

# 介護日誌自動作成のための 介護分野オントロジー生成技術の研究

野亦優<sup>1</sup> 沼尾雅之<sup>1</sup>

**概要:** 現在日本では少子高齢化が急速に進み、それに伴って介護業界への関心が高まっている。介護業界の問題の一つに間接業務の多さが挙げられる。職員同士の申し送り事項の伝達や、介護サービス利用者家族への連絡手段として介護日誌があるが、この入力自動化できれば、介護施設職員はより良いサービスの提供に集中できるようになると考えられる。本論文では、介護日誌の自動入力を目指して、介護分野に特有のオントロジーの自動構築について検討した。

## Study of the dictionary automatic generation technology Particular to care field of ontology

YU NOMATA<sup>1</sup> MASAYUKI NUMAO<sup>1</sup>

### 1. はじめに

近年日本では高齢化が急速に進み、2060年には高齢化率は40%近い水準になると推計されている[1]。このことから、様々な分野で介護業界への関心は高まっていると考えられる。しかし、介護現場が直面している課題は多岐に渡る[1,2]。

その課題の一つとして、介護日誌の作成等、間接業務の多さが挙げられる[3,4]。文献[3]によれば、通所介護における業務負担割合として、直接介護(レクリエーションや入浴手伝いなど)38.3%に対し、送迎や介護日誌の記録、申し送り等の間接業務が41%を占める。ここでの介護日誌とは、介護施設の職員が、他の職員・利用者家族への申し送り事項や、利用者に対し行ったサービス、利用者の様子を記したものである。これらの情報は職員間、利用者家族との情報伝達手段、あるいは有事の際の法的証拠として扱われるため必要不可欠であるが、実際に介護を行う度に書き込むのは非効率的である。後でまとめて書き込む場合は、職員の記憶に頼ることになり、それは正確性に欠けてしまう上に、伝達事項の記載忘れなど、致命的なミスにつながる可能性もある。介護を行ったその時にボイスメモを残せたり、あるいはセンサー等で介護施設利用者のある動作を感知できた瞬間に自動で日誌に書き込めたりするシステムがあれば、間接業務の負担が改善されるのではないかと期待する。

そこで本研究では、介護日誌の自動生成を目標として、介護分野に特有のオントロジーを自動作成することを目的とした。今回は介護学会で発表された論文をテキストマイニ

ングすることで、介護に関係するもの・ことのオントロジーを自動生成するシステムを提案する。

このシステムの応用例として、以下のユースケースを想定している。

1. 介護士が利用者 A に対し更衣の介助を行った際、「Aさん」「着替え」「シャツ」等の入力キーワードから、「Aさんのシャツの着替えを介助した」といった記録を生成する。
2. (施設にあるあらゆる人、ものがセンサーで感知できる前提で) ADL 認識への応用。例えば、「被介護者(人)」が「洗面所(場所)」にいて、「食事後(時間)」に、「歯磨き粉(もの)」「歯ブラシ(もの)」が観測されたら、その被介護者は「歯磨き」をしている、と認識される。

ユースケース 1 の場合、必要とされるオントロジーは、介助の「着替え」という動作に対して「どの衣服(上衣/下衣/全部...)」の「どこまで介助したか(襟・裾を整える/途中まで引き上げる/全介助...)」などを考えている。

ユースケース 2 の場合、ある日常生活動作 ADL (Activities of Daily Living) に対して関連するものやことを想定している。歯磨きの例として、歯磨粉や歯ブラシ、時間や洗面所を挙げている。

### 2. 関連研究

福原らは、情報推薦を用いた高齢者介護施設向け申し送

<sup>1</sup> 電気通信大学大学院情報理工学研究所

り業務支援システムを開発した[5]。このシステムは、職員一人一人が携帯できるよう、スマートフォンに実装されたアプリケーションで、写真や音声メモを用いて記録が行える。その記録は申し送り・業務遂行情報データベースに保存され、他の職員からも閲覧できる。また、一度入力した申し送り事項については、システムが推薦するため、再び同じ内容を入力する手間がかからない。課題としては、情報推薦を行うためには、事前に手動で業務遂行情報(主介護者や介護サービスに使用した道具等)を登録しておく必要がある点が挙げられる。

