

# 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

峯崎 智裕<sup>1</sup> 松木 萌<sup>1,a)</sup> 井上 創造<sup>1,b)</sup>

受付日 2016年12月23日, 採録日 2017年7月4日

**概要:** 本論文は, 複数介護施設における事故報告書を使用して, テキストマイニングおよび機械学習により事故の起こりうる状況を分析した結果について述べる. まず, テキストデータと多変量データが混合したデータに対して, テキストマイニングおよびランダムフォレスト法を組み合わせて, 事故の結果に対して影響する要因を抽出する方法を提案した. テキストデータについては分かち書きを行い単語文書行列を構成し, 単語についてクラスタリングを行った結果を元の多変量データと結合し, 事故の結果として目的変数を設定しランダムフォレスト法を2段階で適用しながら重要な単語クラスタおよび重要な単語と属性を抽出し, 決定木によって可視化するという手法である. この手法を, 介護施設から得られた5,189件の事故事例について適用したところ, (1) 表皮剥離または出血, 原因不明のあざが発見されて, 初めて事故が起きたことを介護士が把握していること, (2) リハビリやレクリエーション, 移乗を始めとする行動をとる場合, 転倒・ずり落ちによって打撲につながる可能性があること, (3) 共有スペース・個室両方での誤嚥が起こりうること, (4) 無断外出につながるいくつかの行動があること, が分かった.

**キーワード:** テキストマイニング, ランダムフォレスト, 決定木, 介護事故, 要因分析

## Text Analysis of Incident Reports for Preventing Accidents in Caregiving

TOMOHIRO MINEZAKI<sup>1</sup> MOE MATSUKI<sup>1,a)</sup> SOZO INOUE<sup>1,b)</sup>

Received: December 23, 2016, Accepted: July 4, 2017

**Abstract:** In this paper, we analyze incident report data from several nursing homes for understanding the situations related to the incidents and for reducing future incidents, by utilizing text mining and supervised machine learning. First, we propose a method for extracting important factors for incidents for the mixed data of text and multivariate data, combining word-document matrix, clustering, and the random forest. Next, we applied the method to 5,189 incident reports from the nursing homes, and found several patterns such as how the unknown bruises were discovered, behaviors leading to bruises from falls and falls, situations leading to aspiration, and behaviors leading to unattended going out.

**Keywords:** Text mining, random forest, decision tree, caregiving accidents, causality analysis

### 1. はじめに

近年, 介護分野における事故が問題になっている. たとえば, 実際の介護施設で起きている事故として, 高齢者の転倒や, 介護士が高齢者に間違った薬を配分する, 誤薬の事故がある. このような事故防止の対策には, 要介護 (要

支援) 者に対する介護士の人数を増やすことが考えられる. しかし, 現在, 要介護 (要支援) 認定者数は約 546.1 万人であるにもかかわらず, 介護福祉士の従業者数は約 66.1 万人であり (平成 24 年 9 月), 介護認定者に対して介護士が相対的に少ないことが分かる [3]. さらに, 厚生労働省が発表した需給推計によると, 2025 年度には介護職員が約 253 万人必要になるとされているのに対し, 供給の見込みは約 215 万人であり, およそ 38 万人の介護職員が不足する見込みである [4]. つまり, 事故防止の対策として, 介護士の人数を増やすことは現在でも将来的にも現実的な解決案とは

<sup>1</sup> 九州工業大学  
Kyushu Institute of Technology, Kitakyushu, Fukuoka 804-8580, Japan

a) matsuki@sozolah.jp

b) sozo@mns.kyutech.ac.jp

いえない。

そこで、事故予防のための他のアプローチとして、事故の状況や要因を知るために、テキストやデータを分析することが考えられ、過去にも多くの研究が存在する [7], [8], [9], [10], [11]。たとえば、近年の介護分野における問題として、「ベッドからの転落」、「段差につまずき転倒」などといった事故の割合が大きいが、同じような事故でも、原因や、時間帯や環境といった状況はまったく異なることもある。このような事故が起こる状況を総合的に理解し、同様の状況においてあらかじめ対策をとることは、事故の防止、または軽減のために有用であると考えられる。

本研究では、複数の介護施設から得られた実際の事故事例テキストをテキストマイニングおよび機械学習により分析し、事故の結果ごとにその要因や起こりうる状況を特定することを目的とする。文献 [7] は、事故の状況を調査するためインシデントレポートを開発し、データ収集と分析をしている。しかし、入力しやすいさに焦点をおいた自己開発インシデントレポートは、事故が起きた背景が読み取りづらく事故背景の詳細が分かりにくいと考察されている。本論文では、介護士がふだんからテキストで記録している実データの分析を行う。このデータは複数の介護施設で起きた事故・ヒヤリハットを記載した 5,189 件の事故事例テキストである。

実データの分析は、多変量データであるインシデントレポートとは違いテキスト表現であるため、表現の自由度が高くそのままでは分析しにくいという難しさがある。そこで、本論文は、まず、1. テキスト部分について分かち書きにより単語文書行列を構成し、2. 単語について  $k$ -means 法によってクラスタリングを行う。そうすることで、同じような表現をクラスタでまとめ、分析しやすい状態にした。そして、3. 単語クラスタ-文書行列と元の属性情報を結合して多変量データを構成し、4. いくつかの目的変数を設定し、ランダムフォレスト法により重要な属性と単語クラスタを抽出する。さらに、5. 単語クラスタから元の単語を取り出し、6. 再度属性情報と取り出した単語文書行列にランダムフォレスト法を適用して重要性の高い属性と単語を抽出した。その結果抽出された重要な単語と属性について決定木学習による可視化を行い、重大な事故の起こりうる状況を特定した。

分析の結果、(1) 表皮剥離または出血、原因不明のあざが発見されて、初めて事故が起きたことを介護士が把握していること、(2) リハビリやレクリエーション、移乗を始めとする行動をともなう場合に、転倒・ずり落ちによって打撲につながる可能性があること、(3) 共有スペース・個室両方での誤嚥が起こりうること、(4) 無断外出につながるいくつかの行動があること、といったいくつかのパターンが分かった。

