

## 修士論文の和文要旨

研究科・専攻	大学院 情報理工学研究科 情報・ネットワーク工学専攻 博士前期課程		
氏名	吉澤 斗武	学籍番号	2031159
論文題目	センサーデータに介護記録文書を加味した深層学習による 認知症周辺症状予測		
要旨	<p>近年、日本の高齢化社会が進み、認知症を患う被介護者が急増している。被介護者は認知症の行動・心理病状（Behavioral and psychological symptoms of dementia: BPSD）を生じることがあり、介護者の負担と減らし、被介護者に適切な処置を行うためには、IoT センサーデバイスによる BPSD の早期の自動予測が求められている。近年では、数分前の短期的なセンサーデータをもとに機械学習手法を用いて BPSD を予測する研究が行われている。しかし、被介護者ごとの体調や様子、過去の行動履歴など、センサーデータでは捉えられてない介護者の観点のコンテキスト情報を考慮できないという問題点があった。</p> <p>本研究では、IoT デバイスで取得されるセンサーデータによる短期的な情報と、介護記録文書によるコンテキスト情報の2つの情報を考慮しながら、BPSD を自動予測するための Dual-Encoder モデルを提案する。具体的には、深層学習手法を用いてセンサーデータと介護記録文書の情報をそれぞれ異なる Encoder に入力する。提案手法は介護記録文書を考慮することで被介護者の潜在的な BPSD の起こりやすさを考慮し、precision を高くすることができる。評価実験では提案手法の BPSD の予測精度が既存手法を上回ることを示した。</p>		

修士論文

センサーデータに介護記録文書を加味した  
深層学習による認知症周辺症状予測

電気通信大学大学院 情報理工学研究科

情報・ネットワーク工学専攻

学籍番号 2031159

吉澤 亜斗武

主任指導教員 植野 真臣 教授

指導教員 川野 秀一 准教授

2022年2月4日



# センサーデータに介護記録文書を加味した 深層学習による認知症周辺症状予測

吉澤 亜斗武

## 概要

近年、日本の高齢化社会が進み、認知症を患う被介護者が急増している。被介護者は認知症の行動・心理病状（Behavioral and psychological symptoms of dementia; BPSD）を生じることがあり、介護者の負担と減らし、被介護者に適切な処置を行うためには、IoT センサーデバイスによる BPSD の早期の自動予測が求められている。近年では、数分前の短期的なセンサーデータをもとに機械学習手法を用いて BPSD を予測する研究が行われている。しかし、被介護者ごとの体調や様子、過去の行動履歴など、センサーデータでは捉えられてない介護者の観点のコンテキスト情報を考慮できないという問題点があった。

本研究では、IoT デバイスで取得されるセンサーデータによる短期的な情報と、介護記録文書によるコンテキスト情報の2つの情報を考慮しながら、BPSD を自動予測するための Dual-Encoder モデルを提案する。具体的には、深層学習手法を用いてセンサーデータと介護記録文書の情報をそれぞれ異なる Encoder に入力する。提案手法は介護記録文書を考慮することで被介護者の潜在的な BPSD の起こりやすさを考慮し、precision を高くすることができる。評価実験では提案手法の BPSD の予測精度が既存手法を上回ることを示した。

# 目次

<b>1</b>	<b>序論</b>	<b>8</b>
<b>2</b>	<b>関連研究</b>	<b>12</b>
2.1	時系列予測の深層学習手法	12
2.1.1	LSTNet	12
2.1.2	Transformer	14
2.2	文書ベクトルの推定手法	15
2.2.1	Doc2Vec	15
2.2.2	BERT	18
<b>3</b>	<b>提案手法</b>	<b>19</b>
3.1	センサーデータの計測手法	19
3.2	介護記録文書	20
3.3	BPSD 予測のための Dual-Encoder モデル	22
3.3.1	提案手法における LSTNet の出力	23
3.3.2	提案手法における TransformerEncoder の出力	24
<b>4</b>	<b>評価実験</b>	<b>25</b>
4.1	実データの前処理	25
4.2	実験設定	27
4.3	実験結果及び考察	28
4.3.1	センサーデータのための手法選択	29
4.3.2	提案手法の評価実験	30
4.3.3	介護記録文書の有効性の分析	31
4.3.4	特徴量のアブレーション実験	32
<b>5</b>	<b>結論</b>	<b>34</b>
	謝辞	35
	使用したツール	36

目次	5
----	---

---

<b>付録</b>	<b>37</b>
A ハイパーパラメータの設定 . . . . .	37
B under-sampling の割合と精度 . . . . .	38
C Precision-Recall Curve . . . . .	39
<b>参考文献</b>	<b>40</b>

## 表目次

1.1	センサーデータを用いた BPSD 予測の先行研究と本研究の比較 . . . . .	10
3.1	BPSD の種類の一覧 . . . . .	21
3.2	生活支援記録法 F-SOAIP による観点 . . . . .	22
3.3	BPSD とラベル付けされた介護記録文書の例 . . . . .	22
4.1	データセット (分単位) . . . . .	26
4.2	データセット (日単位) . . . . .	26
4.3	SudachiPy による介護記録文書の統計量 . . . . .	27
4.4	センサーデータのみを用いた BPSD 予測結果 . . . . .	29
4.5	TransformerEncoder による window-size $w$ の変化による精度変化 . . . . .	29
4.6	介護記録文書を考慮した提案手法による分単位の BPSD 予測結果 . . . . .	30
4.7	+TransformerEncoder(doc2vec,DBoW) における $w_{doc}$ の変化による精度変化 . . . . .	30
4.8	介護記録文書のみを用いた日単位での BPSD 予測結果 . . . . .	31
4.9	TransformerEncoder(doc2vec, DBoW) による日単位での BPSD 予測におけ る $d_{doc}$ の変化 . . . . .	32
4.10	+TransformerEncoder(doc2vec,DBoW) における特徴量のアブレーション . . . . .	33
A.1	設定されたハイパーパラメータ値 . . . . .	37
A.2	提案手法の設定されたハイパーパラメータ値 . . . . .	38
B.1	+TransformerEncoder(doc2vec,DBoW) における $u$ の変化による精度変化 . . . . .	39

## 図目次

2.1	LSTNet の概要図 . . . . .	12
2.2	Transformer の Encoder の概要図 . . . . .	14

---

2.3	Distributed Bag of Words(DBoW) の概要図 . . . . .	16
2.4	Distributed Memory(DM) の概要図 . . . . .	17
2.5	Bidirectional Encoder Representations from Transformers(BERT) の概要図	18
3.1	センサーの写真 <sup>1)</sup> . . . . .	19
3.2	センサーデータの計測手法の概要図 <sup>1)</sup> . . . . .	20
3.3	提案手法の Dual-Encoder モデルの概要図 . . . . .	23
C.1	+TransformerEncoder(doc2vec,DBoW) モデルにおけるデータ 1 の PR 曲線	40
C.2	+TransformerEncoder(doc2vec,DBoW) モデルにおけるデータ 2 の PR 曲線	40



# 1 序論

日本の高齢者人口の割合は2021年で29.1%と過去最高になり<sup>2)</sup>、2025年には65歳以上の5人に1人が認知症を有すると推計されている<sup>3)</sup>。認知症を有する被介護者は失禁、異常行動、不安、アパシーや焦燥などの認知症の行動・心理病状 (Behavioral and psychological symptoms of dementia;BPSD)<sup>4)</sup>を生じることがある。BPSDの被介護者に対応する介護職員や看護師の負担を軽減し、被介護者に適切な対処を行うためには、早期にBPSDを予測することが重要になってくる。

BPSDの要因は体調や薬剤投与などであるとされており<sup>5)</sup>、近年では簡便にBPSDを予測する方法としてIsmailら(2017)が mild behavioral impairment checklist<sup>6)</sup>、藤生ら(2019)がBPSD気づき質問票57項目版<sup>7)</sup>を開発している。これらは、BPSDになる要因として介護者が判断できる項目を列挙しているが、BPSDは複数の要因が複雑に影響して生じるため、BPSDの予兆を介護者が完全に判断するのは困難である。

そこで近年では、IoTセンサーデバイスで観測したセンサー値を用いてBPSDを自動予測する研究が注目されている<sup>8-18)</sup>。例えば、Gongら(2015)<sup>13)</sup>は、IoTセンサーデバイスを用いてアルツハイマー型認知症を有する12名を対象にBPSDの種別とされる失禁と焦燥の関係を調査している。具体的には、約5日~1ヶ月分の12人について、ベッドに取り付けたセンサーとオムツ内に取り付けた湿度計から失禁の回数を、焦燥の判定を手首につけた加速度センサーから焦燥の回数を調査した。しかし、Gongらの研究では被介護者の人数が少ないことや期間が短いこと、焦燥の評価基準が明確でないといった問題点があった。

一方、Valemboisら(2015)<sup>14)</sup>は被介護者183名(内認知症126名)を対象にIoTセンサーデバイスを用いてアクチグラフによる活動量(運動量、睡眠時間、睡眠回数)を8~10日間分収集し、アパシー、不安、焦燥、異常行動を専門家が介護者への質問によって評価するNeuropsychiatric inventory (NPI)<sup>19)</sup>の結果との関係を分析した。その結果、アパシー、異常行動において活動量との有意な因果関係をKruskal-Wallis検定によって示した。

また、Nesbittら(2018)<sup>16)</sup>は8名の被介護者に対しスマートウォッチによる加速度センサー値と心拍数、スマートフォンによる音声データについて4時間分を収集した。介護者

が BPSD の焦燥に関する 29 項目の行動障害の出現頻度を 7 段階で評価する Cohen-Mansfield Agitation Inventory (CMAI)<sup>20)</sup> を実施し、加速度センサー値によって測定された上肢の活動と焦燥の関連を示した。

しかし、これまでの BPSD 予測の研究では NPI や CMI による主観評価に基づいており、BPSD の予兆から BPSD を自動で予測するシステムではなかった。そこで Khan ら (2019)<sup>17)18)</sup> は 2 名の被介護者に対し 4 つのセンサーを用いて加速度、心拍数、皮膚温度、皮膚電気活動を最大 15 日間分収集し、1~5 分前のセンサー値を入力として、Support Vector Machine(SVM)<sup>21)</sup>、Random Forest(RF)<sup>22)</sup> を用いて焦燥の予測を行った。しかし被介護者が少人数であることや、機器のトラブルにより、分単位で収集されたデータ数は約 6000 個と小規模であった。さらに、BPSD の種別を「焦燥」のみに限定している。これらの先行研究の問題点は山上ら (2021 年)<sup>23)</sup> が指摘している。また、これらの研究では、被介護者ごとの体調や様子、過去の行動履歴などの、センサーデータでは捉えられてない介護者の観点のコンテキスト情報を考慮できないという問題点もある。

本研究では、IoT デバイスで取得されるセンサーデータによる情報と、介護記録文書によるコンテキスト情報を考慮した BPSD の自動予測を行う。具体的には、深層学習手法を用いて異なる 2 種の情報をそれぞれ別の Encoder に入力する Dual-Encoder モデルを提案する。センサーデータの Encoder には Rcurrent Neural Networks (RNN)<sup>24)</sup> を用いた多変量時系列予測モデルである LSTNet<sup>25)</sup> と、自然言語処理や時系列予測の分野で用いられている attention 機構<sup>26-28)</sup> をもつ Transformer<sup>29)</sup> を用いた。

提案手法はセンサーデータのみを用いる場合と比較して precision の向上が期待できる。例えば、センサーデータで騒音が大きいと BPSD が起きる場合があったとしても、健常者が BPSD を生じないように被介護者も必ずしも BPSD を生じるとは限らない。このようにセンサーデータのみを用いた BPSD 予測では precision が低い可能性がある。これを解決するにはセンサーデータでは捉えられない被介護者の BPSD の潜在的な起こりやすさを考慮する必要がある。直近の介護記録文書を考慮することで precision を高くすることが期待できる。

先行研究と本研究の比較を表 1.1 に示す。人数は Valembois ら (2015) より少ないが、データ収集期間が既存研究よりも長期間である。センサー値の数及び BPSD の種別は既存研究の中で最も多く、機械学習手法だけでなく深層学習手法による時系列予測モデルを用いた。また、既存研究が 2 分割交差検証に対し、人数が多いことで 5 分割交差検証を可能にし、再現性の高い結果を提示した。さらにセンサーデータと介護記録文書の両方を加味した研究はこれまでには見当たらない。

