福祉支援施設の支援記録を利用したインシデント発生検出の一手法

松本　典久† 　上野　史†† 　太田　学††

† 岡山大学大学院自然科学研究科　〒 700–8530 岡山県岡山市北区津島中 3–1–1
†† 岡山大学学術研究院自然科学学域　〒 700–8530 岡山県岡山市北区津島中 3–1–1
E-mail: †piia6ope@s.okayama-u.ac.jp, ††{uwano, ohta}@okayama-u.ac.jp

あらまし　高齢者や障がい者に対する福祉支援施設では入所者に対して支援員の数は限られているため，入所者が起こすトラブル，すなわちインシデントの発生を予測して未然に防ぐことが重要である．しかし，インシデントの発生を事前に察知することはベテランの支援員でも難しい．そこで，本研究ではその第一歩として福祉支援施設の支援記録からインシデントの発生を検出するための手法を提案し，実験により有効性を検証する．具体的には過去の支援記録から日付情報や支援記録の本文などを利用するモデルを，Bidirectional Encoder Representations from Transformers (BERT) を用いて作成しインシデントの発生を検出する．10 人の支援記録を用いた実験の結果，提案手法は他の機械学習手法と比べて高い F 値を示した．また，支援記録は記録ごとで扱うより日付ごとで扱う方が高い F 値を示した人物が多く，支援記録の本文に加えて日付情報や記入者情報を与えることで F 値が向上しうることがわかった．

キーワード　インシデント検出，支援記録，BERT，記録本文，福祉支援施設

1 はじめに

近年，高齢者や障がい者に対する介護や支援は広く求められており，高齢者や障がい者に対する福祉支援施設の重要性が増している．しかし，福祉支援施設では入所者の人数に対して支援員の数は限られており，支援には限度がある．また，入所者がインシデントを起こすとそれに対応する必要が追加で発生するため，入所者のインシデントは少ないことが望ましい．したがって，入所者が起こすインシデントを予見し未然に防止することが求められる．しかし，インシデントの発生を察知することはベテランの支援員でも困難である．

そこで本研究では，インシデント発生予測を目指す．しかし発生予測は確定していない未来を当てる必要があり困難なタスクである．そこで発生予測に向けた第一歩として，福祉支援施設の支援記録からインシデントの発生を検出する [1]．具体的には，対象データを福祉支援施設の支援記録とし，記録本文に加えて日付データなどを入力としたインシデント発生の分類器を作成して，対象の支援記録がインシデントか否かを分類することでインシデントの発生を検出する．その際，対象とするインシデントは，他人への暴力など自分以外を対象に害を与える他害とする．これは入所者によるインシデント発生の傾向を知るための一助となり，インシデント発生予測の実現可能性を探ることにもつながる．また，検出には自然言語処理モデルの一つである Bidirectional Encoder Representations from Transformers (BERT) [1] を利用する．BERT の特徴の一つは文脈を考慮することが可能なことである．これにより，対象となる支援記録の文脈と日付などの付加情報を学習させることで，インシデントを判定する．

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された．（研 2104-005）

本稿の構成を述べる．第 2 節では，関連研究を紹介する．第 3 節では，インシデント予測と本稿で使用するモデルについて説明する．第 4 節では，評価実験として支援記録分類実験の内容と結果を示し，考察を述べる．第 5 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

BERT は Devlin らが提案した自然言語処理技術であり，双方向 Transformer [5] というニューラルネットワークを利用した言語モデルである [1]．これは大規模テキストコーパスから事前学習を行ったモデルをファインチューニングにより個々のタスクに合わせて学習することで，汎用的に活用できる．

特徴としては，同一のモデル内で双方向から単語の周囲の文脈を学習できることが挙げられる．単語列に対する双方向での学習は，単方向での学習をそのまま双方向に適用すると学習時に予測すべき単語を先読みするため実現できなかったが，BERT では事前学習の際に，Masked Language Model (MLM) と Next Sentence Prediction (NSP) を用いた学習により，予測する単語を先読みすることを防いでいる．MLM は入力シーケンスにある単語の 15%を [MASK] トークンに置き換え，[MASK] トークンに置き換えられた単語を予測するタスクであり，NSP は文のペアを受け取り，ペアにおいて 2 つ目の文が元の文章において後続の文になっているかを予測するタスクである．

BERT は公開時，General Language Understanding Evaluation (GLUE) ベンチマーク [2] の 8 つのタスク，Stanford Question Answering Dataset (SQuAD) [3] の v1.1 と v2.0，Situations With Adversarial Generations (SWAG) [4]，の合計 11 の自然言語処理タスクで最高記録を達成した．

表 1　支援記録の例

利用者コード	入力者コード	処理日付	記録	インシデントラベル
012345	111111	yyyy/mm/dd	8:30 奇声あり 8：40 窓を叩く音があり、窓を割る。その後、職員の髪を引っ張る。 すぐにリスペリドンを服用し、その後食事を提供すると落ち着かれる。	器物破損#施設備品 暴力#施設の利用者
012346	111111	yyyy/mm/dd	食堂に出られることは多いが、他者にちょっかいを出すことが多い。	なし
	222222		園内活動：壁紙を剥がす行為が多い。感情の起伏も激しい。	器物破損#施設備品