まとめると、本論文の貢献は以下のとおりである。

- (1) 複数の介護施設からの実際の事故事例テキストを使用して、テキストマイニングおよび機械学習により事故の起こりうる状況と傾向を分析した。これは我々が知る限り初めての試みである。
- (2) テキストデータと多変量データを混合したデータに対して、テキストマイニングおよび変数の重要度を測ることができるランダムフォレスト法を組み合わせて、事故の結果に対して大きく影響した要因を抽出する方法を提案した。
- (3) その結果、重大な事故につながる状況を抽出することができた。このことは、今後普及が見込まれる IoT (Internet of Things) を始めとするユビキタスコンピューティング分野において、重点をおいてセンシングすべきコンテキストを人間のテキスト記録から抽出する試みとしてもこの分野に資する。

## 2. 背景と関連研究

介護分野における「ベッドからの転落」「階段につまずき転倒」といった転倒転落や薬の飲み間違い（誤薬）といった事故は、その後本人が健康状態を回復できない可能性もあり、本人の人生に関わる大問題ともなりうる。

理想的には、被介護者を世話する介護士の数を増やせば介護事故は減少すると考えられるが、日本では今後、少子高齢化が進み、労働人口が減少するため、現実的な解決策にはならない [3]。そのため、情報技術をはじめとする様々なアプローチによって、介護士の数を増やさなくとも事故を軽減する方策が期待される。

たとえば、種々のセンサを施設に設置して、転倒などの異常が発生したら検知する異常検出技術は存在する [17], [18]。しかし、これらは主に異常が発生した後に、もしくはその直前に検出することを目指すものであり、事故が起こったときの介護士や被介護者、環境に関する総合的な状況を分析し、予防につなげるものではない。

患者の将来予測の研究として、たとえば長期的な生活モニタリングや医療応用などにおいて、文献 [16], [19] の研究もある。介護施設でも、このような種々のデータを用いて事故を予防する取り組みが重要であろう。

介護施設における事故には、介護者や被介護者、そして環境や時間帯といった様々な要因の影響が考えられ、同じような事故でも原因はまったく違うこともある。このような状況を理解し予防するために、現場で記録されている事故報告書のような事故事例テキストをテキストマイニングや機械学習により分析し、主となる要因を突き止めることは、今後似たような事故が起こりうる状況において先んじて対策を取ることに繋がると考えられる。

介護現場の事故は、医療現場の事故データとは違い、厚生労働省の集計した、広く公開したデータが存在しないため、全国的にデータが統一されていない。そのため、三菱

総合研究所は、介護施設の事故データを全国的に集めることを目的とし、収集したデータの分析を行っている [8], [9]. データは1年間で、889市町村の施設から28,158件の事例データを収集した. 分析の結果、1年間で、1施設33.2件の事故が起きており、転倒事故が一番多いということが分かった.

文献 [10], [11] は、介護施設の「転倒・転落」に焦点を当てた、データの分析を行っている. 文献 [10] では、転倒における43件の多変量型事例データから重要な因子を主成分分析で抽出している. 文献 [11] では、1,039件の多変量事例データをクロス集計しトイレ内など方向転換が多い場所で事故が多発していることが分かった. このように、転倒や転落に焦点をおいた事例分析を行う研究は多くある. 文献 [7] は介護施設における事故を抽象的に把握するために、インシデントレポートを自身で開発し、16施設で1,039件のデータを収集しクロス集計を行い事故の要因を分析している. インシデントレポートの項目は22項目であるが、事故の背景の把握が不十分であると考察している.

一方、テキストマイニング技術は、経営学、医歯薬看護学、工学、経済学など幅広い分野で使われている [15]. たとえば、文献 [12] は、自然文の構造化を行い胸部CT検査の検査レポートの内容を分析し類似記載を特定している. 文献 [13] では、質問紙調査を用いて、業務で発生するエラー内容を分析するため、形態素解析し、形容詞、名詞のみを抽出し、Ward法を用いて階層クラスタリングを行った. その結果、業務改善や知見に関する新たな発見があり、活躍している. このように、自然文のデータを分析することはとても有用であることが分かる.

本研究では、介護士が記載した自然文と多変量データが結合された5,189事例を用いて介護事故事例の分析を行い、自然文と多変量事故データを組み合わせることで、事故の結果系の変数を目的変数とした教師あり学習を行うことで、いままで介護の事故に関して分からなかった、事故につながる、事故要因や状況を把握することを志向している.

事故という結果に対して、その要因を探るには、事故を

目的変数、要因の候補を説明変数とした多変量分析や教師あり機械学習が有用であるが、テキストデータはそのままの状態では多変量分析や教師あり機械学習を行うことはできない. 本研究では、複数の介護施設から得られた実際の事故事例テキストを、多変量データとテキストデータを統合分析する手法を用いて分析し、「事故の種類」「診断」「重症度」といった結果を目的変数としてその要因を探るというアプローチで分析を行う.

機械学習では一般に、特徴量空間の次元が大きくなると、それに伴って空間の容積が指数関数的に増えてしまい、サンプルを指数関数的に増やさない限り推定が難しくなるという次元の呪いの問題がある. そのため、次に行く機械学習の前段階で、特徴量データの次元を削減しておくことが、精度向上のためには有用である. 自動的な変数選択手法は数多く提案されており [20], 大きく分けると、変数の組合せをいくつか調べるラッパー法 [21] (重回帰分析で用いられるステップワイズ法も含まれる), 相互情報量などを用いて変数を選別するフィルタ法 [22], それらを組み合わせる方法 [23] に分かれる. このなかで、ランダムフォレスト、ブースティングといったアンサンブル学習には、アルゴリズム内でどの変数が推定に寄与したかという重要度を測定するメカニズムを持っているものもあるため、これを用いて重要度の大きい変数のみを残す方法も有効である. 我々は、このような次元削減の手法を応用して、目的変数に対する各説明変数の重要度を知ることができるというアプローチを用いた. 介護事故事例テキストを多変量化したデータにそのアプローチを適用する点が新しい.

