

ポスター

ポスター8

医療データ分析4（データ・テキストマイニング）

2018年11月23日(金) 17:00 ～ 18:00 K会場(ポスター、HyperDemo) (2F 多目的ホール)

[2-K-4-2] 介護ライフログ知識抽出のためのテキストマイニング

○串間 宗夫¹, 山崎 友義¹, 酒田 拓也¹, 川野 光一¹, 松尾 亮輔¹, 小川 泰右¹, 荒木 賢二¹, 近藤 千博² (1.宮崎大学医学部
附属病院医療情報部, 2.宮崎市立田野病院)

目的：本研究は、分析ツールとして Total Environment for Text Data Mining (TETDM：テキストデータマイニングのための統合環境) を用いて、介護ライフログに適用することによって、介護記録の分析を行い、その結果について介護行為を可視化する手法、あるいは介護行為を検証する手法を実現する。方法：分析対象として、M市介護老人保健施設Sに記録された、161件の介護ライフログを介護福祉士業務内容とそれに関連する記述を特定するために、様々な意図・文脈を含む記録記述文から分析対象とするデータを取り上げ、解析した。結果：分析として、主な頻出語を抽出し、(1)主役(主語としての頻度が最も高い語彙)、(2)主題(記録のキーワード)、(3)最重要文(一貫性)、(4)主語含有率、(5)主語がある文の割合、(6)不適切表現数、(7)長文の数、(8)総合評価について図表化した。この中で、介護福祉士が業務内容に関連するキーワードとして、トイレ、排尿、コール、午前、車椅子、パット、誘導、介助、様子、入所、移譲、入浴、訪室、夜間、が抽出された。文章要約・主題観点用のキーワードとしては、衣服、着脱、必要、介護、依存、姿勢、意欲が抽出されている。本研究の結果から、介護福祉士の業務の多忙さや、業務に及ぼす大きな要因があると考えられる。また、介護福祉士の業務を基本的に支持しているものと考えられる。このように、介護記録の特徴が定量的に示されることがわかる。まとめ：本研究は、介護ライフログより、テキストマイニングの手法を用いて、介護福祉士が認識する業務内容とそれに関する状態について検討を試みた。これらの分析結果から、介護における中心的な課題が抽出され、諸課題の全体構造等の各介護の役割などが明らかにされた。このことは、多岐にわたっている介護福祉士の業務内容の明確化の手がりになると考えられる。

介護ライフログ知識抽出のためのテキストマイニング

串間 宗夫^{*1}、山崎 友義^{*1}、酒田 拓也^{*1}、川野 光一^{*1}、松尾 亮輔^{*1}、小川 泰右^{*1}、
近藤 千博^{*2}、荒木 賢二^{*1}、

^{*1} 宮崎大学医学部附属病院医療情報部、

^{*2} 宮崎市立田野病院

Text mining for knowledge extraction of nursing care life log

KUSHIMA Munuo ^{*1}, TOMOYOSHI Yamazaki ^{*1}, TAKUYA Sakata ^{*1}, KOICHI Kawano ^{*1},
RYOSUKE Matuo ^{*1}, TAISUKE Ogawa ^{*1}, KAZUHIRO Kondo ^{*2}, KENJI Araki ^{*1}

^{*1} Medical Information, University of Miyazaki Hospital,

^{*2} Miyazaki City Tano Hospital

In this research, we analyze nursing care record by applying TETDM and KH Coder as an analysis tool to nursing care life log. For the analysis results, realize a method to visualize nursing care actions and a method to verify nursing care actions.

For the subjects of analysis, we use 161 nursing care life logs recorded in M City Long-term Care Health Facility S. In this research, we analyze nursing care records including various intentions and contexts in order to identify descriptions related to work contents of nursing care workers.

