値は，データの正解値を表している，つまり，正解値と推定値がともに 0 であるサンプル数は w_0 個，ともに 1 のデータ数は w_1 個としている．また，z_0 は正解値は 0 であるのに，1 と推定したデータ数，z_1 は正解値は 1 であるのに，0 と推定したデータ数を表している．BCR を用いた理由としては，サンプル数に影響されない精度であるため，今回のサンプル数に偏りのあるデータの評価に適している．この評価指標は，文献 [20] や文献 [21] にも用いられている．

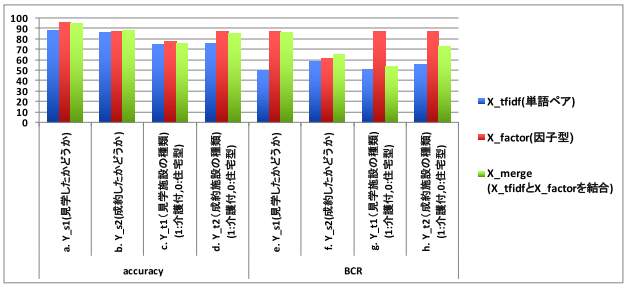


図 2 3 重交差検証で各分類器を評価した結果（x 軸において，a〜d は正解率を，e〜h は BCR，青色が説明変数 X_{tfidf}，赤が説明変数 X_{factor}，緑色が X_{merge} を表す．横軸は目的変数と，精度指標を表す．縦軸は精度（%）を表す）

Fig. 2 The result of the triple cross-validation evaluation (In the x-axis, from “a” to “d” mean accuracy, from e to h mean BCR, blue means X_{tfidf}, red means X_{factor} and green means X_{merge}. The horizontal axis shows the objective variable and the accuracy index. The vertical axis shows accuracy (%).).

以上の評価指標を用いて，**図 2** に，各分類器の精度を示す．この評価方法は 3 重交差検証を行った．図 2 の x 軸において，a〜d は正解率を，e〜h は BCR を用いた精度となっており，目的変数を横軸にし，棒グラフの色を説明変数で表している．

また，データ数の偏りによる考察ができるように，以下に図 2 の基になったコンフュージョンマトリックスを示す（**表 4**，**表 5**，**表 6**）．

この結果から，正解率はすべて 70%以上であり，(a)「見

© 2018 Information Processing Society of Japan

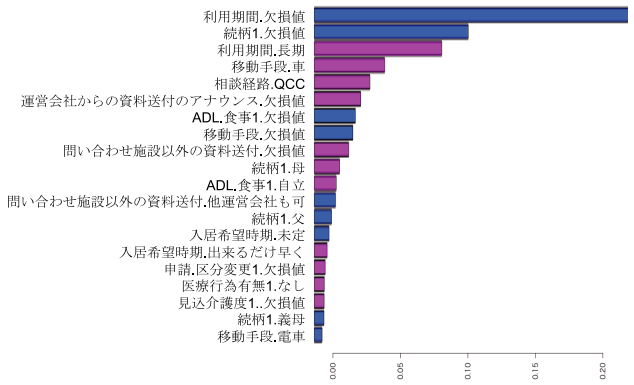


図 3 「見学したかどうか」に対して X_{atr} を説明変数としたときの重要因子と相関を表した図（目的変数：見学したかどうか，説明変数：その他の属性項目 X_{atr}，縦軸：属性名，横軸：重要度，棒グラフのピンク色の属性が正の相関示し，青色の属性は負の相関であることを示している）

Fig. 3 The important factors and correlations when X_{atr} is used as an explanatory variable to whether observation was done (objective variable: that whether the observation was done, explanatory variable: other attribute items X_{atr}, vertical axis: attribute name, horizontal axis: importance, pink attribute of the bar graph shows positive correlation, blue attribute shows. negative correlation).

学するかどうか」の正解率は X_{atr} を説明変数とした場合，96.8%と高い精度である．しかし，BCR による評価によると，X_{tfidf} を説明変数とした場合の精度が下がってしまっている．この原因として，表 4 を見ると，偏った推定をしていることが分かる．つまり，相談経緯の特徴量ベクトルから，(a)「見学するかどうか」の分類はほとんど「見学しない」と推定されてしまう傾向にあるということが分かる．また，目的変数が (b)「成約するかどうか」と (d)「見学施設は介護付き（0）か住宅型（1）か」の場合の精度に注目したとき，3 つの説明変数のすべての BCR 精度が低いことが分かる．

また，説明変数が X_{merge} の場合と，それぞれの X_{tfidf} と X_{atr} との比較をすると，X_{tfidf} か X_{merge} のどちらかが精度が高いという結果になっている．つまり，X_{tfidf} を説明変数より，X_{atr} の属性が分類しやすいということが分かる．特に，(a)「見学するかどうか」や (d)「成約施設の種別」に関して，X_{atr} から推定することは難しいということが分かる．

したがって，重要因子項目の可視化分析は，X_{tfidf} と X_{atr} の 2 つの説明変数のみでの結果を用いる．

4.2 見学したかどうか

見学したかどうかを分類する分類器を分析した．

4.2.1 結果

見学したかどうかを目的変数とし，説明変数は X_{atr} と X_{tfidf} を用いて重要属性の可視化分析を行い，各結果を**図 3** と**図 4** に示す．この結果は，説明変数の文字が多く，図に全体を表すと文字が見えないので，議論で必要な上位 20 項目のみを示す．

