

福祉支援施設の支援記録からのインシデントの予兆検出手法の改良

山本 隆太[†] 上野 史^{††} 太田 学^{††}

[†] 岡山大学大学院環境生命自然科学研究科 〒700-8530 岡山市北区津島中 3-1-1

^{††} 岡山大学学術研究院環境生命自然科学学域 〒700-8530 岡山市北区津島中 3-1-1

E-mail: [†]p0pm8dxf@s.okayama-u.ac.jp, ^{††}{uwano, ohta}@okayama-u.ac.jp

あらまし 上野らは、福祉支援施設の入所者が起こすトラブル（インシデント）の予兆検出のため、施設の支援記録本文を用いた BERT によるインシデントの予兆検出手法を提案した。また実験により、支援記録に含まれる日付などの付随情報を使用することで分類精度が向上する可能性も示唆した。しかし、入所者毎に個別に予兆検出器を学習したため、インシデントをあまり起こさない入所者に対して検出精度が悪かった。そこで本研究では、他の入所者の支援記録を学習に使用した分類器による実験を行い、入所者本人の支援記録のみで学習した分類器と精度を比較する。さらに本研究では、付随情報を利用することで分類精度が向上する可能性について検討した。実験の結果、分類対象者を除く全ての入所者の支援記録で学習する汎用分類器の精度が最もよく、施設 A で平均 F 値 0.790、施設 B で平均 F 値 0.753 となった。またその汎用分類器にインシデントの有無の付随情報を利用した場合、さらに精度が向上し施設 A で平均 F 値 0.825、施設 B で平均 F 値 0.772 となった。

キーワード インシデント予兆検出, BERT, 付随情報, 医療・ヘルスケア, 福祉支援施設

1 はじめに

近年、高齢者や障がい者が入所する福祉支援施設の需要が高まってきている。福祉支援施設では、入所者の人数に対する職員の数は限られており、支援には限度がある。そのため、入所者による器物破損や他者への傷害などといったインシデントが施設内で発生すると、業務の妨げとなり、入所者は十分な支援が受けられなくなる。しかし、インシデントを事前に察知することはベテランの職員でも難しいため、入所者が起こすインシデントの予兆を発見し未然に防止することは重要な課題である。

本研究では、福祉支援施設の支援記録からインシデントの予兆を検出する¹。本研究では、実際に現場で働く施設の職員から見て、支援記録本文に利用者の衝動性が高い状態、あるいはネガティブな心理状態が現れているものをインシデントの予兆と定義している。本研究で利用する手法は、上野らが [1] で提案したものであり、上野らは福祉支援施設の利用者 10 人の支援記録を用いて、各入所者のデータのみで個別に学習した分類器の予兆検出性能を検証した。予兆の検出は、予兆の有無を示すラベルを用いた 2 クラス分類により実現しており、その際上野らは入力として、支援記録の記録本文のみでなく支援記録に含まれる日付などの付随情報を使用することで、分類精度が向上する可能性を示唆している。本研究では、上野らの手法を予兆の程度に応じた 4 値分類に拡張し、予兆を検出する。具体的には、支援記録の本文に加えて日付データなどの付随情報も利用した分類器を作成し、分類器に入力された支援記録を「予兆なし」「予兆（弱）」「予兆（強）」「予兆（両方）」の 4 クラスに分類することで予兆の検出を実現する。4 値分類への拡張により、施設

の職員が予兆の程度に応じてより適切な対応ができるようになる。検出には、自然言語処理モデルの 1 つである Bidirectional Encoder Representations from Transformers (BERT)[2] を利用した上野らの予兆検出器を拡張して利用する。上野らの研究では、分類器の学習に分類対象となる入所者の記録しか用いておらず、作成した分類器は、他の入所者の記録分類には向かない。そのため本研究では、他の入所者の支援記録を学習に使用した分類器による実験を行い、学習に入所者本人の記録しか用いない場合と精度を比較し、分類器の汎用性について考察する。さらに本研究では、分類精度向上のため、付随情報の有効な利用方法を検討する。

本稿の構成を述べる。2 節では関連研究を紹介する。3 節では、本実験で使用する支援記録と付与されたインシデントと予兆のラベルについて説明する。4 節では、本稿で用いるインシデントの予兆検出手法を説明する。5 節で実験について説明し、結果と考察を述べる。6 節では、まとめと今後の課題について述べる。

2 関連研究

2.1 BERT

Bidirectional Encoder Representations from Transformers (BERT)[2] は、Devlin らが提案した自然言語処理モデルであり、双方向 Transformer[3] というニューラルネットワークを利用している。BERT は大量の教師なしデータによる事前学習を行ったモデルを、比較的少量の教師ありデータでファインチューニング (FT) して個々のタスクに合わせて学習する。これにより事前学習済みモデルをそのまま適用するよりも性能の向上が期待できる。

BERT の特徴として、Transformer を利用し、文章を文頭

1：本研究は、岡山大学の研究倫理審査専門委員会の承認を得て実施した。(研 2104-005)

と文末の双方向から学習することで、文脈を考慮できることが挙げられる。BERT ではこの双方向からの学習を、Masked Language Model (MLM) と Next Sentence Prediction (NSP) の2つのタスクを行うことで可能としている。MLM では、まず入力文章を単語ごとにトークンに分割し、その中からランダムに選ばれた 15 % のトークンを確率的に別のトークンで置き換える。具体的には、選択された 15 % のトークンのうち、その 80 % を [MASK] という特殊なトークンに、10 % をランダムな別のトークンに置き換え、残りの 10 % はそのままのトークンとする。その後 [MASK] トークンに置き換えられたトークンを予測し学習するのが MLM のタスクである。NSP は2つの入力文を受け取り、そのペアにおいて2つ目の文が1つ目の文の後続の文になっているかを予測するタスクである。具体的には、文の片方を 50 % の確率で他の文に置き換え、それらが隣り合っているか隣り合っていないか判別することで学習する。BERT は発表時に、General Language Understanding Evaluation (GLUE) ベンチマーク [4] の8つのタスク、Stanford Question Answering Dataset (SQuAD)[5] の v1.1 と v2.0, Situations With Adversarial Generations (SWAG)[6] の計 11 の自然言語処理タスクで最高精度を記録した。

2.2 インシデント検出

介護や福祉、医療分野におけるインシデントやトラブルへの対応は大きな医療事故や介護事故を防ぐためにも重要であり、事例の分析や予防手法の提案、支援システムの開発など様々な研究がされている。

例えば田中らは、大量に蓄積されている医療事故情報やインシデント情報をトラブルの未然防止に活用するための仕組みとして、機械学習を利用した階層型クラスタリングの方法を提案した [7]。具体的には、まず初めに日本医療機能評価機構が WEB 公開している医療事故データベースから必要情報をダウンロードして、ヒヤリハットデータベースを作成した。その後、それらに階層型クラスタリングを適用し、ヒヤリハットデータベース全体のクラスタ構造を各事例の類似性という視点で階層的に表示することで、インシデントをいくつかのグループに分類することを可能とした。このような自動分類により必要なクラスタを深堀できるなど、各医療機関でのインシデントへの事前対応が可能になることが期待できると報告されている。