2.2　文書分類に関する研究

大友らは，いじめ表現辞書を作成することで，ネットいじめの自動検出を行った [6]．いじめ表現辞書とは SO-PMI 値をいじめ度として付与した単語辞書であり，いじめ度が高いほどその単語がいじめに関連する可能性が高いという意味をもつ．複数の機械学習手法に対し，いじめ表現を含む複数の特徴量を与えたところ，最大で検出精度の F 値 0.92 を記録した．また，ほかの特徴量に加えていじめ表現を使用することで精度が向上したと報告されている．

酒井らは，ディープニューラルネットワークを用いて怒り感情が含まれた日本語文書を検出する手法を提案した [7]．まず怒りの種類を 5 種類に分類し，さらにそれらを明示的怒りと暗示的怒りに分類した．そして，文書を構成する個々の文の怒り感情を Convolutional Neural Network (CNN) を用いた分類器により分類し，それらを統合して文書の怒り分類とした．その結果，提案した 3 種類の分類手法の中で，二段階検出が精度 0.38 で最大となった．

また我々は，BERT を用いて Twitter における煽りツイートの検出を行った [8] ．まず対象とするツイートをリプライツイートに限定し，リプライ元との組を作った．それを用いて BERT の学習済みモデルをファインチューニングしてツイート間の関係性を学習させた．その結果，煽りツイートの検出精度の F 値が最大で 0.72 となった．

本研究は，文書からイベントを自動検出するという点では [6] と類似しているが，大友らは BERT を使用していない．またニューラルネットワークを用いて検出する点では [7] と類似しているが，個々の文の判定を統合して文章全体の判定をする酒井らの手法と異なり，最初から文章全体の判定を行う点で異なる．さらに BERT を用いて記録文書からある種のインシデントの発生を検出するという点では [8] の研究と類似している．しかし，分類対象とする文書の本文のみを利用する [8] の手法に対し，本研究では文書に付随する他の情報も用いて検出を行うという点が異なる．また文書の本文に関して, [8] の手法ではリプライツイートとそのリプライ元のペアとなる文章を利用するのに対し，本研究では用いる文書間にそのようなペアとなるような関係性はない．

3　インシデントの検出手法

3.1　支 援 記 録

支援記録の例を表 1 に示す．この支援記録は表 1 に示す通り，記録本文以外はフォーマットが決まっており，記録本文のみ自由記述となっている．表 1 では，2 つ目の記録は普通の文章の記述だが，1 つ目の記録は「8:30」「８：４０」のように自由な形式で時間情報が記述されている．3 つ目の記録は「園内活動：」のような見出しがつけれている．このように記録本文は文字数が 0 文字以上という点以外に決まりがなく，多いものでは 1 つの記録本文が 1000 文字以上となることもある．本研究で支援記録は各支援記録を記録ごとに用いる記録単位と，同一日付のものをまとめて日付ごとに用いる日付単位の 2 種類で使用する．

表 2　インシデントラベル

大分類	対象	ラベル名
器物破損	自分の所有物	器物破損#自分の所有物
	他利用者の所有物	器物破損#他利用者の所有物
	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
	施設の利用者	暴力#施設の利用者
	その他 （訪問者，配達員，etc...）	暴力#施設外の人

3.2　インシデントの定義

本研究で検出するインシデントは，暴力や器物破損のような他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検出しない．また，インシデントは表 2 のような分類に従ってラベルが付与される．実際に発生したインシデントの例としては，自分の衣類を破る，他者の作品を壊す，他者に掴みかかる，などが挙げられる．

インシデントは複数種類存在し，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．ただし本研究では，各記録はインシデントラベルの有無，または表 2 に示した大分類で分類する．インシデントラベルの有無の場合,「インシデント」「非インシデント」の 2 クラス分類を，大分類を用いる場合,「暴力のみ」「器物破損のみ」「暴力かつ器物破損」「インシデントなし」の 4 クラス分類を行う．

3.3　検出モデル

図 1 にインシデントの検出を行うモデルの概略図を示す．これは自然言語処理における事前学習モデルの一つである BERT を利用しており，E_i が入力系列を表し，T_i が出力系列を表している．また，Trm は Transformer [5] を表している．Transformer は Attention 機構を使用したニューラル機械翻訳モデルである．

