MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Detection of adverse event signals with severity grade classification from cancer patient narrative

Satoshi NISHIOKA[1], Masaki ASANO[1], Shuntaro YADA[2], Eiji ARAMAKI[2], Hiroshi YAJIMA[3], Hayato KIZAKI[1] and Satoko HORI[1]

1 Division of Drug Informatics, Keio University Faculty of Pharmacy, Tokyo, Japan, 2 Nara Institute of Science and Technology, Nara, Japan, 3 Mediaid Corporation

Laboratory HP

https://keio-di.jp/en/

COI statement

- HY is the chief executive officer of Mediaid Corporation, which operates the internet patient community, LifePalette.
- The other author(s) declare no competing interests.

Background & Aim

- Early detection of adverse drug reactions (ADRs) is important for better outcome of anticancer treatment
- ADR signals derived from patients' complaints could be overlooked in assessments by physicians [1, 2]
- Patient narratives can be useful for safety events detection[3], but no studies have been employed focusing on Adverse Event (AE) severities

Aim to develop NLP models for AE signal detection from cancer patient narratives, focusing on severity grades which may highlight immediate needs for intervention

[1] E. Basch, The missing voice of patients in drug-safety reporting., N. Engl. J. Med. 362 (2010) 865–869. doi:10.1056/NEJMp0911494.
[2] L. Liut, et al, Clinicians versus patients subjective adverse events assessment: based on patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE)., Qual. Life Res. an Int. J. Qual. Life Asp. Treat. Care Rehabil. 29 (2020) 3009–3015. doi:10.1007/s11136-020-02558-7.
[3] J.-Y. Lee, et al, The Use of Social Media in Detecting Drug Safety–Related New Black Box Warnings, Labeling Changes, or Withdrawals: Scoping Review, JMIR Public Heal. Surveill. 7 (2021) e30137. doi:10.2196/30137.

Methods (1/4)

Data source

- Blog posts written in Japanese in a web community, LifePalette
- The data source consisted of 13,570 posts written by 289 users in 2008 to 2014
- A total of 2,272 posts by breast cancer patients were used for this study

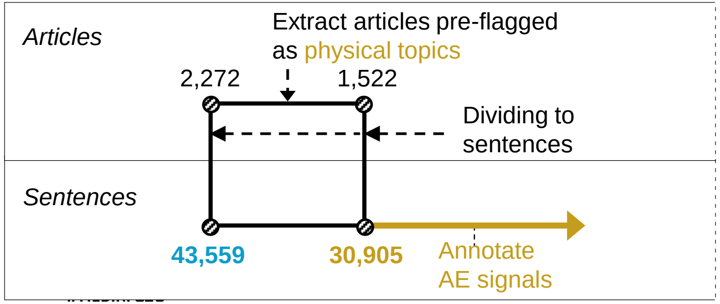
Procedure (overview)

1. Pre-processing
2. Annotation of AE mentions
3. Training and evaluation of deep learning (DL) models

Methods (2/4)

1. Pre-processing

- Divided 2,272 posts into sentences by open-source of "ja_sentence_segmenter"[4]
- After removal of duplicated sentences, obtained 43,559 sentences
- Using the 1,522 posts pre-flagged as physical topics in a preceding study[5], 30,905 sentences were annotated for AE signals



[4] a_sentence_segmenter. https://github.com/wwwcojp/ja_sentence_segmenter.
[5] T. Watanabe, et al, Extracting Multiple Worries from Breast Cancer Patient Blogs Using Multilabel Classification with the Natural Language Processing Model Bidirectional Encoder Representations from Transformers: Infodemiology Study of Blogs, JMIR Cancer. 8 (2022) 1–9. doi:10.2196/37840.

Methods (3/4)

2. Annotation of AE mentions

- Two annotators (SN & MA) annotated each sentence as Grade ≥ 2, 1, or 0 (ie, no AE) per guideline
- Confirmed sufficient concordance level between the two annotators with randomly extracted 50 posts
  - Kappa coefficient at the article level was 0.819 and 0.609 for Grade ≥ 2 and ≥ 1 respectively
- SN completed the rest of annotation