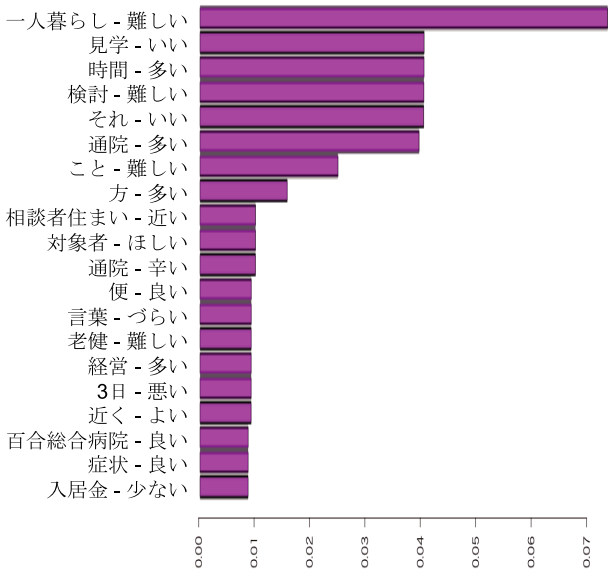


図 4 「見学したかどうか」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図（目的変数：見学したかどうか，説明変数：「相談経緯」テキストから作成したペア単語行列 X_{tfidf}，縦軸：属性名，横軸：重要度，棒グラフのピンク色の属性が正の相関示し，青色の属性は負の相関であることを示している）

Fig. 4 The important factors and correlations when X_{tfidf} is used as an explanatory variable to whether observation was done (objective variable: that whether observation was done, explanatory variable: word pairs matrix X_{tfidf} made from "Consultation background" text, vertical axis: attribute name, horizontal axis: importance, pink attribute of the bar graph shows positive correlation, blue attribute shows. negative correlation).

図 3 は「見学したかどうか」に対して X_{atr} を説明変数としたときの重要因子と相関を表した図である．

表 3 の結果から，上位の属性の属性値が欠損値であるもの多く，負に相関することが分かる．また，「利用期間」属性に注目すると，属性値が「欠損値」は負に相関するが，「長期」となっているものは正に相関することが分かる．また，「続柄 1」属性に注目すると，属性値が「欠損値」「父」「義母」に関しては負に相関するが，「母」は正に相関することが分かる．そして，「移動手段」属性に注目すると，属性値が「欠損値」「電車」の場合は，負に相関し，「車」の場合は正に相関することが分かる．

図 4 は「見学したかどうか」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図である．図 4 と図 3 ともに，縦軸：属性名，横軸：重要度，棒グラフのピンク色の属性が正の相関を示し，青色の属性は負の相関であることを示している．

表 4 において，ほとんどの重要因子が正に相関すること

© 2018 Information Processing Society of Japan

が分かる．重要な属性値として，「1 人暮らし–難しい」や「見学–いい」「時間–多い」「検討–難しい」の単語ペアがある．「1 人暮らし–難しい」や「検討–難しい」は相談者の相談内容と推測できる．また，「見学–いい」「時間–多い」などは相談者の希望であると推測できる．特に，「見学–いい」という属性値は重要度が高いことは納得できる．

4.2.2 考察

「見学するかどうか」を目的変数とした場合の考察をまとめる．表 3 の結果から，見学をしない人は欠損値が多い傾向にあることが分かった．また，見学をする人は施設の利用期間を長期に考えている人や移動手段が車である傾向にあることが分かった．これらの知見から，コールセンタは，より情報を聞き出すことで進捗を進ませ，契約に近づけることが可能であると思われる．提案して，チャットボットで相談者と気軽な質問対応をすることがあげられる．そうすることで，コールセンタにとって人件費削減，データ収集などのいい点がある．また，図 2 の結果より，見学するかどうかは属性情報から高い精度で判断できることが分かっている．そこで，今回生成した関数モデル F を用いて AI チャットボットとし，「見学をする意思があるか」を自動で判断できる可能性がある．また，移動手段が車である場合は，見学に行く傾向であることが分かった．このことより，送迎バスなどの見学を促す仕組みがあれば，見学者は増加すると考えられる．

また，図 4 の結果から，相談経緯から，「1 人暮らし–難しい」という単語ペアが現れると見学する傾向にあることが分かった．しかし，相談経緯に「1 人暮らし–難しい」の現れるサンプルデータの「追客備考」テキスト項目を見てみると，その後成約しない傾向にあることが分かった．その原因は，入居が本人の意思でないことであった．

4.3 成約したかどうか

成約したかどうかを分類する分類器を分析した．

4.3.1 結果

成約したかどうかを目的変数とし，説明変数は X_{atr} と X_{tfidf} を用いて重要因子可視化分析を行い，各結果を**図 5** と**図 6** に示す．この結果は，説明変数の文字が多く，図に全体を表すと文字が見えないので，議論で必要な上位 20 項目のみを図 5 に表す．

図 5 は「成約したかどうか」に対して X_{atr} を説明変数としたときの重要因子と相関を表した図である．

表 5 の結果から，全体的に入居者の状態を表す属性が重要な因子であることが分かる．たとえば，「介護度」「現在地」「認知症有無」「ADL. 入浴 1」が入居者の状態を表す属性に該当する．この中で，「介護度」属性に注目すると，すべての属性値で正に相関することが分かる．また，「現在地」属性に注目すると，「在宅」「入院中」が成約の正に相関することが分かる．また，「入居希望時期」属性に注目

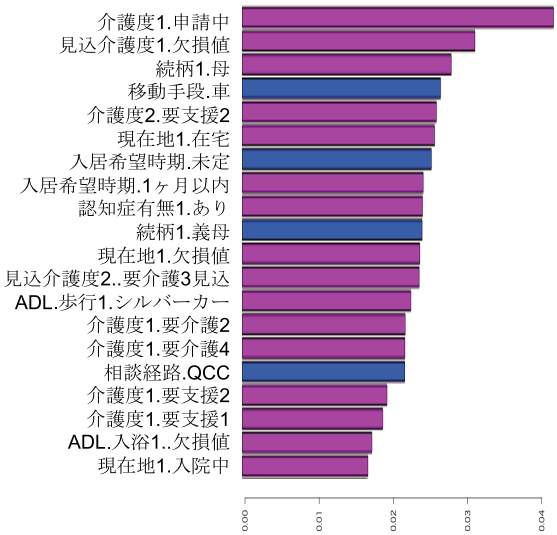


図 5 「成約したかどうか」に対して X_{atr} を説明変数としたときの重要因子と相関を表した図（目的変数：成約したかどうか，説明変数：その他の属性項目 X_{atr}，縦軸：属性名，横軸：重要度，棒グラフのピンクが正の相関示し，青色の因子は負の相関であることを示している）

Fig. 5 The important factors and correlations when X_{atr} is used as an explanatory variable to whether contract was signed (objective variable: that whether contract was signed, explanatory variable: other attribute item X_{atr}, vertical axis: attribute name, horizontal axis: importance, pink attribute of the bar graph shows positive correlation, blue attribute shows. negative correlation).