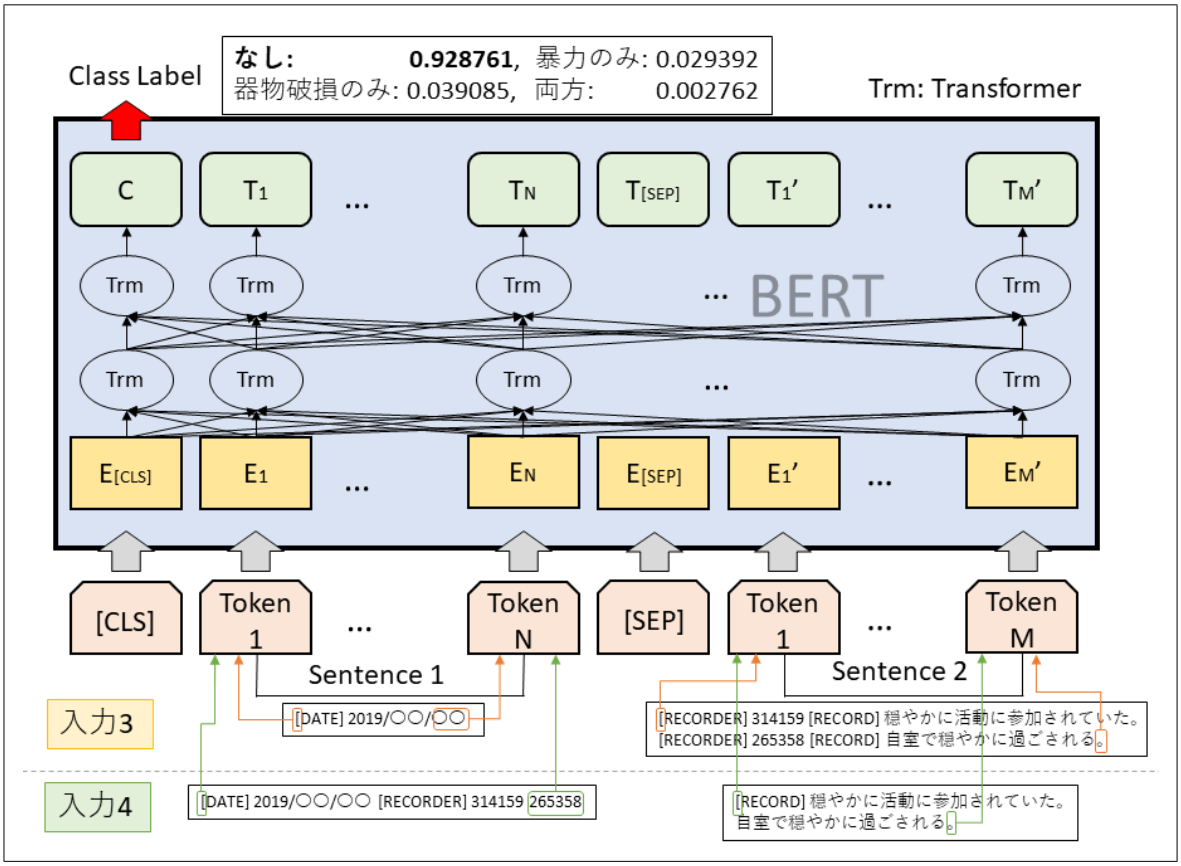

図 1　インシデントの発生検出モデルの概略図

表 3　モデルへの入力

	Sentence1	Sentence2
入力 1	記録本文	なし
入力 2	処理日付	記録本文
入力 3	処理日付	入力者コードと記録本文
入力 4	処理日付と入力者コード	記録本文

表 4　モデルへの入力例

記録本文	[RECORD] record1 record2
処理日付	[DATE] date
入力者コードと記録本文	[RECORDER] recorder1 [RECORD] record1 [RECORDER] recorder2 [RECORD] record2
処理日付と入力者コード	[DATE] date [RECORDER] recorder1 recorder2

BERT では特に Self-attention を利用して単語の重みを決定する．また本研究では入力の形式を実験において比較するため，入力の種類については後で説明する．

このモデルは入力として 1 文または 2 文を受け付ける．入力が 1 文の場合は 2 文目は None となる．None は値が存在しないことを意味する特殊な定数である．入力された文はまずトークンに分割され，文の切れ目の目印として 1 文目の頭に [CLS] トークンが，1 文目の終わりに [SEP] トークンが付与されたのち，2 文が連結される．次にトークンごとに埋め込み表現に変換され，Transformer 層で双方向での学習が行われ，結果が出力層に出力される．出力層の出力のうち，先頭のトークンにはクラスの識別結果が格納される．

BERT のモデルによる出力結果は各ラベルに対して 0 から 1 の範囲で正規化された確率であり，その総和は 1 になる．そのため，2 クラス分類では 2 つのラベルに対する確率が，4 クラス分類では 4 つのラベルに対する確率がそれぞれ出力され，入力に対する各ラベルの確率が最大のクラスに分類する．本研究では，入力を正しくインシデントに分類できればその発生を検出できたことになる．

本研究では分類器は事前学習済みモデルをファインチューニングすることで作成する．その際，支援記録は事前に利用者コードで分け，利用者ごとに分類器を作成する．また事前学習済みモデルとしては，BERT 日本語 Pretrained モデル [2]を利用する．

モデルに対する入力は複数用意し，実験により比較する．本研究では表 3 に示す 4 種類をそれぞれ与えて比較する．

図 1 の入力例は日ごとにまとめた記録の入力 3 が上の例，その入力 4 が下の例である．なお種類を区別するため支援記録の各項目の頭に [DATE] や [RECORDER] のように文字列ラベルを付与する．また，入力 3 の Sentence2 や入力 4 の Sentence1 のように 1 文に 2 つの項目を入力する場合は，文字列ラベルを付与後に単純に結合して入力する．

ここで、一日に二つの支援記録が含まれる日を例にして入力を説明する．このとき処理日付を date，一つ目の支援記録の

2：http://nlp.ist.i.kyoto-u.ac.jp/index.php?ku

入力者コードを recorder1，記録本文を record1，二つ目の支援記録のそれをそれぞれ recorder2， record2 とする．この時，表 3 に示す各入力は表 4 のようになる．記録ごと，または一つの支援記録のみが含まれる日の場合は record2 と recorder2 が消え，三つ以上の支援記録が含まれる日の場合は record3 と recorder3 以後が追加される．