このような介護事故事例テキストに対して自然文と多変量データを組み合わせて教師あり学習を行うアプローチは、我々の知る限り存在しない.

### 3. 事故事例テキスト

本章では、分析に使用した事故事例テキストについて述べる.

本研究では、ある介護サービス事業者が全国に持つ80の

#### 事故集計

平成〇年 〇月

施設名: 〇〇館

担当者: 〇〇, 〇〇, 〇〇, 〇〇

日	曜日	発生時間	氏名	年齢	介護度	認知度	利用区分	事故歴	事故の種類	発生場所	目的	詳細	重症度	診断	備考
1	11月	9時30分	〇〇〇〇様	82	介護1	中	特定	あり	原因不明のあざ・外傷	脱衣所	その他	1階浴室にて脱衣介助の際、右上腕部、右大腿部に直径10cm程度の内出血を発見する。上腕部は黄色に変色している部分が見られることから数日経過しているものと思われる。発見後ナースに報告する。	軽度	あざ	
2	11月	6時30分	〇〇〇〇様	86	介護1	軽	特定	なし	転倒・ずり落ち	居室(入口)	排泄	訪室時「5時半頃トイレに行こうと思って尻餅をついた」と話される。「腰が痛い」との事で確認するが外傷見られず。BP 139/69 67 KT35.8°C看護に報告する。	重症	骨折	後日、整形外科を受診し骨折との診断を受ける。
3	3日	17時40分	〇〇〇〇様	90	介護1	なし	ショートステイ	なし	その他	その他	その他	夕食後の薬に本人様よりタの薬に足りない物があるとの訴えあり、確認すると夕食後のファリール1mgが無い事に気づく。持参薬に1日分余分な物があった為、そちらを内服して頂く。	なし	その他	

図 1 事故報告書の集計の例

Fig. 1 An example of accident reports.



介護施設で起きた、事故・ヒヤリハットを記載した事故報告書を用いた。集計された事故報告書の例を図 1 に示す。

この報告書には事故の内容を記述した日時、年齢、介護度、認知度、事故歴、場所、目的、事故の種類、診断、重症度といった属性情報の他に、詳細や備考の欄に事故詳細テキストも記載されている。

各属性の取りうる値を表 1 に示す。このなかで、「利用区分」属性は、法的に定められた介護施設の種類である。また、「目的」属性は、利用者が行っていた行動の種類を表す。「診断」属性は、事故の結果の怪我や病気などの、スタッフによって判断された診断結果が記録される。「重症度」属性は、事故が起きた結果の深刻さの度合いを表す指標で、「重度」「軽度」および、問題のないほどの軽微な「なし」の三段階で示されている。

使用したデータは、事故件数 5,189 件であり、重度 1,134

表 1 各属性の取りうる値  
Table 1 Values of each attribute.

属性	取る値
年月日、曜日	年月日、曜日
発生時間	時刻
年齢	数字
氏名	氏名
施設名	施設名
介護度	「介護 1」「介護 2」「介護 3」「介護 4」「介護 5」「介護 6」「自立」「要支援 1」「要支援 2」
認知度	「なし」「軽」「重」「中」
利用区分	「グループホーム」「ショートステイ」「デイサービス」「ホームヘルプ」「ホテル」「住宅型」「小規模多機能」「特定」「特定 1 階」「特定 2 階」「特定 3 階」「特定 4 階」「特定 5 階」「特定 6 階」
事故歴	「あり」「なし」
発生場所	「1F フロアー」「2F エレベーター」「2F フロアー」「エレベーター前」「その他」「デイルーム」「トイレ」「フロアー」「ホール」「リハビリ室」「屋外」「居室」「居室 (トイレ)」「居室 (ベッド周囲)」「居室 (上記外)」「居室 (入口)」「共用トイレ」「玄関」「事務所」「自席横」「車輜 (破損事故)」「食堂」「静養室」「洗濯場」「洗面台」「脱衣所」「浴室」「冷倉庫前」「廊下」
目的	「その他」「リハビリ」「レク」「移乗」「移動」「臥床」「外出」「活力朝礼」「起立」「更衣」「行事」「車いす駆動」「車移動」「食事」「食前薬」「整容」「清掃」「洗濯」「朝食後薬」「入浴」「排泄」「服薬」「歩行」
事故の種類	「その他」「異食・誤飲」「火傷」「介護中のあざ・外傷」「感染症」「原因不明のあざ・外傷」「誤薬」「誤薬・誤配」「誤嚥」「自殺」「車両事故」「溺水」「転倒・ずり落ち」「転落」「暴力」「無断外出」
診断	「あざ」「その他」「火傷」「外傷・痛みなし」「誤嚥性肺炎」「骨折」「擦り傷」「出血」「心疾患」「診断」「打撲」「脳疾患」「肺炎」「表皮剥離」
重症度	「なし」「軽度」「重症」

件 (21.9%)、軽度 2,428 件 (46.8%)、なし 1,627 件 (31.4%) であった。また、事故の種類は、全部で 16 種類あり、そのうち「転倒・ずり落ち」が 2,986 件あり、一番サンプル数が多かった。

#### 4. 属性情報とテキストデータを統合した事故状況の分析方法

本章では、本論文で用いた、属性情報とテキストデータを統合して事故の起きやすい状況を分析をする手法を述べる。図 2 はその一連の流れを示す。テキストを文書単語行列として変数選択したいが、そのままと高次元化するため、いったん単語ベクトルをクラスタリングすることで次元を削減してから変数選択し、重要なクラスタ (変数) をもとの単語ベクトルにもどして再度変数選択する、というアプローチを取る。