すると，属性値が「未定」であると，負に相関し，「1 カ月以内」だと，正に相関することが分かる．また，「入居希望時期」属性に注目すると，「できるだけ早く」「1 カ月以内」は正の相関であり，「2 カ月以上」「3 カ月以上」は負に相関することが分かる．図 5 の結果から，「続柄 1」属性に注目すると，「義母」「義父」「本人」は負に相関し，「母」「父」「祖母」「祖父」は正に相関することが分かる．

図 6 は「成約したかどうか」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図である．図 6 と図 5 ともに，縦軸：属性名，横軸：重要度，棒グラフのピンク色の属性が正の相関を示し，青色の属性は負の相関であることを示している．

図 6 の結果から，すべての重要因子が正の相関を示していることが分かる．重要属性を見ると，「介護–難しい」「2 人きり–難しい」「身寄り–ない」などの相談内容が上位にきていることが分かる．また，「相場–高い」「資金–厳しい」「入居金–少ない」などのお金の相談が多いことが分かる．

4.3.2 考察

表 5 の結果から，上位に「介護度」属性が存在し，すべての属性値に対して正に相関することから，介護度は入居時に必要な情報であると考えられ，実際に介護度を知っておく必要がある．さらに，介護度が低いから入居できないということはないことも示している．介護度について，詳

© 2018 Information Processing Society of Japan

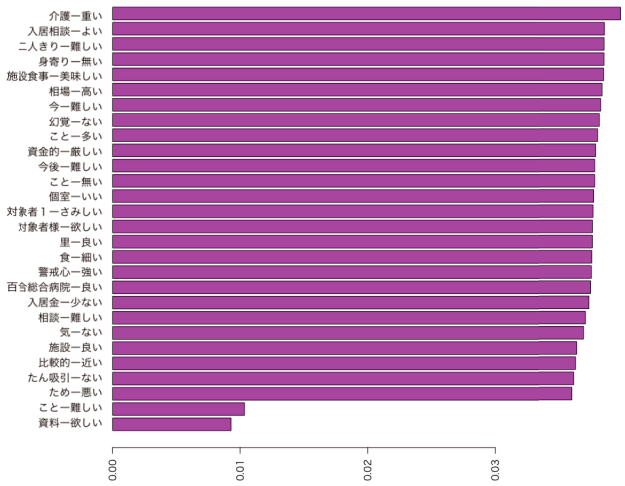


図 6 「成約したかどうか」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図（目的変数：成約したかどうか，説明変数：その他の属性項目 X_{tfidf}，縦軸：属性名，横軸：重要度，棒グラフのピンク色の属性が正の相関を示し，青色の属性は負の相関であることを示している）

Fig. 6 The important factors and correlations when X_{tfidf} is used as an explanatory variable to whether contract was signed (objective variable: that whether contract was signed explanatory variable: word pairs matrix X_{tfidf} made from "Consultation background" text, vertical axis: attribute name, horizontal axis: importance, pink attribute of the bar graph shows positive correlation, blue attribute shows. negative correlation).

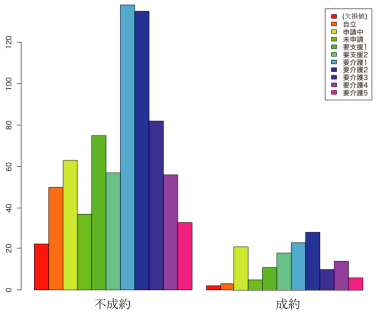


図 7 介護度と成約したかどうかの比較図（横軸は成約したかどうか，縦軸はサンプル数，棒グラフの色は介護度）

Fig. 7 Comparative chart of care degree and whether contract signed was done (the horizontal axis shows whether contract signed was done the vertical axis shows the number of samples, and the color of the bar graph shows the degree of care).

しく考察するため，今回のデータから**図 7** のような，比較図を作成した．横軸は成約したかどうか，縦軸はサンプル数，棒グラフの色は介護度を示している．「介護度」属性が自立と判定されても入居している人は存在することが分かった．そして，自立の人で成約した人の割合は 5%いることが分かった．この割合は，申請中が一番高く 25%であり，自立の確率は低いが，受け入れてくれる介護施設は存在している．この知見は，今後の相談者や入居者の介護施設を考えるタイミングのアドバイスになると考えられる．

「現在地」属性に注目すると，「在宅」「入院中」が正に相関することから，家での 1 人暮らしが寂しい，もしくは，退院後の場所として介護施設に入居したいという背景がうかがえる．実際に「相談経緯」テキスト項目から，「退院後入居したい」や「1 人暮らし」などの単語が含まれると成約する対象者でいることが分かった．