表 1.1 センサーデータを用いた BPSD 予測の先行研究と本研究の比較

	人数	期間	センサー値の数	BPSD の種別	手法	評価方法	介護記録文書の考慮
Gong ら	12	11 名が約 5 日 1 名が約 1 ヶ月	4 (音声はノイズで解析困難)	焦燥, 失禁	センサー値を基に判定	失禁と焦燥の 前後関係をカウント	No
Valembois ら	183	8~10 日	3	アパシー, 不安, 焦燥, 異常行動	NPI など	Kruskal-Wallis 検定	No
Nesbitt ら	8	4 時間の測定を 2 回	4 (音声はノイズで解析困難)	焦燥	CMAI など	センサー波形と比較	No
Khan ら	2	1 名が 15 日 もう 1 名が 13 日	4	焦燥	機械学習 (SVM, RF)	AUC (2 分割交差検証)	No
本研究	30	約 4 ヶ月 最小: 22 日 最大: 123 日 平均: 81 日	9	すべて (18 種類+その他)	機械学習 (SVM, RF, LR) 深層学習 (LSTNet, Transformer)	AUC, F1 (5 分割交差検証)	Yes

本研究のデータ及び実験設定の特徴は以下のとおりである。

- (1) センサーデータによる情報のみならず、介護記録文書によるセンサーデータでは捉えられない介護者の観点からの情報を加味した。
- (2) BPSD の種類を限定しない、包括的（幻聴、妄想、暴言、暴行・暴力、周徘徊、無断外出・夕暮れ症候群、性的不適切行動・逸脱行為、常同行為、脱抑制、易怒声・興奮、収集、大声、うつ、アパシー、無反応・無関心、不安、不潔行為、介護への抵抗・拒否、その他）な予測を可能にした。
- (3) 介護記録文書の入力とは昨日以前のものを使用し、BPSD の予測は 10 分以内に生じるか否かを予測するため、実現場で対応可能な状況を想定している。
- (4) 9 つのセンサー値を用いており、既存研究より最も多い特徴量を用いた。
- (5) センサーデータは、3 つ介護施設から最大 30 人を対象に約 4 ヶ月間収集され、分単位で収集されたデータの規模は 300 万以上と、機械学習を用いた既存研究よりも遥かに大規模である。
- (6) 介護記録文書は生活支援記録法 F-SOAIP に則っており、BPSD の教師ラベルは国立研究開発法人日本医療研究開発機構 (Japan Agency for Medical Research and Development;AMED) の専門家によって信頼性のあるラベル付けがされている。

実験結果では、センサーデータに介護記録文書を加えて深層学習することで予測精度が向上することを示す。またセンサーデータの入力時系列長を大きくすると recall が高くなり、介護記録文書を入力する時系列長を大きくすると precision が高くなることを示す。

ただし、これまでも、医療文書を用いた認知症及び BPSD の研究が知られている<sup>30)31)32)</sup>。Rumshisky ら (2016)<sup>30)</sup> は電子健康記録をもとに Latent Dirichlet Allocation(LDA)<sup>33)</sup> から得られるトピック (不安, 自殺, 下痢など) を用いて, 被介護者が 30 日以内に再入院するかを SVM で予測している。Topaz ら (2020)<sup>31)</sup> は在宅被介護者の入院及び救急搬送率と在宅介護者による臨床記録との関係性を 1 年間分のデータについて調査し, 認知症の診断を受けていない在宅被介護者でも「記憶障害」「不安」「抑うつ気分」「無気力」の症状が 1 つでも記録されている場合には入院及び救急搬送率が増加することを示した。また Al-Harrasi ら (2021)<sup>32)</sup> は認知症診断時の記録文書にある運動的特徴 (身震い, パーキンソン歩行など) から 24 の特徴量を抽出して, 8 年間分のデータについて被介護者の入院期間及び生存率を Cox 回帰予測を行い, 身震いについて有意性を示した。

しかし、これらの研究は BPSD を予測するのではなく、BPSD を特徴量として入院するかなどを予測するタスクであり、本研究とは異なるタスクである。

## 2 関連研究

本章では、提案手法に用いる先行研究として主に時系列データの予測に用いられる深層学習手法と文章の特徴量の推定手法を説明する。

### 2.1 時系列予測の深層学習手法

#### 2.1.1 LSTNet

RNN を用いた多変量時系列予測モデルとして、多変量の入力に対し畳み込みニューラルネットワーク (Convolutional Neural Network; CNN)<sup>34)</sup> で複数の重み行列を用いることで変数の相互関係を捉える LSTNet<sup>25)</sup> が提案されている。図 2.1 に LSTNet の概要図を示す。

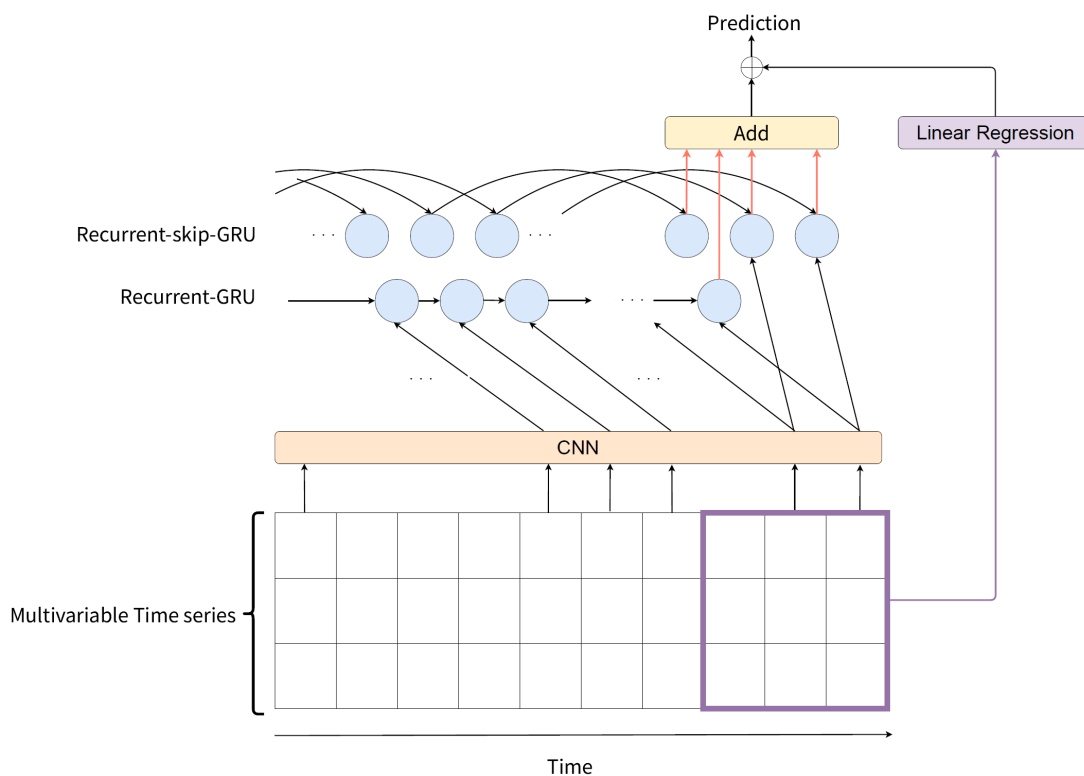


図 2.1 LSTNet の概要図

LSTNet では時刻  $t$  における  $n$  変量のセンサー値の入力  $\mathbf{x}(t) \in \mathbb{R}^n$  とし、行列  $X(t) = [\mathbf{x}(t-\omega+1)^T, \dots, \mathbf{x}(t)^T] \in \mathbb{R}^{n \times \omega}$  を考える。  $\omega \in \mathbb{R}$  はハイパーパラメータである。畳み込み層では  $d_c$  個の重み行列とバイアスを用いて、隠れ状態ベクトル  $\tilde{\mathbf{x}}_t = [\tilde{x}_1(t), \dots, \tilde{x}_{d_c}(t)] \in \mathbb{R}^{d_c}$  を以下で計算する。

$$\tilde{x}_k(t) = \text{RELU}(W_k * X(t) + b_k) \quad (2.1)$$

$$= \text{RELU}\left(\sum_i^n \sum_j^\omega W_{k,i,j} X(t)_{i,j} + b_k\right) \quad (2.2)$$

ここで、  $W_k \in \mathbb{R}^{\omega \times n}$ 、  $b_k \in \mathbb{R}$  は  $k$  番目の重み行列、バイアスであり、  $\text{RELU}(x) = \max(0, x)$  である。

次に RNN の一種である Gated recurrent unit (GRU)<sup>35)</sup> を 2 つ用いて時系列情報を伝播する。 2 つの GRU はラグ次数  $p$  が異なり、異なる時間間隔の情報を考慮することができる。ここではラグ次数  $p = 1$  の通常の GRU を Recurrent-GRU、ラグ次数  $p$  の GRU を Recurrent-skip-GRU とよぶ。

### (1) Recurrent-GRU

1 つ目の GRU では隠れ次元  $R \in \mathbb{R}$  として重み行列  $W_{xi} \in \mathbb{R}^{d_c \times R}$ 、  $W_{hi} \in \mathbb{R}^{R \times R}$ 、バイアス  $\mathbf{b}_i \in \mathbb{R}^R$  により潜在変数  $h_t$  を求める。ただし  $i$  は添字  $i = r, u, c$  を表す。

$$\mathbf{r}_t = \sigma(\tilde{\mathbf{x}}_t W_{xr} + \mathbf{h}_{t-1} W_{hr} + \mathbf{b}_r) \quad (2.3)$$

$$\mathbf{u}_t = \sigma(\tilde{\mathbf{x}}_t W_{xu} + \mathbf{h}_{t-1} W_{hu} + \mathbf{b}_u) \quad (2.4)$$

$$\mathbf{c}_t = \text{RELU}(\tilde{\mathbf{x}}_t W_{xc} + \mathbf{r}_t \odot (\mathbf{h}_{t-1} W_{hc}) + \mathbf{b}_c) \quad (2.5)$$

$$\mathbf{h}_t = (\mathbf{1} - \mathbf{u}_t) \odot \mathbf{h}_{t-1} + \mathbf{u}_t \odot \mathbf{c}_t \quad (2.6)$$

ここで  $\sigma$  はシグモイド関数、  $\odot$  はアダマール積、  $\mathbf{1}$  は全要素 1 の  $R$  次元ベクトルである。  $\mathbf{u}_t$  は式 (2.6) において  $\mathbf{h}_t$  の  $\mathbf{c}_t$  の重み付けを表す変数であり、  $\mathbf{r}_t$  は式 (2.5) において  $\mathbf{c}_t$  を求める際に、時刻  $t-1$  の情報をもつ  $\mathbf{h}_{t-1}$  をどの程度考慮するかを決める変数である。

### (2) Recurrent-skip-GRU

もう 1 つラグ次数  $p$  をもつ Recurrent-skip-GRU を考える。簡単のため 1 つ目の GRU と同じ文字を用いる。まず、Recurrent-skip-GRU では異なる隠れ次元  $S \in \mathbb{R}$  を設定する。

$$\mathbf{r}_t = \sigma(\tilde{\mathbf{x}}_t W_{xr} + \mathbf{h}_{t-p} W_{hr} + \mathbf{b}_r) \quad (2.7)$$

$$\mathbf{u}_t = \sigma(\tilde{\mathbf{x}}_t W_{xu} + \mathbf{h}_{t-p} W_{hu} + \mathbf{b}_u) \quad (2.8)$$

$$\mathbf{c}_t = \text{RELU}(\tilde{\mathbf{x}}_t W_{xc} + \mathbf{r}_t \odot (\mathbf{h}_{t-p} W_{hc}) + \mathbf{b}_c) \quad (2.9)$$

$$\mathbf{h}_t = (\mathbf{1} - \mathbf{u}_t) \odot \mathbf{h}_{t-p} + \mathbf{u}_t \odot \mathbf{c}_t \quad (2.10)$$

Recurrent-GRU で得られた  $\mathbf{h}_t$  を  $\mathbf{h}_t^R \in \mathbb{R}^R$ , skip-GRU で得られた  $\mathbf{h}_t$  を  $\mathbf{h}_t^S \in \mathbb{R}^S$  とする. 重み行列を  $W^R \in \mathbb{R}^{D \times R}$ ,  $W_i^S \in \mathbb{R}^{D \times S}$  (ただし  $i = 0, \dots, p-1$ ), バイアスを  $b \in \mathbb{R}^D$  として隠れ次元  $D \in \mathbb{R}$  の隠れベクトル  $\mathbf{h}_t^D \in \mathbb{R}^D$  を計算する.

$$\mathbf{h}_t^D = W^R \mathbf{h}_t^R + \sum_{i=0}^{p-1} W_i^S \mathbf{h}_{t-i}^S + b \quad (2.11)$$

次に  $\mathbf{x}(t-q+1)$  から  $\mathbf{x}(t)$  までの  $q$  個を入力する線形回帰を考える. 重み係数  $W_k^{ar} \in \mathbb{R}$  (ただし  $k = 0, \dots, q-1$ ), バイアス  $\mathbf{b}^{ar} \in \mathbb{R}^n$  として  $\mathbf{h}_t^L \in \mathbb{R}^n$  を計算する.

$$\mathbf{h}_t^L = \sum_{k=0}^{q-1} W_k^{ar} \mathbf{x}(t-k) + \mathbf{b}^{ar} \quad (2.12)$$

$D = n$  とすると以下の式で出力  $\hat{\mathbf{y}}_t \in \mathbb{R}^n$  が求まる.

$$\hat{\mathbf{y}}_t = \mathbf{h}_t^D + \mathbf{h}_t^L \quad (2.13)$$

$\hat{\mathbf{y}}_t$  は入力系列  $\mathbf{x}(1) \sim \mathbf{x}(t)$  を入力した際の出力であり, 時刻  $t+1$  における  $n$  変量の回帰予測値である.  $\hat{\mathbf{y}}_t$  と実際の値の二乗誤差を損失関数として確率的勾配降下法で学習する.