As a result of analyzing, we extracted the main frequent words, the protagonist (the vocabulary with the highest frequency as the subject), the subject (record keyword), the most important sentence (consistency), Subject content rate, percentage of sentences with subject, inappropriate number of expressions, number of long sentences, overall evaluation.

Based on these analysis results, central issues in nursing care were extracted, and the role of nursing care such as overall structure of various subjects was clarified. Furthermore, from the results of the analysis, it is thought that clarification of the work content of care workers will be handled.

Keywords: Medical information, Text mining, Nursing care life log, Total Environment for Text Data Mining (TETDM), KH Coder.

1. 緒論

近年、膨大なテキストデータの中から得られるデータの取捨選択や、テキストの内容を素早く解析し理解するための環境が医療介護現場でも強く望まれている。

介護記録のテキストデータは、被介護者の施設サービス利用内容記録と介護者の観察記録を統合したテキスト記録であり、他職種への連携・伝達や介護者間での被介護者の状態を把握できる手段として使用されている。また、介護業務の効果的運用と改善、介護業務従事者の教育・研修のため、介護記録の二次的利用法の開発が現場の介護職員から強く望まれているが、この課題を解決する具体的な検討は少ない。

介護施設での介護記録に蓄積されたデータの二次的利用法の開発では、テキストデータの量が膨大であり、データを体系的に整理することが大きな障害であった。この障害を克服し、膨大なテキストデータから、上記課題解決に利用できる知識を獲得する手法として、テキストマイニング技術が注目されている[1,2,3,4]。

本研究では、ライフログの定義が漠然としているが、本研究が対象としているのは、業務記録であって、一般的なライフログとは異なる範囲のものと捉えられる場合もある。一般的に、ライフログとは、人間の生活・行い・体験を、映像・音声・位置情報などのデジタルデータとして記録する技術、あるいは記録自体のことである。本研究では、介護現場で記録されたテキストデータを利用している。

2. 目的

本研究は、分析ツールとして Total Environment for Text Data Mining (TETDM : テキストデータマイニングのための統合環境) [5]と KH Coder を用いて、介護ライフログに適用することによって、介護記録の分析を行い、

その結果について介護行為を可視化する手法、あるいは介護行為を検証する手法を実現する。

さらに、その手法を発展させることで、介護現場にフィードバックし介護記録を介護士間で共有し、介護行為の改善等に活用可能にすることを目指している。

3. 先行研究

介護記録については、様々な先行研究が行なわれている。

柴田ら[6]の研究では、福祉生活に関する記録について論じられており、利用者側の負担や介護記録から情報を収集・分析して、利用者が生きがいを感じたり、思い出として残したりするような日記として出力するシステムを提案している。

福田ら[7]の研究では、介護等業務で発生した業務文脈や従業員スキル・考え方、顧客状況など主観を含むさまざまな業務情報を記録するデータ基盤であるデータ利用取り組みについて報告されている。

倉ら[8]の研究では、調査票の自由記述文(テキストデータ)より、テキストマイニングの手法を用いて、介護支援専門員が認識する時間を要する業務内容とそれに関連する心理状態についての検討が試みられている。本研究では、実際の介護現場での介護記録を直に分析し介護現場にフィードバックし介護行為に活用する点が特徴としてある。

4. 介護記録

介護記録は、被介護者の24時間、365日の様子を介護者が記載し介護者のサービス内容記録としても利用されている。記録だけが目的でなく、書かれた内容を他者に伝え、あるいはデータを蓄積・分析し、それを活用して

より良いケアにつなげることを目的としている。

以下に、記載例を示す。

記載例：「(夜間状況)入浴日のため、午前入浴を行う。入浴後デイルームにて水分摂取時、A子さんは、お風呂は気持ちよかったですか?と尋ねると、ああ～、気持ち良かったよ。やっぱり風呂はいいな、と話される。」