「入居希望期間」属性に注目すると，「2 カ月以上」を境に契約する傾向かどうかが分かれている．これは，入居者は 1 カ月以内で決めたいという差し迫った状態だと，入居を 1 カ月の間で決めてしまう傾向にあると考えられる．実際にデータの調査を行うと，「入居希望時期」属性が「1 カ月以内」「できるだけ早く」としているサンプルの成約率は 8%であったのに対して，それ以外のサンプルの成約率は 2%と約 4 倍も違うことが分かった．また，「現在地」属性の考察と「入居希望期間」属性を合わせると，「入院期間が終わるから，1 カ月以内で入居したい」や，「1 人暮らしで心配だから，できればすぐ入居させたい」という背景が考えられる．他の介護入居に関する調査 [22] では，十分な準備期間がないままに入居検討を迫られることが多い，ということが分かっている．つまり，もっと介護施設検討時期を設けることが，介護施設入居後も安心した生活がすごせると思われる．

また，「続柄」属性が，「義母」「義父」「本人」は負に相関し，「母」「父」「祖母」「祖父」は正に相関することから，「続柄」によっても成約に影響があることが分かる．この違いを分析し，提示すること，もしくは類似の相談者状況を提示することで，成約への意識の変化があるかもしれない．たとえば，「続柄」が本人である場合，成約した相談内容の提示をすることで，「同じような人がいる」という安心感から，成約に至ることが可能性としてはある．

また，図 6 の結果から，「介護–難しい」「2 人きり–難しい」「身寄り–ない」などの相談内容をする人の契約率が高いことが分かる．特に，「2 人きり–難しい」「身寄り–ない」といった人間関係に悩む背景を持つ相談者が介護施設を必要としていることが分かる．

4.4 見学施設はどのようなタイプか

見学施設の種別が「介護付き（0）」か「住宅型（1）」かを分類する学習器を分析した．

4.4.1 結果

見学施設の種別が「介護付き（0）」か「住宅型（1）」か，を目的変数とし，説明変数は X_{atr} と X_{tfidf} を用いて重要因子可視化分析を行い，各結果を**図 8** と**図 9** に示す．この結果は，説明変数の文字が多く，図に全体を表すと文字が見えないので，議論で必要な上位 20 項目のみを図 8 に示す．

図 8 は，「見学した施設が「介護付き」か「住宅型」か」

© 2018 Information Processing Society of Japan

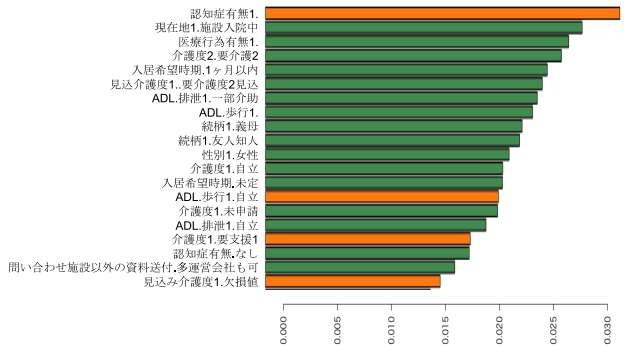


図 8 「見学した施設が「介護付き」か「住宅型」か」に対して X_{atr} を説明変数としたときの重要因子と相関を表した図（目的変数：見学した施設が「介護付き」か「住宅型」か，説明変数：X_{atr}，縦軸：属性名，横軸：重要度，棒グラフの黄色が「介護付き」の相関を示し，緑色が「住宅型」の相関を示す）

Fig. 8 The important factors and correlations when X_{atr} is used as an explanatory variable to the type of observation facility which can be either “with nursing care” or “residential type” (objective variable: facility type either “with nursing care” or “residential type”, explanatory variable: other attribute item X_{atr}, vertical axis: attribute name, horizontal axis: importance, the yellow color of the bar graph shows the correlation of “with nursing care”, and the green color shows the correlation of “residential type”).

に対して X_{atr} を説明変数としたときの重要因子と相関を表した図である．

表 8 の結果から，入居対象者の状態を表す属性が上位に多く存在することが分かる．特に，「認知症の有無」に注目すると，属性値が「欠損値」だと介護付きに相関があり，「なし」だと住宅型に相関があることが分かる．ほかにも「介護度」や「ADL」という単語を含む属性が重要であることが分かる．「介護度 1」に注目すると，要支援は介護付き，自立は住宅型に相関があることが分かる．

図 9 は，「見学した施設が「介護付き」か「住宅型」か」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図である．図 8 と図 9 ともに，縦軸を項目名，横軸を重要度，棒グラフの黄色が「介護付き」の相関を示し，緑色が「住宅型」の相関を示す．

図 9 の結果から，まず，「費用」について悩んでいる傾向が見える単語ペアが 4 件存在する．それらの相談者は「介護付き」を選択する傾向にあることが分かる．そして，「入居者の状態」について，介護が必要であることが想定できる単語ペアが 7 件存在し，「介護付き」を見学していることが分かる．また，「距離」について悩んでいると想定できるのは 3 件存在することが分かった．

4.4.2 考察

一般的に，介護度が自立の人は「住宅型」を選択することが分かった．つまり，介護を重要視し，それに対応する施設を見学しているように思える．しかし，介護度による

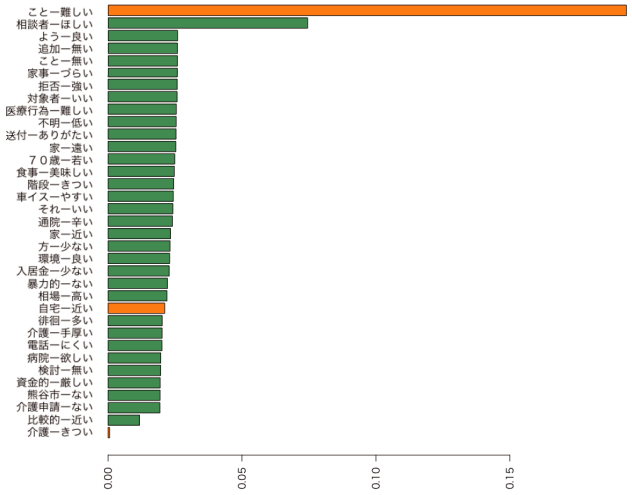


図 9 「見学した施設が「介護付き」か「住宅型」か」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図（目的変数：見学した施設が「介護付き」か「住宅型」か，説明変数：X_{tfidf}，縦軸：属性名，横軸：重要度，棒グラフの黄色が「介護付き」の相関を示し，緑色が「住宅型」の相関を示す）

Fig. 9 The important factors and correlations when X_{tfidf} is used as an explanatory variable to the type of observation facility, either “with nursing care” or “residential type” (objective variable: which is the observation facility “with nursing care” or “residential type”, explanatory variable: word pairs matrix X_{tfidf} made from “Consultation background” text, vertical axis: attribute name, horizontal axis: importance, the yellow color of the bar graph shows the correlation of “with nursing care”, and the green color shows the correlation of “residential type”).