### 2.1.2 Transformer

近年, 自然言語処理の分野では Vaswani ら<sup>29)</sup> が attention 機構のみの深層学習である Transformer を提案した. Transformer における attention では入力情報同士の関連性を計算し, 文章内の単語の関係を attention 内で推定できるようになったので精度が向上した. 図 2.2 に Transformer の Encoder の概要図を示す.

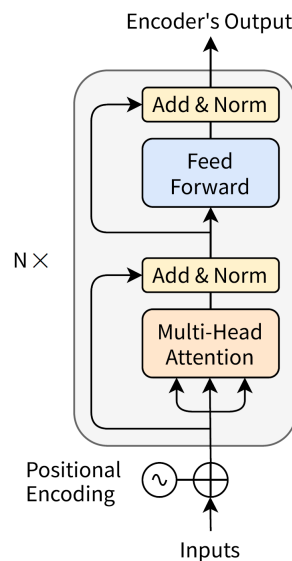


図 2.2 Transformer の Encoder の概要図

Transformer は Encoder と Decoder の 2 つの機構からなるが、本研究では提案手法に用いる Encoder 内の self-attention についてのみ説明する。時刻  $t$  において  $n$  変数の入力  $\mathbf{x}(t) \in \mathbb{R}^n$  とする。時刻  $t = 1$  から  $t = t$  に入力を行列  $X = [\mathbf{x}(1)^T, \dots, \mathbf{x}(t)^T] \in \mathbb{R}^{n \times t}$  で表す。  $l$  層への入力行列を  $X^{(l)} \in \mathbb{R}^{n \times t}$  とし、  $X^{(0)} = X$  とする。重み行列  $W^Q \in \mathbb{R}^{n \times d_k}$ ,  $W^K \in \mathbb{R}^{n \times d_k}$ ,  $W^V \in \mathbb{R}^{n \times d_v}$  を考え、アテンション  $\alpha \in \mathbb{R}^T$  を計算し  $X^{(l+1)}$  を求める。ただし  $d_k, d_v$  は隠れ次元数である。

$$Q = W^Q X^{(l)} \quad (2.14)$$

$$K = W^K X^{(l)} \quad (2.15)$$

$$V = W^V X^{(l)} \quad (2.16)$$

$$\alpha = \text{Softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) \quad (2.17)$$

$$X^{(l+1)} = \alpha V \quad (2.18)$$

ただし  $\text{Softmax}_i(\mathbf{x}) = \frac{e^{x_i}}{\sum_j e^{x_j}}$  である。これを層数  $L$  だけ繰り返した後、Decoder に入力して attention を計算することで最終的な出力系列  $Y \in \mathbb{R}^{m \times d_v}$  を得る。ただし  $m$  は出力の次元数で、語彙数である。学習はクロスエントロピーを損失関数として確率的勾配降下法を用いる。

## 2.2 文書ベクトルの推定手法

### 2.2.1 Doc2Vec

自然言語処理の分野において Le ら (2014)<sup>36)</sup> は深層学習を用いた文書ベクトルの推定手法として Doc2Vec を考案した。文書ベクトルとは文の特徴量を連続量の複数の変数の組で表したものである。Doc2Vec には、文に含まれる単語の有無を基にした Distributed Bag of Words version of Paragraph Vector (DBoW) と語順を考慮した Distributed Memory Model of Paragraph Vectors (DM) の 2 つの文書ベクトルの推定方法がある。以下で DBoW と DM について詳しく説明する。

#### (1) Distributed Bag of Words (DBoW)

Distributed Bag of Words (DBoW) について説明する。図 2.3 に DBoW の概要図を示す。



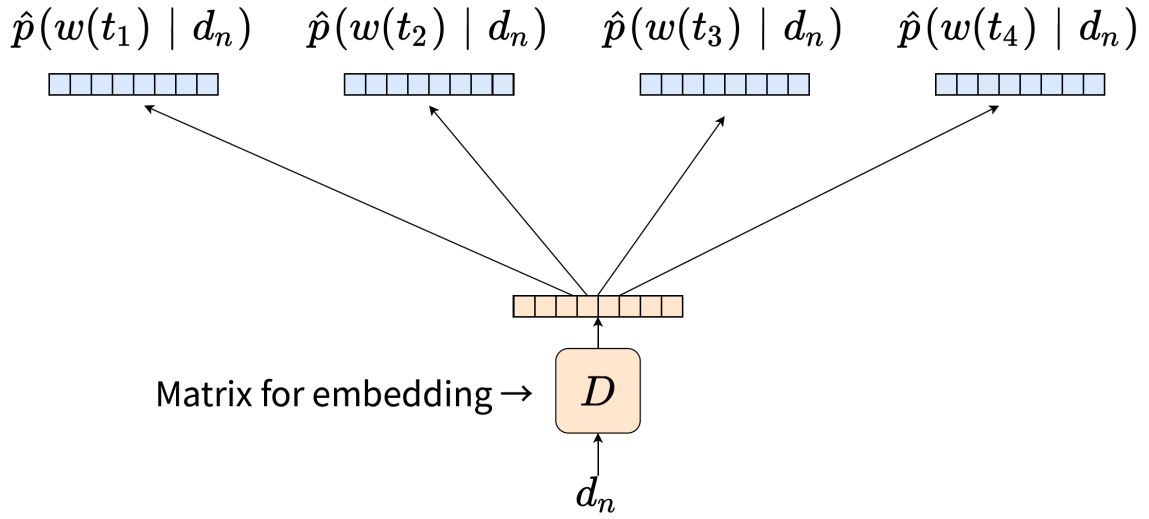


図 2.3 Distributed Bag of Words(DBoW) の概要図

$N$  個の文書はそれぞれ文書 ID をもち、文書 ID が  $n$  の文書を変数  $d_n$  と表す。  $d_n$  の one-hot ベクトル  $\mathbf{x}_{d_n} \in \{0, 1\}^N$  から埋め込み行列  $D \in \mathbb{R}^{N \times m}$  によって得られる  $m \in \mathbb{R}$  次元の埋め込みベクトル  $\mathbf{e}_{d_n} \in \mathbb{R}^m$  を以下の式より計算する。

$$\mathbf{e}_{d_n} = \mathbf{x}_{d_n} D \quad (2.19)$$

さらに語彙数  $V \in \mathbb{R}$  として重み行列  $W^O \in \mathbb{R}^{m \times V}$  として出力  $\hat{\mathbf{p}}_{d_n} \in \mathbb{R}^V$  を得る。

$$\mathbf{y}_{d_n} = \mathbf{e}_{d_n} W^O \quad (2.20)$$

$$\hat{\mathbf{p}}_{d_n} = \text{Softmax}(\mathbf{y}_{d_n}) \quad (2.21)$$

ここで文書  $d_n$  に含まれる語彙からサンプリングされた語彙を表す集合を  $\mathcal{W}$ 、  $\mathcal{W}$  のそれぞれの要素がもつ語彙 ID の集合を  $\mathcal{C}$  とする。このとき  $|\mathcal{W}| = |\mathcal{C}|$  はコンテキストサイズと呼ばれるハイパーパラメータである。サンプリングされた語彙がそれぞれ独立であることを仮定して、同時確率分布  $p(\mathcal{W}|d_n)$  を求める。

$$p(\mathcal{W}|d_n) = \prod_{i \in \mathcal{C}} \hat{p}_{d_n, i} \quad (2.22)$$

同時確率分布  $p(\mathcal{W}|d_n)$  を最大化するように以下の目的関数  $L$  を確率的勾配降下法で学習し、誤差逆伝播法により埋め込み行列  $D$  が学習され、文書ベクトルが推定される。

$$L = -\log p(\mathcal{W}|d_n) \quad (2.23)$$

## (2) Distributed Memory(DM)

次に Distributed Memory(DM) について説明する． 図 2.4 に DM の概要図を示す．

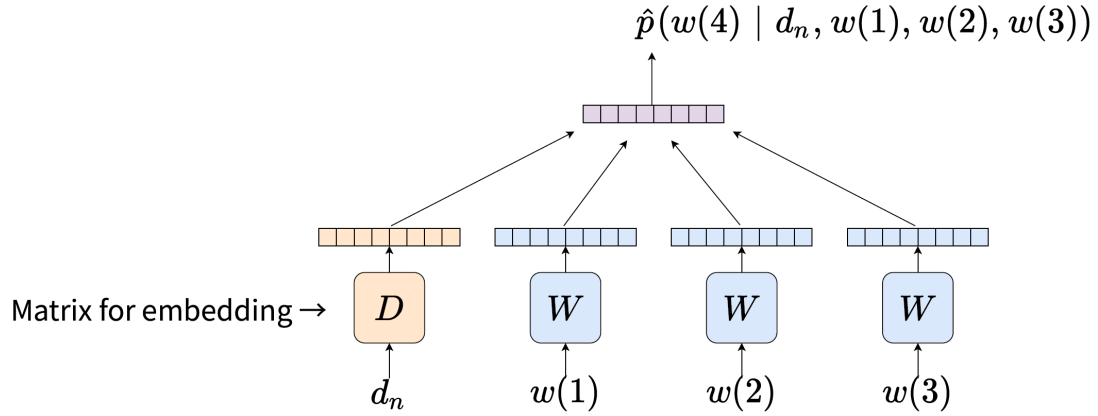


図 2.4 Distributed Memory(DM) の概要図

DM では文書 ID  $n$  の他に文書  $d_n$  の文章の先頭から  $|\mathcal{C}|$  番目までの単語をサンプリングして Doc2Vec に入力し，  $|\mathcal{C}| + 1$  番目の単語を予測する．先頭から位置  $t \leq |\mathcal{C}|$  番目の語彙 ID  $c \in \mathcal{C}$  の単語を  $w_c(t) \in \mathcal{W}$  と表す．  $w_c(t)$  の one-hot ベクトル  $\mathbf{x}_{w_c(t)} \in \{0, 1\}^V$  と，埋め込み行列  $W \in \mathbb{R}^{N \times m}$  によって得られる埋め込みベクトル  $\mathbf{e}_{w_c(t)} \in \mathbb{R}^m$  から以下の式より計算する．

$$\mathbf{e}_{w_c(t)} = \mathbf{x}_{w_c(t)} W \quad (2.24)$$

$$\mathbf{h}_{d_n, \mathcal{W}} = \mathbf{e}_{d_n} + \sum_{t=1}^{|\mathcal{C}|} \mathbf{e}_{w_c(t)} \quad (2.25)$$

最後に語彙数  $V \in \mathbb{R}$  として重み行列  $W^O \in \mathbb{R}^{m \times V}$  として出力  $\hat{\mathbf{p}}_{d_n, \mathcal{W}} \in \mathbb{R}^V$  を得る．

$$\mathbf{y}_{d_n, \mathcal{W}} = \mathbf{h}_{d_n, \mathcal{W}} W^O \quad (2.26)$$

$$\hat{\mathbf{p}}_{d_n, \mathcal{W}} = \text{Softmax}(\mathbf{y}_{d_n, \mathcal{W}}) \quad (2.27)$$

文書  $d_n$  の先頭から  $T = |\mathcal{C}| + 1$  番目の単語を語彙 ID  $i$  として  $w_i(T)$  と表すと，  $\hat{\mathbf{p}}_{d_n, \mathcal{W}}$  の  $i$  番目の要素が  $p(w_i(T) | d_n, \mathcal{W})$  となる．  $T$  番目の単語の予測確率  $p(w_i(T) | d_n, \mathcal{W})$  を最大化するように以下の目的関数  $L$  を確率的勾配降下法で学習し，誤差逆伝播法により埋め込み行列  $D$  が学習され，文書ベクトルが推定される．

$$L = -\log p(w_i(T)|d_n, \mathcal{W}) \quad (2.28)$$

## 2.2.2 BERT

Doc2Vec では文書から単語をサンプリングして文書ベクトルを推定したため、文全体の特徴量が捉えら得ないという欠点があった。そこで、自然言語処理の分野において Devlin ら (2018)<sup>37)</sup> は文の前後双方向からの attention を推定する Bidirectional Encoder Representations from Transformers(BERT) を提案した。これにより単語の前後関係が重要な自然言語処理の分野において高い予測精度を示した。図 2.5 に BERT の概要図を示す。