- (1) 事故詳細テキストについて分かち書きを行い、毎回の報告を 1 つの文書と見なした単語文書行列を作る。
- (2) 単語文書行列の単語ベクトルを  $k$ -means 法を用いてクラスタリングを行うことで、単語をいくつかの単語クラスタに分ける。
- (3) 単語文書行列の単語を単語クラスタごとにまとめた単語クラスタ-文書行列と元の属性情報を結合し、各文書をサンプルとした多変量データと見なす。
- (4) 多変量データに対して目的変数を設定し、ランダムフォレスト法を用いて、重要度の高いクラスタを抽出する。
- (5) (1) の単語文書行列から、(4) で重要度の高かったクラスタを構成する単語ベクトルを取り出し、再度属性情報と結合し多変量データを構成する。
- (6) (5) の多変量データについて再度 (4) の目的変数につ

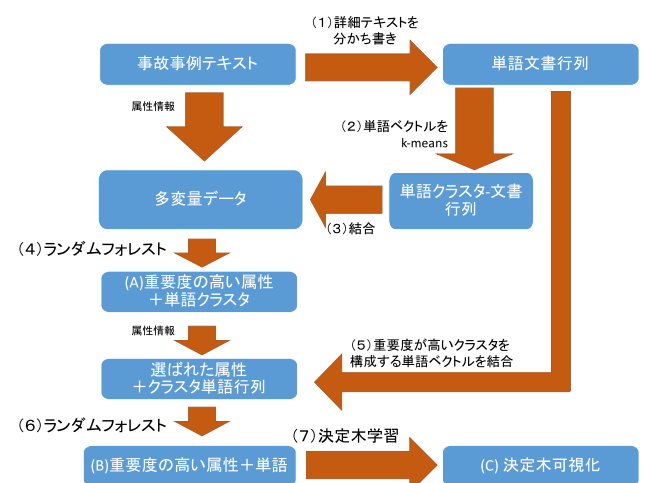


図 2 事故事例テキストから、事故の起きやすい状況を分析する一連の流れ。図中の (A), (B), (C) を 5 章の結果に示す

Fig. 2 Overview of analyzing the situations of accidents from text data. (A), (B), (C) in this figure are shown in Section 5.

いてランダムフォレスト法を適用し、重要度の高い属性と単語を抽出する。

(7) (6)で抽出された重要度の高い説明変数と(4)の目的変数を用いて、決定木で可視化する。

以下では、その詳細を述べる。

#### 4.1 単語行列の作成とクラスタリング

事件事例テキストから、事故詳細部分を取り出し、形態素解析ツール MeCab を用いて分かち書きを行う。そして、各文書に現れる単語数を値として単語・文書行列を作成する。このとき、分かち書きされた単語数が2,701個であり、 $2,701 \times 5,189$ の次元数の大きい単語文書行列となってしまうため、以下の問題が発生した。

- 行列が大きすぎて、分析しにくい。
  - 人が作成した文章であるため、同じ表現の単語がある。
- 2つ目の問題点は、たとえば、分かち書きした単語の中に「ベット」と「ベッド」があった。これは、同じ意味を表しているが、そのまま分析を行ってしまえば、違う単語として扱われ、重要度が1/2ずつに分散してしまう。

そこで、単語列の次元圧縮を行うことで、上記の問題点を解決する。次元圧縮を行うために、 $k = 50$ としてk-means法によるクラスタリングを行い、似た特徴の単語を似た意味の単語と見なす。その結果、 $50 \times 5,189$ の単語クラスタ文書行列となった。kの値はいくつか試行したが、小さくすると後段のランダムフォレスト法で重要な変数が抽出されず、大きくしすぎると要素が1つのクラスタばかりが発生したため、試行錯誤のうえ50と設定した。

#### 4.2 ランダムフォレスト法と決定木

クラスタリングした上記の単語クラスタを元に、単語クラスタ-文書行列を再度構成し、それと事故報告書の「曜日」「介護度」「認知度」「利用区分」「事故歴」「発生場所」「目的」「事故の種類」「診断」「重症度」の属性情報を結合し、多変量データを作成する。このなかで、結果系、つまり事故の結果分かると考えられる変数は、

- 「事故の種類」：起こった事故の種類
- 「診断」：事故の結果、スタッフによってなされた診断
- 「重症度」：事故の結果の深刻さの度合い

の3つである。したがって、この多変量データから上記3つの属性をそれぞれ目的変数と設定し、それ以外を説明変数とした。そして、それぞれの目的変数に対してランダムフォレスト法を適用し、機械学習を行った。

ランダムフォレスト法に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。

的に理解しやすい。

つまり、(1)により説明変数から「重症度」などの目的変数を精度よく推定しながら、(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、ランダムフォレスト法においては、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 [24]。

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。

### 5. 分析結果

本章では、4章で述べた属性情報とテキストデータの統合分析手法を3章で紹介したデータに適用した結果を述べる。

#### 5.1 属性と単語クラスタへのランダムフォレスト法の適用

4章で述べたとおり、今回は目的変数として「事故の種類」「診断」「重症度」の3つの属性を設定した。以下、各目的に属性と単語クラスタへランダムフォレスト法を適用した結果(図2の(A))を述べる。

「診断」を目的変数とした場合の推定精度は63.2%だった。このときの各説明変数の重要度が0.7%以上のものを表2に示す。ただし、単語数が非常に多かった10位の単語クラスタについてはその他の単語と考え、省略した。

表2 「診断」を目的変数として単語クラスタと属性を推定したときの重要度

Table 2 Importances of word-clustered explanatory variables when the objective variable is “Diagnosis”.

説明変数	変数重要度 (%)
1 重症度	10.30
2 「処置 剥離 出血 表皮」	5.20
3 事故の種類	4.40
4 「部 受診 骨折 大腿 日 病院」	1.70
5 目的	1.40
7 発生場所	1.00
8 「等 発赤 対応 頭部 腫脹 移乗 打撲 行い 疼痛」	0.70
9 「左」	0.70
10 利用区分	0.70

表 3 「事故の種類」を目的変数として単語クラスと属性を推定したときの重要度

Table 3 Importances of word-clustered explanatory variables when the objective variable is “Type of accident”.

説明変数	変数重要度 (%)
1 目的	4.50
2 重症度	4.30
3 「転倒」	3.40
4 診断	3.40
5 「薬」	3.20
6 「処置 剥離 出血 表皮」	2.20
7 発生場所	2.00
8 「訪室」	1.30
9 「の 発見」	0.70
10 介護度	0.70

表 4 「重症度」を目的変数として単語クラスと属性を推定したときの重要度

Table 4 Importances of word-clustered explanatory variables when the objective variable is “Severity”.