ものだと推定精度が高いと思われるが，図 2 によると推定結果は 60%周辺で，あまり高いとはいえない．この原因として考えられるのは，入居できる施設は変わるわけではなく，「住宅型」も「介護付き」も入居可能な介護度は制限がないところがほとんどである．したがって，介護がまだそこまで必要でない高齢者は将来的に「介護付き」に入居することも考えられる．一方で，介護付きより，住宅型のほうが，入居金が安いなどお金的な面や，図 9 の結果にも表れている，「距離」の条件も考慮すると，介護度が高くても「住宅型」に入居することもありうる．つまりは，「介護度」以外の面から施設種別を選択していることが考えられる．

4.5 成約施設はどのようなタイプか

成約施設の種別が「介護付き（0）」か「住宅型（1）」かを分類する学習器を分析した．

4.5.1 結果

成約施設の種別が「介護付き（0）」か「住宅型（1）」か，を目的変数とし，説明変数は X_{atr} と X_{tfidf} を用いて重要因子可視化分析を行い，各結果を**図 10** と**図 11** に示す．

図 10 は，「成約した施設が「介護付き」か「住宅型」か」

© 2018 Information Processing Society of Japan

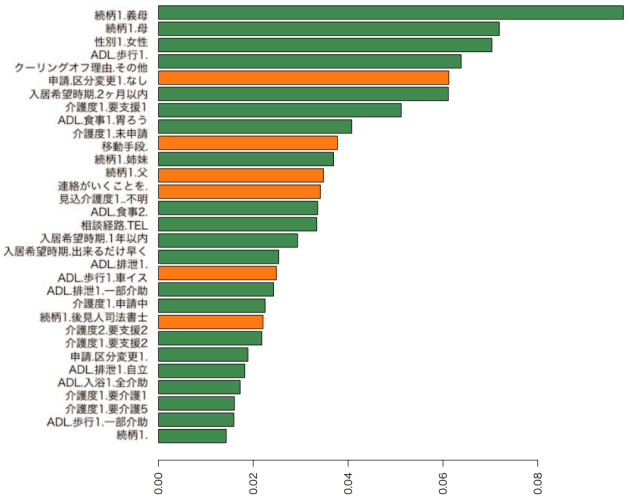


図 10 「成約した施設が「介護付き」か「住宅型」か」に対して X_{atr} を説明変数としたときの重要因子と相関を表した図（目的変数：成約した施設が「介護付き」か「住宅型」か，説明変数：X_{atr}，縦軸：項目名，横軸：重要度，棒グラフの黄色が「介護付き」の相関を示し，緑色が「住宅型」の相関を示す）

Fig. 10 The important factors and correlations when X_{atr} is used as an explanatory variable to the type of contract signed facility “with nursing care” or “residential type” (objective variable: which is the contract signed facility “with nursing care” or “residential type”, explanatory variable: other attribute item X_{atr}, vertical axis: attribute name, horizontal axis: importance, the yellow color of the bar graph shows the correlation of “with nursing care”, and the green color shows the correlation of “residential type”).

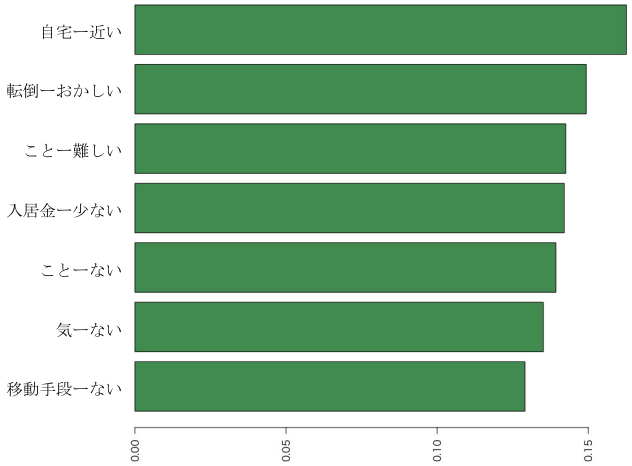


図 11 「成約した施設が「介護付き」か「住宅型」か」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図（目的変数：成約した施設が「介護付き」か「住宅型」か，説明変数：X_{tfidf}，縦軸：項目名，横軸：重要度，棒グラフの黄色が「介護付き」の相関を示し，緑色が「住宅型」の相関を示す）

Fig. 11 The important factors and correlations when X_{tfidf} is used as an explanatory variable to the type of contract signed facility “with nursing care” or “residential type” (objective variable: which is the contract signed facility “with nursing care” or “residential type”, explanatory variable: word pairs matrix X_{tfidf} made from “Consultation background” text, vertical axis: attribute name, horizontal axis: importance, the yellow color of the bar graph shows the correlation of “with nursing care”, and the green color shows the correlation of “residential type”).