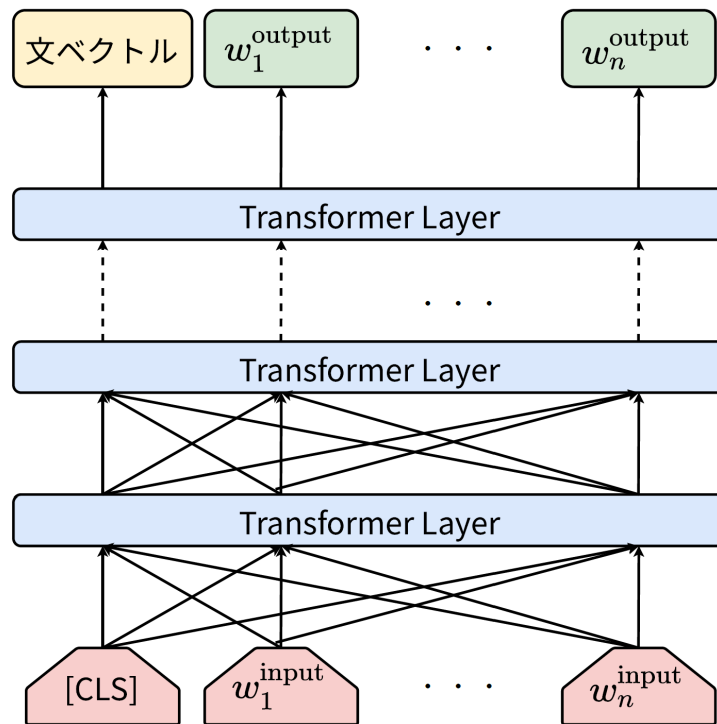


図 2.5 Bidirectional Encoder Representations from Transformers(BERT) の概要図

BERT の入力文の先頭には CLS トークンと呼ばれる特殊記号を入力し、それに対応する出力として CLS ベクトルを出力する。CLS ベクトルは文全体の特徴量を表す文書ベクトルとして使われる。BERT を分類タスクとして fine-tuning する際は、CLS ベクトルを更に分類用のニューラルネットワークに入力することでクラスを予測し、誤差逆伝播法で BERT を再学習し、CLS ベクトルを推定する。

## 3 提案手法

### 3.1 センサーデータの計測手法

本研究では3つの日本の介護施設 A,B,C において、A では9人、B では27人、C では29人の被介護者を対象に20週間分（約4ヶ月分）のセンサーデータを取得する。センサーの写真を図 3.1 に、センサーデータの計測手法の概要図を図 3.2 に示す。図センサーデータは2種類あり、環境センサーにより騒音、照度、湿度、温度、気圧、CO2濃度を測定し、ベッド型センサーにより睡眠状態、心拍数、呼吸数を測定し、合計で9つのセンサー値を取得する。なお、環境センサーは秒単位、ベッド型センサーは分単位で測定されている。各センサー値は1週間ごとに1つのファイルに出力され、被介護者1人当たり最大で180ファイルのデータが計測期間内で取得される。



(a) 生体センサーの写真



(b) 環境センサーの写真

図 3.1 センサーの写真<sup>1)</sup>

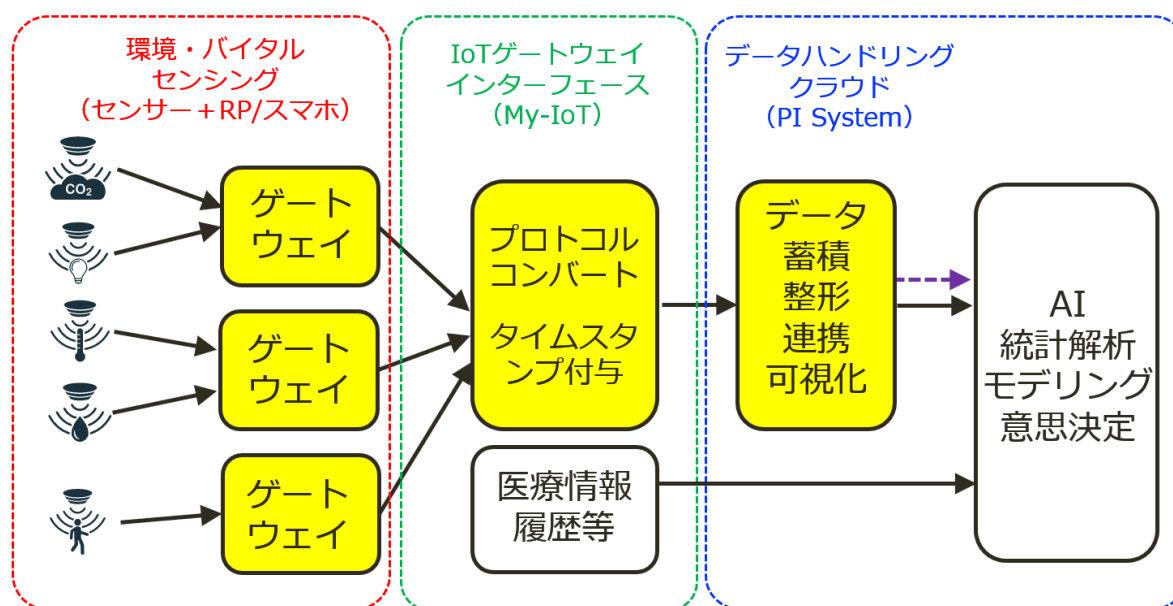


図 3.2 センサーデータの計測手法の概要図<sup>1)</sup>

## 3.2 介護記録文書

本データセットにおける BPSD の教師ラベル付けは、介護記録文書を基に AMED の専門家がラベル付けを行った。介護記録文書は介護者が被介護者に対し 1 日当たり平均 7.8 回、生活支援記録法 F-SOAIP に則って記録されたものである。F-SOAIP は表 3.2 に示す観点別に記録する方法である。専門家は表 3.1 に示す 18 種類とその他の BPSD について教師ラベル付けを行う。実際に BPSD と判定された介護記録文書の例を表 3.3 に示す。

表 3.1 BPSD の種類の一覧

	種類	具体的な症状
1	幻聴	実際にはないものが見えたり・聞こえたりす
2	妄想	盗られたという・嫉妬する・別人という
3	暴言	他者を傷つけるような乱暴な言葉を発する 言いがかりをつける・罵る
4	暴行・暴力	他者に乱暴な行いをする
5	周徘徊	うろうろする・目的もなく歩く 不安そうに歩き回る・同じところを歩き回る
6	無断外出・夕暮れ症候群	家/施設から出たがる・家に帰りたがる
7	性的不適切行動・逸脱行為	他者への性的に不適切な行為・性的関係強要
8	常同行為	こだわって同じ行為を何度も繰り返す
9	脱抑制	我慢できない、衝動的に行動する
10	易怒声・興奮	怒りっぽい・破る・壊す・物を投げる
11	収集	物をため込む
12	大声	大声をあげる・鳴声が続く・さけぶ
13	うつ	悲観的で気分が落ち込んでいる
14	アパシー	やる気がない、自分からは動かない
15	無反応・無関心	声掛けに反応が無い・興味を示さない
16	不安	心配ばかりする
17	不潔行為	尿や便で汚す、何日も入浴しない
18	介護への抵抗・拒否	介護拒否・食事拒否
19	その他	その他

表 3.2 生活支援記録法 F-SOAIP による観点

略称	名称	日本語名称	具体的な内容
F	Focus	着眼点	ニーズ・気がり・場面の表現
S	Subjective Data	主観的情報	被介護者の言葉
O	Objective Data	客観的情報	介護者の観察から得られた情報 環境・経過など
A	Assessment	アセスメント	介護者の判断・解釈
I	Intervention/Implementation	介入・実施	介護者の対応・支援・介護など
P	Plan	計画	当面の対応予定

表 3.3 BPSD とラベル付けされた介護記録文書の例

F-SOAIP の種別	介護記録文書	BPSD の種別
O	納得せずフロア内を徘徊している。	5
S	「帰んだ。」	6
O	夕食後より眠前薬の催促頻回に聞かれる。その都度声かけ。	8
O	急にホールから大きな怒鳴り声がある。	10
O	ベッド上にて泣いている。	13
S	もう、しんどいね。	14
O	下衣尿失禁の為更衣行方。抵抗強く見られた。	17

### 3.3 BPSD 予測のための Dual-Encoder モデル

Khan ら (2019)<sup>18)</sup> は数分前のセンサー値をもと機械学習手法 (SVM, RF) を用いて BPSD を予測した。しかしセンサーデータでは捉えられない介護者による情報が考慮されない問題点があった。一方で, Ismail ら (2017)<sup>6)</sup> と藤生ら (2019)<sup>7)</sup> は介護者による情報が BPSD の予測に有用であることを示した。

そこで本研究では, センサー値に加え, 介護記録文書の両方の情報を入力とした深層学習手法による BPSD 予測手法を提案する。概要図を図 3.3 に示す。提案手法では数分前のセンサー値の他に, 被介護者の昨日以前の直近の介護記録文書を複数入力する。介護記録文書を用いることで, センサー値では捉えられない介護者の観点からの情報を考慮し, 被介護者の潜在的な BPSD の起こりやすさを考慮することができる。具体的には, 数分前

のセンサー値を既存の時系列予測モデルである LSTMNet や Transformer の Encoder に入力するだけでなく、介護記録文書の時系列のための Encoder を用意した Dual-Encoder を用いて予測する。深層学習手法による文書ベクトル生成には Doc2Vec 及び BERT を用いた。

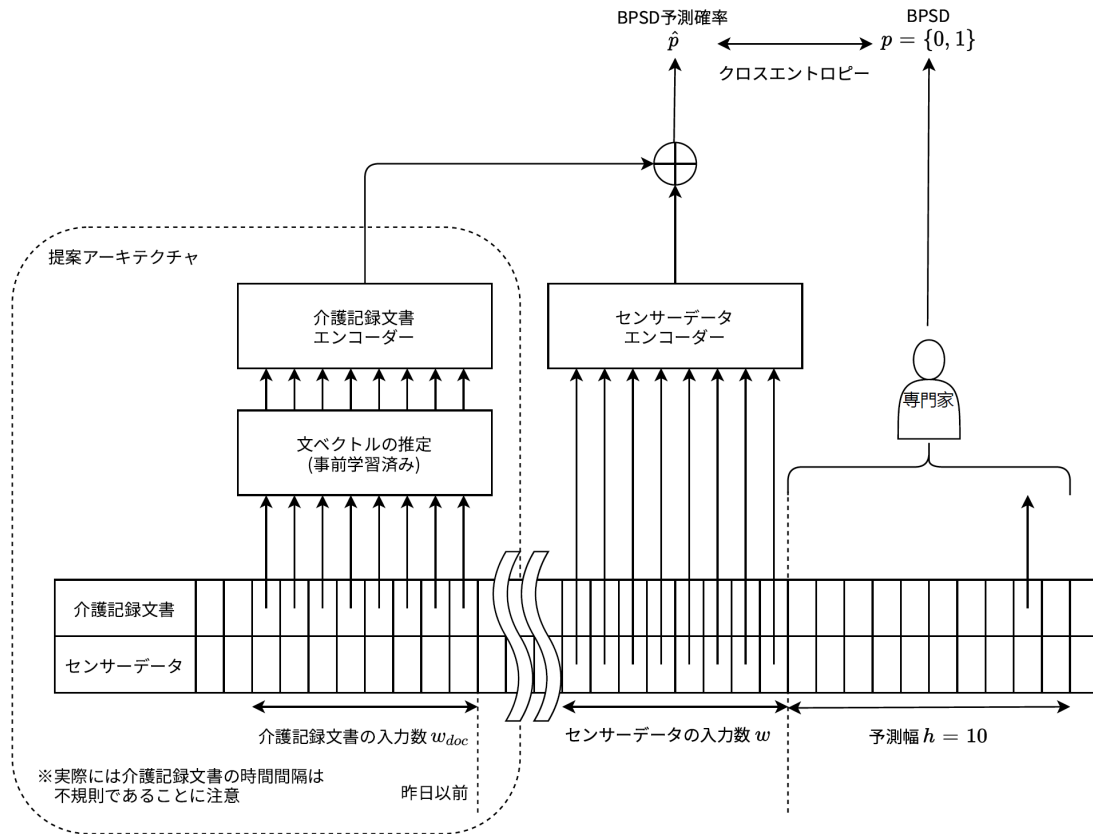


図 3.3 提案手法の Dual-Encoder モデルの概要図

### 3.3.1 提案手法における LSTMNet の出力

本研究では、BPSD になるかならないかを 2 値で予測する分類タスクであるため、LSTMNet の出力の次元数は  $D=2$  とする。また、式 (2.12) で表される線形回帰は以下のように変更する。

時刻  $t$  における  $n$  変数の入力  $\mathbf{x}(t) \in \mathbb{R}^n$  とし、 $\mathbf{x}(t-q+1)$  から  $\mathbf{x}(t)$  までの  $q$  個を考える。重み係数  $W_k^{ar} \in \mathbb{R}^{D \times n}$  (ただし  $k = 0, \dots, q-1$ )、バイアス  $\mathbf{b}^{ar} \in \mathbb{R}^D$  として  $\mathbf{h}_t^L \in \mathbb{R}^D$  を計算する。

$$\mathbf{h}_t^L = \sum_{k=0}^{q-1} W_k^{ar} \mathbf{x}(t-k) + \mathbf{b}^{ar} \quad (3.1)$$



式 (2.11) から得られる  $\mathbf{h}_t^D$  より以下の計算をする.

$$\hat{\mathbf{y}}_t = \mathbf{h}_t^D + \mathbf{h}_t^L \quad (3.2)$$

$$\hat{p}_{t,i} = \text{Softmax}_i(\hat{\mathbf{y}}_t) \quad (3.3)$$

$\hat{p}_{t,i}$  は時刻  $t$  においてクラス  $i$  である予測確率であり, BPSD である予測確率である. 式 (3.3) について真値とのクロスエントロピーを計算し, 確率的勾配降下法で学習する.