4 評 価 実 験

4.1 実験の概要

4.1.1 実験に用いるデータ

実験には福祉支援施設に入所する男性 6 名，女性 4 名，計 10 名分の支援記録を用いる．実験では人物ごとにファインチューニングおよび分類を行う．また，記録単位と日付単位をそれぞれ分けて実験する．

各利用者ごとの支援記録は，表 5 のような内訳である．表中の「両方」とは，対象データに対して暴力と器物破損のインシデントラベルがどちらも付与されていることを意味する．「なし」はいずれのインシデントラベルも付与されていないことを意味する．

「あり (割合)」とは記録件数，日数それぞれの全データ中の「あり」とその割合である．「あり」はインシデントラベルのいずれかが付与されていることを意味し，表中の「暴力のみ」「器物破損のみ」「両方」の 3 つの値の和である．これより，いずれの人物もデータの 6 割以上が「なし」に分類されていることがわかる．したがってこの実験で扱うのは不均衡なデータである．一方，支援記録を記録単位ではなく日付単位で扱うことで「あり」の割合の値が増加し，この偏りは緩和される．

4.1.2 評価実験の内容

評価実験として提案モデルを 4.1.1 項で示したデータを用いて学習し，その性能を評価する．提案手法では，全データの 7 割をファインチューニング，1 割を検証，2 割をテストに用いる．

本稿では以下の (I)–(IV) の実験を行う．

(I) 提案手法と他の手法との性能比較 (入力 1)

(II) 提案手法による記録ごとの分類と日ごとの記録の分類の比較 (入力 1)

(III) 2 クラス分類と「4 クラス分類を 2 クラス分類に変換する場合」の比較 (入力 1)

(IV) 入力を変更した場合の比較

まず実験 (I) では，提案手法の有効性を確認するために提案手法以外の機械学習手法と性能を比較する．実験 (II) では，提案手法のモデルに記録本文のみを使用した場合の記録ごとの記録の分類と日ごとにまとめた記録の分類の結果をそれぞれ示し比較する．実験 (III) では，2 クラス分類と 3.2 節で示したインシデントの大分類を用いた 4 クラス分類を「インシデント」と「非インシデント」の 2 クラスに変換した場合を比較する．実験 (IV) では，表 3 のそれぞれの入力を比較する．

4.1.3 実験 (I) の比較対象の手法

性能比較のために scikit-learn[3] の以下の分類器を利用する．

- 非線形 SVM
- k-近傍法
- ランダムフォレスト

上記の手法を用いて日本語文を処理するため，入力テキストは Bag of Words (BoW) でベクトルに変換する．BoW は文書に単語が含まれている頻度のみを考えて，単語の順序は考慮しないモデルである．BoW に対して TF-IDF による重みづけを行ったのち，LSI(Latent Semantic Indexing) により，300 次元まで次元圧縮する．文の単語単位での分かち書きには MeCab[4] を，ベクトル変換，重みづけ，次元圧縮には gensim[5]を利用した．また，非線形 SVM，k-近傍法，ランダムフォレストの分類器のパラメータは下記の通り設定した．なお，記載のないパラメータはすべてデフォルトのままとした．

非線形 SVM 独自の設定として，正規化項の係数 C=10，rbf カーネルを使用し，カーネルの係数 γ=0.1 とした．k-近傍法の設定として，k=5 とし，近傍点の評価に距離も考慮する．ランダムフォレストの設定として，決定木作成の評価指標はジニ係数，作成する決定木は 10，クラスごとに重みは自動的に補正されるようにした．また，$random\_state$=1，n_jobs=2 とした．

各分類器は，ラベル付けした支援記録データを無作為に分割し，全データの 8 割を学習，2 割をテストに用いる．このテストデータは，提案手法の評価に用いるものと同一である．出力は,「インシデント」「非インシデント」の 2 クラスであり，これを評価に用いる．

4.1.4 評 価 指 標

モデルの評価には，正解率 (Accuracy)，再現率 (Recall)，適合率 (Precision)，F 値 (F-measure) を利用する．それぞれの値を求める式は以下のようになる．

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{F-measure} = \frac{2\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \tag{4}$$

なお，

- TP: インシデントに分類されたインシデント記録
- FP: インシデントに分類された非インシデント記録
- FN: 非インシデントに分類されたインシデント記録
- TN: 非インシデントに分類された非インシデント記録

である．

F 値は再現率と適合率の調和平均である．なお，再現率はすべてのインシデントの記録に対し，分類器がインシデントに分類したものの割合であり，適合率は分類器がインシデントに分