に対して X_{atr} を説明変数としたときの重要因子と相関を表した図である．

図 11 は，「成約した施設が「介護付き」か「住宅型」か」に対して X_{tfidf} を説明変数としたときの重要因子と相関を表した図である．図 10 と図 11 ともに，縦軸を項目名，横軸を重要度，棒グラフの黄色が「介護付き」の相関を示し，緑色が「住宅型」の相関を示す．

図 10 の結果から，上位に「続柄」属性が多く，次に「介護度」や「ADL」のような入居者の状態を表す属性が多いことが分かる．「続柄」属性に注目すると，「続柄. 義母」「続柄 1. 母」「続柄. 姉妹」は住宅型を成約し，「続柄 1. 後見人司法書士」は介護付きに契約する傾向にあることが分かる．また，介護付きに成約する傾向にある属性に注目すると，「移動手段」「ADL 食事 2」「連絡が行くことを」などの属性が欠損値であることが分かる．

図 11 の結果から，「自宅–近い」「入居金–少ない」などの施設の希望や「転倒–おかしい」という入居者の問題点と考えられる単語ペアがあることが分かる．

4.5.2 考察

図 10 の結果から，女性が「住宅型」に成約する傾向があることが分かる．厚生労働省の調査によると，女性の健康寿命と寿命の差は男性の健康寿命と寿命の差より 3 歳ほど長いことが平成 22 年の調査で分かっている [23]．つまり，介護が必要な長さが女性のほうが長いということである．しかし，実際には介護付きではなく，住宅型に相関がある．その理由を「追客備考」テキスト項目から調査した．その結果，女性の施設入居の決め手は，「部屋のきれいさ」や「対応がいい」など直接的に感じる印象で決定しているからであるといえる．

5. 考察まとめ

前章の各節で行った考察をここでまとめ，そのまとめた結果から新たな考察をする．

まず，コールセンタにとって，有用な情報と考えられるのは，

- 見学するかどうかは，属性情報から高い精度で判定することができ，見学しないと判定される要因因子は，欠損値であるため，AI チャットボットのような，相談者が気軽に情報を提供しやすいシステムの導入が望まれる．
- 「2 人きり–難しい」「身寄り–ない」という相談背景から成約する傾向があることが分かった．

© 2018 Information Processing Society of Japan

次に，相談者や入居者に有用な情報であったのは，

- 介護度を調べることは入居時に必要な情報であるため，事前に知っておくのがよい．
- 介護度が「自立」の場合でも介護施設に入居している人は，5%存在する．
- 準備期間が 1 カ月以内に迫って入居をしている人と 2 カ月以上のゆとりをもって入居する人の成約確率は 4 倍もの差がある．
- 「入院期間が終わりそうだから」や「1 人暮らしになって心配だから」と入居を考えると，入居検討準備期間が短い
- 成約するかどうかは「続柄」属性が重要であることが分かった．

最後に，介護施設にとって，有用な情報であったのは，

- 「介護度」や「住宅型」の違いの面から考えて，見学施設の決定を行うことより，お金や距離的な面で選択する傾向にある人が多い．
- 女性が住宅型に決める傾向にあるが，施設の機能ではなく，「きれいさ」や「印象」が決め手になっている．
- 移動手段が車であると見学に行く傾向にあるため，見学の際の送迎のアクセスの改善が進捗に効果的かもしれない．

これらの一環の考察として，「続柄」属性や「介護度」属性が成約に重要であることが多いことが分かった．これらのことから，「続柄」では「本人」「義母」「義父」などが成約に負の相関であったことから考察すると，相談できるコミュニティを作り相談し合うことで，同じような人がいる安心感や悩みの軽減ができる可能性がある．施設決定の決め手として，「施設種別の傾向分析」からともに「距離」が重要であることがあげられた．これは，入居条件が家族が住んでいる場所と近い場所にある施設で決めていると考えられる．しかし，そのような条件は施設の制限が強制されてしまう．今回の属性情報として，「交通手段」情報があるため，車で行くと簡単な範囲や，電車で行くと簡単な範囲という制限を広げる必要がある．そのため，施設探しとして，距離制限を交通手段によって変わるシステムの導入が望まれる．

また，今回の分析から，入居後のデータ分析では分かりえなかった知見も得られた．「2 人きり–難しい」「身寄り–ない」という相談背景から成約を考えているケースがあることや「自立」でも入居している人は 5%いるということから，これから「介護」の必要になった時期ではなく，早めの入居を考える人が増えてくることが考えられる．しかし，介護度が「自立」で入居可能な施設は少なく，公的な施設だとほとんどない [24]．

今回は，介護度や認知症の有無から，成約か見学をするかの将来予測を試みたが，コールセンタの業務では，顧客満足度が大切で，顧客の喜怒哀楽を読み取ることが大切な指標である．最近では，テキスト文から感情分析をする API もある [25]．つまり，喜怒哀楽を認識する感情分析をすることは，コールセンタ業務改善に有用と考えられ，今後取り組むべき課題の 1 つである．

また，今回の知見と考察から得られたこととして，突然の 1 人暮らしや健康状態の悪化による，早急な入居が多いということが分かった．この急いだ入居は，経済的な準備不足や施設の検討が不十分であるという問題が考えられる．近年では，自宅においてセンサを用いた高齢者の 1 人暮らしを見守るシステムの研究がある [26]．これらのシステムにより，自宅の見守りシステムなどの出力から介護レベルを測定し，介護施設に入居する必要があるかどうかを予知できる可能性がある．この見守りシステムからの出力には，前述の AI チャットボットを自宅に導入し，そこから会話で得られる情報も含まれる．この研究の方向も，介護分野における重要な研究になるといえる．

本稿における分析やその結果の解釈における注意点としては，本分析では，真の相関と偽相関を区別するところまでは立ち入っていない点である．結果につながる真の原因ではないのに相関があるように見える，偽相関の要因に対して制御をかけても，結果が分析結果どおりに影響するとはいえない．真の相関と偽相関を区別するには，また介護分野の専門家が入念にこれらの結果を精査することが必要であるが，少なくとも今回の知見は真の原因の候補を列挙できたといえる．