介護記録文書を考慮する際は, 介護記録文書のエンコーダである LSTM の出力  $\hat{\mathbf{y}}_t^{doc}$  を用いて以下のように  $\hat{p}_{t,i}$  を求める.

$$\hat{p}_{t,i} = \text{Softmax}_i(\hat{\mathbf{y}}_t + \hat{\mathbf{y}}_t^{doc}) \quad (3.4)$$

### 3.3.2 提案手法における TransformerEncoder の出力

前述の通り, 本研究は 2 値の時系列分類タスクのため  $L$  層の Encoder の出力  $X^{(L)} \in \mathbb{R}^n$  に対し, 出力の次元数  $d_o = 2$  として重み行列  $W^O \in \mathbb{R}^{d_o}$  として以下の計算を行う.

$$\tilde{\mathbf{y}}_t = \frac{1}{t} \sum_{i=1}^t X_i^{(L)} \quad (3.5)$$

$$\hat{\mathbf{y}}_t = W^O \tilde{\mathbf{y}}_t \quad (3.6)$$

$$\hat{p}_{i,t} = \text{Softmax}_i(\hat{\mathbf{y}}_t) \quad (3.7)$$

$\hat{p}_{i,t}$  は時刻  $t$  においてクラス  $i$  である予測確率であり, BPSD である予測確率である. 式 (3.7) について真値とのクロスエントロピーを計算し, 確率的勾配降下法で学習する.

介護記録文書を考慮する際は, 介護記録文書のエンコーダである TransformerEncoder の出力  $\hat{\mathbf{y}}_t^{doc}$  を用いて以下のように  $\hat{p}_{t,i}$  を求める.

$$\hat{p}_{t,i} = \text{Softmax}_i(\hat{\mathbf{y}}_t + \hat{\mathbf{y}}_t^{doc}) \quad (3.8)$$

## 4 評価実験

### 4.1 実データの前処理

被介護者のデータは最大で180ファイルのデータを取得するが、機器のトラブルや被介護者の状況によりファイルが欠損しており、欠損が多い被介護者のデータでは適切なパラメータ推定が行えない。そこで本研究では(1)欠損したファイル数が36未満(2割未満)の19人のみ用いたデータ、(2)欠損したファイル数が100未満の30人のみを用いたデータの2つのデータで実験を行った。環境データは秒単位、ベッド型センサーは分単位で取得されているため、環境データについては1分間の平均をとることで時間間隔を揃えた。センサーデータが出力したファイルの中に存在する欠損値については、被介護者ごとに当日の平均値で欠損値処理を行った。

本研究では、BPSDを予測するための機械学習手法として、SVM, RandomForest, LogisticRegressionを用いた。それぞれの手法において時系列を考慮するために、時刻 $t$ のセンサー値 $s(t)$ として1分前との差分 $s_{diff}(t)$ と、それを1分前の値で割った変化率 $s_{pct.change}(t)$ を入力として加えた。

$$s_{diff}(t) = s(t) - s(t-1) \quad (4.1)$$

$$s_{pct.change}(t) = \frac{s(t) - s(t-1)}{s(t-1)} \quad (4.2)$$

よって本研究では、9つのセンサー値から27の特徴量を生成し、 $n = 27$ として時刻 $t$ の $n$ 変量の入力 $\mathbf{x}(t) \in \mathbb{R}^n$ を考えた。

介護記録文書についても以下の前処理を行った。

- スペース, タブ, 改行の除去
- 全角・半角の統一
- 3回以上の繰り返し表現の除去

- 桁区切りの除去と数字を 0 に置換
- 記号の除去 (、。「」は除く)

実用上の観点から、現在から直後（1 分後）に起きる PBSD の予測をしても、現場の介護者はすぐに対応できない。そこで本研究ではセンサーデータを用いた実験では、10 分以内に BPSD が生じる場合を正例とし、データセットの詳細を表 4.1 に示した。また、本研究では介護記録文書のみでの BPSD の予測も行い、この場合は介護記録文書の時間間隔が不規則であるため、翌日に BPSD が生じる場合を正例とする日単位の予測を行った。日単位のデータセットの詳細を表 4.2 に示した。また介護記録文書を SudachiPy<sup>38)</sup> (辞書は sudachidict\_core を使用) の A,B,C の 3 つの分割単位で分かち書きした場合の語彙数、単語数を表 4.3 に示す。

表 4.1 データセット（分単位）

	人数	期間	データ数 (分単位)	負例数	正例数 (10 分以内に BPSD が生じる)	負例数/正例数
データ 1	19	最小：43 日 最大：123 日 平均：97 日	2,558,658	2,552,082	6,576	388
データ 2	30	最小：22 日 最大：123 日 平均：81 日	3,365,784	3,357,649	8,135	413

表 4.2 データセット（日単位）

	人数	期間	データ数 (日単位)	負例数	正例数 (翌日に BPSD が生じる)	負例数/正例数
データ 1	19	最小：43 日 最大：123 日 平均：97 日	1,859	1,464	395	3.71
データ 2	30	最小：22 日 最大：123 日 平均：81 日	2,459	1,982	477	4.16

表 4.3 SudachiPy による介護記録文書の統計量

F-SOAIP の種別	データ	文書数	語彙数			単語数		
			A	B	C	A	B	C
O	データ 1	12833	1685	1689	1689	642058	636781	636028
	データ 2	21317	1823	1827	1827	962076	954183	953128
S	データ 1	1793	981	983	983	57928	57722	57691
	データ 2	3003	1195	1197	1197	105401	105022	104952

## 4.2 実験設定

本実験では4つの実験を行った。1つ目はセンサーデータのための手法選択として機械学習手法である RandomForest, SVM, LogisticRegression と深層学習手法である LSTNet, TransformerEncoder で BPSD の予測を分単位で行い、深層学習手法の方が予測精度が高いことを示す。2つ目は提案手法の評価実験として、センサーデータに介護記録文書を考慮するために提案手法の LSTNet+LSTM 及び TransformerEncoder+TransformerEncoder の Dual-Encoder モデルで BPSD の予測を分単位で行い、提案手法の方が予測精度が高いことを示す。また、介護記録文書は分単位といった細かい粒度よりも、より大きな粒度である日単位の予測で有効であると考えられ、日ごとの予測がセンサーデータの分単位の予測を補助していると考えられる。そこで3つ目に介護記録文書の有効性の分析として、介護記録文書のみで翌日に BPSD が起きるかかどうかといった日単位での予測を実験し、高い予測精度であることを示す。4つ目は最後の実験として、センサーデータの特徴量を1つ除去した場合の予測した場合に AUC がどれだけ低下するかを確認する、特徴量のアブレーション実験を行った。評価指標として AUC, accuracy, precision, recall, F1 を用いた。ただし、データセットが不均衡データであるため、AUC を重視して評価する。

本実験では5分割交差検証を行い、訓練データ：検証データ：テストデータ=6：2：2で分割した（19人データの場合は12：3：3，30人データの場合は18：6：6）。訓練データ、検証データ、テストデータはそれぞれ平均値と標準偏差を求め標準化した値を用いた。検証データは深層学習手法における early-stopping<sup>39)</sup> で使い、最大300epochとして検証データが40回改善されなかった場合に学習を終了した。

センサーデータの入力時系列の長さ（window-size）を  $w$  とし、 $w = [2, 4, 8, 16]$  と指数関数的に増やし、付録 A に示す各モデルのハイパーパラメータともにグリッドサーチを

行った。

本実験で用いるデータセット（分単位）の規模は大きく、かつ正例が少数な不均衡データである。このため、学習時間の長期化や precision, recall, F1 が 0 になりモデル間を比較する指標が少なくなる問題がある。そこで訓練データの負例の数を減らす under sampling を行った。モデル間を比較するために、訓練データの負例を減らす割合を 200 分の 1 に統一した。一方でデータセット（日単位）は不均衡の度合いが小さいので、介護記録文書のみを用いた日単位の BPSD の予測では under-sampling を行わなかった。

また、各モデルについて以下の実験設定を行った。

- SVM で AUC を求める際には probability calibration として訓練データについて分離超平面からの符号付き距離についてロジスティック回帰で学習し、確率値を推定した。同様の手法として Platt Scaling<sup>40)</sup> があり、広く知られている一般的な方法である。
- doc2vec では表記ゆれに頑健なトークナイザーである SudachiPy（辞書は sudachi-dict\_core を使用）の C モードを用いて分かち書きを行い、DM, DBoW とともに比較した。事前学習では全文書なかで 10 回以下の出現する語彙を無視し、コンテキストサイズを 8, epoch 数を 30 とし、文書ベクトルの次元数  $d_{doc} = [16, 32, 64, 128]$  で確率的勾配降下法で学習し、BPSD の予測においてグリッドサーチを行った。
- BERT ではモデル及びトークナイザーとして Hugging Face 社の Transformers の cl-tohoku/bert-base-japanese-whole-word-masking を用い、fine-tuning の有無で比較した。fine-tuning する際は分類タスクを行い、検証データを用いて最大 30epoch とし 10 回改善されなくなるまで AdamW<sup>41)</sup> で学習した。BERT は事前学習モデルが 768 次元であるため、768 次元固定の文書ベクトルが推定される。
- 事前学習及び BPSD の予測において、考慮する介護記録文書は生活支援記録法 F-SOAIP による分類の内、数が多くかつ被介護者の直接的な情報が記述されている「O」「S」のみを使用した。昨日以前の直近の文書ベクトルを考慮する数  $w_{doc} = [2, 4, 8, 16]$  としてグリッドサーチし、学習には Adam<sup>42)</sup> を用いた。

### 4.3 実験結果及び考察

本研究ではデータ 1(19 人)とデータ 2(30 人)に対し、AUC, Accuracy, precision, recall, F1 で評価した。ただし 5 分割交差検証による評価指標の平均を表示するため、precision,

recall から算出される F1 が必ずしも一致していないことに注意されたい。

### 4.3.1 センサーデータのための手法選択

表 4.4 センサーデータのみを用いた BPSD 予測結果

手法	データ 1					データ 2				
	auc	acc	precision	recall	F1	auc	acc	precision	recall	F1
RandomForest	0.6387	<b>0.9847</b>	<b>0.0091</b>	0.0485	<b>0.0128</b>	0.6478	<b>0.9800</b>	0.0053	0.0369	0.0088
SVM	0.6586	0.8362	0.0056	0.3577	0.0108	0.6345	0.8433	0.0046	0.3032	0.0090
LogisticRegression	0.6627	0.8848	0.0051	0.2418	0.0091	0.6813	0.9190	<b>0.0064</b>	0.1995	<b>0.0123</b>
LSTNet	0.6730	0.8350	0.0057	0.3533	0.0106	0.6805	0.8703	<b>0.0064</b>	0.2873	<b>0.0123</b>
TransformerEncoder	<b>0.6817</b>	0.7595	0.0048	<b>0.4829</b>	0.0095	<b>0.6833</b>	0.7969	0.0051	<b>0.4088</b>	0.0100

