

様々な災害ドメインのクロノロジーに対する優先度推定

孝壽真治¹ 竹内孔一²渡邊暁洋³ 平山隆浩⁴ 中尾博之⁵¹ 岡山大学大学院環境生命自然科学研究科 ² 岡山大学学術研究院環境生命自然科学学域³ 兵庫医科大学危機管理医学講座 ⁴ 岡山大学学術研究院医歯薬学域⁵ 大阪行岡医療大学医療学部理学療法学科

s-kouju@s.okayama-u.ac.jp takeuc-k@okayama-u.ac.jp

ph-akihiro@s.okayama-u.ac.jp ce-hirayama@okayama-u.ac.jp

nakaonakaokobe@yahoo.co.jp

概要

災害対策本部では連携機関からの報告や要請をもとに、その優先度に応じて医療リソースの分配を決定している。これらの要請などは時系列形式の文書としてまとめたクロノロジーと呼ばれる形式で記録される。本研究では、クロノロジーの優先度を推定する機械学習モデルを構築し、学習データとは異なる災害で作成されたクロノロジーにおける優先度の推定精度を検証する。実験の結果、GPTNeoX-3.6bを利用したモデルが、学習データとは異なる災害のクロノロジーの優先度推定において、特に有効であることが明らかとなった。

1 はじめに

災害が発生すると、災害対策本部には連携する病院や診療所と様々な情報のやり取りを行う。これらの情報には、各施設の被災状況などの報告と、災害対策本部に対する物資や保険医療に関する支援の要請が含まれる。一方で、災害対策本部が提供できるスタッフや物資には限界があるため、これらのリソースを適切に配分する必要がある。そのため、災害対策本部に届く大量の情報を整理して混乱を防ぎ、連携機関への指示やリソースの分配を適切に行うために、クロノロジーと呼ばれる経時活動記録を作成している。

クロノロジーには災害対策本部と連携機関との間のやり取りを時系列順に全て記録する。そのため、クロノロジーの中には人命に関わるような対応の優先度が高いものと、単なる報告のように災害対策本部が対応する必要性が薄く優先度が低いものが混在している。そのため、災害対策本部の職員はこれら

のクロノロジーの優先順位を判断して災害対応を行っている。

そこで、クロノロジーの優先度を高・中・低の3段階で定義し、その優先度を機械学習モデルで推定することで、災害対策本部職員を補助し、負担を軽減して災害対策に専念できるように研究を行う。本研究では特に、特定の災害のクロノロジーで学習したモデルが、他の災害のクロノロジーに対してどの程度優先度を推定できるかを検証する。実験の結果、GPTNeoX-3.6bを利用したモデルが、学習データとは異なる災害のクロノロジーの優先度推定において、特に有効であることが明らかとなった。

2 関連研究

災害時における文書処理に関連した研究では実システムと結びつけた研究が展開されている。クロノロジーのように時系列で記録する方法をベースに災害情報を共有するシステム「災害ネット」[1]、災害時における Twitter などを含む SNS の情報を活用するシステム「DISAANA」[2]¹⁾、「高度自然言語処理プラットフォーム」[3]が提案されている。一方で、後者の2件は SNS の文書を利用しており、クロノロジーではない。また、前者の研究はクロノロジーを利用してはいるが、その優先度は推定していないため本研究とは異なる。

さらに、文書の分類の際には文書をベクトル化してニューラルネットワークを用いて識別する手法が近年は一般的となっている。文書のベクトル化手法としては、文書を語彙空間でベクトル化する BoW や、Transformer[4]の Encoder 部分をベースにする BERT[5]、Decoder 部分をベースにする GPT-3[6]や、

1) 現在 DISAANA は 2023 年 12 月 28 日に終了した。

オープンソースモデルの GPT-NeoX-20B[7] がある。

クロノロジーの優先度を上記のような文書分類の手法を用いて推定する研究 [8, 9] がある。これらの研究では、GPT-NeoX を用いたモデルが優先度の推定において高い性能を示している。また、2 番目の性能を示すモデルが優先度高では BoW、優先度中では BERT と逆転するという特徴が見られることが示されている。一方で、これらの研究では西日本豪雨に対するクロノロジーデータのみを使用しており、学習データとは異なる災害で作成されたクロノロジーに対するモデルの有用性や、豪雨災害以外の災害ドメインにおける有効性については研究がなされていない。

そこで本研究では、先行研究 [9] で提案したモデルを基にして、西日本豪雨のクロノロジーを用いて学習を行い、モデルを構築する。そして、豪雨災害ドメインで学習したモデルが異なる災害ドメインにおいても有効であることを明らかにする。