実験に協力していただいた介護施設では、介護記録記載の質向上に向けた取り組みとしては、介護記録主任が記載者本人に記載内容についての感想をフィードバックしているが、施設全体としての取り組みでなく、また、統一した記載手順もなく、改善手法を考える手段もないという現実がある。

分析対象として、M市介護老人保健施設Sに記録された、161件の介護ライフログを介護福祉士業務内容とそれに関連する記述を特定するために、様々な意図・文脈を含む記録記述文から分析対象とするデータを取り上げ、解析した。

各記録者は、異なっており複数の介護者の記録が混在しており、任意に選択した3例の被介護者についての介護記録を用いている。各事例1日分の記録者は単一である。記録回数については、1日分を1行として1被介護者に対して1ファイル作成している。

5. 解析方法

人手では扱えないような膨大な量のデータに対して、数値を分析対象とする量的分析とテキストを分析対象とする質的分析の両方を行うことが考えられる。本研究では、テキストマイニングを用いてテキストから知識抽出を行う。

5.1 テキストマイニング

テキストマイニングとは、形式化されていないテキストデータ(通常の自然文)を単語などに分割(形態素解析)し、その出現頻度や相関関係などをデータマイニングの技術を使って解析することで、一定の知見や発想を得るテキストデータ分析技術の総称をいう。

また、構造化・非構造化を問わず多様なデータに適用可能であり、他の方法では知りえない傾向や誰も気付かなかった知識を抽出することができる。

図1にテキストマイニング分析の流れ図を示す。

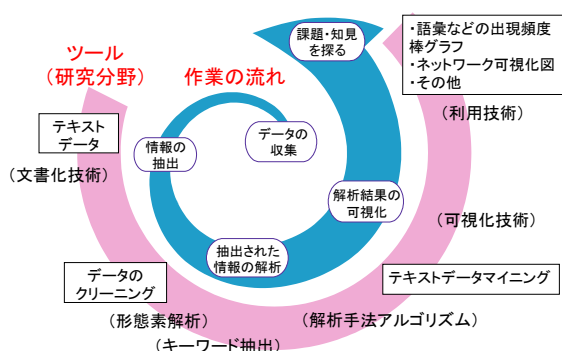


図1 テキストマイニング分析の流れ図

5.2 TETDMとKH Coder

TETDMは、複数のテキストマイニング技術を柔軟に組み合

わせて使える統合環境の構築を目指し、TETDM サイト上で公開されており、誰でも自由に使えるようになっている[9]。TETDMはライトウェイトなツールを提供するプラットフォームに位置づけられる。

データ分析作業を本格的かつ定期的に行うことが決定しているような場合には、商用の高機能・高価なツールの導入も検討に値するが、本研究のように有効な分析方法も定まっておらず、試験的に分析を行う段階には、TETDMの適用が適していると考えられる。独立した複数のパネル内に、処理ツールと可視化ツールを1つずつペアとしてセットすることで動作する[10,11]。

一方、KH Coderは、TETDM同様、テキストマイニングのためのフリーソフトウェアである。各種の検索を行えるほか、どんな言葉が多く出現していたのかを頻度表から見る事ができる。さらに多変量解析によって、一緒に出現することが多い言葉のグループや、同じ言葉を含む文書のグループを見ることで、データ中に含まれるコンセプトを探索できる。

また一部の文書群に注目した場合に、その文書群に特に多く出現する言葉をリストアップすることで、その文書群の特徴を探索できる。あるいは分析者が指定した基準によって、文書の分類を自動的に行うこともできる等の特徴がある。

6. 結果

介護記録テキストデータを入力として、テキスト分析とこれに関連する解析結果を以下に示す。複数の結果図パネルを用いて、介護福祉士の興味に応じた出力結果をインタラクティブに取得できる環境となっている。