説明変数	変数重要度 (%)
1 診断	12.90
2 事故の種類	7.00
3 「部 受診 骨折 大腿 日 病院」	4.00
4 「処置 剥離 出血 表皮」	2.20
5 「薬」	1.40
6 目的	1.20
7 利用区分	0.80
8 発生場所	0.80
9 「外傷」	0.70

表 2 より、「重症度」属性、「処置 剥離 出血 表皮」単語クラス、「事故の種類」属性、「部 受診 骨折 大腿 日 病院」単語クラスの順に重要度が高いことが分かる。「処置 剥離 出血 表皮」は主に症状を表す単語クラスで「部 受診 骨折 大腿 日 病院」もまた症状や病院にかかったことに関するクラスといえるだろう。

「事故の種類」を目的変数とした場合の推定精度は、76.0%だった。このときの各説明変数の重要度が0.7%以上のものを表 3 に示す。

表 3 より、「重症度」属性、「転倒」単語クラス、「診断」属性、「薬」単語クラス、「処置 剥離 出血 表皮」単語クラス、「発生場所」属性、「訪室」単語クラスの順に重要度が高いことが分かる。「転倒」、「処置 剥離 出血 表皮」のように事故の内容に関するクラスである。

「重症度」を目的変数とした場合の推定精度は、80.1%だった。このときの各説明変数の重要度が0.7%以上のものを表 4 に示す。ただし、単語数が非常に多かった10位の単語クラスについてはその他の単語と考え、省略した。

表 4 より、「事故の種類」属性、「部 受診 骨折 大腿 日 病院」単語クラス、「処置 剥離 出血 表皮」単語クラス

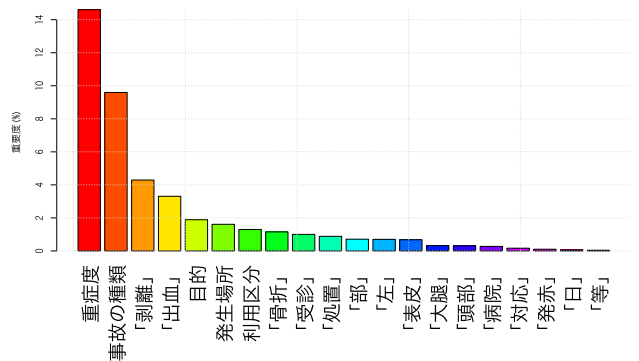


図 3 「診断」を目的変数としたときの属性と単語の重要度  
Fig. 3 Importances of attributes and words when the objective variable is “Diagnosis”.

タ、「薬」単語クラス、「目的」属性、の順に重要度が高いことが分かる。「処置 剥離 出血 表皮」のように事故の内容に関するクラスとともに、「薬」という誤薬に直結する単語クラスも抽出された。

### 5.2 属性と重要単語ベクトルへのランダムフォレスト法の適用

次に、上記で得られた重要な単語クラスについて、属性と元の単語ベクトルへランダムフォレスト法を適用した結果(図 2 の(B))を述べる。(A)の結果選択する単語クラスの基準としては、表 2~表 4 に現れた重要度が0.7%以上の単語クラスを抽出した。

図 3 は、「診断」を目的変数、属性と重要単語を説明変数としてランダムフォレスト法を行ったときの重要度である。このときの推定精度は、63.1%だった。単語の重要度に着目すると、「剥離」「出血」「骨折」「受診」「部」「左」「表皮」の順に重要度が高い。診断に直結するような症状を表す単語が上位にくることが分かる。ここで、単語「部」は、元のテキストにおいて「頭部」、「頸部」、「臀部」、「大腿部」といった体の部位を表す単語の一部であることが多かった。単語「左」に関して、「左腕」、「左足」、「左肩」、「左手」といった体の左側部位を表す単語の一部となっていた。

図 4 は、「事故の種類」を目的変数、属性と重要単語を説明変数としてランダムフォレスト法を行ったときの重要度である。このときの推定精度は、74.5%だった。単語の重要度に着目すると、「転倒」「薬」「訪室」「剥離」「発見」「出血」の順に重要度が高い。「転倒」や「薬」のような事故の種類に直結する単語が上位にくる。また、「訪室」が比較的上位にきているのが興味深い。「訪室」をもとのテキストで確認したところ、介護士が何かの目的で「訪室」し、事故を発見したと書かれている。つまり、部屋に入った介護士が事故を発見することが事故の種類に関係するといえる。

図 5 は、「重症度」を目的変数、属性と重要単語を説明

変数としてランダムフォレスト法を行ったときの重要度である。このときの推定精度は、82.1%だった。単語の重要度に着目すると、「受診」「病院」「剥離」「薬」「出血」「外傷」の順に重要度が高い。「受診」「病院」といった単語を元のテキストで確認すると、重症度が高く病院の受診につ

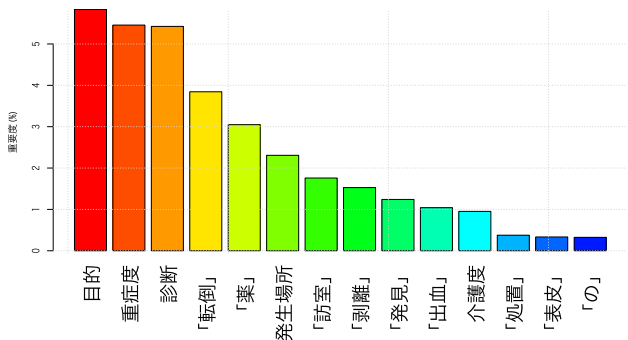


図 4 「事故の種類」を目的変数としたときの属性と単語の重要度  
Fig. 4 Importances of attributes and words when the objective variable is “Type of accident”.

