

令和 6 年 6 月 7 日現在

機関番号：84420

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K07206

研究課題名（和文）患者レジストリを二次利用した潜在的患者ニーズの自動抽出方法の検討

研究課題名（英文）Investigation of automated method of identifying latent needs by secondary use of a patient registry

研究代表者

種村 菜奈枝（Tanemura, Nanae）

国立研究開発法人医薬基盤・健康・栄養研究所・国立健康・栄養研究所 食品保健機能研究部・室長

研究者番号：50790630

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究では、2件、AI技術開発を行い、論文にて公表した。  
(1) 類似した単語（日本語）を自動集約するための同義語辞書の構築と適応：発話者の潜在ニーズを予測するにあたり、ニューラルネットワークを用いて単語をベクトル変換する手法であるWord2Vecモデルを用いた同義語辞書の適応が機械学習モデルの精度改善に寄与した。(2) Zスコアを用いた特徴量選択によるモデル精度向上に向けた有用性評価：発話者の潜在ニーズを予測するための機械学習モデルの構築において、新手法(Zスコアを用いた特徴量の選択)の有用性を評価した結果、高水準のモデル精度を維持したことから、適切に情報除去を行い予測処理量削減に貢献した。

研究成果の学術的意義や社会的意義

- 学術的意義：発話者の潜在ニーズを口語テキストから予測するにあたり、Word2Vecモデルを用いた同義語辞書の適応、又はZスコアを用いた特徴量選択技術が、機械学習モデルの精度に寄与した。
- 社会的意義：本研究で開発したAI技術により、一般市民の潜在的なニーズを自動抽出の上、政策等の意思決定の場に「ボイス」として反映させるための社会システムへの活用等が今後、期待される。

研究成果の概要（英文）：In this research, two AI-related technological developments were conducted.  
(1) Prediction and Visualization of Latent Needs: Improving the accuracy of machine learning models using the Word2Vec model: This study examined how to improve the accuracy of the model using the Word2Vec model, that used a neural network to transform words into vectors. The adaptation of the synonym dictionary using the Word2Vec model can improve the accuracy of the model.(2) Extracting the latent needs of dementia patients and caregivers from transcribed interviews in Japanese: an initial assessment of the availability of morpheme selection as input data with Z-scores in machine learning: A new scheme based on Z-score adaptation for machine learning models was developed to predict the latent needs of dementia patients and their caregivers by extracting data from interviews in Japanese.

研究分野：ヘルスコミュニケーション

キーワード：ヘルスコミュニケーション 患者市民参画 潜在ニーズ 口語テキスト 機械学習モデル Word2Vecモデル Zスコア 同義語辞書

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

### 1. 研究開始当初の背景

近年、一般者も政策決定の場のみならず、創薬開発といった様々な意思決定の場での市民参画が進んでおり、一般者の「個人」のニーズを社会システムへ迅速に反映することが重要である。

しかし、当事者の語りの文脈から行間を読むといったコミュニケーションを特徴とする我が国における「ハイコンテキスト文化」を考慮した場合、顕在ニーズのみならず、潜在ニーズも踏まえた当事者の「ボイス」を社会形成へ反映させるための技術的な検討が必要であると考えた。

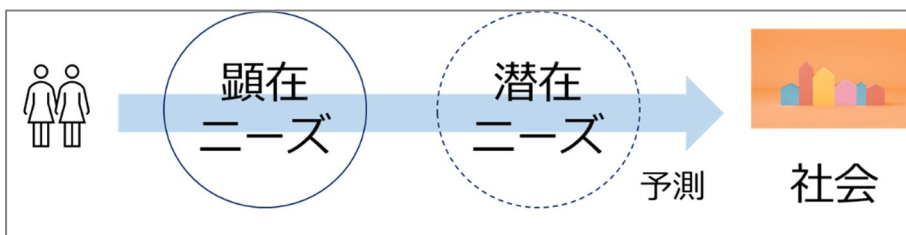


図. 当事者の語りの文脈からの潜在ニーズの社会形成プロセスへの反映

### 2. 研究の目的

本研究では、患者や介護者等へのインタビュー等で取得した口語テキストから一般者の「個人」の潜在ニーズを予測するため、インタビューにおける語りの文脈から潜在ニーズの自動抽出を効率的に行うための人工知能の開発を行うにあたって、次の2つの技術的なテーマを検討した。

検討した技術的なテーマ

類似した単語（日本語）を自動集約するための同義語辞書の構築と適応：掲載論文 (1)

Zスコアを用いた特徴量選択によるモデル精度向上に向けた有用性評価：掲載論文 (2)