表 4.5 TransformerEncoder による window-size  $w$  の変化による精度変化

$w$	データ 1					データ 2				
	auc	acc	precision	recall	F1	auc	acc	precision	recall	F1
2	0.6720	<b>0.8227</b>	0.0049	0.3533	0.0087	0.6693	<b>0.8348</b>	0.0048	0.3230	0.0095
4	0.6762	0.7725	<b>0.0050</b>	0.4521	<b>0.0097</b>	0.6711	0.8130	<b>0.0054</b>	0.3818	<b>0.0106</b>
8	<b>0.6817</b>	0.7595	0.0048	<b>0.4829</b>	0.0095	<b>0.6833</b>	0.7969	0.0051	0.4088	0.0100
16	0.6630	0.7623	0.0046	0.4647	0.0090	0.6608	0.7466	0.0044	<b>0.4432</b>	0.0087

センサーデータのみを用いた BPSD 予測精度の比較結果を表 4.4 に示す。RandomForest は Accuracy が高い一方で recall が低く負例のみを予測しており、データ 1 では比較手法の中で最も AUC が低くなった。SVM 及び LogisticRegression はデータ 1 では深層学習手法 (LSTNet, TransformerEncoder) に劣るが、データ 2 では LogisticRegression は LSTNet と precision, F1 が同程度で、AUC については LogisticRegression が LSTNet より上回った。これは LSTNet は RNN であるため過去の情報を忘却するが、数分前という短期的なセンサーデータの入力についてはデータの忘却が不適切であることを示している。これに対し、過去の忘却を組み込んでいない attention 機構をもつ TransformerEncoder がデータ 1 とデータ 2 で AUC 及び recall が最も高い値を示した。

表 4.5 より、グリッドサーチで最適化された TransformerEncoder の入力時系列の長さ (window-size)  $w$  は  $w = 8$  であり、BPSD を予測するうえで過去の情報は短く過ぎても長すぎても AUC が低下することがわかった。またデータ 2 においては  $w$  が大きいほど Accuracy が低く、recall が高い。これは  $w$  が大きいほど BPSD の予兆のダイナミクスを捉えられる一方で、負例を誤って正例と分類されやすくなっていると考えられる。

## 4.3.2 提案手法の評価実験

表 4.6 介護記録文書を考慮した提案手法による分単位の BPSD 予測結果

手法	データ 1					データ 2				
	auc	acc	precision	recall	F1	auc	acc	precision	recall	F1
LSTNet	<b>0.6730</b>	0.8350	0.0057	<b>0.3533</b>	0.0106	<b>0.6805</b>	0.8703	<b>0.0064</b>	<b>0.2873</b>	<b>0.0123</b>
+LSTM (doc2Vec, DBoW)	0.6344	0.8739	0.0044	0.2134	0.0084	0.6735	0.8904	0.0053	0.2187	0.0103
+LSTM (doc2Vec, DM)	0.6616	0.8524	0.0050	0.3125	0.0094	0.6642	0.8902	0.0051	0.2304	0.0099
+LSTM (bert)	0.6622	0.8828	<b>0.0068</b>	0.2719	<b>0.0116</b>	0.6645	<b>0.9594</b>	0.0062	0.096	0.0114
+LSTM (bert_finetuning)	0.6638	<b>0.9363</b>	0.0065	0.1783	<b>0.0116</b>	0.6478	0.9589	0.0045	0.0772	0.0084
TransformerEncoder	0.6817	0.7595	0.0048	<b>0.4829</b>	0.0095	0.6833	0.7969	0.0051	0.4088	0.0100
+TransformerEncoder (doc2Vec, DBoW)	<b>0.7019</b>	0.8053	<b>0.0051</b>	0.4133	<b>0.0099</b>	<b>0.7084</b>	<b>0.8453</b>	<b>0.0060</b>	0.3417	<b>0.0116</b>
+TransformerEncoder (doc2Vec, DM)	0.6816	<b>0.8257</b>	0.0050	0.3746	0.0097	0.6914	0.8059	0.0054	0.4206	0.0106
+TransformerEncoder (bert)	0.6890	0.7965	0.0044	0.4040	0.0086	0.6813	0.7799	0.0048	0.4033	0.0094
+TransformerEncoder (bert_finetuning)	0.6845	0.7725	0.0045	0.4492	0.0084	0.6872	0.7471	0.0051	<b>0.4678</b>	0.0100

表 4.7 +TransformerEncoder(doc2vec,DBoW) における  $w_{doc}$  の変化による精度変化

$w_{doc}$	データ 1					データ 2				
	auc	acc	precision	recall	F1	auc	acc	precision	recall	F1
2	0.6859	0.8023	0.0051	<b>0.4213</b>	0.0100	0.6863	0.8259	0.0055	<b>0.3797</b>	0.0107
4	<b>0.7019</b>	0.8053	0.0051	0.4133	0.0099	0.7066	0.8450	0.0061	0.3786	0.0121
8	0.6926	0.8139	<b>0.0054</b>	0.4111	<b>0.0105</b>	<b>0.7084</b>	0.8453	0.0060	0.3417	0.0116
16	0.6347	<b>0.8198</b>	0.0042	0.3534	0.0080	0.6941	<b>0.8599</b>	<b>0.0067</b>	0.3109	<b>0.0125</b>

介護記録文書を考慮した提案手法による分単位の BPSD 予測精度の結果を表 4.6 に示す。+の後に追加された手法は介護記録文書の時系列を考慮する Encoder を示し、括弧内は文書ベクトルの推定手法である。

TransformerEncoder に doc2vec の DBoW を考慮すると、データ 1 とデータ 2 の両方において最も AUC が高くなった。この理由については次の 4.3.3 節での介護記録文書のみを用いた実験で議論するように、語彙数が少なく簡潔で短い文書が多いため低次元の

文書ベクトルでかつ最もシンプルな手法である DBoW が最も AUC を向上させると考えられる。

また，介護記録文書を考慮する数  $w_{doc}$  を変化による精度変化を表 4.7 に示す．データ 1 とデータ 2 で， $w_{doc}$  が大きくなると Accuracy が高く，recall が低い．これは介護記録文書を考慮することで，誤って BPSD と予測される場合を軽減していることを示す．ただし，BPSD の正例の予測については昨日以前の介護記録文書よりも，数分前のセンサーデータの方に直接的なダイナミクスが現れるので文書を考慮しない場合よりも recall が低下すると考えられる．

### 4.3.3 介護記録文書の有効性の分析

表 4.8 介護記録文書のみを用いた日単位での BPSD 予測結果

手法	データ 1					データ 2				
	auc	acc	precision	recall	F1	auc	acc	precision	recall	F1
LSTM (doc2vec,DBoW)	<b>0.6877</b>	0.7697	0.3876	0.1516	0.1656	<b>0.6898</b>	0.7728	<b>0.4526</b>	<b>0.0821</b>	<b>0.0817</b>
LSTM (doc2vec,DM)	0.6330	0.7593	<b>0.4064</b>	<b>0.1587</b>	<b>0.1823</b>	0.6462	0.7873	0.3113	0.0580	0.0752
LSTM (bert)	0.6247	0.7616	0.0485	0.0327	0.0390	0.6592	0.7894	0.0494	0.0384	0.0432
LSTM (bert_finetuning)	0.6314	<b>0.7789</b>	0.3333	0.0313	0.0570	0.6524	<b>0.8037</b>	0.2333	0.0057	0.0112
TransformerEncoder (doc2vec,DBoW)	<b>0.6958</b>	0.7335	0.3092	<b>0.3734</b>	<b>0.3010</b>	<b>0.7024</b>	0.7719	0.4096	<b>0.2857</b>	<b>0.3105</b>
TransformerEncoder (doc2vec,DM)	0.6445	0.7371	<b>0.4373</b>	0.1813	0.1820	0.6698	<b>0.7981</b>	<b>0.5014</b>	0.0481	0.0819
TransformerEncoder (bert)	0.6151	0.7056	0.2518	0.3524	0.2794	0.6633	0.7820	0.3420	0.0870	0.0918
TransformerEncoder (bert_finetuning)	0.6214	<b>0.7601</b>	0.2850	0.2805	0.2410	0.6732	0.7167	0.2181	0.2452	0.2008

介護記録文書を考慮した提案手法による日単位の BPSD 予測精度の結果を表 4.8 に示す．手法の括弧内は文書ベクトルの推定手法である．

LSTM と TransformerEncoder とともに doc2vec の DBoW で推定された文書ベクトルを入力した時が最も AUC が高くなった．特に TransformerEncoder に DBoW のベクトルを入れた場合が最も recall が高くなり，F1 も最も高くなった．この理由として，介護記録文書は生活支援記録法 F-SOAIP に則って内容ごとに分類され，主観的情報「O」には相槌や挨拶，客観的情報「S」には「消灯」「トイレ」「入眠中」など些細なことも詳細に記録し



表 4.9 TransformerEncoder(doc2vec, DBoW) による日単位での BPSD 予測における  $d_{doc}$  の変化

$d_{doc}$	データ 1					データ 2				
	auc	acc	precision	recall	F1	auc	acc	precision	recall	F1
16	0.5380	0.7434	0.2445	0.2147	0.2063	0.6779	0.7557	0.3204	0.2439	0.2506
32	0.5986	<b>0.7646</b>	<b>0.5819</b>	0.2280	0.2693	0.6384	<b>0.7949</b>	<b>0.4619</b>	0.2343	0.2876
64	<b>0.6958</b>	0.7335	0.3092	<b>0.3734</b>	<b>0.3010</b>	<b>0.7024</b>	0.7719	0.4096	<b>0.2857</b>	<b>0.3105</b>
128	0.6414	0.7299	0.4634	0.2999	0.2572	0.5802	0.7605	0.3584	0.1862	0.2142

である。このため LSTM は、過去の情報を忘却する性質をもつため、多くの介護記録文書を用いても、BPSD の予兆に重要でない情報が混在することによって、本来予測に必要なデータを忘却している可能性が高いと考えられる。一方で TransformerEncoder は BPSD の予兆を忘却せずに recall が高くなったと考えられる。

また、TransformerEncoder(doc2vec, DBoW) における文書ベクトルの次元  $d_{doc}$  の変化を表 4.9 に示す。これより、データ 1 とデータ 2 の両方で  $d_{doc} = 64$  で最も AUC が大きくなることが分かり、それ以上の次元では AUC が低下することが分かる。また、介護記録文書の語彙数は表 4.3 より語彙数が 2000 以下と少なく、高次元の文書ベクトルで AUC が低下する原因だと考えられる。BERT は固定の 768 次元の高次元の文書ベクトルを推定するため、BERT よりも doc2vec の DBoW の方が AUC が高くなったと考えられる。さらに、生活支援記録法 F-SOAIP に則って簡潔に短く記述されている場合が多いので、DM よりも単語の有無のみを考慮するシンプルな文書ベクトルの推定法である DBoW の方が精度が高くなったと考えられる。

#### 4.3.4 特徴量のアブレーション実験

AUC が最も高くなった+TransformerEncoder(doc2vec,DBoW) モデルについて、入力するセンサー値を 1 つ除去（アブレーション）して、精度が下がる度合いを表 4.10 に示す。除去された特徴量を「マイナス特徴量名」で表し、除去をする前の評価指標との差で表す。また、赤字が最大値、青字が最小値を表す。AUC の昇順で表しているため、表の上部の特徴量ほど重要であることを表す。

表 4.10 より、データ 1、データ 2 の両方で騒音 (sound) と睡眠状態 (SleepWaveform-Value) を除去すると AUC が低下しており、BPSD の予測に特に重要な特徴量であることがわかる。また recall も低下する一方で precision はほとんど変わらないことから、BPSD の予兆を表す一方で BPSD でない時との差別化が困難であることを表している。例えば、

騒音がある場合に BPSD が起こることもあれば、必ずしも BPSD が起こるとは限らないことを示している。

また、データ 1, データ 2 の両方で気圧 (pressure) や心拍数 (HeartRate), 呼吸数 (RespirationRate) は除去すると AUC が高くなった。特に気圧はデータ 2 では除去した方がすべての指標が高くなるため、BPSD を予測する要因ではないと考えられる。心拍数・呼吸数に関しては被介護者によって異なる場合が多いため、テストデータを別の被介護者とする本研究の実験設定では除去すると recall が下がるものの、その他の指標においては上がったと考えられる。