峯崎らは、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って事故の起こりうる状況を分析した [8]。テキストデータ部分について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で重要な単語クラスタ及び重要な単語を抽出し、決定木学習によって可視化した。その結果、1. 出血は目に見える外傷が発見されて介護士が初めて気づく、2. 移乗などの際に転倒による打撲が発生する可能性がある、3. 食事時に誤嚥が起こりうる、4. 無断外出につながるいくつかの行動がある、などのいくつかの事故パターンを発見した。

中沢は、高齢者介護施設において、施設職員の支援や負担軽減を目的として、夜間徘徊者の発見、通報、声掛け、足止めを

総合的に実行する見守りロボットを開発した [9]。ロボットに搭載された赤外線カメラにより施設廊下を監視し、HOG 特徴量とフレーム間差分法による動体検知を併用することで徘徊者を検出した。また、その後施設職員へ無線連絡をすると同時に、ロボットが徘徊者に近づき、職員の音声で呼びかけることで職員が到着するまでの時間稼ぎが可能なシステムを構築した。

青木らは、介護 AI 記録分析ツール “けあらば”を開発した [10]。これは同社の開発した介護 AI 入力予測ツールによって作成された介護記録を対象とした記録分析ツールであり、施設利用者の日常生活における基本動作の評価値を推定することを可能とした。今後は介護記録から利用者の事故予兆や状況悪化の予測を行うことを目指していると報告されている。

上野らは、福祉支援施設で問題となるインシデントの予兆検出のために、BERT を利用した支援記録を予兆の有無で分類する手法を提案した [1]。具体的には、支援記録の記録本文とその付随情報を入力として、予兆の有無の2クラス分類を行うことで予兆の検出を実現している。実験では、利用者 10 人の支援記録を用いて提案する予兆検出器の精度を検証し、平均で 0.87 の F 値を確認した。また、8 人の利用者は付随情報を活用することで検出精度が向上することを確認した。

本研究は、支援記録を利用して施設利用者の介助の助けとする点では [10] と類似しているが、[10] は文例を選択することで介護記録が生成できる入力支援ツールがあるため、入力形態が統一化されている記録を利用している。一方本研究では、文章の記述形式といったものが定まっていない自由記述形式の記録を利用する。また、画像処理による支援システムを構築する [9] に対し、本研究では主に自然言語処理によって支援記録の文章を解析して用いる。さらに、[7] や [8] が介護施設や医療現場の、事故やインシデントの記録を分析するのにに対し、本研究では記録からインシデントの予兆の検出をする。

3 支援記録と付与ラベル

3.1 支援記録

本研究で用いる支援記録は、福祉支援施設の職員が各入所者に対して実施した支援について記録したものである。記録が書かれるタイミングは様々であり、支援の直後に書かれたものや業務終了時に一日分の支援をまとめて記録したものなどがある。また支援記録の本文は、文の長さや文字数、記述形式といったものが定まっていない自由記述形式の文章となっている。使用する支援記録の例を表 1 に示す。表 1 の通り、支援記録には先に説明した自由記述の本文以外にも複数の情報が含まれている。入所者 ID は施設入所者を判別するための 6 桁の数値である。処理日付はその記録の入力された年月日を意味する。入力者 ID は支援をして記録を書いた支援者を判別するための 6 桁の数値である。インシデントラベルはその記録に含まれるインシデントの情報を、予兆ラベルはその記録に含まれる予兆の情報をそれぞれ表している。これらのラベルは元の支援記録には含まれておらず、後述の基準に従って人手で付与されたものを本研究で使用している。

表 1 支援記録とラベルの例

入所者 ID	入力者 ID	処理日付	記録本文	予兆ラベル	インシデントラベル
000111	000011	2020/04/12	寝ていることが多い。奇声もなく静かに過ごされている。	予兆なし	なし
000111	000012	2020/04/12	作業：朝礼を終え、イオン水を飲んだあとに職員への他害がある。クールダウンまでは早かったが、原因は不明である。	予兆（強）	暴力#職員
000122	000011	2020/04/13	舎活動：居室で穏やかに過ごされている。（14：30）大きな声が聞こえ、見に行くと笑っている。	予兆（弱）	なし

表 2 インシデントとそのラベル

種類	対象	ラベル
器物破損	自分の所有物	器物破損#自分の所有物
	他入所者の所有物	器物破損#他入所者の所有物
	職員の所有物	器物破損#職員の所有物
	施設の所有物	器物破損#施設備品
	その他	器物破損#その他
暴力	職員	暴力#職員
	施設の入所者	暴力#施設の入所者
	その他	暴力#その他

3.2 インシデントの定義とインシデントラベル

本研究では、周囲に大きな損害を与える状況をインシデントと定義する。周囲に大きな損害を与える状況の例としては、暴力全般や水や唾をかけられるなどの職員や他の入所者が不快な思いをするものや、器物破損のような修復コストがかかるものなどが挙げられる。なお本研究ではインシデントは、暴力や器物破損が自己以外の対象に害を与える他害に限定する。そのため、自傷行為のような自身に害を与えるものはインシデントとしていない。この定義に従って、インシデントには表 2 のようなラベルが付与されている。一つの記録に対して複数のインシデントラベルが付与されることもある。このインシデントは福祉支援施設での支援業務を妨げるものであり、現場の施設職員の意見を参考に定められたものである。また本研究では、ここに定めたインシデントの種類を、予兆検出の分析には用いるが、インシデントの分類は行わない。

3.3 予兆の定義と予兆ラベル

本研究では、支援者である施設職員の判断に基づいて、支援記録本文に施設入所者の衝動性が高い、あるいはネガティブな心理状態が現れているものを予兆と定義する。予兆ラベルには「予兆（弱）」「予兆（強）」の 2 種類がある。「予兆（弱）」は、衝動性が高い、あるいはネガティブな心理状態が現れている状態、「予兆（強）」は衝動性あるいはネガティブな感情を抑えることができず、結果として周囲へ影響が出てしまっている状態と定義している。「予兆（弱）」の例としては、自傷行為、情緒が乱れている、大声を発する、などがあげられ、「予兆（強）」の例としては、他者への暴力、器物破損などのインシデントや、物に当たる行為などが挙げられる。一方で、やる気がない、暴言、騒ぐ、などはそれだけでは予兆とはせず、記述内容の文脈によって予兆か予兆でないかを判断している。例えば、気分が悪い、つらそうな表情、などは精神面の話であれば「予兆（弱）」であるが、体調面の話であればこれらは「予兆なし」となる。

また一つの記録に対して「予兆（弱）」と「予兆（強）」の予兆ラベルが同時に付与されることもあり、本研究においてはそういった記録を「予兆（両方）」として扱う。本研究では、この予兆ラベルを用いて、支援記録を「予兆なし」「予兆（弱）」「予兆（強）」「予兆（両方）」の 4 クラスへ分類を行う。

4 インシデントの予兆検出

4.1 インシデントの予兆の検出手法

図 1 に使用する予兆検出器の概略を示す。この検出器ではまず、入力された記録本文を BERT により文書表現ベクトルに変換する。また、支援記録から支援者 ID や日付情報などの本文以外の付随情報を取り出し、付随情報ベクトルに変換する。次に、これらの文書表現ベクトルと付随情報ベクトルを入力とする分類器によって支援記録の予兆を、「予兆なし」「予兆（弱）」「予兆（強）」「予兆（両方）」の 4 クラスに分類する。その際に使用する分類器は図 1 の右側に示す 4 層のニューラルネットワークである。この各層は全結合層であり、左から順に 256 次元、128 次元、32 次元、4 次元ベクトルへと線形変換される。活性化関数にはシグモイド関数、出力層には softmax 関数を用いている。損失関数には交差エントロピー誤差を採用し、Adam[11]によって最適化する。分類器のエポック数は 20、バッチサイズは 32 とする。