### 3. 研究の方法

類似した単語（日本語）を自動集約するための同義語辞書の構築と適応

この技術検討では、口語テキストから発話者の潜在的ニーズを予測するための機械学習モデル構築、及びニューラルネットワークを用いて単語をベクトル変換する手法である「Word2Vecモデル」を用いて、機械学習モデルの精度改善を検討した。

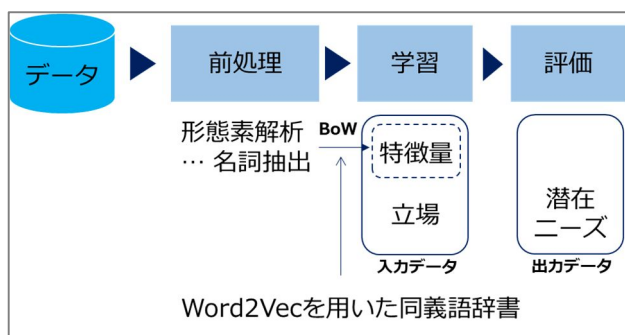


図. Word2Vec モデルを用いた機械学習モデル構築

予備検討では、機械学習モデルの精度比較を行い、最適なモデルを選択した。本調査では、Word2Vec モデルを用いて同義語辞書を作成し、この辞書を使用して同一の特徴量に変換し学習を行う新手法を検討し、新手法の適応の有無でモデルの精度比較を行った。

類似した単語を自動集約するための「同義語辞書」の作成方法は、前後に出現する単語の分布傾向が近い単語同士は同義語の関係にあると仮定した上で、口語テキストに対して Word2Vec を用いて、文脈情報から単語同士の意味的な距離関係を学習させ、単語をベクトル化し、ワード法で近い意味の単語をクラスタリングすることで「同義語辞書」を作成した。

### Z スコアを用いた特徴量選択によるモデル精度向上に向けた有用性評価

発話者の口語テキストには、潜在ニーズの有無に関わらず、共通して出現しやすい話し言葉由来の単語が多数含まれる傾向に着目した。そこで、本研究は、発話者の潜在的ニーズ自動抽出のための機械学習モデル構築において、Z スコアによる潜在ニーズを含む形態素を考慮した特徴量選定の新手法を考案した。