今回の知見から，2 章で述べた介護分野における問題の「突然の退去命令」について考えた際，結果から突然の 1 人暮らしや健康状態の悪化による準備不足が原因にあると推測される．そのため，準備不足にならないためには，健康管理を日々行うことや，見守りシステムの導入が望まれる．また，「介護士による虐待」などの他の介護分野問題を解決するための考察をするには，まだ課題が残る．なぜなら，今回のデータは入居前の傾向を調べることが目的で，入居後にトラブルが起きるかどうかの予測はさらなるデータが必要である．そこで，安価で手軽にデータ収集できるクラウドソーシング技術を活用することで，十分なデータが集まる可能性がある [27]．

6. まとめ

本稿では，介護施設紹介コールセンタの記録データをアンサンブル学習法を用いて，分析を行った．分析方法は，重要属性を棒グラフで，属性と目的変数の相関を色で，同時に可視化することで，考察を容易にした．また，説明変数を 1 と 0 の 2 値にしたこと，テキスト部分をペア単語としたことの 2 点も理解を容易にした点である．

分析内容は，(a)「見学したかどうか」，(b)「成約したかどうか」，(c)「見学する種別が介護付きか住宅型か」，(d)「成約する種別が介護付きか住宅型か」を目的変数とし，1.

© 2018 Information Processing Society of Japan

相談経緯テキスト文から生成した形容詞特徴量ベクトル郡と，2. その他の属性郡，2 つの説明変数集合を用いた．そうすることで，この 4 つの目的変数に対する相談者と入居者の傾向を分析し，相談者と入居者，コールセンタ，介護施設の 3 者それぞれに有用な知見を取得する．

そこで，本稿では，介護施設に入居する前のサービスである介護施設紹介コールセンタのデータを分析し，施設入居前の状況を分析する．このデータは，相談者 1 名に対し 1 行，「性別」や「介護度」のような属性を列とした，7,685 行 106 列のビッグデータである．このデータを分析することは，コールセンタ，介護施設，相談者と入居者の 3 人の登場人物にとって有用な知見が得られることが期待される．たとえば，「見学，成約する人はどのような人か，どのような傾向にあるのか」を分析することで，コールセンタにとっては，より良いサービスの改善につながる可能性がある，というような知見が取得できる．そこで本研究では，介護施設紹介コールセンタのデータを分析し，介護施設，相談者や入居者，コールセンタに有用な知見を取得する．分析方法は，アンサンブル学習を用いて，説明変数の属性値を 2 値化し，属性の重要度と目的変数に対する正負の相関を同時に可視化した．

結果，コールセンタにとって，有用な知見として，欠損値が多い相談者は見学に行かない傾向にあること，移動手段が車だと見学に行く傾向にあることが分かった．相談者や入居者に有用な知見であったのは，介護度が「自立」でも入居している人は 5%いること，準備期間が 1 カ月以内に迫って入居をしている人と 2 カ月以上のゆとりをもって入居する人の成約確率は 4 倍もの差があること，などがあげられる．介護施設にとって，有用な情報であったのは，「2 人きり–難しい」「身寄り–ない」という相談背景から成約する傾向があることが分かったこと，「介護度」や「住宅型」の違いの面から考えて，見学施設の決定を行うことより，お金や距離的な面で選択する傾向にある人が多いことなどがあげられる．

それらの結果から，コールセンタのシステムとして AI チャットボットシステムの導入の可能性や，施設探しのマップ機能として交通手段で変化する機能の可能性があると期待された．

今回は，実際に介護施設入居後にトラブルがあったのかどうか分からないという課題があるため，今回分かった知見をもとに介護現場へのアンケートを作成し，安価で手軽にデータ収集できるクラウドソーシングでデータを活用することが重要であると考えた．

謝辞 本研究は株式会社 LIFULL（LIFULL Lab）と九州工業大学の共同研究によるものである．データの分析作業や得られた知見の有用性の検証にあたっては，株式会社 LIFULL senior の水崎 創介氏，泉 雅人氏，望月 義隆氏，北山 典子氏，薄葉 元紀氏の協力および助言を得た．