3：https://scikit-learn.org/stable/

4：https://taku910.github.io/mecab/

5：https://radimrehurek.com/gensim/

表 5　支援記録の内訳

人物	性別	記録件数						日数					
		暴力のみ	器物破損のみ	両方	あり (割合)	なし	合計	暴力のみ	器物破損のみ	両方	あり (割合)	なし	合計
A	女性	100	5	1	106(4.39%)	2310	2416	94	5	1	100(7.23%)	1284	1384
B	女性	188	13	3	204(5.93%)	3235	3439	157	11	3	171(12.35%)	1214	1385
C	女性	289	311	30	630(19.09%)	2670	3300	186	223	59	468(36.11%)	828	1296
D	女性	35	30	22	87(4.53%)	1834	1921	29	27	22	78(10.44%)	669	747
E	男性	154	1	1	156(4.11%)	3641	3797	142	1	1	144(7.57%)	1759	1903
F	男性	49	3	0	52(1.57%)	3271	3323	46	3	0	49(2.56%)	1862	1911
G	男性	84	13	2	99(2.92%)	3288	3387	77	12	2	91(4.76%)	1822	1913
H	男性	5	10	0	15(0.44%)	3425	3440	5	10	0	15(0.79%)	1897	1912
I	男性	58	13	2	73(1.91%)	3742	3815	54	12	3	69(3.61%)	1842	1911
J	男性	4	2	0	6(0.16%)	3873	3879	4	2	0	6(0.31%)	1904	1910

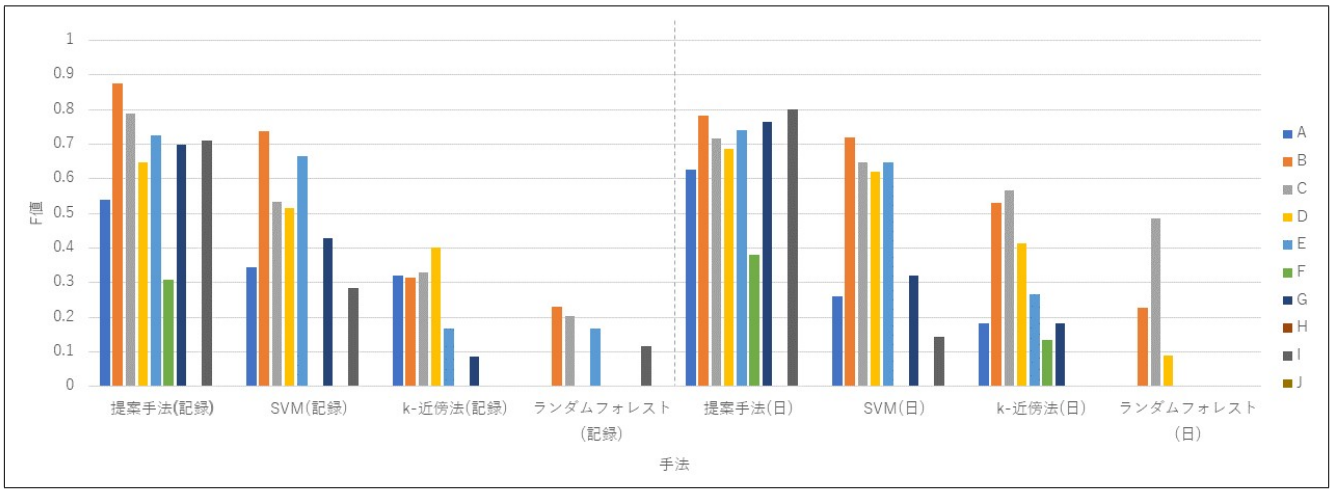

図 2　提案手法と比較対象の手法の F 値

類したデータのうち，実際にインシデントの記録であるものの割合である．

インシデントの検出という観点では，再現率はどれだけインシデントを見逃さないか，適合率はどれだけ誤検出を防げるかを示しており，F 値はそれらを総合的に評価した指標となる．

また，インシデントと非インシデントの両方に係る正解率は，インシデントの正解率が低くても非インシデントの正解率が高い場合，値は大きくなる．そのため，データが非インシデントに偏っている本研究ではインシデント検出の評価指標としてあまり有効とは言えない．

4.2　実 験 結 果

4.2.1　実験 (I): 提案手法と他の手法との比較

2 クラス分類の結果の F 値を図 2 に示す．図は横軸が手法，縦軸が F 値，各棒は各人物の分類結果を表す．この図よりまず人物 H と J はいずれの分類手法でも一切分類ができていないことがわかる．これは表 5 に示すように H と J はインシデントの割合が 1%未満と特に少ないためであると言える．

各分類手法について見てみると，ランダムフォレストはほとんどの人物に対して結果が出ておらず，k-近傍法は F 値が提案手法に比べてかなり低い．SVM は一部の人物では提案手法に近い F 値を示しているものの，10 人すべての人物について提案手法を上回ったものは一つもない．したがって，いずれの人物に対しても提案手法が最も高い F 値となったことがわかる．よって，他の手法と比較して提案手法を用いる有効性が確認されたと言える．

4.2.2　実験 (II): 記録の分類と日ごとの記録の分類の比較

BERT を用いる提案手法による 2 クラス分類の結果を表 6 に示す．表中の「–」は評価指標の算出式の分母の値が 0 となり値が存在しなかったことを意味する．よって，H と J の 2 人は全くインシデント検出ができていないことがわかる．そのため以下では H と J を除く 8 人に関して比較する．また，表 6 の数値の下線は F 値に関して，記録の分類と日ごとの記録の分類で比較した場合に値が大きい方を意味する．