その後、新手法の有無別に潜在ニーズを自動抽出する機械学習モデルの精度比較を行うことで、新手法の有用性を評価した。

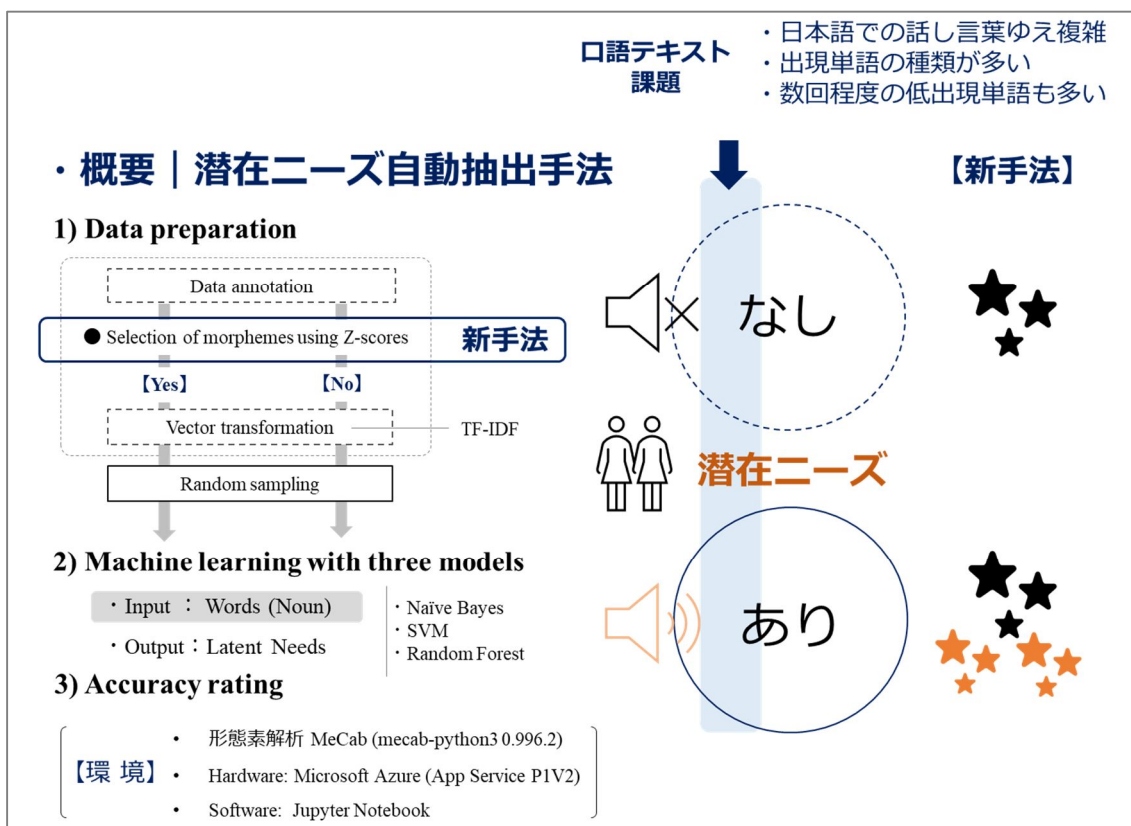


図. 発話者の潜在的ニーズ自動抽出のための機械学習モデル構築

## 4. 研究成果

### (1) 結果の概要

類似した単語（日本語）を自動集約するための同義語辞書の構築と適応

予備検討でのモデル選定実験の結果、モデル精度は、比較検討した 3 つの木構造モデルの中

でも xgboost で F 値 0.54 と最も高かった。本調査では、モデル精度は新手法（同義語辞書）適応ありで F 値 0.61、新手法（同義語辞書）適応なしで F 値 0.54 であり、Word2Vec モデルを用いた同義語辞書の適応が機械学習モデルの精度改善に寄与した。

機械学習モデルのひとつである「xgboost」とは、勾配ブースティングと呼ばれるアンサンブル学習と決定木を組み合わせた手法で非常に高い汎化能力があり、一般的に決定木よりも予測精度が高いと言われている。そのため、本研究においても xgboost で最も予測精度が高かったと思われる。Word2Vec を用いた同義語辞書の適応が精度改善に寄与しており、本アプローチの有用性を確認することができた。

#### Z スコアを用いた特徴量選択によるモデル精度向上に向けた有用性評価

新手法適応別の機械学習モデル精度は、新手法の適応ありの場合/サポートベクターマシン (SVM): F 値 0.81、ランダムフォレスト (RF): F 値 0.77、ナイーブベイズ (NB): F 値 0.75、新手法の適応なしの場合/SVM: F 値 0.79、RF: F 値 0.81、NB: F 値 0.73 であった。

新手法を用いた SVM と NB にて F 値がやや向上し、SVM で最も高精度であった。ナイーブベイズは、データの特徴が独立で相互に無相関であるという前提で機能する。そのため、この仮定が非現実的であることを考慮すると、SVMの方が予測に適していると考えられた。

新手法の適応によって、モデル精度の向上は認められなかったが、新手法の適応がない場合と同等のモデル精度であった。これらのことから、新手法は、適切に情報の除去をすることで、データ処理量の削減に貢献しつつもモデル精度に影響を与えないことが考えられた。

今後、機械学習実行前の処理にて、Z スコアを用いた特徴量選択技術の利活用が期待される。

#### (2) 今後の展望

我々の取り組みによって、語りの文脈から潜在ニーズの自動抽出が可能となった。一般的に、少数の声「ボイス」は、社会システムへ反映されづらいこともあるが、我々を取り巻く生活環境を踏まえ、社会システムをより良くしていくためには、当事者の「ボイス」が重要である。

例えば、創薬のみならず医療や看護、介護の現場の他、様々な情報ニーズに対する対象者の潜在ニーズの自動抽出ならびに可視化技術の利活用が考えられる。早期に潜在的なニーズをリアルタイムに検知することで、意味解釈が難しい発話者の語りの文脈から潜在ニーズを察知できる可能性があり、サービスやモノと最終利用者とのニーズのギャップを最小化させることが可能となり、より良いサービスやモノの社会受容に寄与することが、今後、期待されると考える。

#### <文献> 掲載論文 2件

- (1) 種村菜奈枝, 町井 湧介, 佐々木剛, 他.: 情報処理学会 論文誌デジタルプラクティス 4(3): 69-73, 2023.
- (2) Tanemura, N., Sasaki, T., Miyamoto R. et al.: BMC Medical Informatics and Decision Making 23(203), 2023.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 2件）

|   |                       |
|---|-----------------------|
| 1. 著者名<br>種村菜奈枝, 佐々木剛, 佐藤淳子, 神林崇          | 4. 巻<br>50(6)         |
| 2. 論文標題<br>ナルコレプシー患者が抱えている顕在または潜在ニーズの実態調査 | 5. 発行年<br>2021年       |
| 3. 雑誌名<br>臨床精神医学                          | 6. 最初と最後の頁<br>615-622 |
| 掲載論文のDOI (デジタルオブジェクト識別子)<br>なし            | 査読の有無<br>有            |
| オープンアクセス<br>オープンアクセスではない、又はオープンアクセスが困難    | 国際共著<br>-             |