図 1 の右にある分類器の 1 層目では、文書表現ベクトルを 256～0 次元、付随情報ベクトルを 0～256 次元の全結合層へ入力して、ベクトルの次元を変換し、2 つのベクトルを合わせて 256 次元のベクトルとしている。また各層間では Dropout[12]を実行している。Dropout はニューラルネットワークの過学習防止のための手法の一つで、各層ごとに一定確率でランダムにノードの出力を 0 にすることで無効化して学習を進める。

分類器の出力結果は 0 から 1 の範囲で正規化された各ラベルに対する確率であり、その総和は 1 になる。本稿で行う 4 クラス分類では「予兆なし」「予兆（弱）」「予兆（強）」「予兆（両方）」の 4 つのクラスに対する確率が出力され、確率が最大のクラスに分類する。

図 1 の左の BERT の部分は、事前学習済みモデルを支援記録を用いてファインチューニングし分類に利用する。また本研究では、BERT 日本語 Pretrained モデル²を利用する。

また本研究で入力として使用する付随情報は表 3 に示す 5 種類である。日付情報として利用する月、日、曜日はそれぞれ 12 次元、31 次元、7 次元、入力者 ID は 100 次元、インシデント

2 : https://nlp.ist.i.kyoto-u.ac.jp/?ku_bert_japanese

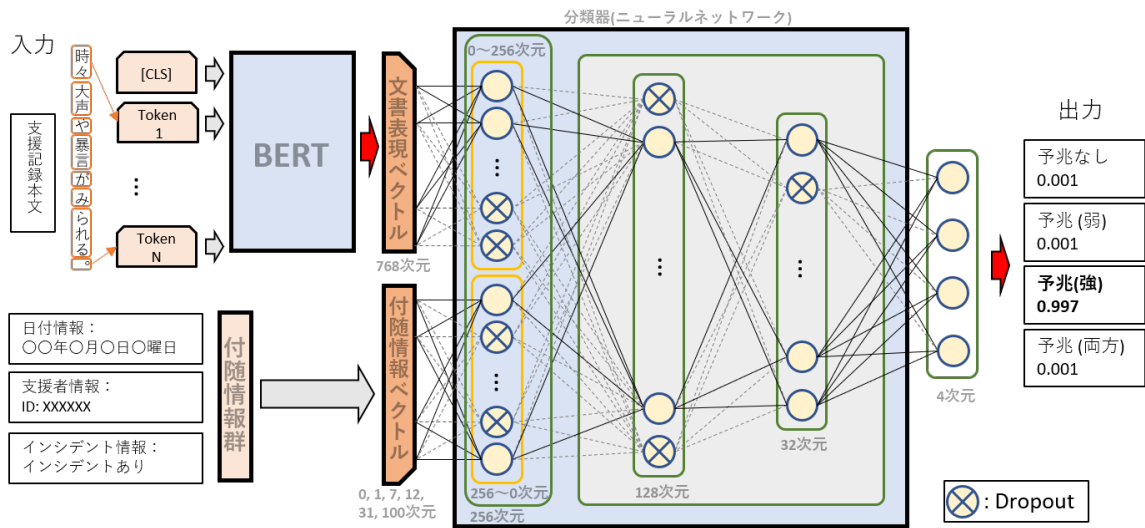


図 1 予兆検出器の概略図 [1]

表 3 付随情報と次元

付随情報	ベクトルの次元
月	12 次元
日	31 次元
曜日	7 次元
入力者 ID	100 次元
インシデントの有無	1 次元

の発生の有無は 1 次元の One-hot ベクトルで表現する。

4.2 個人専用分類器と汎用分類器

上野らは、BERT のファインチューニングに分類対象となる入所者本人の記録のみを用いたため、本稿ではこの分類器を個人専用分類器と呼ぶ。本稿では分類器の汎用性について考察するため、この個人専用分類器の他に、他の入所者の支援記録をファインチューニングに用いた汎用分類器を使用する。汎用分類器は、分類対象者と同施設に入所している他の入所者全員の支援記録で学習をする汎用分類器（同施設）、分類対象者とは別の施設に入所している入所者全員の支援記録で学習をする汎用分類器（他施設）、分類対象者を除く全ての入所者の支援記録で学習をする汎用分類器（本人以外）の 3 つを使用する。

5 インシデントの予兆分類実験

5.1 実験に用いるデータ

インシデント予兆分類実験には、2 つの福祉支援施設の入所者 10 名ずつ、計 20 名の支援記録を用いる。この支援記録には、3 節で説明した「予兆なし」「予兆（弱）」「予兆（強）」の 3 つの予兆ラベルが付与されており、「予兆（弱）」と「予兆（強）」の両方のラベルが付与されている記録は「予兆（両方）」ラベルが付与されているものとして扱う。これらの支援記録は図 1 の BERT のファインチューニングや、分類器の学習に利用する。表 4 に支援記録の統計情報を示し、図 2 に入所者ごとの支援記録の各予兆ラベルの割合をまとめる。表 4 には、入所者ごとに、

各予兆ラベルの付与されている記録件数と全記録に占める割合を示している他、インシデントラベルが付与されている記録の件数とその全記録に占める割合についても記している。

5.2 実験内容

本研究では、以下の (1),(2) について実験する。

- (1) 汎用分類器による予兆分類
- (2) 付随情報を利用した予兆分類

まず実験 (1) では、他の入所者の支援記録を学習に用いた汎用分類器により予兆分類をし、その精度の評価、考察をする。この際、4 節で述べた個人専用分類器と汎用分類器（同施設）、汎用分類器（他施設）、汎用分類器（本人以外）の 3 つの汎用分類器を使用し分類実験をする。

個人専用分類器は、分類対象の入所者本人のデータを用いて五分割交差検定により分類精度を求め、その他 3 つの汎用分類器は五分割交差検定を行わず、各支援記録集合で学習した分類器に対して、テスト用データとなる分類対象者の記録全件分類精度を求める。

実験 (2) では、日付などの付随情報を利用して予兆分類をし、その精度を評価する。上野らの作成した付随情報を入力とする分類器は、入力の文書表現ベクトルと付随情報ベクトルをそれぞれ 128 次元のベクトルに変換し、利用している。実験 (2) では、このベクトルの次元を変更して、付随情報の有効性を比べる。具体的には、分類器 1 層目の 256 次元のうちの文書表現ベクトルと付随情報ベクトルの重み割合を、16:0~0:16 まで変化させ、その精度を比較する。重み割合が 16:0 の場合は文書表現ベクトルのみ、0:16 の場合は付随情報ベクトルのみを利用した分類となる。

また実験 (2) では、付随情報ベクトルの次元に文書表現ベクトルの次元を合わせた実験も行う。つまり、各付随情報ベクトルの次元はそのまま、文書表現ベクトルを付随情報ベクトルの次元まで削減し入力とする。例えば、付随情報が「入力者 ID」の場合は、文書表現ベクトルと付随情報ベクトルがそれぞれ