記録と日でまとめた記録の分類の結果を比較する．まず正解率は，記録と比較して日ごとになると 8 人全員で低くなった．これは非インシデントの正解率の影響が大きく，記録単位から日付単位にすることでデータ中の非インシデントの割合が小さくなるためである．再現率は，B，C の 2 人は記録に比べて日で低下し，A は変化なし，D，E，F，G，I の 4 人は上昇している．適合率は，B，C，D，E の 4 人は記録に比べて日で低下し，A，F，G，I の 4 人は上昇している．その結果，F 値は B，C の 2 人は記録に比べて日で低下し，A，D，E，F，G，I の 6 人は上昇した．このことから，8 人それぞれの F 値を比較すると日ごとの記録の方が記録よりも良い結果を示していると

表 6　記録の分類と日ごとの記録の分類

人物	記録の分類				日ごとの分類			
	正解率	再現率	適合率	F 値	正解率	再現率	適合率	F 値
A	0.9656	0.5000	0.5882	0.5405	0.9579	0.5000	0.8333	0.6250
B	0.9840	0.8478	0.9070	0.8764	0.9474	0.7105	0.8710	0.7826
C	0.9260	0.7459	0.8349	0.7879	0.8137	0.7381	0.6966	0.7168
D	0.9665	0.5217	0.8571	0.6487	0.9295	0.6000	0.8000	0.6857
E	0.9786	0.6364	0.8400	0.7241	0.9687	0.6800	0.8095	0.7391
F	0.9866	0.2500	0.4000	0.3077	0.9662	0.3333	0.4444	0.3810
G	0.9809	0.7143	0.6818	0.6977	0.9792	0.7222	0.8125	0.7647
H	0.9971	0.0000	–	–	0.9974	0.0000	–	–
I	0.9880	0.6875	0.7333	0.7097	0.9870	0.7692	0.8333	0.8000
J	1.0000	0.0000	–	–	0.9974	0.0000	–	–

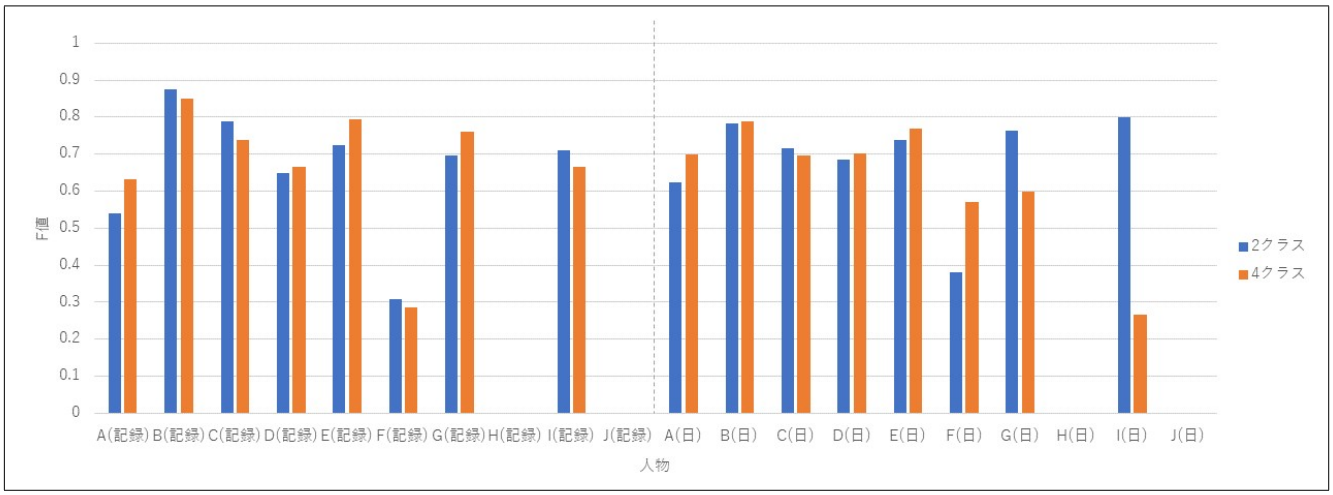

図 3　2 クラス分類と 4 クラス分類の F 値

言える．

4.2.3　実験 (Ⅲ): 2 クラス分類と 4 クラス分類の比較

2 クラス分類は「インシデント」と「非インシデント」の分類である．4 クラス分類は,「暴力のみ」「器物破損のみ」「両方」「非インシデント」の 4 クラスに分類した後,「非インシデント」以外の 3 クラスをまとめ 2 クラスに変換する結果を図 3 に示す．横軸は人物，縦軸は F 値，棒は各分類手法の結果を表す．人物 H と J については 4 クラス分類でも全く検出できなかった．

図 3 より元々の 2 クラス分類と 4 クラスを 2 クラスに変換した場合のどちらが有効かは人物によって異なり，あまり大きな差は見られない．一方，人物 F，G，I の 3 人の日ごとの記録の 4 クラス分類を 2 クラスに変換した場合には大きな差が見られる．4 クラス分類の場合，4 つのクラスに対するそれぞれの確率の値が最大のクラスに分類するため，複数ラベルが近い確率を出力することも多い．表 5 より人物 F，G，I は全く検出ができていない人物 H と J に次いでインシデント割合が小さい．そのため，不均衡データにより学習が安定しておらず，その結果として分類の正誤が大きく変動した可能性がある．

4.2.4　実験 (Ⅳ): 入力を変更した場合の比較

入力として表 3 に示した 4 種類の入力を用い，それぞれの分類結果を比較する．結果を図 4 に示す．横軸が人物，縦軸が F 値，棒は各入力の結果を表す．人物 H と J はすべての入力においてインシデントを全く検出できず，F 値が算出できなかった．