可視化図を確認して、介護福祉士が望む大切な事柄であるかどうかの判断に利用する。

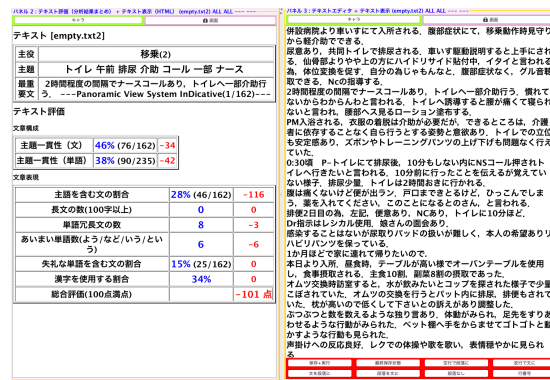


図2 テキスト評価(分析結果まとめ)

図2に、介護記録の文字列入力テキストデータの一部を示す。表示されているテキストを入力テキストとしてファイルに保存し、全ツールが処理を再実行している。

主役、主題、最重要文、一貫性、主語含有率、不適切表現数、長文数、失礼文数、漢字の割合を表示している。分析として、主な頻出語を抽出し、主役(主語としての頻度が最も高い語彙)、主題(記録のキーワード)、最重要文(一貫性)、主語含有率、主語がある文の割合、不適切表現数、長文の数、総合評価について総合評価をしている。記録の主役として、移乗、主題として、トイレ、午前、排尿、介助、コール、一部、ナース、最重要文として、2時間程度の間隔でナースコールあり、トイレへ一部介助行、として抽出されている。



図3 介護記録語彙情報まとめ

図3では、トイレが最も多く出現し、介護施設で業務内容に関連する頻出キーワードとして、トイレ、排尿、コール、午前、車椅子、パット、誘導、介助、様子、入所、移譲、入浴、訪室、夜間、が抽出された。単語とその品詞番号、出現頻度、出現文数、出現セグメント数を出力し、文章で使われている単語の情報を確認することができる。

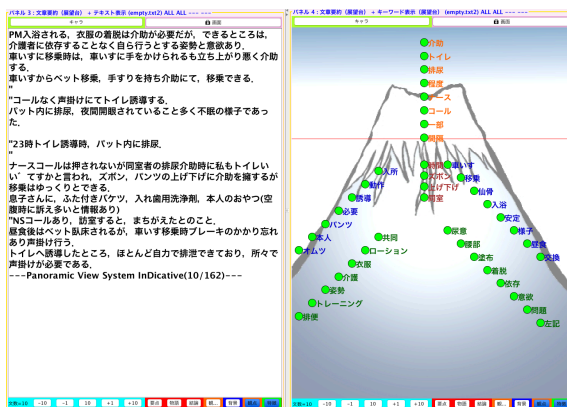


図4 文章要約

図 4 に、指定した数の重要文を抽出し、キーワード表示を示す。観点語：要約する際の観点を表す単語、背景語：文章の背景となる高頻度の単語、特徴語：文章の特徴を表す、観点語との結びつきが強い単語を示す。山の上部にある線より上にある、介助、トイレ、排尿、程度、ナース、コール、一部、間隔をキーワードとして配置され、指定した数の重要文を抽出している。

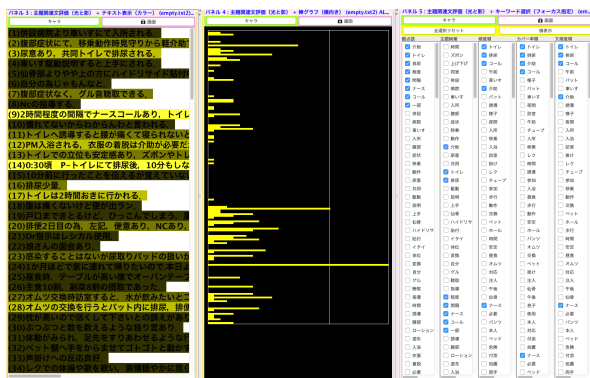


図 5 主題関連文評価

図 5 に、主題関連文評価を示す。各文の主題との関連度を計算し、テキスト表示、棒グラフ、キーワード選択し、文章中の各文の主題関連度を評価して文をハイライトしている。