表 4.10 +TransformerEncoder(doc2vec,DBoW) における特徴量のアブレーション

手法	データ 1					手法	データ 2				
	auc	acc	precision	recall	F1		auc	acc	precision	recall	F1
TransformerEncoder +TransformerEncoder (doc2Vec, DBoW)	0.7019	0.8053	0.0051	0.4133	0.0099	TransformerEncoder +TransformerEncoder (doc2Vec, DBoW)	0.7084	0.8453	0.0060	0.3417	0.0116
-SleepWaveformValue	-0.0353	+0.0532	0.0000	-0.1472	0.0000	-temperature	-0.1787	+0.0132	-0.0009	-0.0756	-0.0017
-sound	-0.0239	+0.0592	+0.0003	-0.1318	+0.0006	-sound	-0.0304	+0.0192	-0.0006	-0.0602	-0.0011
-humidity	-0.0041	+0.0729	+0.0008	-0.1488	+0.0014	-SleepWaveformValue	-0.0106	+0.0329	-0.0001	-0.0772	-0.0003
-pressure	+0.0030	+0.0772	+0.0013	-0.1275	+0.0025	-light	-0.0035	+0.0372	+0.0004	-0.0559	+0.0008
-temperature	+0.0057	+0.0657	+0.0008	-0.1215	+0.0015	-CO2	-0.0008	+0.0257	-0.0001	-0.0499	-0.0002
-HeartRate	+0.0071	+0.0775	+0.0013	-0.1204	+0.0025	-RespirationRate	+0.0006	+0.0375	+0.0004	-0.0488	+0.0008
-CO2	+0.0076	+0.0665	+0.0013	-0.0966	+0.0025	-humidity	+0.0011	+0.0265	+0.0004	-0.0250	+0.0008
-light	+0.0101	+0.0540	+0.0014	-0.0645	+0.0028	-pressure	+0.0036	+0.0140	+0.0005	+0.0071	+0.0011
-RespirationRate	+0.0104	+0.0697	+0.0018	-0.0864	+0.0036	-HeartRate	+0.0039	+0.0297	+0.0009	-0.0148	+0.0019

## 5 結論

本研究では、IoT デバイスで取得されるセンサーデータによる情報と、介護記録文書によるセンサーデータでは捉えられない情報の 2 つの情報を考慮しながら、BPSD を自動予測するための Dual-Encoder モデルを提案した。

既存研究では、被介護者ごとの体調や様子、過去の行動履歴など、センサーデータでは捉えられてない介護者の観点のコンテキスト情報を考慮できないという問題点があった。

そこで本研究では、深層学習手法を用いることで介護記録文書を文書ベクトルで表現し、センサーデータの時系列を考慮する Encoder と介護記録文書の時系列を考慮する Encoder の 2 つを用いた Dual-Encoder モデルを提案した。提案手法は介護記録文書によるコンテキスト情報を考慮することで被介護者の潜在的な BPSD の起こりやすさを考慮し、precision を高くすることができる。

評価実験では、センサーデータに介護記録文書を加えて深層学習することで予測精度が向上することを示した。また、センサーデータの入力時系列長を大きくすると recall が高くなり、介護記録文書を入力する時系列長を大きくすると precision が高くなることを示した。この結果より、介護記録文書による介護者の観点からの情報は BPSD の予測に非常に有効であることがわかった。

本研究の主な有用性は 1) センサーデータによる情報のみならず、介護記録文書によるセンサーデータでは捉えられない情報を加味した。2) BPSD の種類を限定しない、包括的 (幻聴, 妄想, 暴言, 暴行・暴力, 周俳, 無断外出・夕暮れ症候群, 性的不適切行動・逸脱行為, 常同行為, 脱抑制, 易怒声・興奮, 収集, 大声, うつ, アパシー, 無反応・無関心, 不安, 不潔行為, 介護への抵抗・拒否, その他) な予測を可能にした。

今後の課題として、BPSD の種別ごとの予測や不均衡データのアプローチを用いた精度向上が挙げられる。

## 謝辞

研究室に配属されてからこれまで多大なるご指導，ご協力いただきました植野真臣教授に深く感謝致します。研究面のみならず，基礎知識を身に着けるための習慣や思う存分研究に打ち込める環境を与えてくれました。共同研究に置きましたは，南泰浩教授，内海彰教授には多くの助言をして頂いたことに深く感謝いたします。またデータを収集及び提供していただいた国立研究開発法人日本医療研究開発機構 (Japan Agency for Medical Research and Development;AMED) の専門家及び介護施設関係者の皆様には深く感謝いたします。副指導教官である川野秀一准教授には，論文の推敲に辺り助言を頂いたことに深く感謝いたします。そして植野研究室の皆様には深く感謝いたします。偉大な先輩方，ひたむきな後輩達から刺激を受けていました。そして何より同期には，研究の相談やアドバイス，日常生活面においても支えられました。このような環境に身を置くことができたことに感謝いたします。

また，本研究は「BPSD 予測・予防により介護負担を軽減する認知症対応型 AI・IoT サービスの開発と実装」プロジェクトの一環であり。一部の実験は，電気通信大学 人工知能先端研究センター (AIX) の計算機を利用して実施したものです。

## 使用したツール

機械学習手法 (SVM, RF, LR) には scikit-learn の実装を用いた。

[1] scikit-learn

<https://scikit-learn.org/stable/>

介護記録文書の前処理には neologdn を用いた。

[1] neologdn

<https://github.com/ikegami-yukino/neologdn>

Doc2Vec は gensim の実装を用い、トークナイザーは表記ゆれに強い SudachiPy を用いた。

[1] gensim

<https://radimrehurek.com/gensim/>

[2] SudachiPy

<https://github.com/WorksApplications/SudachiPy>

BERT のモデル及びトークナイザーは Hugging Face 社の Transformers の cl-tohoku/bert-base-japanese-whole-word-masking を用いた。

[1] Transformers (cl-tohoku/bert-base-japanese-whole-word-masking)

<https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

本実験は BERT では pytorch, それ以外は tensorflow を用いて Docker Hub にある以下のリポジトリ及びタグで実験を行った。

[1] pytorch/pytorch:1.5-cuda10.1-cudnn7-devel

[2] tensorflow/tensorflow:2.0.0-gpu-py3-jupyter

# 付録

## A ハイパーパラメータの設定

センサーデータのみを用いた手法と文書ベクトル推定手法のハイパーパラメータを表 A.1 に示す。また、提案手法のセンサーデータに介護記録文書を考慮した BPSD 予測モデルについて、最も AUC が高くなった TransformerEncoder+TransformerEncoder(doc2Vec, DBoW) のハイパーパラメータを表 A.2 に示す。表に記載されていないパラメータは各ツールのデフォルト設定を用いている。

表 A.1 設定されたハイパーパラメータ値

手法	パラメータ名	グリッドサーチの範囲	固定パラメータ	ベストパラメータ	
				データ 1	データ 2
RandomForest	n_estimators	64,128,256		128	256
	bootstrap	True, False		True	True
	min_samples_leaf	1, 3, 5, 7		1	3
	criterion		gini	gini	gini
	w	2,4,8,16		4	4
svm	C	1,0.1, 0.01, 0.001, 0.0001		0.0001	0.0001
	kernel		rbf	rbf	rbf
	gamma		scale	scale	scale
	w	2,4,8,16		4	8
LogisticRegression	C	1,0.1, 0.01, 0.001, 0.0001		0.001	0.001
	penalty		L2	L2	L2
	solver		newton-cg	newton-cg	newton-cg
	w	2,4,8,16		4	4
LSTNet	CNNKernels	0,1,2,3,4		1	2
	skips	0,1,2,3,4		3	2
	highways	0,1,2,3,4		1	2
	GRUUnits	32,64,128		32	128
	SkipGRUUnits	2,4,8		2	8
	dropouts	0.3,0.4,0.5		0.5	0.5
	CNNFilters	32,64,128		32	32
	bs		2048	2048	2048
Transformer	w	2,4,8,16		4	2
	hopping_nums	1,2,3		1	2
	hidden_dims	32, 64, 128		32	32
	dropouts	0.3,0.4,0.5		0.5	0.5
	bs		2048	2048	2048
Doc2Vec	w	2,4,8,16		8	8
	window		8	8	8
	min_count		10	10	10
BERT	workers		4	4	4
	bs		32	32	32
BERT	lr		2.00E-05	2.00E-05	2.00E-05

表 A.2 提案手法の設定されたハイパーパラメータ値

手法	パラメータ名	グリッドサーチの範囲	固定パラメータ	ベストパラメータ	
				データ1	データ2
TransformerEncoder +TransformerEncoder (doc2Vec, DBoW)	hopping_mums	1,2,3		1	2
	hidden_dims	32, 64, 128		32	32
	dropouts	0.3,0.4,0.5		0.5	0.5
	bs		2048	2048	2048
	w	2,4,8,16		8	8
	w_doc	2,4,8,16		4	8
	d_doc	16,32,64,128(doc2vec)	769(BERT)		32

## B under-sampling の割合と精度

本研究で用いるデータは表 4.1 に示すように、不均衡データである。このため、すべてのデータを用いた場合、AUC, Accuracy を除く precision, recall, F1 が 0 になり、モデル間で比較が困難になる。そのため本研究では under sampling という訓練データの負例をランダムにサンプリングして減らす手法を用いた。モデル間を比較するために負例のデータを 200 分の 1 ( $u = 0.005$ ) に統一してサンプリングした。

ここでは、データ 1 において、サンプリングされた負例の割合  $u$  を変化した場合の精度の変化を、AUC が最も高くなった+TransformerEncoder(doc2vec,DBoW) モデルについて表 B.1 に示す。

表 4.1 より、 $u = 0.005$  で AUC が最も高くなり、 $u$  が高すぎても低すぎても AUC が低下することが分かった。また precision と recall のトレードオフの関係性を確認し、実現場での需要に答えて柔軟に対応できることを示した。データ 2 でも同様の結果が得られた。

表 B.1 +TransformerEncoder(doc2vec,DBoW)における  $u$  の変化による精度変化

u	データ 1				
	auc	acc	precision	recall	F1
0.010	0.6705	0.9017	0.0059	0.2168	0.0111
0.009	0.6888	<b>0.9081</b>	<b>0.0074</b>	0.2598	<b>0.0138</b>
0.008	0.6808	0.8951	0.0067	0.2595	0.0116
0.007	0.6866	0.8610	0.0061	0.3485	0.0119
0.006	0.6888	0.8199	0.0051	0.3891	0.0099
0.005	<b>0.7019</b>	0.8053	0.0051	0.4133	0.0099
0.004	0.6838	0.7750	0.0044	0.4184	0.0086
0.003	0.6869	0.6859	0.0045	0.5766	0.0089
0.002	0.6944	0.5503	0.0041	0.7175	0.0082
0.001	0.6762	0.3377	0.0035	<b>0.8505</b>	0.0069

## C Precision-Recall Curve

表 4.6 より, AUC が最も高くなった+TransformerEncoder(doc2vec,DBoW) モデルの precision はデータ 1 で 0.0051, データ 2 で 0.0060 であった. また表 B.1 より, under-sampling の割合  $u$  を変化させても, precision は 0.0074 が最大であり, 低いことがわかる.

そこで, +TransformerEncoder(doc2vec,DBoW) モデルについて 5 分割交差検証の各モデルのテストデータについて PR 曲線 (Precision-Recall Curve) を図示し, 正例と負例を判別する予測確率の閾値を変更した場合の precision と recall の関係について確認した. データ 1 の PR 曲線を図 C.1 に, データ 2 の PR 曲線を図 C.2 に示す. 交差検証の  $n$  番目のテストデータに対する PR 曲線を「CV $n$ 」と表している.

図 C.1 より, CV3 において precision=0.015 で recall=0.1, precision=0.010 で recall=0.200 となっている. 図 C.2 では, precision=0.015 以上になる閾値はない一方で, CV5 では precision=0.010 で recall=0.500 となった. しかし, precision が低いモデルも存在し, 人数が少ないため, 訓練データとテストデータの乖離が確認できる.