金竹らは、介護記録に出現すべき情報となるオントロジーを作成した[6]。これにより、介護におけるある状況において必要な入力情報の種類が特定できるようになる。著者らは介護施設に勤める職員のインタビューからこれらのオントロジーを構築した。ここでは例として、状況・排泄におけるオントロジーを図1に示す。

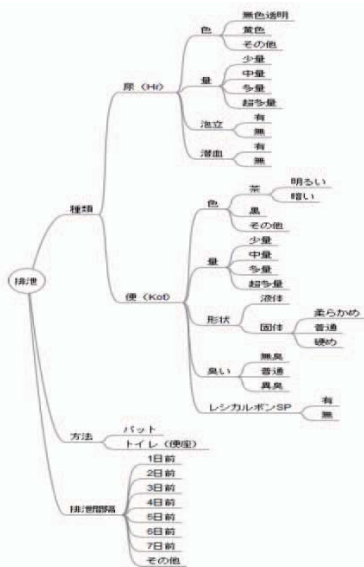


図1、金竹らが構築した排泄に関するオントロジー[6]

### 3. 介護オントロジーの作成

本章では、介護日誌の自動生成に対する課題とその解決手法の提案について述べる。

#### 3.1 介護日誌自動生成システムの要件

はじめにで触れた通り、介護施設職員の間接業務の負担が、本来主軸である介護サービスへ注力する時間を妨げている。特に介護現場での日誌の作成において注目されていることが、ペーパーレス化である[3]。現状では、多くの施設で職員が手書きで日誌を書き、更にその内容を利用者家族への連絡帳等にまた手書きで転記するという無駄が残されている。そこで厚生労働省は介護分野についても ICT 化を推進している[7]。

更に、昨今では介護施設に外国人技能実習生を積極的に

受け入れる潮流になりつつある[8]。そこで問題になるのが、難解な専門用語である。「褥瘡」等の専門用語を多く含む文章や会話のやり取りが申し送りの際に発生するが、実習生の多くがそれらを即時に理解することは困難である。

そこで、介護日誌自動生成システムの要件として、以下3点を考えた。

- I. 可能な限り手動入力の必要がない
- II. 専門用語への対応
- III. 用語の更新に強い

#### 3.2 課題解決のための提案

3.1で想定した課題のうち、Iについては音声やプルダウンメニューを導入した入力画面を実装することで負担の軽減に繋がると考えた。本節では、極力人手での入力を避けるため、システム側で設定すべき情報の選定と、専門用語・新語へのアプローチについて検討する。

介護記録において必要な情報は、「誰が、いつ、どこで、どのような介助を受け、何が観察されたか」といった、基本的に5W1Hで表せるものである。そこで what に当たる、施設利用者の行動を中心にテキストマイニングを行い、関連するもの・ことのオントロジーを組み立てれば、入力画面の構築に有用ではないかと考えた。ここでテキストマイニングを用いる利点は、文献[6]でも触れられているように、日誌に何を記録するかは介護士によるところが大きいいため、頻繁に入力される情報を客観的、統計的に明らかにできるからである。

専門用語への対応としては、次のように対策できる。まず、専門用語とは名詞の連続である(e.g. 「褥瘡」は「褥」と「瘡」)から、特定の名詞の組み合わせが何度も出現するならば、一つの単語として扱うよう辞書を作成する。そしてその辞書を使用してテキストマイニングを行い、類義語抽出を行うことで、「褥瘡」であれば「床ずれ」のように比較的平易な表現に置き換えられると考えた。