3 データ

本研究で使用するクロノロジーとその優先度、構築したデータセットについて説明する。

3.1 クロノロジー

本研究では、実際の災害において作成されたクロノロジーを使用する。クロノロジーは、災害発生時に災害対策本部でやり取りされる情報に対して、日時・発信者・受信者・報告内容などの項目をホワイトボードに記録したものであり、これは表 4 のような形式で電子化されている。

3.2 優先度

先行研究 [8, 9] と同様に災害医療のトリアージの考え方を参考に重要度と緊急度の 2 つの観点から優先度を分類している。重要度は生命の危機に関することであれば高いとし、緊急度は時間的猶予が無い場合に高いとする。そして優先度は、重要かつ緊急度が高いクロノロジーを優先度を高とし、どちらか一方でも高いものを中、どちらも低いものを低として分類する。この基準をもとに、2 人のアノテーターが同じクロノロジーに対して優先度を付与し、2 人の判定の平均をそのクロノロジーの優先度としている。

3.3 データセット

先行研究 [9] では西日本豪雨 (2018 年) におけるクロノロジーを用いて実験を行った。本研究では、新たに熊本豪雨 (2020 年)、熊本地震 (2016 年)、北海道胆振東部地震 (2018 年) の 3 つのクロノロジーデータを用いてデータセットを構築する。各クロノロジーにおける優先度の内訳とデータの件数は表 1 である。

表 1 実験に使用するデータの優先度の内訳

優先度	西日本豪雨	熊本豪雨	熊本地震	北海道胆振東部地震
低	4084 (0.906)	1272 (0.943)	4951 (0.892)	1719 (0.883)
中	377 (0.084)	71 (0.053)	534 (0.096)	209 (0.107)
高	48 (0.011)	6 (0.004)	67 (0.012)	18 (0.009)
合計	4509	1349	5552	1946

* データの件数 (割合) の形式で記載する

なお、実験ではクロノロジーの内容の項目部分のみを使用するため、実験用のクロノロジーデータセットはクロノロジーの内容のテキストとアノテーターが付与した優先度ラベルの組である。

4 検証モデル

本研究では、先行研究 [9] の提案モデルを基にした以下の 4 つのモデルを構築する。

4.1 BoW を利用したモデル

先行研究 [9] と同様に 3 層ニューラルネットワークを用いる。中間層のユニット数を入力層に近い方から順に、300, 100 とし、出力層のユニット数は 3 とする。入力層のユニット数は BoW の語彙数と同じだが、本研究では先行研究と比較して語彙数が増加している。これは、先行研究では西日本豪雨のみのデータを扱っていたのに対して、本研究では複数の災害ドメインのデータを扱うためである。このモデルは実験結果の表では BoW と表記する。

4.2 BERT を利用したモデル

本研究では、東北大学が作成した事前学習済み BERT²⁾ を利用したモデルを 2 種類構築する。1 つ目は、先行研究 [9] と同様に [CLS] トークンに対応する埋め込みを BERT の最終層から 4 層使用するモデルである。このモデルではクラス分類用の層と BERT の最終 4 層のパラメータのみを更新して学習

2) <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking> 2024/1/5 アクセス

を行う。このモデルは実験結果の表では BERT-last4 と表記する。

2つ目は、BERT の全ての”query”, ”key”, ”value” の全結合層に対してランク 8 の LoRA[10] モジュールを導入し、LoRA チューニングを行うモデルである。クラス分類には最終層の [CLS] トークンに対応する埋め込みを用いる。このモデルは実験結果の表では BERT-LoRA と表記する。

4.3 GPTNeoX-3.6b を利用したモデル

先行研究 [9] と同様に、rinna 株式会社が作成した事前学習済み GPTNeoX-3.6b³⁾ を利用したモデルを構築する。GPTNeoX-3.6b の全ての”query_key-value” の全結合層に対してランク 8 の LoRA モジュールを導入し、LoRA チューニングを行うモデルである。クラス分類には最終層の [EOS] トークンに対応する埋め込みを用いる。このモデルは実験結果の表では GPTNeoX-3.6b と表記する。

5 実験

本実験では、西日本豪雨のデータを用いて学習を行い、西日本豪雨以外の豪雨災害や異なる災害ドメインである地震のデータにおけるモデルの有効性について検証する。

5.1 実験設定

本実験では、4章の4つのモデルに対して、西日本豪雨を学習に用いてモデルを構築し、その他のデータセットに対する優先度の推定精度を検証する。

学習に用いた西日本豪雨のデータセットのうち、全体の $\frac{1}{6}$ を開発データとして利用する。残りのデータを学習データとテストデータに用いて 5-Fold Cross Validation を行う。また、学習に使用していない熊本豪雨、熊本地震、北海道胆振東部地震のデータセットについては、全てのデータをテストデータとして使用する。