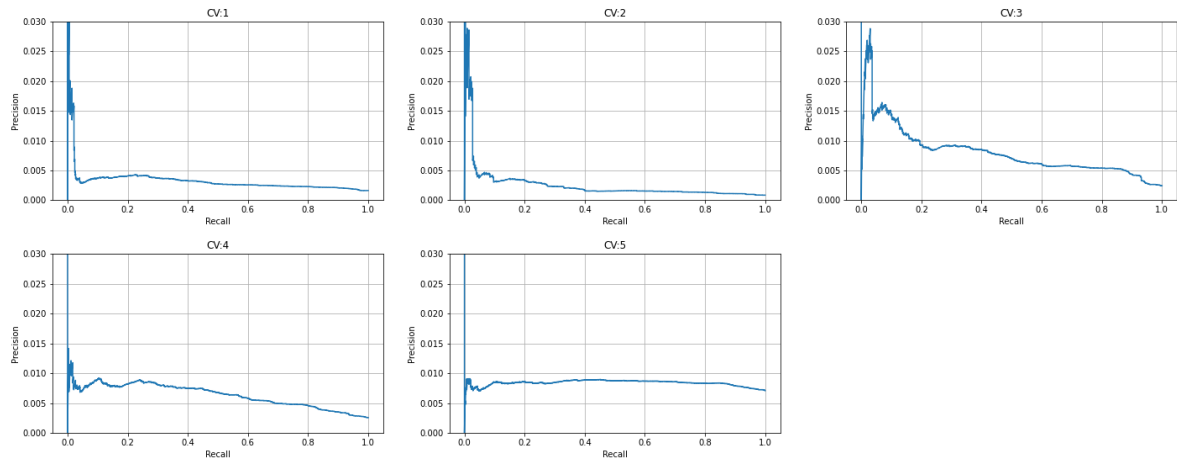


図 C.1 +TransformerEncoder(doc2vec,DBoW) モデルにおけるデータ 1 の PR 曲線

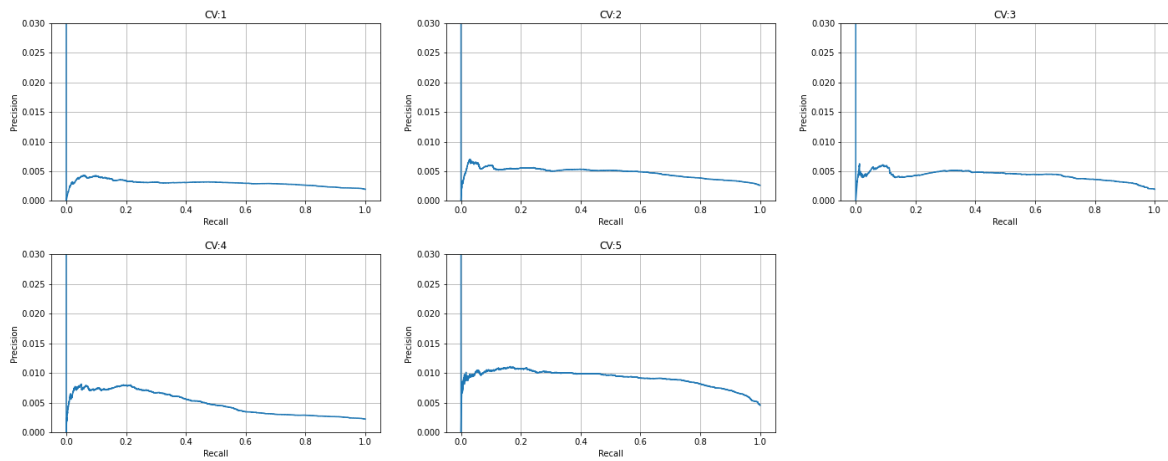


図 C.2 +TransformerEncoder(doc2vec,DBoW) モデルにおけるデータ 2 の PR 曲線

## 参考文献

- 1) 横川慎二. データ基盤整備と応用. BPSD 予測・予防により介護負担を軽減する認知症対応型 AI・IoT サービスの開発と実装第 2 回外部評価委員会, 2021 年 3 月 19 日.
- 2) 総務省統計局. 統計トピックス no.129 統計からみた我が国の高齢者 - 「敬老の日」にちなんで -. <https://www.stat.go.jp/data/topics/pdf/topics129.pdf>, 2021-09-19.
- 3) 内閣府. 平成 29 年版高齢社会白書. [https://www8.cao.go.jp/kourei/whitepaper/w-2017/html/gaiyou/s1\\_2\\_3.html](https://www8.cao.go.jp/kourei/whitepaper/w-2017/html/gaiyou/s1_2_3.html), 2017.
- 4) Sanford I. Finkel, Jorge Costa e Silva, Gene Cohen, Sheldon Miller, and Norman Sartorius. Behavioral and Psychological Signs and Symptoms of Dementia: A Consensus Statement on Current Knowledge and Implications for Research and Treatment. *International Psychogeriatrics*, Vol. 8, No. S3, pp. 497–500, May 1997. Publisher: Cambridge University Press.
- 5) 山口晴保. BPSD の定義、その症状と発症要因. 認知症ケア研究誌, Vol. 2, pp. 1–16, 2018.
- 6) Zahinoor Ismail, Luis Agüera-Ortiz, Henry Brodaty, Alicja Cieslak, Jeffrey Cummings, Corinne E. Fischer, Serge Gauthier, Yonas E. Geda, Nathan Herrmann, Jamila Kanji, Krista L. Lanctot, David S. Miller, Moyra E. Mortby, Chiadi U. Onyike, Paul B. Rosenberg, Eric E. Smith, Gwenn S. Smith, David L. Sultzer, and Constantine Lyketsos. The Mild Behavioral Impairment Checklist (MBI-C): A rating scale for neuropsychiatric symptoms in pre-dementia populations. *Journal of Alzheimer's disease : JAD*, Vol. 56, No. 3, pp. 929–938, 2017.
- 7) 藤生大我, 内藤典子, 滝口優子, 伊東美緒, 山上徹也, 山口晴保. BPSD 予防をめざした「BPSD 気づき質問票 57 項目版 (BPSD-NQ57)」の開発. 認知症ケア研究誌, Vol. 3, pp. 24–37, 2019.
- 8) Laura M. Ghali, Robert W. Hopkins, and Peter Rindlisbacher. Temporal shifts in peak daily activity in Alzheimer's disease. *International Jour-*

- nal of Geriatric Psychiatry*, Vol. 10, No. 6, pp. 517–521, 1995. .eprint:  
<https://onlinelibrary.wiley.com/doi/pdf/10.1002/gps.930100612>.
- 9) Guy Nagels, Sebastiaan Engelborghs, Ellen Vloeberghs, Debby Van Dam, Barbara A. Pickut, and Peter Paul De Deyn. Actigraphic measurement of agitated behaviour in dementia. *International Journal of Geriatric Psychiatry*, Vol. 21, No. 4, pp. 388–393, April 2006.
  - 10) LuAnn Etcher, Ann Whall, Rahul Kumar, Davangere Devanand, and Vikram Yera-gani. Nonlinear indices of circadian changes in individuals with dementia and aggression. *Psychiatry Research*, Vol. 199, No. 1, pp. 77–78, August 2012.
  - 11) Azziza Bankole, Martha Anderson, Tonya Smith-Jackson, Aubrey Knight, Kyunghui Oh, Jeff Brantley, Adam Barth, and John Lach. Validation of Noninvasive Body Sensor Network Technology in the Detection of Agitation in Dementia. *American Journal of Alzheimer's Disease & Other Dementias*® , Vol. 27, No. 5, pp. 346–354, August 2012. Publisher: SAGE Publications Inc.
  - 12) Amber Knuff, Roxanne H. Leung, Dallas P. Seitz, Luljeta Pallaveshi, and Amer M. Burhan. Use of Actigraphy to Measure Symptoms of Agitation in Dementia. *The American Journal of Geriatric Psychiatry*, Vol. 27, No. 8, pp. 865–869, August 2019.
  - 13) Jiaqi Gong, Karen Moomaw Rose, Ifat Afrin Emi, Janet P. Specht, Enamul Hoque, Dawei Fan, Sriram Raju Dandu, Robert F. Dickerson, Yelena Perkhounkova, John Lach, and John A. Stankovic. Home wireless sensing system for monitoring nighttime agitation and incontinence in patients with Alzheimer's disease. In *Proceedings of the conference on Wireless Health*, pp. 1–8, Bethesda Maryland, October 2015. ACM.
  - 14) L. Valembois, C. Oasi, S. Pariel, W. Jarzebowski, Carmelo Lafuente-Lafuente, and Joel Belmin. Wrist Actigraphy: A Simple Way to Record Motor Activity in Elderly Patients with Dementia and Apathy or Aberrant Motor Behavior. *The journal of nutrition, health & aging*, Vol. 19, , May 2015.
  - 15) Basel Kikhia, Thanos G. Stavropoulos, Stelios Andreadis, Niklas Karvonen, Ioannis Kompatsiaris, Stefan Sävenstedt, Marten Pijl, and Catharina Melander. Utilizing a Wristband Sensor to Measure the Stress Level for People with Dementia. *Sensors (Basel, Switzerland)*, Vol. 16, No. 12, p. 1989, November 2016.

- 16) Christianne Nesbitt, Ajay Gupta, Shubham Jain, Kurt Maly, and Hamid Reza Okhravi. Reliability of Wearable Sensors to Detect Agitation in Patients with Dementia: A Pilot Study. In *Proceedings of the 2018 10th International Conference on Bioinformatics and Biomedical Technology, ICBBT '18*, pp. 73–77, New York, NY, USA, May 2018. Association for Computing Machinery.
- 17) Shehroz S. Khan, Tong Zhu, Bing Ye, Alex Mihailidis, Andrea Iaboni, Kristine Newman, Angel He Wang, and Lori Schindel Martin. DAAD: A Framework for Detecting Agitation and Aggression in People Living with Dementia Using a Novel Multi-modal Sensor Network. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 703–710, November 2017. ISSN: 2375-9259.
- 18) Shehroz S. Khan, Sofija Spasojevic, Jacob Nogas, Bing Ye, Alex Mihailidis, Andrea Iaboni, Angel Wang, Lori Schindel Martin, and Kristine Newman. Agitation Detection in People Living with Dementia using Multimodal Sensors. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3588–3591, Berlin, Germany, July 2019. IEEE.
- 19) J. L. Cummings, M. Mega, K. Gray, S. Rosenberg-Thompson, D. A. Carusi, and J. Gornbein. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology*, Vol. 44, No. 12, pp. 2308–2314, December 1994.
- 20) Jiska Cohen-mansfield, Marcia S. Marx, and Alvin S. Rosenthal. A Description of Agitation in a Nursing Home. *Journal of Gerontology*, Vol. 44, No. 3, pp. M77–M84, May 1989.
- 21) Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273–297, September 1995.
- 22) Leo Breiman. Random Forests. *Machine Learning*, Vol. 45, No. 1, pp. 5–32, October 2001.
- 23) 山上徹也, 山口晴保. 情報通信技術を活用したBPSD 予防・予測システム開発の現状と課題. *認知症ケア研究誌*, Vol. 5, pp. 8–15, 2021.

- 24) Ronald J. Williams and David Zipser. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, Vol. 1, No. 2, pp. 270–280, June 1989.
- 25) Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18*, pp. 95–104, New York, NY, USA, June 2018. Association for Computing Machinery.
- 26) Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- 27) Yoon Kim, Carl Denton, Luong Hoang, and Alexander M. Rush. Structured Attention Networks. *CoRR*, Vol. abs/1702.00887, , 2017. arXiv: 1702.00887.
- 28) Thang Luong, Hieu Pham, and Christopher D. Manning. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412–1421, Lisbon, Portugal, September 2015. Association for Computational Linguistics.
- 29) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., 2017.
- 30) A. Rumshisky, M. Ghassemi, T. Naumann, P. Szolovits, V. M. Castro, T. H. McCoy, and R. H. Perlis. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational Psychiatry*, Vol. 6, No. 10, pp. e921–e921, October 2016. Number: 10 Publisher: Nature Publishing Group.
- 31) Maxim Topaz, Victoria Adams, Paula Wilson, Kyungmi Woo, and Miriam Ryvicker. Free-Text Documentation of Dementia Symptoms in Home Healthcare: A Natural Language Processing Study. *Gerontology and Geriatric Medicine*, Vol. 6, p. 2333721420959861, January 2020. Publisher: SAGE Publications Inc.

- 32) Ahmed M. Al-Harrasi, Ehtesham Iqbal, Konstantinos Tsamakis, Judista Lasek, Romyne Gadelrab, Pinar Soysal, Enno Kohlhoff, Dimitrios Tsipsios, Emmanouil Rizos, Gayan Perera, Dag Aarsland, Robert Stewart, and Christoph Mueller. Motor signs in Alzheimer’s disease and vascular dementia: Detection through natural language processing, co-morbid features and relationship to adverse outcomes. *Experimental Gerontology*, Vol. 146, p. 111223, April 2021.
- 33) David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, Vol. 3, No. Jan, pp. 993–1022, 2003.
- 34) Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, November 1998. Conference Name: Proceedings of the IEEE.
- 35) Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pp. 103–111, Doha, Qatar, October 2014. Association for Computational Linguistics.
- 36) Quoc Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 1188–1196. PMLR, June 2014. ISSN: 1938-7228.
- 37) Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- 38) Kazuma Takaoka, Sorami Hisamoto, Noriko Kawahara, Miho Sakamoto, Yoshitaka Uchida, and Yuji Matsumoto. Sudachi: a Japanese Tokenizer for Business. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).

- 
- 39) N. Morgan and H. Boursard. Generalization and Parameter Estimation in Feed-forward Nets: Some Experiments. In *Advances in Neural Information Processing Systems*, Vol. 2. Morgan-Kaufmann, 1990.
  - 40) John C. Platt. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*, pp. 61–74. MIT Press, 1999.
  - 41) Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.
  - 42) Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.