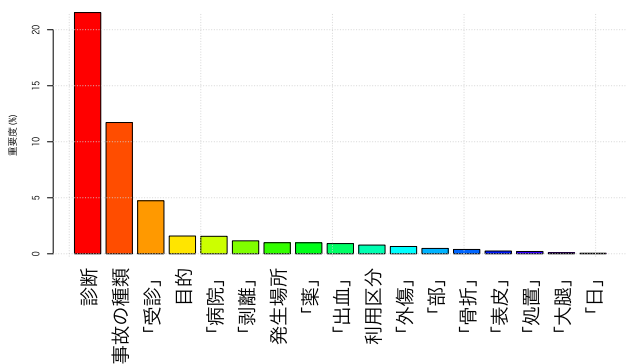


図 5 「重症度」を目的変数としたときの属性と単語の重要度  
Fig. 5 Importances of attributes and words when the objective variable is “Severity”.

ながった場合が多かった。また、「薬」は誤薬にという重症度の高い事故につながる単語であり、「剥離」「薬」「骨折」といった単語も軽度の重症度の事故につながる単語である。

5.3 属性と重要単語の決定木による可視化

最後に、上記で得られた重要な属性と単語について、決定木学習を行って複数の条件の可視化をした結果（図 2 の (C)）を述べる。

図 6 は、重要度の高い属性「目的」「発生場所」「利用区分」および単語ベクトル「剥離」「出血」「骨折」「受診」「処置」「部」「左」「表皮」を説明変数として選択し、「診断」を目的変数として決定木学習を行った結果である。

図より、単語「剥離」が現れると出血が多い。単語「出血」や単語「骨折」が現れると当然出血や骨折が多い。それ以外の場合、目的が「リハビリ、レク、移乗、移動、臥床、外出、活力朝礼、起立、更衣、車いす駆動、車移動、整容、朝食後薬、排泄、服薬、歩行」のときに、単語「受診」があれば打撲、「その他、行事、食事、食前薬、清掃、選択、入浴」のときには、発生場所によって痛みのない外傷かその他になるが、痛みのない外傷などの問題の少ない事故が多く、また推定精度も低い。

これらのことから、

- 出血：単語「剥離」や単語「出血」が現れる場合
- 骨折：単語「骨折」が現れる場合
- 打撲：それ以外の場合、目的が「リハビリ、レク、移乗、移動、臥床、外出、活力朝礼、起立、更衣、車いす駆動、車移動、整容、朝食後薬、排泄、服薬、歩行」のときに、単語「受診」がある場合
- 外傷・痛みなし、その他：その他の場合

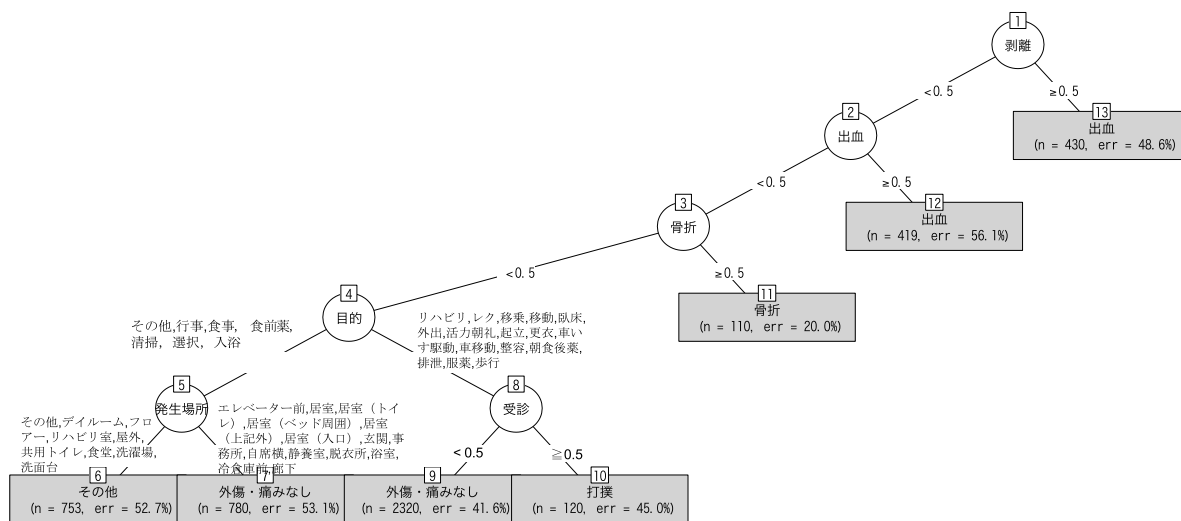


図 6 「診断」を目的変数とした決定木  
Fig. 6 Decision tree when the objective variable is “Diagnosis”.







分)および単語ベクトル「受診」「病院」「剥離」「薬」「出血」「外傷」を説明変数として選択し、「重症度」を目的変数として決定木学習を行った結果である。

図より、単語「薬」が現れたり単語「受診」が現れると重症につながり、そうでない場合には単語「剥離」や単語「出血」が現れると高い確率で軽度につながる。そうでない場合には、単語「外傷」が現れれば、なし(低い重症度)、現れなければ軽度につながるが、推定精度は低い。

これらのことから、

- 重症：単語「薬」や単語「受診」が現れる場合
  - 軽度：単語「剥離」や単語「出血」が現れる場合またはそれ以外で単語「外傷」が現れない場合
  - なし：上記以外で単語「外傷」が現れる場合
- といった状況で発生しやすいことが分かる。

## 6. 考察

以下では、目的変数として設定した「診断」「事故の種類」「重症度」についてそれらの要因となる状況を考察する。

### 6.1 「診断」を目的変数として分析を行った場合

「剥離」とは、ここでの意味は「表皮剥離」のことであり、テキスト上では、「剥離」や「表皮剥離」と表される。これは擦りむいた傷のようなものを表す。「剥離」を含む文書を見ると、剥離を発見し、事故の詳細を調査するといった流れが多い。また、「出血」も同様に事故調査のきっかけになる要因であることが分かった。つまり、表皮剥離や出血の原因となる事象については、事故時の状況の特定ができていないことが分かった。6.3節では、剥離や出血は、軽度の事故だと分類されているが、軽度とはいえ防止できることが望ましい。今後、この事故発生時点から過去に遡って生活パターンや行動を記録やセンシングし、事故が起きるタイミングを特定していく作業が必要である。