この時、マイニング対象となるコーパスに、随時新しく作成された介護日誌を追加すれば、新語にも対応できる可能性がある。

#### 3.3 提案の構成

被介護者の行動とそれに関わる物事の抽出について、構成を図2に示す。

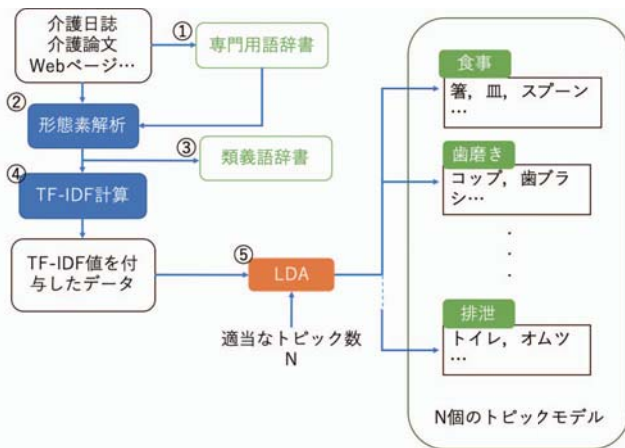


図2, 提案の構成

介護日誌や、介護についての学会論文、クローリングした Web ページ等をコーパスとする。そのコーパスに対し、まず専門用語の抽出を行い、TermExtract [9] を利用して独自の辞書 kaigo.dic を自動作成する。これは、日本語文章は英文とは違い、単語毎の区切りがないため、テキストマイニングを行う際は、事前に文を分かち書きしておく必要がある。しかし、デフォルト辞書では分割されてしまう単語があるので、それらの言葉を一単語として扱えるようにするためである。この辞書を Mecab [10] で指定することで、例えば「認知症」が「認知」「症」と分けられてしまう事象を防ぐことができる。

図3は、今回生成した介護専門用語辞書 kaigo.dic の一部抜粋である。専門用語の横にある 1285 の数字は一般名詞を表す ID, 更にその隣の-3200 などのマイナス付き数字は、Mecab で解析を行う際に最長マッチングをするため、より文字数の多い方を優先するためのコスト値である。今回収録した単語は 23,415 語だった。

高齢者,1285,1285,-2078,名詞,一般,*,*,*,*,高齢者,-,-
対象者,1285,1285,-2078,名詞,一般,*,*,*,*,対象者,-,-
介護職員,1285,1285,-3200,名詞,一般,*,*,*,*,介護職員,-,-
在宅生活,1285,1285,-3200,名詞,一般,*,*,*,*,在宅生活,-,-
参加者,1285,1285,-2078,名詞,一般,*,*,*,*,参加者,-,-
介護度,1285,1285,-2078,名詞,一般,*,*,*,*,介護度,-,-
利用者家族,1285,1285,-4472,名詞,一般,*,*,*,*,利用者家族,-,-
介護老人保健施設,1285,1285,-9050,名詞,一般,*,*,*,*,介護老人保健施設,-,-
認知症高齢者,1285,1285,-5878,名詞,一般,*,*,*,*,認知症高齢

図3, kaigo.dic の内容(一部抜粋)

よって、例えば「認知症高齢者」が文書中にあった場合、3文字の「認知症」、「高齢者」よりコストの小さい6文字の「認知症高齢者」が優先されるので、登録した専門用語

が分かち書きされることはない。

また、本来であれば辞書作成の際、読み仮名も登録すべきだが、今回の目的では読み仮名が必要となる場面がないので省略した。

次に、データについて先に作成した専門用語辞書を用いて形態素解析を行い、分かち書き及び品詞情報を付加する。

その形態素解析結果に対し、word2vecを行い、コーパス中の全単語について類義語の集合、類義語辞書を作成しておく。

類義語辞書とは別に、ここまでの前処理が施されたデータに対し、TF-IDF 値を計算し、単語ごとの重みを求める。TF-IDF とは、TF (Term Frequency)と IDF (Inverse Document Frequency)の二つの概念を組み合わせたものである。TF は各文書においてその単語が出現した回数を意味する。IDF はある単語が多く文書に出現するなら低い値を、逆に少数の文書にしか出現しないなら高い値を示すものである。TF-IDF はこれらを掛けることで、その単語の登場頻度が高いほど、また、その単語の文書毎の偏りがあるほど、大きい値を得ることができる。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