Item	Definition
AE signal criteria	Any untoward medical occurrence in a patient regardless of the cause (i.e., includes not only potential ADRs, but also events caused by primary disease, surgery or other physical events such as falling down)
- Grade ≥ 2	- Sentences reveal clear limitations in Activities of Daily Living (ADL) (i.e., describing concrete limitations to patients' ADL or physiological functions such as "I cannot do [something] due to [symptom]", or events not seen in a healthy condition and hindering ADL [e.g., intolerable pain, severe fatigue/nausea/rush/diarrhea, vomiting, fever ≥ 38.0 degrees C, appetite loss, insomnia etc.])
- Grade 1	- Sentences not falling into Grade ≥ 2 (e.g., tolerable pain, mild fatigue/nausea/ itchiness/constipation, fever < 38.0 degrees C, impaired taste etc.) - All alopecia related sentences, because interpreting limitations to ADL arising from mentions of alopecia was infeasible in this study
Exclusion criteria (reason to exclude)	➢ Changes in mental state, such as worry or anxiety (because we could not judge whether they are AEs or just non-morbid daily emotional changes) ➢ Mentions for recovery from AEs (when there are no on-going subjective symptoms) ➢ Descriptions explaining about illness, future possibilities, or speculations (there are no actual events happening) ➢ Reference to news articles etc. (there are no actual events happening) ➢ Test results or findings without subjective symptoms (where no report is included from patients themselves)

Methods (4/4)

3. Training and evaluation of DL models

- DL models used for this study: **BERT, ELECTRA, and T5**
- **2 classification tasks** for Grade ≥ 2/1 or not
  - ➢ Sentence level: used **BERT and ELECTRA** for such short token input i.e., ≤ 512 tokens
  - ➢ Article level: used **all the three models including T5**
    - For BERT and ELECTRA, the highest grade of sentences in an article classified by sentence level task was treated as the grade of the article in this article level task
- Used 80% as training dataset and 20% as test dataset
- Performance was evaluated with **precision, recall and F1 score**

Results (1/3)

	Grade ≥ 2	Grade ≥ 1
# of articles (% in source)	191 (8.4%)	893 (39.3%)
# of sentences (% in source)	428 (1.0%)	2,843 (6.5%)
Examples of sentences	• I've been down sick and unable to stay awake. • When I stood up to go to the bathroom I was attacked by severe vomiting. • Mouth and face are both numb (BIRI-BIRI*).....can't hold a pen firmly.	• Still feeling heavy with general fatigue. • Aches as the anesthesia wears off. • Can't stop hair loss... • My mouth felt bumpy (BOKO-BOKO*) • It's coming!!!! It is side effect...

Underline describes clear limitation on their ADLs, * indicates onomatopoeic expression in Japanese.

Results (2/3)

Performance scores in the AE **sentence** classification task

Target and model	Precision	Recall	F1
Grade ≥ 1			
BERT	0.581	0.650	0.614
ELECTRA	0.574	0.735	**0.645**
Grade ≥ 2			
BERT	0.250	0.500	**0.333**
ELECTRA	0.246	0.500	0.330

Results (3/3)

Performance scores in the AE **article** classification task

Target and model	Precision	Recall	F1
Grade ≥ 1			
BERT*	0.658	0.794	0.720
ELECTRA*	0.632	0.845	0.723
T5	0.854	0.849	**0.852**
Grade ≥ 2			
BERT*	0.316	0.710	0.438
ELECTRA*	0.288	0.667	0.402
T5	0.541	0.526	**0.533**

* For BERT & ELECTRA, the highest grade of sentences in an article classified by sentence level task was treated as the grade of the article

Discussion

- **T5 showed the best F1 scores** in the article task, which were 0.852 and 0.533 for Grade ≥ 1 and Grade ≥ 2 respectively
  - ➢ Full context in each article enabled the DL model to interpret AE signals precisely
- **Error analysis** for Grade ≥ 2 false positives
  - Most were Grade 1 mentions, which may harbor potential Grade ≥ 2 events
  - The others were mentions without AE signals describing
    - impaired ADL because of non-AE events, or
    - the time course of clinical treatment in detail
  - ➢ To assess the impact of such false positives, prospective study with the AE signal detection system will be required

Conclusion

DL models were developed to detect AE signals from cancer patient narratives with the highlight of severity grade

This suggests that

- Applying these DL models to patient-generating text data could be a new scheme to monitor AE signals occurring outside of clinics
- The scheme may help to connect patients with medical professionals by detecting early AE signals for high-severity events, contributing to earlier intervention and better anti-cancer treatment outcomes

Acknowledgements

- This work was supported by JSPS KAKENHI Grant Number 21H03170 and JST, CREST Grant Number JPMJCR22N1, Japan
- Preliminary work by Ms. T. Watanabe at our laboratory, and discussions with Dr. K. Kawakami (Cancer Institute Hospital, Japanese Foundation for Cancer Research, Japan) inspired us to plan this study