|   |                 |
|---|-----------------|
| 1. 著者名<br>Tanemura Nanae, Sasaki Tsuyoshi, Miyamoto Ryotaro, Watanabe Jin, Araki Michihiro, Sato Junko, Chiba Tsuyoshi  | 4. 巻<br>23      |
| 2. 論文標題<br>Extracting the latent needs of dementia patients and caregivers from transcribed interviews in Japanese: an initial assessment of the availability of morpheme selection as input data with Z-scores in machine learning | 5. 発行年<br>2023年 |
| 3. 雑誌名<br>BMC Medical Informatics and Decision Making   | 6. 最初と最後の頁<br>- |
| 掲載論文のDOI (デジタルオブジェクト識別子)<br>10.1186/s12911-023-02303-3  | 査読の有無<br>有      |
| オープンアクセス<br>オープンアクセスとしている (また、その予定である)  | 国際共著<br>-       |

|   |                     |
|---|---------------------|
| 1. 著者名<br>Nanae Tanemura, Yusuke Machii, Tsuyoshi Sasaki, Junko Sato, Michihiro Araki, Tsuyoshi Chiba | 4. 巻<br>4(3)        |
| 2. 論文標題<br>発話者の潜在ニーズ予測とその可視化 Word2Vecモデルを用いた機械学習モデルの精度改善に関する検討  | 5. 発行年<br>2023年     |
| 3. 雑誌名<br>情報処理学会論文誌 デジタルプラクティス  | 6. 最初と最後の頁<br>69-73 |
| 掲載論文のDOI (デジタルオブジェクト識別子)<br>なし  | 査読の有無<br>有          |
| オープンアクセス<br>オープンアクセスとしている (また、その予定である)  | 国際共著<br>-           |

〔学会発表〕 計4件（うち招待講演 2件/うち国際学会 0件）

|  |
|--|
| 1. 発表者名<br>種村菜奈枝, 町井湧介, 佐々木剛, 荒木通啓, 佐藤淳子, 千葉剛                              |
| 2. 発表標題<br>口語テキストでの発話者の潜在ニーズ予測とその可視化 - Word2Vecモデルを用いた機械学習モデルの精度改善に関する検討 - |
| 3. 学会等名<br>ヘルスコミュニケーションウィーク2022  |
| 4. 発表年<br>2022年  |

|  |
|--|
| 1. 発表者名<br>種村菜奈枝   |
| 2. 発表標題<br>S3 ナルコレプシー患者の声を反映した 医療環境の実現に向けて：ナルコレプシー患者やその家族が抱えている潜在的ニーズの実態調査 |
| 3. 学会等名<br>第30回日本臨床精神神経薬理学会（JSCNP2020）（招待講演）                               |
| 4. 発表年<br>2021年  |

|  |
|--|
| 1. 発表者名<br>佐々木剛  |
| 2. 発表標題<br>S3 ナルコレプシー患者の声を反映した 医療環境の実現に向けて：臨床医として開発医師として |
| 3. 学会等名<br>第30回日本臨床精神神経薬理学会（JSCNP2020）（招待講演）             |
| 4. 発表年<br>2021年  |

|  |
|--|
| 1. 発表者名<br>種村 菜奈枝, 佐々木 剛, 宮本 遼太郎, 渡邊 仁, 佐藤 淳子, 荒木 通啓, 千葉 剛                     |
| 2. 発表標題<br>口語テキストに基づく認知症患者や介護者の潜在ニーズ自動抽出の検討 -Zスコアを用いた特徴量選択によるモデル精度向上に向けた有用性評価- |
| 3. 学会等名<br>第33回 日本臨床精神神経薬理学会学術集会   |
| 4. 発表年<br>2023年  |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

|               | 氏名<br>(ローマ字氏名)<br>(研究者番号) | 所属研究機関・部局・職<br>(機関番号) | 備考 |
|---------------|---------------------------|-----------------------|----|
| 研究<br>分担<br>者 | 佐藤 淳子                     | 慶應義塾大学・薬学部（芝共立）・客員教授  |    |
|               | (Sato Junko)              |                       |    |
|               | (10231341)                | (32612)               |    |

6. 研究組織（つづき）

|       | 氏名<br>(ローマ字氏名)<br>(研究者番号)                           | 所属研究機関・部局・職<br>(機関番号)                     | 備考 |
|-------|---|---|----|
| 研究分担者 | 漆原 尚巳<br><br>(Hisashi Urushihara)<br><br>(10511917) | 慶應義塾大学・薬学部（芝共立）・教授<br><br><br><br>(32612) |    |
| 研究分担者 | 佐々木 剛<br><br>(Sasaki Tsuyoshi)<br><br>(90507378)    | 千葉大学・医学部附属病院・准教授<br><br><br><br>(12501)   |    |

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
|         |         |