$$idf_i = \log \frac{|D|}{|\{d: d \ni t_i\}|} \quad (2)$$

$n_{i,j}$ : 文書  $d_j$  における単語  $t_i$  の出現回数

$|D|$ : 総文書数

$|\{d: d \ni t_i\}|$ : 単語  $t_i$  を含む文書数

最後にこれらの前処理を施した論文に対し、トピックモデル LDA を適用し、トピックを生成する。

こうして生成したトピックから、介護分野特有のオントロジーを生成し、入力キーワードから介護者・被介護者の行動認識や介護内容の入力事項の類推に利用する。また、類義語辞書は、出力した文に介護士にとっての未知語があった場合、言い換え候補を探す際に用いる。

## 4. 実験

今回はオントロジー構築の前段階として、トピックモデルを利用して日常生活動作(ADL)のモデルを作成する実験を行った。ADL として、「歯磨き」、「入浴」、「食事」、「排泄」に関するトピックモデルを構築し、関連キーワード(表1)の入力に対し、対応するトピックの出力が得られるようになることが目標である。

まず全論文データから専門用語辞書を作成した。次いでその専門用語辞書をデフォルト辞書と組み合わせて形態素解析を行い、論文データを分かち書きにした上に品詞情報を付加した。ここで品詞が名詞、形容詞、副詞であるもの

を抽出し、各単語の TF-IDF 値を計算した。そして TF-IDF 値の計算結果が付与されたデータに対し LDA でトピックモデルを構築した。最後に、それぞれの ADL に関するキーワード群を入力して、各トピックに対する類似度を求めた。

#### 4.1 使用コーパスについて

実験に使用したコーパスは、2015 年度と 2018 年度の全国介護老人保健施設大会で発表された論文である。全国介護老人保健施設大会とは、公益社団法人全国介護老人保健施設協会が主催している大会で、老健施設のサービスの発展を目的として開催されている学会である。2015 年度と 2018 年度のセッション数は図 3 のようになっている(縦軸がカテゴリ名、横軸が論文件数、青棒が 2015 年度、オレンジ色の棒が 2018 年度)。論文数は、口頭、ポスター発表合わせて 2015 年度は 1,306 件、2018 年度は 1,099 件だった。また、カテゴリ「R4」とは全国介護老人保健施設協会が定めた、在宅復帰を利用者の目的としたアセスメントからケアプラン、ケア実績、評価の構築に主眼を置き、チーム介護の効率を上げるためのアセスメント方式のことである。

「ケアマネジメント」は 2015 年度のみ、「人材」カテゴリは 2018 年度のみのカテゴリだった。

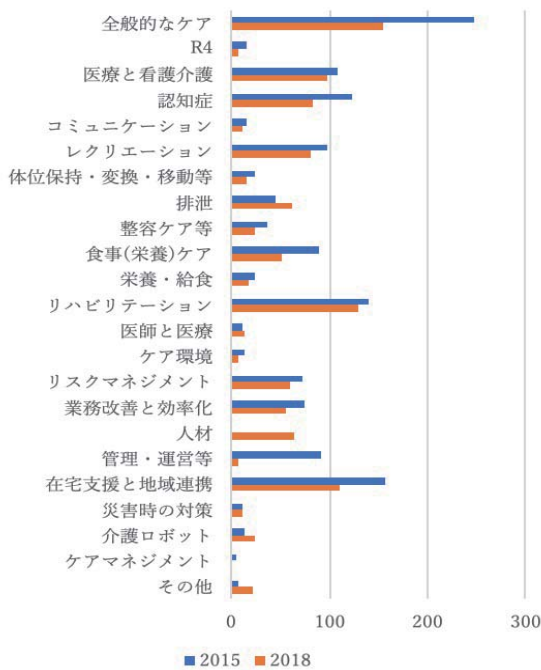


図 4, カテゴリ毎の論文件数

#### 4.2 トピックモデルを用いた ADL 認識

今回は ADL として、「歯磨き」、「入浴」、「食事」、「排泄」に関するトピックモデルを構築し、関連キーワード(表 1)の入力に対し、対応するトピックの出力が得られるようになることが目標である。