5.2 学習設定

クロノロジーの優先度ラベルは、高・中・低の3段階のクラスなので、順序回帰の手法を用いて学習する。順序回帰にはいくつかの手法が存在するが [11]、本研究では損失関数の目標出力にソフトラベルを与える手法 [12] を用いた。よって、学習時の損

失関数の目標出力には、正解クラス \bar{k} とあるクラス k ($k = 1, 2, \dots, K$) との距離を $|\bar{k} - k|$ として、式 (1) で求められる表 2 のようなソフトラベルを与える。

$$d_k = \frac{\exp(-|\bar{k} - k|)}{\sum_{i=1}^K \exp(-|\bar{k} - i|)} \quad (1)$$

表 2 各優先度における目標出力

優先度	ハードラベル	ソフトラベル
低	[1, 0, 0]	[0.665, 0.245, 0.090]
中	[0, 1, 0]	[0.212, 0.576, 0.212]
高	[0, 0, 1]	[0.090, 0.245, 0.665]

全てのモデルにおいて、学習時の損失関数には高・中・低それぞれに対して $4 \cdot 2 \cdot 1$ の重みを与えて学習した。

また、学習では以下の停止条件で EarlyStopping を適用する。

- 開発データの優先度高・中の F 値の平均が 10 回更新されなかった場合に停止
- 開発データに対する優先度高・中の F 値の平均が最良の epoch の学習結果を利用する

なお、学習はミニバッチサイズ 16 で実施し、全ての実験において EarlyStopping により学習が停止した。

5.3 実験結果

本実験の評価指標には F 値を用い、実験結果として 5-Fold の各 F 値を平均して百分率表記にした値を表 3 に示す。

表 3 実験 1：豪雨ドメインの結果

モデル	西日本豪雨	熊本豪雨	熊本地震	北海道胆振東部地震
優先度：高				
BoW	35.40	10.44	10.82	7.22
BERT-last4	27.85	13.47	20.80	12.89
BERT-LoRA	37.53	14.18	19.06	12.97
GPTNeoX-3.6b	35.19	20.19	23.59	17.99
優先度：中				
BoW	40.90	26.39	40.20	41.38
BERT-last4	39.40	23.04	39.88	38.13
BERT-LoRA	39.59	25.91	41.16	37.76
GPTNeoX-3.6b	43.90	33.12	48.09	46.13
優先度：低				
BoW	91.85	91.93	90.48	89.62
BERT-last4	93.63	92.22	92.68	91.40
BERT-LoRA	93.28	92.63	92.31	90.99
GPTNeoX-3.6b	94.16	93.38	93.75	91.22

3) <https://huggingface.co/rinna/japanese-gpt-neox-3.6b> 2024/1/5 アクセス

表 3 の実験結果では、西日本豪雨データセットの優先度高は BERT-LoRA の F 値が最も高く、北海道胆振東部地震データセットの優先度低は BERT-last4 の F 値が最も高い。それ以外の全ての結果は GPTNeoX-3.6b の F 値が最も高い。

6 考察

西日本豪雨は学習に用いたデータのため、そのテストデータに対する推定精度は、全てのモデルにおいて他の災害のデータと比べて顕著に高くなっている。しかし、今回の実験では西日本豪雨の優先度高の推定結果において、GPTNeoX-3.6b よりも BERT-LoRA の方が高く、BoW について GPTNeoX-3.6b は 3 番目となっている。先行研究 [9] とは異なり、優先度高における BoW の性能が GPTNeoX-3.6b よりも僅かに高いが、これは学習データ、開発データ、テストデータの割合を変更したことに加えて、他の災害のデータを扱うために BoW の語彙を拡張したことが影響していると考えられる。また、優先度高において、BERT-last4 の推定精度が低い点は先行研究の結果と同様である。

一方で着目すべき点は、BERT-LoRA の結果が西日本豪雨の優先度高において、BERT-last4 と比べて顕著に向上している点である。LoRA の適用によってさらに学習パラメータ数を抑えつつ、最終 4 層のみを学習させた際よりも結果が向上している。このことから、BERT はクロノロジーの優先度高の推定において BoW や GPTNeoX-3.6b と同等以上の能力を有しているが、ファインチューニングの手法によっては、BERT-last4 のようにその性能を十分に発揮できない場合があると考えられる。

さらに、本実験において学習に使用したクロノロジーは西日本豪雨のものである。そのため、同じ豪雨災害である熊本豪雨の方が地震のデータに比べて推定精度高くなると予測が立つが、実際には優先度高と中の結果ではほぼ全てのモデルにおいて熊本地震の推定精度が高い。例外として、北海道胆振東部地震の優先度中の BoW の結果だけは熊本地震よりも高いが、熊本豪雨よりも北海道胆振東部地震の結果が高いため、豪雨ドメインとは異なる地震ドメインの方が推定精度が高いと言える。また、西日本豪雨以外のクロノロジーの優先度高と中においては、GPTNeoX-3.6b が常に高い精度となっており、他のモデルと比較して特に汎化性能が高いことがわかる。一方で、BoW の優先度高は、学習を行った西