また、打撲については、リハビリやレクリエーション、移乗といった、利用者が移動や行動をともなうような目的の場合に、「受診」をともなう打撲が発生している。6.2節でも同様にリハビリや移乗といった行動をともなう際に転倒・ずれ落ちが発生することから、これらの行動の際に転倒・ずれ落ちによって、受診をともなう打撲が発生することが考えられる。つまり、リハビリやレクリエーション、移乗を始めとする行動にともなう場合、転倒・ずれ落ちによって打撲につながる可能性があることが分かった。今後、利用者がこれらの行動を行う際の事故のタイミングを分析していくことによって、行動ごとにどのタイミングでスタッフが介助したり声掛けをすることで事故を予防できるかを分析することが求められる。

### 6.2 「事故の種類」を目的変数として分析を行った場合

転倒・ずれ落ちについては、リハビリや移動といった被

介護者の行動をともなう場合や、場所が居室全般や廊下において頻繁に起きている。これにより、つまり場所と行動の種類が特定できれば、その危険性をある程度予測することができる可能性がある。今後、場所と行動をセンシングまたは認識できるシステムと組み合わせて、転倒・ずれ落ちの危険予知や事前通報ができれば有用であると考えられる。

原因不明のあざ・外傷については、6.1節で述べたように、剥離や出血によって気づくことが多く、予防にはさらなる生活パターンの分析が必要である。

誤嚥については、当然ながら食事中に起こりうる。元のテキストを見てみると、共有スペースで食事中にスタッフが目を放した間に起きる誤嚥と、自室内で何か食べ物を口にした際に誤嚥し、ナースコールなどでスタッフが気づく場合に分かれた。つまり、誤嚥の事故を防ぐためには、共有スペースと自室内の両方のパターンを考える必要があることが分かった。この両方について対策が必要であるが、どちらも食事や嚥下行動の認識ができるようなセンシング技術を開発すれば、これらの予防や検知が期待できる。

無断外出については、レクリエーションやレク、更衣、入浴といった目的で居室から外に出た場合に、玄関や屋外に出て無断外出してしまう場合があることが分かる。無断外出については、介護度や認知症といった個人差が関係してくることが考えられる。ちなみに、無断外出の事故の中で認知症(認知度が軽、中、または重)の割合は96.1%と、全体の事故の中での認知症の割合85.8%より10%程度高かった。つまり、無断外出を予防するためには入居者の行動を予測することが考えられるが、認知症患者に焦点をあてることがより重要だと分かった。

さらには個人差と合わせて、無断外出を予測するアルゴリズムを予測できる可能性を探ることが期待される。

### 6.3 「重症度」を目的変数として分析を行った場合

重症度については単語「薬」や単語「受診」が現れれば重症となり、単語「剥離」や単語「出血」が現れれば、または単語「外傷」が現れない場合に軽度となる。これらの結果から、重症度は、テキストに現れる「薬」「受診」「剥離」「出血」といった単語から重症度のある程度推定できることが分かる。6.1節からは、リハビリやレクなどの行動の際に打撲して受診するケースが多く、また6.2節からは「薬」が現れると誤薬につながる。つまり打撲や誤薬が重症度の高い事故につながる事が分かる。

また、単語「剥離」や単語「出血」が現れれば、軽度となるが、これは、6.1節から、皮膚剥離から出血が発覚することがあることを述べたように、これらの出血は軽度の事故と見なされることが分かる。

その他の場合は、単語「外傷」が現れる場合は重症度がなし、つまり無視できる程度の事故となる。また6.1節で

は目的がリハビリ、レク、移乗といった行動の場合に受診につながる打撲か、痛みのない外傷かその他に分類されることが分かったが、この場合はそのうち問題が少ない外傷であることが類推される。逆にいえば、6.1節での、打撲につながる行動について対策を講じれば、無視できない打撲への対策は十分である可能性が高い。

今回、多変量データを作成する際、*k*-means法によるクラスタリングを行い、2,701単語を50個のクラスタに分類を行ったが、1つのクラスタに2,400単語近くも分類されているクラスタがあった。これはその他の雑多な単語としてまとめられたものと考えられるが、2,701単語のうち、約2,400単語が1つのクラスタにまとめられると残りの約300単語が49のクラスタに分類されたことになってしまう。この問題の対策としては、事故詳細テキストから単語を抽出する際、複数の文章に含まれている単語のみを抽出することによって、文書間の距離を測りやすくなり解決されると考えられる。今回は2つ以上の詳細テキストに含まれている単語を抽出したので、今後はさらに大きく設定する必要も考えられる。

## 7. まとめ

本研究では、ある介護サービス事業者の複数の介護施設で起きた事故・ヒヤリハットを記載した事故事例テキストを使用し、テキストマイニングを行った。テキストデータと多変量データが混合したデータの分析手法として、テキストデータ部分については単語文書行列を作成して元のデータと結合し、事故の結果を目的変数としてランダムフォレスト法を2段階で適用しながら重要な単語クラスタおよび重要な単語を抽出し、決定木学習によって可視化した。目的変数として、事故の属性である「診断」「事故の種類」「重症度」に対する事故の属性・単語の重要度を調べ、重要度の高い属性・単語を用い決定木を作成した結果、

(1) 出血という事故については表皮剥離または出血、原因不明のあざが発見されて初めて発覚し、その原因となる事象は特定はできていないこと、(2) リハビリやレクリエーション、移乗を始めとする行動をとともなう場合に、転倒・ずれ落ちによって受診をとともなう打撲の可能性があること、(3) 共有スペースで食事の誤嚥、個室での誤嚥の両方が考えられること、(4) 無断外出につながるいくつかの行動があることといったいくつかのパターンが分かった。