図 4 より，常に有効な入力は明らかではないことがわかる．例えば入力 1 と他の入力を比較しても，いずれかが常に最高の結果を出しているわけではないことがわかる．

4.3　考　　察

4.3.1　分類手法に関する考察

図 2 より，記録本文のみを利用して分類する場合，提案手法は他の手法と比較して有効であると言える．他の機械学習手法の場合，用意したデータのみで学習を行うため少数データでは学習が不十分となり，十分な性能を発揮できないことが多い．一方提案手法に用いる BERT は学習済みモデルをファインチューニングするため，少数データでも事前学習と合わせて高い性能を発揮することができる．そのため，表 5 に示したようにインシデントのデータが少ない本研究では提案手法が有効であった．

4.3.2　記録の分類と日ごとの記録の分類に関する考察

記録の分類と日ごとの記録の分類に関して，表 5，表 6 よりインシデントデータの割合が相対的に多い人物は記録，それ以外の人物は日ごとの分類が良い結果となっていることがわかる．ここで，記録件数と日数では日数の方がインシデントデータの割合が大きいことを踏まえると，本研究のようにインシデントデータが少ない場合，インシデントデータの割合が大きい方が

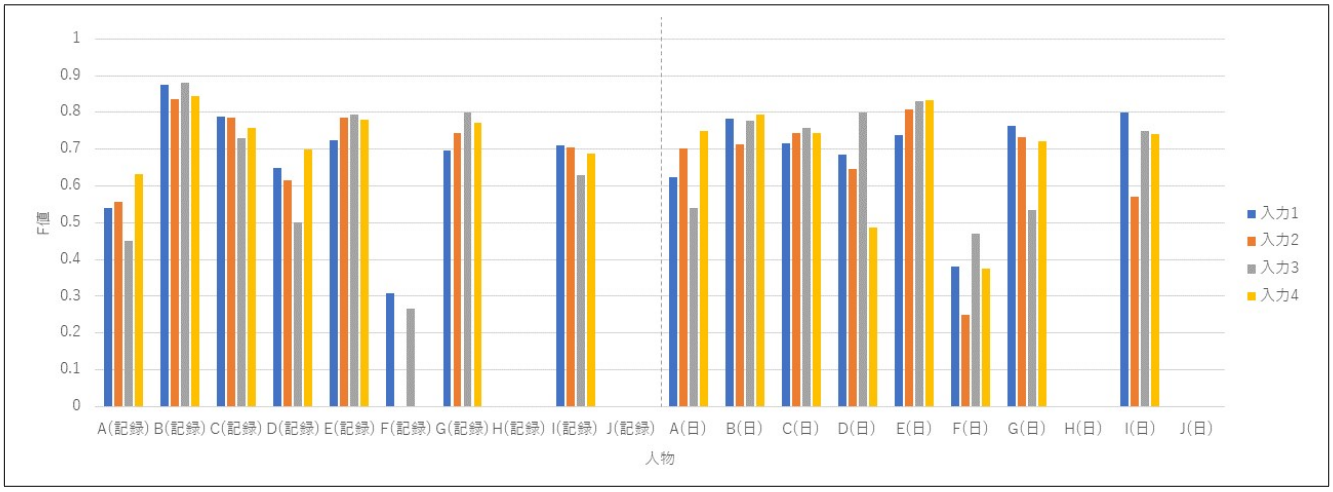

図 4　4 種類の入力に対する F 値

よい結果が出ると言える．また，実際の支援記録には多くのインシデントデータが含まれるとは限らない．そのため，基本的には記録を日ごとで扱った方が分類の F 値が高くなると言える．

4.3.3　分類クラスに関する考察

表 5 に示した，4 クラスのデータ割合に着目する．まず A，B，E，F，G，I の 6 人はインシデントのうち「暴力のみ」の割合が他のインシデントに比べて大きく最大である．このうち G，I を除く 4 人は「暴力のみ」が全インシデントの 9 割以上，G，I は「暴力のみ」の割合が 8 割程度を占めており，2 クラス分類と 4 クラス分類にあまり差がないと言える．

一方，C には「暴力のみ」と「器物破損のみ」が近い割合で含まれ，D は「暴力のみ」と「器物破損のみ」と「両方」のいずれも近い割合となっている．ここで図 3 より，C は通常の 2 値分類が，D では 4 クラスを 2 クラスに変換する分類がそれぞれ F 値が高くなっている．このことから，インシデント種類が偏っていなくても 2 つの分類手法に明確な優劣はないことがわかる．

4.3.4　入力形式に関する考察

入力の形式に関して図 4 の分類結果について考える．記録の分類の場合，入力 1 は人物 C，F，I，入力 3 は人物 B，E，G，入力 4 は人物 A，D に関して，それぞれ全入力の中で F 値が最も大きい．入力 2 はいずれの人物においても F 値が最大の入力となることはなかった．また日ごとの分類の場合，入力 1 は人物 G，I，入力 3 は人物 C，D，F，入力 4 は人物 A，B，E に関して，それぞれ全入力の中で F 値が最も大きい．また入力 2 はいずれの人物においても F 値が最大の入力となることはなかった．