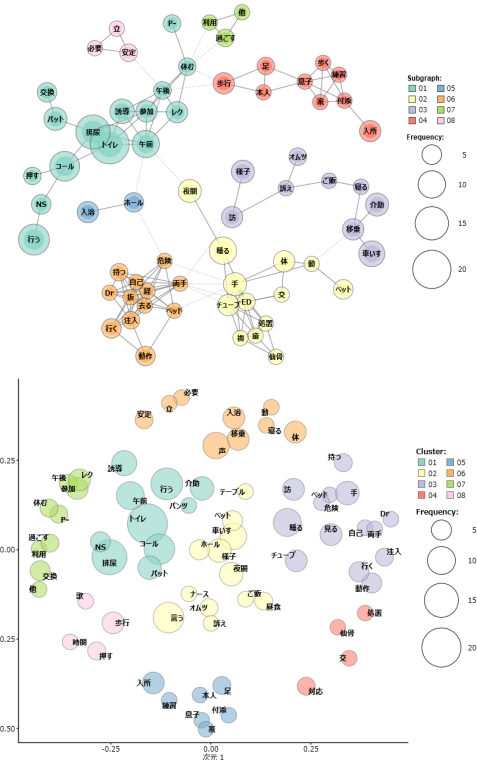


図 6 共起ネットワークと多次元尺度法

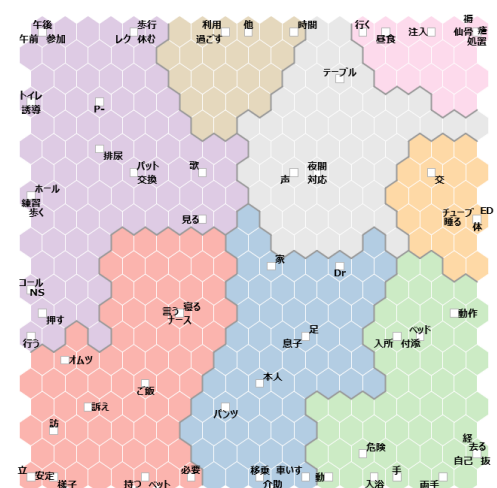


図7 自己組織化マップ

図 6 に、語と語の繋がりをネットワーク図化したものを示す。トイレを中心に考えて、トイレから、排尿、誘導等が共起しており、介護のつながりが中心に見える。更に、強い共起の中心になっているのでキーワードである。関連する項目と組み合わせ配置した図を示す。項目ごとに、語の出現傾向を確認することが可能である。語の布置を見ると、車いす、様子が、ほぼ中央にあり、語の関係性においてどの言葉とも近いと言えるが、感情のみの語で強く結びつく関係にはなっていない。様子とベットという語の近さを確認できる。また、レクという言葉

の近くに午後、参加などの施設内での行事を表す言葉が配置されていることが分かる。

共起ネットワークでは、知識抽出グループとして、5 つに分類された。抽出語のまとまりから、グループ 1 を「トイレ」、グループ 2 を「家族」、グループ 3 を「介助」、グループ 4 を「手技」、グループ 5 を「動作」と解釈した。

図 7 において、自己組織化マップでのクラスターでトイレに注目すると、トイレ、誘導、午後、午前、参加、歩行、レク、休む、パット、交換、ホール、練習、歩く、コール、NS、押す、行う等が形成されていた。

