

豪雨災害におけるクロノロジーの優先度推定に対する LLM を用いた修正手法の検討

孝壽真治¹ 竹内孔一²

渡邊暁洋³ 平山隆浩⁴ 中尾博之⁵

¹ 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、豪雨災害におけるクロノロジーの優先度を高・中・低の3段階で定義し、その優先度を推定するクラス分類モデルを構築する。さらに、その推定結果を大規模言語モデルを用いて修正する手法することで、優先度高事例の検出率がどの程度向上可能かを検証した。実験の結果、優先度高の Recall は向上させることができたが、Precision と優先度中の性能は低下し、これらの改善が今後の課題である。

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及び大規模言語モデルを用いて推定することで、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、ModernBERT[1]を用いて優先度推定用の3クラス分類モデルを構築し、Llama-3-ELYZA-JP-8B[2]を用いてクロノロジーに関する事前知識から推定結果を一部修正することで、予測漏れが特に致命的となる優先度高事例の検出率向上を行う。

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」[3]、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」[4]¹⁾、「高度自然言語処理プラットフォーム」[5]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。

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

表1 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

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[6] の Encoder 部分をベースにする BERT[7]、Decoder 部分をベースにするオープンソースモデルの Llama3[8] 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 [9] がある。上記の研究では、優先度高事例の検出率向上を目的として、優先度のクラス分類モデルを構築する際に、Data Annotation や単純な複製によって学習データの拡張を実施し、性能向上が確認されている。

一方で、上記のようなデータ拡張手法を導入した場合でも、優先度高事例の検出率は依然として低い。そこで、本研究では人間が優先度を判断する際に利用する事前知識を用いて、大規模言語モデルを活用して優先度推定結果を修正し、優先度高事例の検出率向上を目的とする。

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 1 のような形式で電子化されている。

3.2 優先度

先行研究 [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

西日本豪雨 (2018 年) ・ 熊本豪雨 (2020 年) の 2 つの実際に発生した豪雨災害において作成された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 2 である。

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。

表 2 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
低	4084 (0.906)	1272 (0.943)
中	377 (0.084)	71 (0.053)
高	48 (0.011)	6 (0.004)
合計	4509	1349

* データの件数 (割合) の形式で記載する

4 推定結果の LLM による修正

データセットの構築において、アノテーターが優先度を判断する際に利用した参照知識を基に、大規模言語モデルを用いて優先度高に該当するクロノロジーを再検出する。判断の参考とする知識には、例えば以下のようなものがある。

- 病院が断水などして、さらに、なにか助けを求めている場合は高とする
- 透析を必要とする患者の搬送や必要な物資の要求は高とする
- 医療用の水が不足して患者の生命に対してなにか対処が必要な場合は高とする
- 病院など患者がいる建物の建物倒壊が記載されている文章は高とする

これらの判断基準を大規模言語モデルのプロンプ

トに与え、クロノロジーのテキストがこの基準に該当するかを判定させ、クラス分類モデルによる優先度推定結果を修正する。

5 実験

本実験では、西日本豪雨・熊本豪雨のクロノロジーデータセットに対して、ModernBERTを用いた3クラス分類モデルを構築し、このクラス分類モデルの優先度推定結果をLlama-3-ELYZA-JP-8Bにより優先度判定の参照知識を用いて修正することで、優先度高事例の検出率向上が達成できるかを検証する。

5.1 実験設定

本実験では、クラス分類モデルの構築とLLMによる修正の2段階で実験を行う。なお、データセットを学習データ6割、検証データ2割、評価データ2割に分割して実験を実施した。

5.1.1 クラス分類モデルの構築

本実験では、LLM-jpが作成した事前学習済みModernBERT[1]を用いて、優先度推定用の3クラス分類モデルを構築する。ModernBERTのファインチューニングではLoRA[10]を用いて、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。クラス分類では最終層の[CLS]トークンに対応する埋め込みを用いる。

また、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3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3 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・2・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20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

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。

5.1.2 LLMによる修正

クラス分類モデルによる推定結果を修正する大規模言語モデルには、株式会社ELYZAが作成したLlama-3-ELYZA-JP-8B[2]を用いる。

また、修正対象とするクロノロジーは、クラス分類モデルによる推定結果が優先度中であるものとし、参照知識による判定で優先度高に該当するとLLMが判断した場合にのみ、優先度を高に修正する。

5.2 実験結果

ModernBERTによるクラス分類モデルの優先度推定結果を表4に、Llama-3-ELYZA-JP-8Bによる修正後の結果を表5に示す。

表4 ModernBERTによる優先度推定結果

優先度	Precision	Recall	F1-score
高	0.97	0.93	0.95
中	0.43	0.58	0.49
低	0.20	0.20	0.20

表5 Llama-3-ELYZA-JP-8Bによる修正後の結果

優先度	Precision	Recall	F1-score
高	0.97	0.94	0.95
中	0.34	0.28	0.31
低	0.07	0.40	0.12

実験の結果、優先度高のRecallは向上したが、Precisionと優先度中の性能は低下した。

6 考察

Llama-3-ELYZA-JP-8Bによる修正では、優先度高を検出するために優先度高を判断するための参照知識をプロンプトに与えたが、実際のクロノロジー

に判定では、多数の判断基準からなる複雑かつ段階的なものであるため、優先度高の偽陽性が多く発生したと考えられる。そのため、今後は優先度中・低に対する参照知識も追加し、ワークフロー形式で LLM が段階的に処理する機構が必要であると考えられる。

7 おわりに

本研究では、西日本豪雨と熊本豪雨における豪雨災害のクロノロジーデータセットに対して、ModernBERT を用いた3クラス分類モデルを構築した。さらに、このクラス分類モデルの優先度推定結果のうち、優先度中と推定されたクロノロジーに対して、Llama-3-ELYZA-JP-8B を利用して、優先度判定の参照知識を用いた再判定を行うことで、優先度高事例の検出率向上が達成できるかを検証した。実験の結果、優先度高の Recall は向上したが、Precision と優先度中の性能は低下し、これらの改善が今後の課題である。

謝辞

本論文の筆頭著者は、公益財団法人大本育英会様から奨学金を賜り、本研究の遂行に専念することができました。心から感謝申し上げます。

参考文献

- [1] Issa Sugiura, Kouta Nakayama, and Yusuke Oda. *llm-jp-modernbert: A modernbert model trained on a large-scale japanese corpus with long context length*, 2025.
- [2] Masato Hirakawa, Shintaro Horie, Tomoaki Nakamura, Daisuke Oba, Sam Passaglia, and Akira Sasaki. *elyza/llama-3-elyza-jp-8b*, 2024.
- [3] 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.biprogy.com/solution/service/disaster-net.html> 2026/01/08 アクセス).
- [4] NICT. *対災害 SNS 情報分析システム (DISAANA)*, 2020. (https://www.soumu.go.jp/main_content/000672984.pdf 2026/01/08 アクセス).
- [5] 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 2026/01/08 アクセス).
- [6] 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.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. *Bert: Pre-training of deep bidirectional transformers for language understanding*, 2018.
- [8] AI@Meta. *Llama 3 model card*. 2024.
- [9] Shinji Koju, Koichi Takeuchi, Akihiro Watanabe, Takahiro Hirayama, and Hiroyuki Nakao. *Estimating priority of chronology logs in earthquakes and torrential rains*. In **Proceedings of the 2024 8th International Conference on Natural Language Processing and Information Retrieval, NLPPIR '24**, p. 54–61, New York, NY, USA, 2025. Association for Computing Machinery.
- [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.