表 4 支援記録の統計情報

施設	人物	性別	記録件数 (割合)					インシデント件数
			予兆なし	予兆 (弱)	予兆 (強)	予兆 (両方)	合計	
施設A	a	女性	1,519(62.9%)	646(26.7%)	88(3.6%)	163(6.7%)	2,416	106(4.4%)
	b	女性	2,207(64.2%)	999(29.0%)	83(2.4%)	150(4.4%)	3,439	205(6.0%)
	c	女性	1,830(55.5%)	607(18.4%)	503(15.2%)	360(10.9%)	3,300	630(19.1%)
	d	女性	1,510(78.6%)	304(15.8%)	14(0.7%)	93(4.8%)	1,921	87(4.5%)
	e	男性	2,800(73.7%)	832(21.9%)	83(2.2%)	82(2.2%)	3,797	156(4.1%)
	f	男性	2,702(81.3%)	551(16.6%)	40(1.2%)	30(0.9%)	3,323	52(1.6%)
	g	男性	2,564(75.7%)	599(17.7%)	67(2.0%)	157(4.6%)	3,387	99(2.9%)
	h	男性	3,034(88.2%)	367(10.7%)	31(0.9%)	8(0.2%)	3,440	15(0.4%)
	i	男性	3,216(84.3%)	428(11.2%)	94(2.5%)	77(2.0%)	3,815	73(1.9%)
	j	男性	3,320(85.6%)	532(13.7%)	14(0.4%)	13(0.3%)	3,879	6(0.2%)
施設B	k	男性	1,781(90.0%)	63(3.1%)	121(6.0%)	37(1.8%)	2,002	149(7.4 %)
	l	男性	1,927(96.3%)	59(2.9%)	10(0.5%)	7(0.3%)	2,003	11(0.5%)
	m	男性	1,443(72.1%)	157(7.8%)	379(18.9%)	22(1.1%)	2,001	385(19.2%)
	n	男性	1,621(81.0%)	277(13.8%)	78(3.9%)	26(1.3%)	2,002	56(2.8%)
	o	男性	1,667(83.3%)	120(6.0%)	202(10.1%)	14(0.7%)	2,003	174(8.7%)
	p	男性	1,774(88.7%)	168(8.4%)	54(2.7%)	4(0.2%)	2,000	58(2.9%)
	q	男性	1,885(94.3%)	95(4.8%)	17(0.9%)	3(0.2%)	2,000	18(0.9%)
	r	男性	1,546(77.3%)	251(12.5%)	144(7.2%)	62(3.1%)	2,003	206(10.3%)
	s	男性	744(97.4%)	19(2.5%)	0(0.0%)	1(0.01%)	764	1(0.01%)
	t	男性	1,747(87.3%)	91(4.5%)	135(6.7%)	28(1.4%)	2,001	133(6.6%)

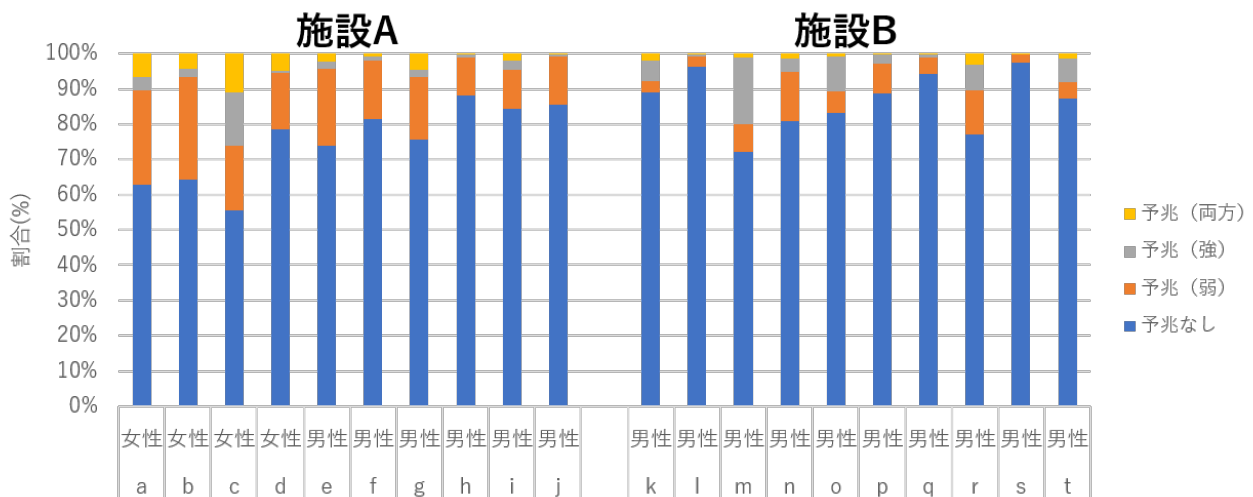


図 2 支援記録の予兆ラベルの割合

100 次元となり、図 1 の分類器 1 層目の次元は 256 次元ではなく 200 次元となる。この際、付随情報が「インシデントの有無」である場合のみ、文書表現ベクトルを付随情報ベクトルに合わせて 1 次元にせず、10 次元に削減して利用する。具体的には、図 1 の右側の分類器の一層目の次元を指定の次元とすることで実現している。これらの付随情報を用いた実験は、先の汎用分類器による予兆分類実験における汎用分類器 (本人以外) を用いて行う。

5.3 評価指標

分類器の評価には、以下に示す再現率 (Recall)、適合率 (Precision)、F 値 (F-measure) を利用する。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{F - measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3)$$

例えば、予兆ラベル「予兆 (強)」についての TP, TN, FP, FN は下記ようになる。

- TP: 「予兆 (強)」だと分類された記録のうち、実際に「予兆 (強)」である記録の件数
- TN: 「予兆 (強)」ではないと分類された記録のうち、実際に「予兆 (強)」でない記録の件数

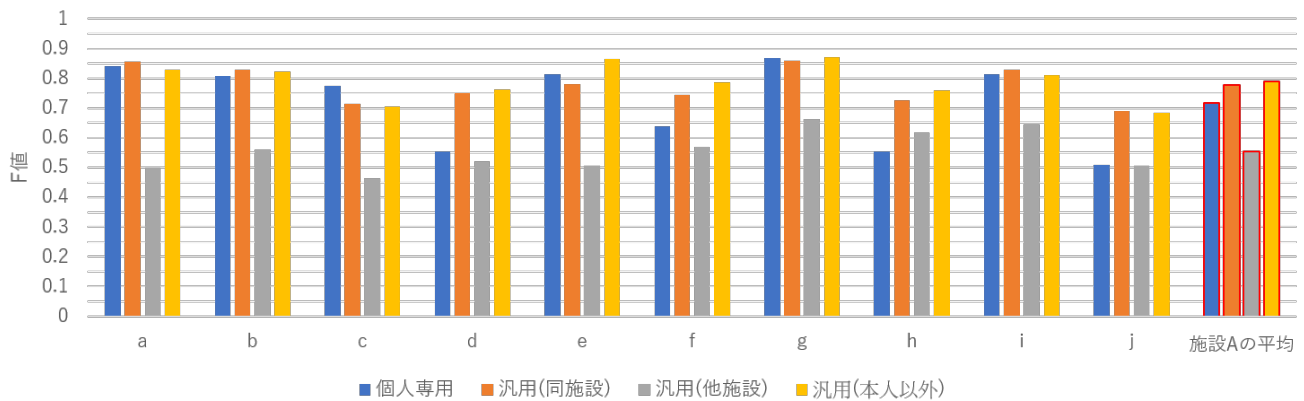


図 3 各入所者の分類器ごとの F 値 (施設 A)