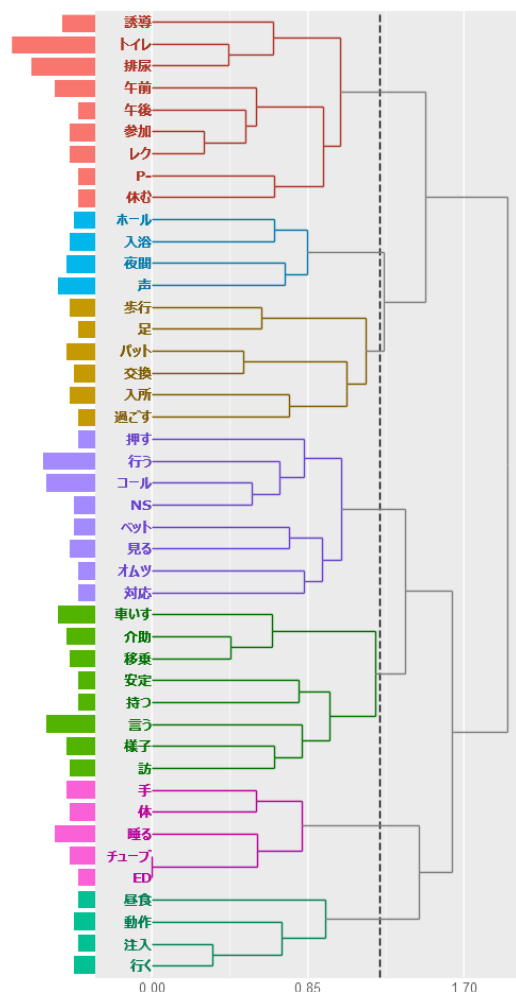


図 8 階層型クラスター分析

図 8 に、抽出した語から、出現パターンが似ている語の組み合わせを階層的に見ることから、クラスター分析を実行するための類似する語を階層構造で分類した図を示す。最小出現数を 4 以上に限定して、階層型クラスター分析を行なった。誘導、トイレ、排尿が同一クラスターに含まれており、その他、車いす、介助、移乗についても、老人ホーム施設に関することも納得できる。このことから、要介護度の数値が高いことが示されている。

クラスター分析では、知識抽出が 5 つのクラスターに分類された。知識抽出のまとまりから、クラスター 1 を「トイレ」、クラスター 2 を「状態」、クラスター 3 を「ベット」、クラスター 4 を「介助」、クラスター 5 を「動作」と解釈した。

図 9 において、出現語の関連を 2 次元のマップで示す。行く、見るという言葉が、全体の中心近くに位置している。利用者さんは、K さんが中心に位置している。K さんは、ベットでのオムツ換えの訴えがあり、また、U さんはチューブ等の処置がされていることがわかる。

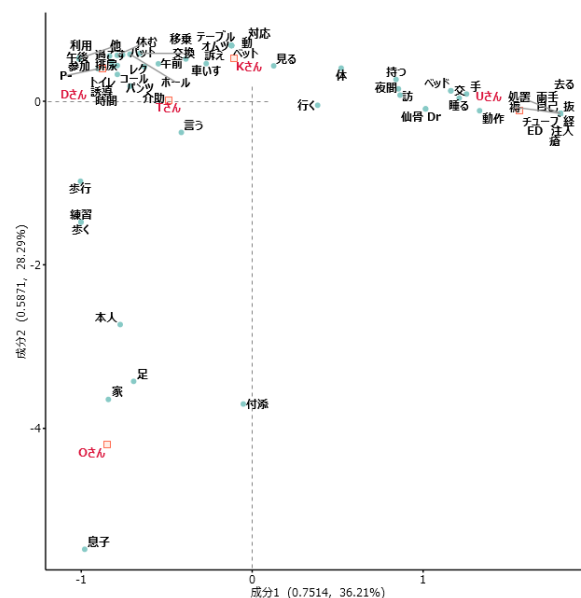


図 9 対応分析

7. 考察

本研究の結果として、抽出された頻出語が本研究のテーマになっている。

介護における基本的な語彙を中心に介護記録がされており、単なる記録やメモとしての性質をもつ反面、一般的な自然言語として記述されていることが確認できることから、他の介護士に説明できる記録文章であると判断できる。