今後データを増やし、施設ごとに、事故の起こりやすい時間帯や場所、周りの人の状況といったコンテキストを詳細に分析していくことで、さらにユビキタスコンピューティング分野におけるコンテキストウェアネス研究と融合していくことが望まれる。

謝辞 本研究の一部は、基盤研究(B)「物理層と意味層の2階層からなるセンサコンテキスト推定技術(研究代表者:井上創造)」による。実験データの提供、そして本研

究に際して、多くのご指摘をいただきました介護施設の皆様に感謝いたします。

## 参考文献

- [1] 平澤桂一, 松村成宗, 金丸直義, 阿部匡伸: 病院・介護施設における転倒転落事故予防システム, NTT技術ジャーナル, pp.32-35 (2008), 入手先 (<http://163.137.191.238/journal/0801/les/jn200801032.pdf>).
- [2] Vikramaditya, J. and Cook, D.J.: Anomaly detection using temporal data mining in a smart home environment, *Methods of information in medicine*, Vol.47, No.1, pp.70-75 (2008).
- [3] 厚生労働省: 介護保険事業状況報告 月報 (暫定版), 入手先 (<http://www.mhlw.go.jp/topics/0103/tp0329-1.html>).
- [4] 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, 入手先 ([http://www.mhlw.go.jp/file/04-Houdouhappyou-12004000-Shakaiengokyoku-Shakai-Fukushikibanka/270624houdou.pdf\\_2.pdf](http://www.mhlw.go.jp/file/04-Houdouhappyou-12004000-Shakaiengokyoku-Shakai-Fukushikibanka/270624houdou.pdf_2.pdf)).
- [5] 林 泰史: 高齢者の転倒防止, 日本老年医学会雑誌, Vol.44, No.5, pp.591-594 (2007).
- [6] 岡部貴博, 吉川大弘, 古橋 武, インシデントレポート解析のための多重接続型階層的テキストマイニング手法の提案, 日本知能情報ファジィ学会ファジィシステムシンポジウム講演論文集, Vol.22, pp.54-54 (2006).
- [7] 三田寺裕治, 赤澤宏平: 介護保険施設における介護事故の発生状況, 社会医学研究, Vol.30, No.2, pp.123-130 (2013).
- [8] 三菱総合研究所: 高齢者介護施設における介護事故の実態及び対応策のあり方に関する調査研究事業 (2009).
- [9] 三菱総合研究所: 高齢者介護施設における介護事故の実態及び対応策のあり方に関する調査研究事業 事故から学ぶ (2010).
- [10] 坂本 望, 森山英樹, 今北英高ほか: 介護老人保健施設痴呆呆専棟における転倒の危険因子, 日本職業・災害医学会会誌, Vol.52, No.3, pp.161-165 (2004).
- [11] 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, 共栄学園短期大学研究紀要, Vol.26, pp.29-38 (2010).
- [12] 二見 光, 山岸宏匡, 川口 修ほか: 構造化技術を用いた読影レポートの類似記載を特定する手法の開発, 日本放射線技術学会雑誌, Vol.66, No.9, pp.1229-1236 (2010).
- [13] 五十嵐博, 福土政広ほか: テキストマイニングを用いた診療放射線技師のヒューマンエラー分析, 日本保健科学学会誌, Vol.13, No.2, pp.59-70 (2010).
- [14] 松村真宏: テキストデータのマーケティングへの活用と課題, A Journal of Japan Industrial Management Association, Vol.18, No.1, pp.32-37 (2008).
- [15] 齋藤朗宏: 日本におけるテキストマイニングの応用, The Society for Economic Studies The University of Kitakyushu Working Paper Series, No.2011-12 (2012).
- [16] Robben, S.: Longitudinal Ambient Sensor Monitoring for Functional Health Assessments: A Case Study, *Proc. 2014 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Adjun. Publ.*, pp.1209-1216 (2014).
- [17] Luštrek, M. Gjoreski, H., Kozina, S., Cvetkoviü, B., Mirchevska, V. and Gams, M.: Detecting Falls with Location Sensors and Accelerometers, 23rd IAAI Conference, pp.1662-1667 (2011).
- [18] Shen, H., Balasubramanian, A., Lamarca, A. and Wetherall, D.: Enhancing mobile apps to use sensor hubs without programmer effort, *UbiComp 2015 - Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Com-*

- put., pp.227–238 (2015).
- [19] Forkan, A.R.M. and Khalil, I.: A probabilistic model for early prediction of abnormal clinical events using vital sign correlations in home-based monitoring, *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp.1–9 (2016).
  - [20] Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, *J. Mach. Learn. Res.*, Vol.3, pp.1157–1182 (2003).
  - [21] Kohavi, R. and John, G.H.: Wrappers for feature subset selection, *Artif. Intell.*, Vol.97, No.1–2, pp.273–324 (1997).
  - [22] Peng, H. Long, F. and Ding, C.: Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.27, No.8, pp.1226–1238 (2005).
  - [23] Somol, P. and Novovi, J.: Flexible-Hybrid Sequential Floating Search in Statistical Feature Selection, *Search*, pp.632–639 (2006).
  - [24] Breiman, L.: Random forest, *Mach. Learn.*, Vol.45, No.5, pp.1–35 (1999).



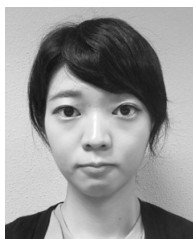
井上 創造 (正会員)

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)。同大学システム情報科学研究院助手, 同大学附属図書館研究開発室助教授(准教授)を経て, 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。現在に至る。この間, 2014年ドイツカールスルーエ工科大学訪問研究員。ユビキタスヘルスケアと行動認識に興味を持つ。IEEE, ACM, 日本データベース学会, 電子情報通信学会, 日本知能情報ファジィ学会, 日本医療情報学会各会員。



峯崎 智裕

1992年生。2015年九州工業大学工学部電気電子工学科卒業。2017年同大学大学院修士課程修了。



松木 萌 (学生会員)

1993年生。2016年九州工業大学総合システム工学科卒業。2017年9月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。センサ行動認識およびテキストマイニングに興味を持つ。