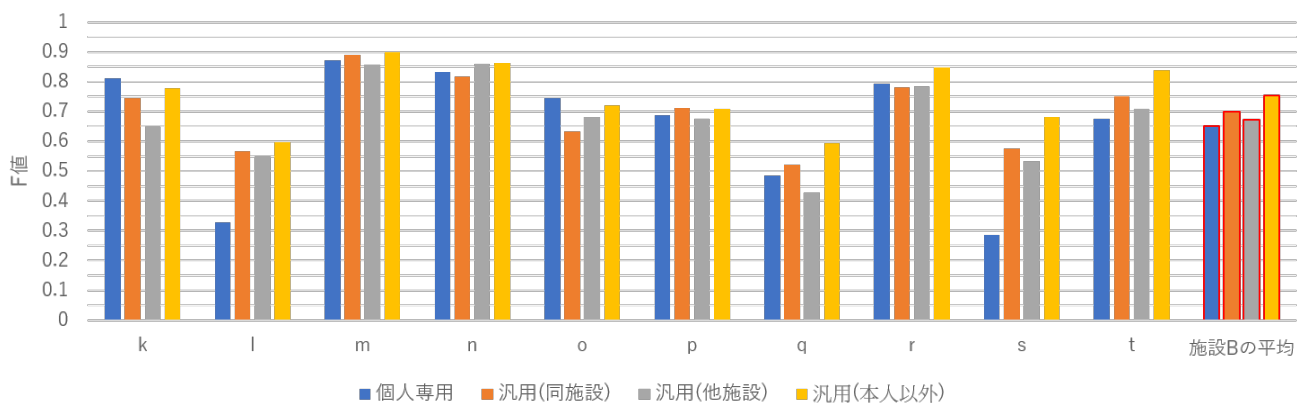


図 4 各入所者の分類器ごとの F 値 (施設 B)

- FP:「予兆 (強)」だと分類された記録のうち,「予兆 (強)」ではない記録の件数
- FN:「予兆 (強)」ではないと分類された記録のうち,「予兆 (強)」である記録の件数

よって再現率は全ての「予兆 (強)」である文章に対し,分類器が正しく「予兆 (強)」と分類したものの割合であり,適合率は分類器が「予兆 (強)」と分類した文章のうち,実際に「予兆 (強)」であるものの割合である。また F 値は再現率と適合率の調和平均となる。

以下の評価実験では,4つの予兆ラベルについて個別にこの値を計算した後,その平均を求める。

5.4 汎用分類器による予兆分類実験

5.4.1 実験結果

汎用分類器による予兆分類結果の F 値を図 3 (施設 A の各入所者)と図 4 (施設 B の各入所者)に示す。

図 3 と図 4 のいずれにおいても,入所者ごとの差異はあるが,各施設の平均で見ると汎用分類器 (本人以外) の精度が最も良く,次いで汎用分類器 (同施設) の精度が良いことが分かる。各入所者について見ると,施設 A の入所者 d, h や施設 B の入所者 l, q のように個人専用分類器の F 値が他の入所者に比べて小さくなっている人物は,個人専用分類器に比べて,汎用分類器 (同施設) や汎用分類器 (本人以外) の F 値が大きいことが

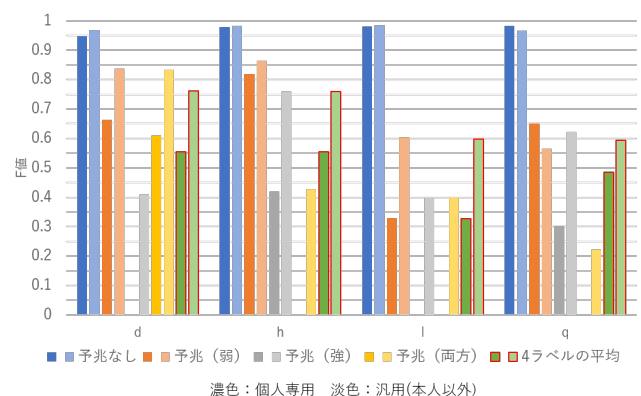


図 5 個人専用分類器と汎用分類器 (本人以外) の F 値の比較

分かる。また,図 3 の施設 A では汎用分類器 (他施設) の F 値が小さいのに対し,図 4 の施設 B の汎用分類器 (他施設) の F 値は,他の分類器とあまり差がないことが分かる。

5.4.2 考察

両施設の平均の F 値で,汎用分類器 (本人以外) 及び,汎用分類器 (同施設) の値が大きいのは,先に述べた施設 A の入所者 d, h や施設 B の入所者 l, q のような人物の分類精度の影響がある。当該入所者の分類結果では,汎用分類器 (本人以外),汎用分類器 (同施設) の F 値が大きい。これらの人物は,表 4 や図 2 の支援記録の統計情報を見ると分かるように,「予兆 (強)」

表 5 「予兆（強）」の少ない入所者の支援記録の分類事例

	記録本文	予兆ラベル	分類結果	
			個人専用	汎用（本人以外）
(1)	朝の着替えの際、自身のパジャマを割く。	予兆（強）	予兆なし	予兆（強）
(2)	脱衣場で〇〇さんの頭を叩く。注意する。 突かれた事に怒った様子。	予兆（両方）	予兆（弱）	予兆（両方）
(3)	朝食後□□さんが食器を片付けないことに 腹を立て食堂内で追いかけて掴みかかろうとする。	予兆（両方）	予兆（強）	予兆（強）

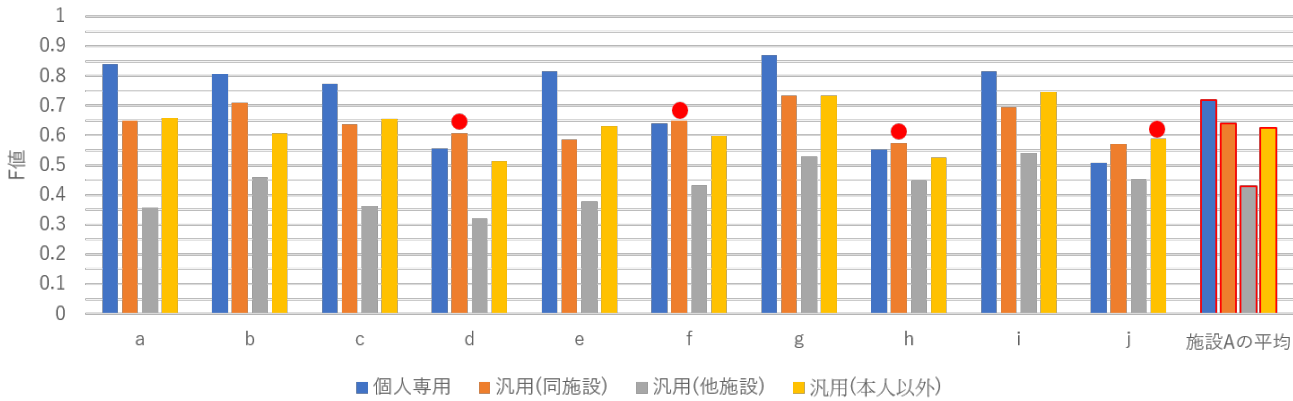


図 6 学習回数を揃えた実験における各入所者の分類器ごとの F 値 (施設 A)