そのため，記録本文に何かの情報を付加しても必ずしも F 値が向上するとは限らないが，日付情報や入力者情報などを付加することで F 値が向上しうることが示唆された．

4.3.5　事 例 分 析

表 7 に実験 (Ⅱ) の記録の分類で誤分類された記録の例を示す．なおこの分類では人物 E の記録本文のみを利用している．表 7 の (1)，(2)，(3) が非インシデントと分類されたインシデントの記録，(4)，(5) がインシデントと分類された非インシデントの記録である．

まず，(1)，(2)，(3) には「服をつかむ」「叩く蹴るの暴行」「手が出る」のような直接的な他害を表す語が含まれているのにも関わらず非インシデントに分類された．しかし，(2) に関しては「とても笑顔」「すぐに笑顔」といったインシデントには似つかわしくない語が含まれているため，それが影響した可能性がある．

次にインシデントに分類された (4)，(5) には「叩かれる」「窓を叩く」など直接的な他害に関する言葉が含まれている．実際には他者からのインシデント被害と器物破損が発生しない程度の叩くではあるが，それが判定に影響した可能性がある．

このように支援記録には，記録本文の文面のみで分類するのは難しい事例が存在する．そのため，精度向上のためには記録本文のみではなく，日付情報や統計的な分析情報などの追加の情報の入力が必要であると考える．

4.3.6　インシデント発生予測に関する考察

本研究の目的の 1 つは，インシデント発生予測の実現可能性を探ることである．このインシデント発生予測は，ある時点までの支援記録からその次の未来のインシデントの発生の有無を予測するタスクである．本稿の実験より支援記録からインシデント発生検出を行う場合，F 値 0.7 程度の性能を発揮することがわかった．一方，インシデント発生予測を行う場合，発生検出より困難であることが予想される．

この要因としては，インシデント発生予測のための手掛かりが検出のための手掛かりよりも不明瞭であることが挙げられる．発生検出の場合，表 7 の支援記録のように「服をつかむ」「叩く蹴るの暴行」「他者に手が出る」などの表現が手掛かりとなり得る．一方発生予測の場合，インシデント発生の予兆となるような手掛かりがあるとは限らず，また手掛かりがあったとしても予兆の検出とそれを利用した発生予測はまた別の問題となる可能性がある．したがって，専門家の力を借りて予兆などについて慎重に検討する必要がある．

表 7　誤分類された人物 E の支援記録の例

	記録	インシデントラベル
(1)	起床時、職員の服をつかんでくる。理由は不明。食堂の方を気にされている。	暴力#職員
(2)	入浴前はとても笑顔だが、なかなか入浴しない。入浴後は、水分の要求があり、職員に叩く蹴るの暴行があった。 しかし、廊下に出ると走って居室に入り、その後すぐに笑顔になる。	暴力#職員
(3)	他者に手が出ることがある。居室に促すと不服そうな顔をされる。	暴力#施設の利用者
(4)	紙に○を描く作業を行う。途中、座席から滑り落ちた際、_____さんに叩かれる場面あったが、気にせず作業される。	なし
(5)	ドライブ：イライラしており、大きな声を出したり、窓を叩くなどの行動が見られた。	なし

5 ま と め

本研究では，インシデントの発生検出のため，BERT を利用した支援記録の分類モデルを提案し，その分類性能を実験により評価した．提案モデルは支援記録の本文や日付情報などを入力として受け取り，入力された記録を分類する．その際，インシデントか非インシデントかの 2 クラス，またはインシデントの種類に応じた 4 クラスに分類する．

実験では提案手法とその他の手法で実際の支援記録を分類して比較した．また，提案手法の入力や利用する情報の種類を変えて分類実験を行った．その結果，提案手法は SVM，k-近傍法，ランダムフォレストのいずれの手法も上回り，提案手法の有効性が確認できた．また，支援記録を記録と日でそれぞれ分類した結果を比較すると，日でまとめる方が分類結果の F 値が高い人物が多かった．また，2 クラス分類と，4 クラスで分類した後で 2 クラスに集約した結果を比較したところあまり差が見られなかった．さらに入力データとして記録本文に別の情報を加えたところ，日付情報と記入者情報を加えることが有効な場合があるとわかった．

今後は，インシデント発生検出精度の向上に取り組みたい．例えば，インシデントの発生間隔などの統計的情報を加えることを検討したい．また，インシデントの発生予測への展開も今後の課題である．

謝　　辞

本研究は株式会社岡山システムサービスとの共同研究である．本研究で使用したデータは同社より提供を受けた．

文　　献

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “Bert: Pre-training of deep bidirectional transformers for language understanding.” arXiv preprint arXiv:1810.04805, 2018.

[2] 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.

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

[4] 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.

[5] 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 30, pp. 5998–6008, 2017.

[6] 大友泰賀，張建偉，中島伸介，李琳．“いじめ表現辞書を用いた Twitter 上のネットいじめの自動検出．”DEIM2020，C7-1，2020.

[7] 酒井優介，藤ノ木太郎，安藤雅洋，湯川高志．“ディープラーニングを用いた暗示的怒りの自動検出手法．”第 33 回人工知能学会全国大会，4M3-J-9-04，2019.

[8] 松本典久，上野史，太田学．“BERT を利用した煽りツイート検出の一手法．”DEIM2021，I14-2，2021.