表 1, 入力するサンプル文字列

ADL	文字列
歯磨き	歯磨き, 義歯, 歯ブラシ, 歯磨き粉, 洗浄剤, 入れ歯
入浴	タオル, 浴槽, 石鹸, 風呂
食事	食器, テーブル, 箸, スプーン, エプロン, とろみ粉
排泄	トイレ, オムツ, パッド, 下剤

そこで、先に示した介護論文のうち、今回指定した ADL に関連するカテゴリ、「一般的なケア」、「排泄」、「整容ケア」、「食事(栄養)ケア」、「栄養・給食」の 5 つのカテゴリに属する論文についてテキストマイニングの対象とした。これは他のカテゴリは ADL との関連が薄いので、解析対象にするとノイズとなる可能性が高いためである。

続いて、対象となる全論文に、作成した専門用語辞書を用いて形態素解析を行い、名詞、形容詞、副詞を抽出した。

さらにカテゴリを一文書として TF-IDF を用いて、抽出した各単語に重みを付与した。この処理により、「介護」や「職員」といった、頻出するが内容の解析には有益ではない単語が重要語として扱われる事態を防いだ。

最後にこれらの前処理を施した論文に対し、LDA を適用した。その結果分類されたトピックとトピック毎の代表語を表 2 に示す。トピック数 N は N=4, 5, ..., 7 まで測定し、最も分類精度の良かった N=6 とした。

表 2, トピックとトピック毎の代表語

トピック番号	代表語
0	口腔ケア, 口腔, 歯科衛生士, ポリ, 仕上げ磨き, 浴槽, 義歯, 口腔清掃, 整容, 風呂
1	経口摂取, 間食事, 食事栄養, ミールラウンド, 口腔, 銀杏, 木犀, チェリー, パレス, ホテル
2	行事食, 給食, ビュッフェ, ルーム, 献立, カルシウム, スピーチロック, 全般, 栄養, チェリー
3	口腔ケア, 口腔, 経口摂取, ポリ, パッド, カルシウム, 歯科衛生士, 銀杏, 仕上げ磨き, 口腔清掃
4	口腔, 口腔ケア, 歯科衛生士, パッド, 木犀, 行事食, チェリー, 浴槽, 経口摂取, カルシウム
5	パッド, 排尿, 排泄ケア, おむつ, 当て方, 削減, 下剤, 駅, 便秘, カンファレンス



このモデルに対し、サンプルとして ADL に関する表 2 に示す文を入力として与えたところ、次の表 3 の結果が得られた。各項目の数字は、それぞれの文章がどのトピックに分類されるかを表した確率である。

表 3, 各 ADL が分類されたトピック

	0	1	2	3	4	5
歯磨き	<b>0.895</b>	0.021	0.021	0.021	0.021	0.021
入浴	<b>0.790</b>	0.042	0.042	0.042	0.042	0.042
食事	0.021	<b>0.895</b>	0.021	0.021	0.021	0.021
排泄	0.034	0.033	0.033	0.033	0.033	<b>0.833</b>

表 1, 2, 3 から、概ねその ADL に関する単語が代表語として挙げられているトピックに分類されていることが分かる。

しかし、歯磨きと入浴が同じトピックに分類されてしまっている。これは介護論文のカテゴリに従えば、どちらも「整容ケア」に属するものなので、トピック分類の際同じものとされてしまったと考えられる。