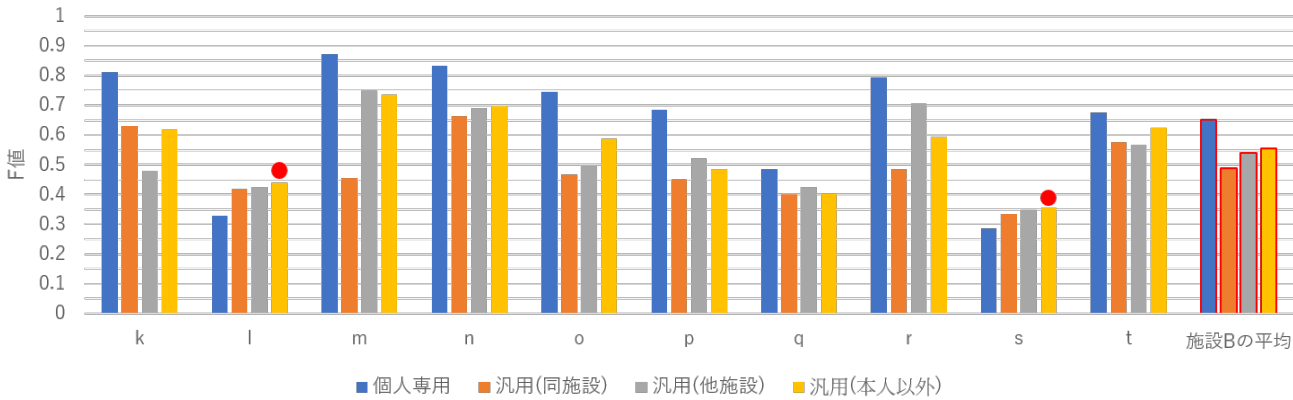


図 7 学習回数を揃えた実験における各入所者の分類器ごとの F 値 (施設 B)

を含む支援記録の割合が他の入所者と比べて小さい。そのため、自身の記録のみでの学習では「予兆（強）」や「予兆（両方）」ラベルの分類が困難である。図 5 に個人専用分類器と汎用分類器（本人以外）の予兆ラベル毎の F 値を示す。色の濃い棒が個人専用分類器、淡い棒が汎用分類器（本人以外）の F 値である。個人専用分類器の当該の入所者の F 値を見ると、「予兆（強）」や「予兆（両方）」ラベルの分類精度が悪いことが分かる。一方、汎用分類器（本人以外）の当該入所者の予兆ラベルごとの F 値は、個人専用分類器と比べて「予兆（強）」や「予兆（両方）」の値が相対的に大きくなっていることが分かる。

本実験において、予兆の少ない入所者について個人専用分類器で誤分類され、汎用分類器（本人以外）では正しく分類された記録本文の例を表 5 の (1)、(2) に、汎用分類器（本人以外）でも誤分類となってしまった例を (3) に示す。表 5 の (1) の例では、「自身のパジャマを割く」という「予兆（強）」である器物

破損の記述があるが、個人専用分類器では「予兆（弱）」に分類されている。(2) の例では、「頭を叩く」という「予兆（強）」の記述と「怒った様子」という「予兆（弱）」の記述があるため「予兆（両方）」であるが、個人専用分類器では、「予兆（弱）」に分類されている。汎用分類器（本人以外）では、このような「予兆（強）」を含む記録の誤分類が少なくなっている。

一方いずれも誤分類となってしまった (3) の例では、「追いかけて掴みかかろうとする」という「予兆（強）」の記述と「腹を立て」という「予兆（弱）」の記述があるため、「予兆（両方）」であるが、分類結果はどちらも「予兆（強）」となった。

また、各汎用分類器で学習に使用する記録の件数を、五分割交差検定を行う個人専用分類器の学習で使用する件数まで削減し、学習回数を揃えた分類実験を行った。分類結果の F 値を図 6（施設 A の各入所者）と図 7（施設 B の各入所者）に示す。図 6 と図 7 のいずれにおいても、各施設の平均を見ると個人専

表 6 施設 A と施設 B の記録本文の統計情報

施設	文字数			「予兆（強）」を含む記録の文字数			人物の名前		「：」区切りの表記	
	最大値	最小値	平均	最大値	最小値	平均	件数	全記録に対する割合	件数	全記録に対する割合
施設 A	1,047	0	56.00	1,047	12	130.73	2,671	8.16% (2671/32,717)	9,567	29.24% (9,567/32,717)
施設 B	809	3	44.80	448	6	53.91	4,689	24.97% (4689/18,779)	188	1.00% (188/18,779)