介護の様子を可視化によって解釈することが可能であり、今回抽出された語彙は、介護記録用辞書を作成する上で妥当なものである。

抽出された語彙の関係を可視化することで、介護要点を特徴づけながら、介護記録方法を標準化できる可能性を示している。更に、本研究より介護記録の電子的介護記録システムを構築できる方向性が示唆できた。

8. 結論

本研究は、介護ライフログより、テキストマイニングの手法を用いて、介護福祉士が認識する知識抽出について検討を試みた。結果として、多岐にわたっている介護福祉士の知識内容の明確化の手がかりになると考えられる。今後は、本研究で明らかにした解析結果による介護記録分析に関する研究をさらに積み重ね、記録に関するデータベースを構築したい。

謝辞

本研究は JSPS 科研費 JP18K11530 の助成を受けたものである。本研究は、宮崎大学医の倫理審査委員会の承認を得ている。関係者各位の協力に感謝する。

参考文献

- 1) Muneo Kushima, Kenji Araki, Tomoyoshi Yamazaki, Sanae Araki, Taisuke Ogawa, Nobol Sonehara, Text Data Mining of Care Life Log by the Level of Care Required Using KeyGraph, The International MultiConference of Engineers and Computer Scientists (IMECS 2017), Hong Kong, March, 2017.
- 2) 串間宗夫、荒木賢二、山崎友義、荒木早苗、小川泰右、曾根原登、Gephi を用いた介護ライフログ病歴解析、2017 年電子情報通信学会総合大会、2017 年 3 月。
- 3) Muneo Kushima, Kenji Araki, Nobol Sonehara, Improvement of QoL by integration and analysis of nursing care life log, and academic medical and health information infrastructure, 2016 JSAI International Symposia on AI, Nov. 2016.
- 4) 串間宗夫、荒木賢二、山崎友義、荒木早苗、曾根原登、介護現場における暗黙知を基にした知識の共有と活用、第 20 回日本医療情報学会春季学術大会、2016 年 6 月。
- 5) テキストデータマイニング、統合環境 TETDM (Total Environment for Text Data Mining) 、
<http://tetdm.jp/pukiwiki/index.php>.
- 6) 柴田邦臣、服部哲、松本早野香、介護に関する記録を利用者の生きがいに結びつける「福祉生活の思い出日記」サイトの構築、情報処理学会研究報告、Vol.2011-GN-79 No.3、2011。
- 7) 福田賢一郎、中島正人、福原知宏、渡辺健太郎、西村拓一、本村陽一、介護現場における情報収集と利用、2015 年度人工知能学会全国大会 (第 29 回) JSAI2015、1K5-NFC-05b-2、2015。
- 8) 兪秀娟、神部智司、笠原幸子、清水由香、岡田進一、白澤政和、介護支援専門員が認識する時間を要する業務内容とそれに関連する心理状態に関する探索的研究—テキストマイニングに基づいた自由記述文からのキーワード抽出—、生活科学研究誌・Vol.12 (2013)、2013。
- 9) 砂山渡、高間康史、西原陽子、徳永秀和、串間宗夫、阿部秀尚、梶並知記、テキストデータマイニングのための統合環境 TETDM の開発、人工知能学会論文誌、Vol.28、No.1、2012。
- 10) 砂山渡、高間康史、徳永秀和、串間宗夫、西村和則、松下光範、北村侑也、統合環境 TETDM を用いた社会実践、人工知能学会論文誌、Vol.32、No.1、pp. NFC-A 1-12、2017。
- 11) 砂山渡、高間康史、西原陽子、徳永秀和、串間宗夫、阿部秀尚、梶並知記、松下光範、Bollegala Danushka, Total Environment for Text Data Mining、人工知能学会論文誌、Vol.26、No.4、pp.483 - 493、2011。