また、トピック 5 以外のトピックに「口腔」、「口腔ケア」といった同じ単語が上位の代表語として入り込んでいる。これはコーパスとした介護論文において、食事関連のケアで、食事そのものより、口腔ケアが大きな比重を占めていたからだと考えた。

上記二つの問題については、コーパスの質を確保し、最適なトピック数を設定できればある程度回避できるようになると考える。

更に、これらのトピックは文書中の単語から機械的に分けたものなので、どのトピックがどの ADL を表すのかは、表 2 のようなテスト用の文を入力して人が確認する必要がある。よって介護日誌に必要な事項を人の方から決め打ちでトピックとして設定できる手法を探ることも今後の課題である。

## 5. まとめと今後の課題

本論文では、介護学会で発表された論文をテキストマイニングすることで、いくつかの ADL についてトピックモデルを構築することができた。

しかしながら、各トピックが表す ADL が重複している、どのトピックがどの ADL を表すのか人が代表語から決定しているなど課題も明らかになった。

コーパスについては、介護論文だけではなく、実際に記入された介護日誌や、web からクロールしたテキストなども加えて拡張させていきたい。また、コーパスの拡張

が成れば、今回対象とした歯磨き、入浴、食事、排泄以外の ADL モデルを作成できると期待する。

トピックモデルの手法も、今回は LDA を用いたが、LDA は教師なし学習で分類するため、その分類されたトピックが何を示すかは結果から人が見極めなければならない問題がある。半教師付き LDA など、教師あり学習を取り入れるなどして、モデルの精度を高めていきたい。

## 参考文献

- [1] 厚生労働省 (2016) 「医療と介護を取り巻く現状と課題等」 [online] <https://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000167844.pdf>
- [2] 厚生労働省 (2014) 「福祉・介護人材確保の現状・課題と対応」 [online] <https://www.mhlw.go.jp/topics/2014/01/dl/tp0120-12-03p.pdf>
- [3] 介護経営の教科書、デイサービスの業務効率化の事例 3 選 [online] <http://kaigonews.net/2016/11/08/通所介護（デイサービス）における業務負担を削/>
- [4] 公益財団法人介護労働安定センター大分支部 (2016) 「介護事業所の介護サービスの質の向上となる業務効率化の取り組みとして」 [online] <https://www.pref.oita.jp/uploaded/attachment/2022475.pdf>
- [5] 福原知宏, 中島正人, 三輪洋靖, 濱崎雅弘, 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 28 巻 6 号 B, pp468-479 (2013)
- [6] 金竹香織, 小林一郎, 橋田浩一, 介護支援システムにおける介護データの構造化と入力インタフェースの開発, 情報処理学会第 76 回全国大会
- [7] 厚生労働省 (2015) 「健康・医療・介護分野における ICT の活用について」 [online] [https://www.mhlw.go.jp/file/06-Seisaku-jouhou-12600000-Seisakutoukatsukan/0000038005\\_1\\_12.pdf](https://www.mhlw.go.jp/file/06-Seisaku-jouhou-12600000-Seisakutoukatsukan/0000038005_1_12.pdf)
- [8] 厚生労働省 (2014) 「介護分野における外国人材に関する諸制度や動向について~技能実習制度など~」 [online] [http://www.fukushihoken.metro.tokyo.jp/kiban/fukushijinzei/taisakusuishin-kikou/3\\_un-ei\\_kyogikai.files/14\\_koroshou1.pdf](http://www.fukushihoken.metro.tokyo.jp/kiban/fukushijinzei/taisakusuishin-kikou/3_un-ei_kyogikai.files/14_koroshou1.pdf)
- [9] 前田朗, 「専門用語 (キーワード) 自動抽出 Python モジュール termextract」 [online] <http://gensen.dl.itc.u-tokyo.ac.jp/pytermextract/>
- [10] 工藤拓, 「MeCab: Yet Another Part-of-Speech and Morphological Analyzer」 [online] <http://taku910.github.io/mecab>