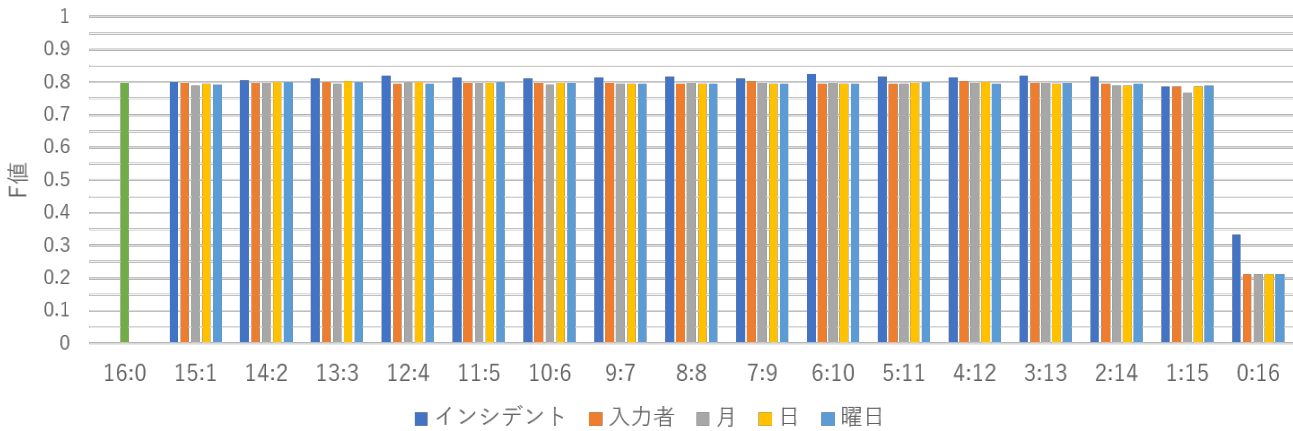


図 8 文書表現ベクトルと付随情報ベクトルの割合と施設 A の入所者の F 値の平均

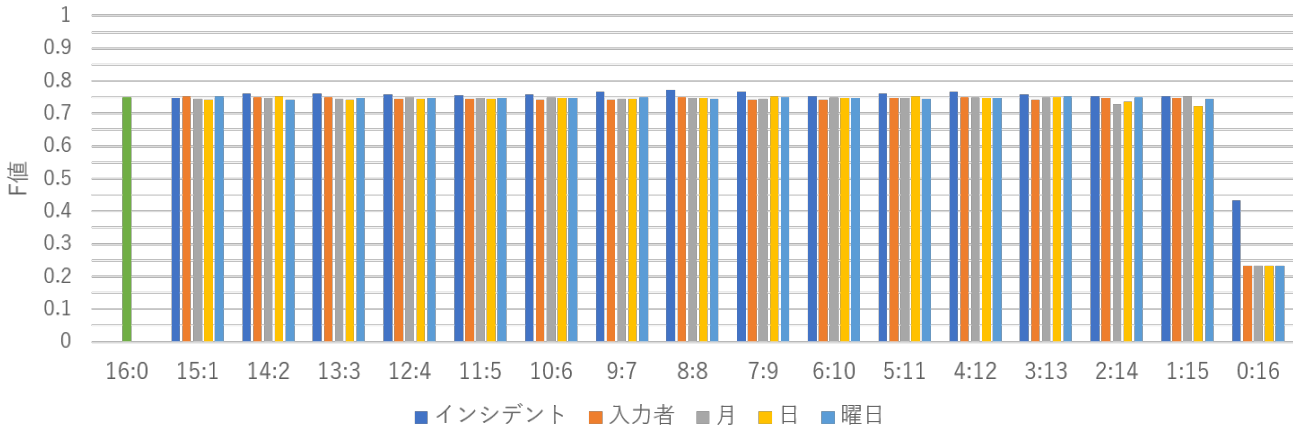


図 9 文書表現ベクトルと付随情報ベクトルの割合と施設 B の入所者の F 値の平均

用分類器の F 値が最大であることが分かり、各入所者についても個人専用分類器の F 値が最大となる人物が殆どである。よって、学習に利用する支援記録の件数が等しい場合は、学習に利用する記録は、本人の記録が良いことが分かる。しかし、図 6 と図 7 でグラフに赤丸を付けた入所者については、個人専用分類器と比べて汎用分類器 (本人以外) や汎用分類器 (同施設) の F 値が大きくなっている。これらの入所者は、先に述べた「予兆（強）」を含む支援記録が他の入所者と比べて少ない人物である。このことから、そのような「予兆（強）」を含む支援記録の少ない人物には、他者の記録を用いた学習も有効である。

また、図 3 と図 4 の比較でも、図 6 と図 7 の比較でも、汎用分類器 (他施設) の分類精度が、両施設で大きく異なっていたため、施設ごとに記録本文の内容に差があると考え、分析を行った。具体的には、両施設の支援記録の文字数の最大値、最小値、平均値を調べ、さらに「〇〇さん」というような、人物の名前を変換した記述や、記録の初めに「舍活動：～」や「作業：～」

といった固有の表現を含む支援記録の件数を調べた。その結果を表 6 に示す。文字数から、施設 A の記録の方が文章が長い傾向にあることが分かる。中でも「予兆（強）」を含む記録の文字数については、施設 A は施設 B と比べ、最大値、平均値共に 2 倍以上の長さとなっている。また、人物名は施設 B の記録本文に多く出現し、「：」区切りの表記は施設 A の記録本文に多く出現しているように、施設間で記録の文章に差があることが分かった。

5.5 付随情報を利用した予兆分類実験

5.5.1 実験結果

図 1 の分類器 1 層目の 256 次元の文書表現ベクトルと付随情報ベクトルの重み割合を変化させた予兆分類実験の結果の F 値を図 8 (施設 A の入所者の平均) と図 9 (施設 B の入所者の平均) に示す。なおこれらの図のグラフの 16:0 は、付随情報を用いない文書表現ベクトルのみで分類した結果の F 値である。

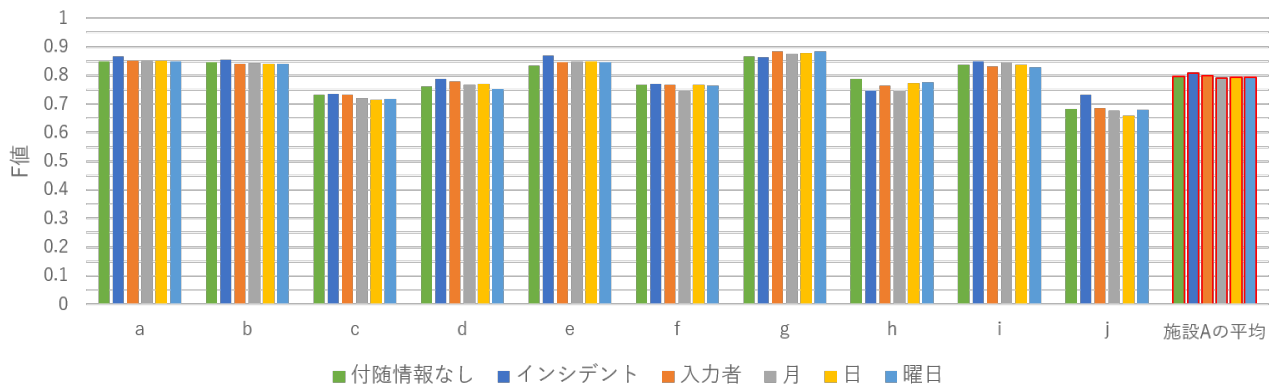


図 10 文書表現ベクトルを付随情報ベクトルに合わせ削減した実験の F 値 (施設 A)

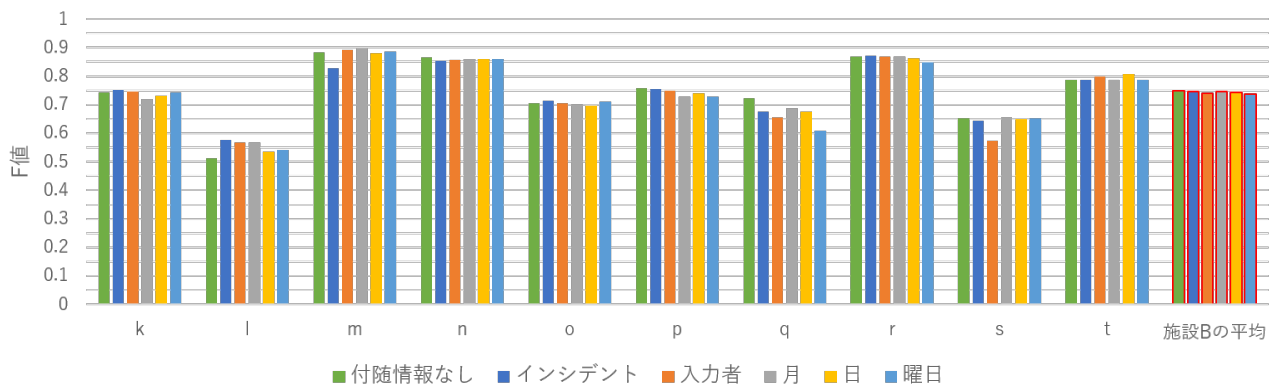


図 11 文書表現ベクトルを付随情報ベクトルに合わせ削減した実験の F 値 (施設 B)