参考文献

[1] 日本創世会議：東京圏高齢化危機回避戦略 (2015)，入手先 ⟨http://www.policycouncil.jp/⟩.
[2] 厚生労働省：2025 年に向けた介護人材にかかる需給推計（確定値）について (2015)，入手先 ⟨http://www.mhlw.go.jp/file/04-Houdouhappyou-12004000-Shakaiengokyoku-⟩.
[3] 長谷川大悟，藤田好彦，坂本晴美，巻 直樹，若山修一，稲田晴彦，奥野純子，柳 久子：介護老人保健施設入所者の転倒発生状況 2，日本転倒予防学会誌，Vol.2, No.3, pp.23–32（オンライン），DOI: 10.11335/tentouyobou.2.3_23 (2016).
[4] 中野一茂：特別養護老人ホームにおける排泄時の転倒・転落について：その防止策の考察 (2010).
[5] 毛利光伸：実用的になった介護ロボット，旭リサーチセンター (2016).
[6] 厚生労働省：第 4 章見守り支援機器（介護施設），入手先 ⟨http://www.mhlw.go.jp/sinsei/chotatu/chotatu/wto-kobetu/2017/02/dl/wt0227-04_08.pdf⟩.
[7] 北九州市ホームページ介護ロボット等導入実証事業ワーキンググループ，入手先 ⟨http://www.city.kitakyushu.lg.jp/ho-huku/18200090.html⟩.
[8] シニアの安心相談室，入手先 ⟨https://www.senior-anshin.com/news/kaigo/20170920/⟩.
[9] 厚生労働省：平成 27 年度高齢者虐待の防止，高齢者の養護者に対する支援等に関する法律に基づく対応状況等に関する調査結果，入手先 ⟨http://www.mhlw.go.jp/file/04-Houdouhappyou-12304500-Roukenkyoku-Ninchishougyakutaiboushitaisakusuishinshitsu/0000155730.pdf⟩.
[10] Breiman, L.: Random Forests, *Mach. Learn.*, Vol.45, No.1, pp.5–32 (online), DOI: 10.1023/A:1010933404324 (2001).
[11] 福田賢一郎，濱崎雅弘，福原知宏，藤井亮嗣，堀田美晴，西村拓一：介護現場における申し送り情報の分析：業務改善に向けて（言語理解とコミュニケーション）–（第 5 回テキストマイニング・シンポジウム），電子情報通信学会技術研究報告 = IEICE technical report：信学技報，Vol.114, No.211, pp.11–16 (2014)（オンライン），入手先 ⟨https://ci.nii.ac.jp/naid/110009950306/⟩.
[12] 那須川哲哉：コールセンターにおけるテキストマイニング（＜特集＞「テキストマイニング」），人工知能学会誌，Vol.16, No.2, pp.219–225 (2001).
[13] 峯崎智裕，松木 萌，井上創造：介護サービス向上に向けた介護事故事例テキストの分析，情報処理学会論文誌，Vol.58, No.10, pp.1701–1711 (2017).
[14] Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System (online), DOI: 10.1145/2939672.2939785 (2016).
[15] 松木 萌，井上創造，清田陽司：介護施設紹介コールセンター記録のアンサンブル学習による傾向分析，pp.648–655 (2017).
[16] 工藤 拓，松本裕治：チャンキングの段階適用による日本語係り受け解析，情報処理学会論文誌，Vol.43, No.6, pp.1834–1842 (2002).
[17] 徳永健伸：情報検索と言語処理（言語と計算），東京大学出版会 (1999).
[18] Minezaki, T., Matsuki, M. and Inoue, S.: 介護サービス向上に向けた介護事故事例テキストの分析，Vol.58189, No.10124, pp.1701–1711 (2017).
[19] Greenwell, B.M.: pdp: An R Package for Constructing Partial Dependence Plots, *The R Journal*, Vol.9, No.1, pp.421–436 (2017) (online), available from ⟨https://journal.r-project.org/archive/2017/RJ-2017-016/index.html⟩.
[20] Inoue, S., Ueda, N., Nohara, Y. and Nakashima, N.: Mo-

© 2018 Information Processing Society of Japan

bile Activity Recognition for a Whole Day: Recognizing Real Nursing Activities with Big Dataset, *Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15*, pp.1269–1280, ACM (online), DOI: 10.1145/2750858.2807533 (2015).
[21] 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)（オンライン），入手先 ⟨https://ci.nii.ac.jp/naid/40021082607/⟩.
[22] 原田泳幸，滝山真也：施設（高齢者向けホーム）の入居者・保証人を対象とした「介護に関する意識調査」，入手先 ⟨https://www.benesse-style-care.co.jp/lab/news/docs/20151111research.pdf⟩.
[23] 厚生労働省：厚生労働省ホームページ，入手先 ⟨http://www.mhlw.go.jp/bunya/kenkou/dl/chiikigyousei_03_02.pdf⟩.
[24] 厚生労働省：介護事業所・生活関連情報検索」「平成 27 年度介護報酬改定に向けて（社）全国有料老人ホーム協会「有料老人ホーム・サービス付き高齢者向け住宅に関する実態調査研究」サービス付き高齢者向け住宅情報提供サービス (2018)，入手先 ⟨http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi_kaigo/kaigo_koureisha/kouhyou/index.html⟩.
[25] Google API, available from ⟨https://cloud.google.com/natural-language/docs/analyzing-sentiment?hl=ja⟩.
[26] 田中　仁，中内　靖：ユビキタスセンサによる独居高齢者見守りシステム（機械力学，計測，自動制御），日本機械学会論文集 C 編，Vol.75, No.760, pp.3244–3252（オンライン），DOI: 10.1299/kikaic.75.3244 (2009).
[27] 鹿島久嗣，梶野　洸：クラウドソーシングと機械学習（＜特集＞知識の転移），人工知能学会誌，Vol.27, No.4, pp.381–388 (2012).



松木 萌　（正会員）

1993 年生．2016 年九州工業大学総合システム工学科卒業．2017 年同大学大学院工学府先端機能システム工学専攻博士前期課程修了．現在，同大学院工学府工学専攻博士後期課程在学中．ユビキタスヘルスケアと行動認識（機械学習）に興味がある．



井上 創造　（正会員）

1997 年九州大学工学部情報工学科卒業．2002 年九州大学大学院システム情報科学研究科博士後期課程修了・博士（工学）．2002 年より同システム情報科学研究院・システム LSI 研究センター助手．2006 年より同附属図書館研究開発室助教授（准教授）．2009 年より九州工業大学大学院工学研究院基礎科学研究系准教授，2018 年より同大学院生命体工学研究科，現在に至る．この間，2009～2014 年九州大学附属図書館特別研究員，2011～2012 年九州大学大学院システム情報科学研究院非常勤講師，2014 年ドイツカールスルーエ工科大学訪問研究員，2015～2016 年九州先端科学技術研究所特別研究員，2017 年より理化学研究所革新知能統合研究センター客員研究員．本会シニア会員．



清田 陽司　（正会員）

株式会社 LIFULL 主席研究員．2004 年京都大学大学院情報学研究科博士課程修了．2004 年から 2012 年まで東京大学情報基盤センター助手・助教・特任講師．2007 年に東京大学発スタートアップとして株式会社リッテルを共同創業し，企業買収により 2011 年より現職．不動産分野における情報推薦や画像処理の研究開発に取り組むとともに，共同研究やデータセット提供を通じたオープンイノベーション推進に従事．博士（情報学）．人工知能学会編集委員会副委員長，Code4Lib JAPAN 共同代表．人工知能学会，言語処理学会，日本図書館情報学会各会員．

© 2018 Information Processing Society of Japan