日本豪雨では高い精度であるのに反して、その他のデータにおいては非常に低くなっており、汎化性能に乏しいことが分かる。

これらの点から、災害の種類の違いは推定精度にあまり影響を及ぼさない可能性が示唆されているが、これについては更なる検証が必要である。なぜなら、GPTNeoX-3.6b が西日本豪雨以外のデータで高い性能を示しており、クロノロジーの優先度に関する一般的な知識を獲得できていると考えられるものの、西日本豪雨の優先度高では BERT-LoRA や BoW に劣っている。この原因として、BERT-LoRA や BoW が学習データに対して過学習を起こしている可能性と、西日本豪雨のクロノロジーが他災害のクロノロジーと比べて特殊な事例が含まれる可能性が考えられる。そのため、西日本豪雨のクロノロジーに特殊な事例が含まれる場合は、災害の種類の違いによる推定精度への影響の評価には更なる検証が必要だといえる。

以上のことから、データセットごとのクロノロジーの特性については、更なる検証が必要であるものの、GPTNeoX-3.6b の汎化性能は他モデルと比較して高いと言える。また、BERT の全層に LoRA を適用して学習することで、BERT の最終 4 層のみを学習させる先行研究の BERT モデルのから、性能の向上が見られた。

7 おわりに

本研究では、西日本豪雨災害のクロノロジーデータを用いてモデルを構築し、西日本豪雨とは別の豪雨災害や、豪雨災害とはドメインの異なる地震のクロノロジーデータに対するモデルの有効性について検証を行った。実験の結果、GPTNeoX-3.6b を利用したモデルが、学習データとは異なる災害のクロノロジーの優先度推定において、特に有効であることが明らかとなった。一方で、各災害におけるデータセットごとのクロノロジーの特性については更なる検証が必要であり、今後の課題である。

謝辞

本研究の一部は厚生労働科学研究費補助金(21IA2401)(数理最適化モデルによる小学校グリッドに基づく他組織連携システム(MACS)の解析)(略称A-MACS)に基づいて実施された。

参考文献

- [1] UNISYS. クロノロジー型危機管理情報共有システム 災害ネット, 2022. (<https://www.unisys.co.jp/solution/biz/disaster-net/> 2022/05/21 アクセス).
- [2] NICT. 対災害 SNS 情報分析システム (DISAANA), 2020. (<https://www.nict.go.jp/resil/> 2023/5/21 アクセス).
- [3] NEC. Twitter 上の災害に関する情報をリアルタイムで解析・可視化する「高度自然言語処理プラットフォーム」, 2020. (https://jpn.nec.com/press/202006/20200626_01.html 2023/05/21 アクセス).
- [4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All You Need. **arXiv preprint arXiv:1706.03762**, 2017.
- [5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018.
- [6] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Nee-lakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.
- [7] Sidney Black, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, Michael Pieler, Usvsn Sai Prashanth, Shivanshu Purohit, Laria Reynolds, Jonathan Tow, Ben Wang, and Samuel Weinbach. GPT-NeoX-20B: An open-source autoregressive language model. In **Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models**, pp. 95–136, virtual+Dublin, May 2022. Association for Computational Linguistics.
- [8] 孝壽真治, 竹内孔一, 渡邊暁洋, 平山隆浩, 中尾博之. 災害医療におけるクロノロジーの優先度識別. 研究報告 情報基礎とアクセス技術 (IFAT), Vol. 2023-IFAT-149 (3), pp. 1–5, 2023.
- [9] Shinji Koju, Koichi Takeuchi, Akihiro Watanabe, Takahiro Hirayama, and Hiroyuki Nakao. Estimating task priority in japanese disaster chronology logs. In **Proceedings of the 2023 7th International Conference on Natural Language Processing and Information Retrieval**, 2023. (In press).
- [10] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. **CoRR**, Vol. abs/2106.09685, , 2021.
- [11] 岡谷貴之. 深層学習改訂第2版. 講談社, 2015.
- [12] Raul Diaz and Amit Marathe. Soft Labels for Ordinal Regression. In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**, 2019.

A 参考情報

表4 電子化されたクロノロジーの例

年月日	時刻	発信者	受信者	内容
2018/7/9	09:50	○大○○さん	岡山○○ Dr.	昨日までの流れをプレゼン 今日の方針を話し合い
2018/7/9	14:15	本部 DMAT ○○○	県 DMAT 本部	DMAT ○○ 1 チーム 2 チーム DMAT を要請
2018/7/9	16:59	○○医療センター	KuDRO 本部	[○○病名の略称] ○○消防救急車で○大へ ○○才女性

* 文章の一部を記号に置き換えている