0:16 はその反対に付随情報のみで分類した結果の F 値である。図 8 と図 9 のいずれにおいても、インシデントの有無以外の付随情報については、ベクトルの重み割合を変更しても、F 値がほとんど変化せず、0:16 においては F 値が大きく下がる。月、日、曜日の付随情報ベクトルのみを使用した場合は、分類結果として「予兆なし」のみしか出力しておらず、入力者 ID の付随情報ベクトルのみを使用した場合もほぼ「予兆なし」のみを出力した。一方、インシデントの有無の付随情報ベクトルを用いた場合は、F 値が大きくなっている場合がある。また、付随情報ベクトルのみを使用する場合においても、他の付随情報と比べると相対的に F 値が大きくなっている。

次に、文書表現ベクトルを付随情報ベクトルの次元まで削減して入力とした予兆分類の結果の F 値を図 10 (施設 A) と図 11 (施設 B) に示す。比較のため付随情報なし 16:0 の F 値も示している。図 10 と図 11 のいずれも、入所者ごとに結果の差異はあるものの、平均でみると付随情報なしの結果とあまり差がないことが分かる。

5.5.2 考 察

図 8 と図 9 より、インシデントの有無以外の付随情報はほとんど有効性が確認できなかった。インシデントの有無については、インシデントの付随情報ベクトルを使用した場合に F 値が 0.03 以上大きくなることのある入所者が施設 A で 10 名中 5 名、施設 B で 10 名中 4 名の計 9 名存在しており、分類精度の向上に有効な場合があった。

図 8、図 9 の予兆分類実験において、インシデントの付随情報ベクトルをベクトル割合 1:15 で使用した場合に分類精度が向上した入所者の分類事例を表 7 に示す。表 7 の (1) が付随情報ベクトルを使用した場合に正しく分類ができるようになった例、(2) が付随情報ベクトルを使用した場合に誤分類となった例である。表 7 の (1) の例では、「他利用者の手をはねたり押して」という「予兆 (強)」の記述があるにもかかわらず、文書表現ベクトルのみの場合では「予兆なし」に分類されているが、インシデントの付随情報ベクトルを用いた場合では正しく分類されている。(1) の支援記録は、「インシデントあり」となっており、正しく分類ができた。また例には示していないが、「予兆 (弱)」のみを含む記録を、文書表現ベクトルのみの場合は「予兆 (両方)」と誤分類、インシデントの付随情報ベクトルを用いた場合は「予兆 (弱)」に正しく分類ができた事例がある。この例の支援記録は、「インシデントなし」となっており、正しく分類ができた。

一方 (2) は誤分類の例で「××さんを引っ張ろうとする」という「予兆 (強)」の記述があり、文書表現ベクトルのみの場合は「予兆 (強)」と正しく分類出来ているが、インシデントの付随情報を使用した場合は「予兆なし」の誤分類となった。この例は、「インシデントなし」の支援記録で、その上 1:15 と付随情報ベクトルの比重が大きいため、「予兆なし」に誤分類した可能性がある。

文書表現ベクトルを付随情報ベクトルの次元まで削減し入力

表 7 インシデントの付随情報ベクトルを利用した分類事例

	記録本文	予兆ラベル	分類結果	
			文書表現ベクトルのみ	付随情報あり (1:15)
(1)	職員 (△△) に笑顔で付いて歩き、他の利用者が 職員の側に来たり手を引いたりすると、他利用者の手をはねたり押して移動させたりしていた。	予兆 (強)	予兆なし	予兆 (強)
(2)	起床し血圧測定を行う。その後××さんを 引っ張ろうとする事が何度もあった。	予兆 (強)	予兆 (強)	予兆なし

とした図 10 と図 11 の予兆分類実験では、入所者毎に差はあったものの、両施設ともに平均の値は、付随情報があってもなくてもあまり差はなかった。そのため、本実験のようにベクトルの次元を揃えるだけでは、分類精度の向上はあまり期待できない。

6 ま と め

本稿では、上野らの提案した福祉支援施設の支援記録からのインシデントの予兆検出器を、本人以外の入所者の支援記録で学習して予兆検出実験を行った。予兆の検出は、支援記録の本文を利用した分類器を作成し、対象の支援記録を「予兆なし」「予兆 (弱)」「予兆 (強)」「予兆 (両方)」の 4 クラスに分類することで実現した。また、この分類に支援記録の本文以外の付随情報が有効かどうか実験した。

分類器の学習に本文以外の入所者の支援記録を利用する実験では、施設 A 入所者の全ラベルの平均 F 値は、個人専用分類器で 0.718、汎用分類器 (同施設) で 0.778、汎用分類器 (他施設) で 0.555、汎用分類器 (本人以外) で 0.790、施設 B 入所者のそれは、個人専用分類器で 0.652、汎用分類器 (同施設) で 0.699、汎用分類器 (他施設) で 0.673、汎用分類器 (本人以外) で 0.753 となり、汎用分類器 (本人以外) の分類精度が最も高かった。また、学習に本人以外の入所者の支援記録を利用した分類器は、予兆やインシデントの少ない入所者の予兆検出に有効であることが分かった。

分類器の学習に付随情報を利用する実験では、インシデントの有無の情報を利用した場合の予兆分類の F 値が最も大きく、施設 A の入所者では文書表現ベクトルと付随情報ベクトルの割合が「6:10」の場合に全ラベルの平均 F 値が 0.825、施設 B の入所者では、ベクトルの割合が「8:8」の場合に全ラベルの平均 F 値が 0.772 だった。インシデントの有無の付随情報は分類精度の向上に有効である可能性がある。

今後は、両施設間の支援記録文章の違いを調査し、分類器の汎用性の向上に有効な文章の特徴の発見を目指す。付随情報については、有用な可能性のあるインシデントの有無などの付随情報の有効な利用方法を検討する。

文 献

[1] 上野史, 松本典久, 太田学, “福祉支援施設の支援記録からの BERT によるインシデントの予兆検出,” 情報処理学会論文誌, Vol. 64, No. 11, pp. 1561–1566, 2023.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “Bert: Pretraining of deep bidirectional trans-

formers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need,” Advances in Neural Information Processing Systems 30, pp. 5998–6008, 2017.

[4] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman, “GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding,” arXiv preprint arXiv:1804.07461, 2018.

[5] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang, “SQuAD: 100,000+ Questions for Machine Comprehension of Text,” arXiv preprint arXiv:1606.05250, 2016.

[6] Rowan Zellers, Yonatan Bisk, Roy Schwartz, and Yejin Choi, “SWAG: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference,” arXiv preprint arXiv:1808.05326, 2018.

[7] 田中健次, 坂東幸一, 津本周作, 佐藤直樹, “医療界でのインシデント情報を活用したリスク対応,” 横幹, Vol. 13, No. 2, pp. 84–90, 2019.

[8] 峯崎智裕, 松木萌, 井上創造, “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,” 情報処理学会論文誌, Vol. 58, No. 10, pp. 1701–1711, 2017.

[9] 中沢信明, “高齢者介護施設における夜間徘徊者の検出,” 社会安全とプライバシー, Vol. 3, No. 2, pp. 8–12, 2019.

[10] 青木三重子, 撫中達司, 豊川絢也, 竹内銀河, “科学的介護を支援する介護 AI 記録分析ツール「けあらぼ」,” 三菱電機技報, Vol. 95, No. 8, pp. 506–509, 2021.

[11] Diederik P Kingma and Jimmy Ba, “Adam: A Method for Stochastic Optimization,” arXiv preprint arXiv:1412.6980, 2014.

[12] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” Journal of Machine Learning Research, Vol. 15, No. 1, pp. 1929–1958, 2014.