MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231032

Rule-Based Natural Language Processing Pipeline to Detect Medication-Related Named Entities: Insights for Transfer Learning

Zoie S. Y. WONG[a,1], Neil WATERS[a], Data Artist Team[b], Nicholas I-Hsien Kuo[c], Jiaxing LIU[d]
[a] *Graduate School of Public Health, St. Luke's International University, OMURA Susumu & Mieko Memorial St. Luke's Center for Clinical Academia, Japan*
[b] *Data Artist, Japan and Dentsu Data Artist Mongol, Mongolia (Members: Amarsanaa Agchbayar, Batmunkh Batsaikhan, Telmuun Enkhbold, Khatanbold Batzorig, Od Ganzorig)*
[c] *Centre for Big Data Research in Health, The University of New South Wales*
[d] *School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, China*
ORCiD ID: Zoie S. Y. WONG https://orcid.org/0000-0003-4499-9779,
Jiaxing LIU https://orcid.org/0000-0002-5184-4313

Abstract. We document the procedure and performance of a rule-based NLP system that, using transfer learning, automatically extracts essential named entities related to drug errors from Japanese free-text incident reports. Subsequently, we used the rule-based annotated data to fine-tune a pre-trained BERT model and examined the performance of medication-related incident report prediction. The rule-based pipeline achieved a macro-F1-score of 0.81 in an internal dataset and the BERT model fine-tuned with rule-annotated data achieved a macro-F1-score of 0.97 and 0.75 for named entity recognition and relation extraction tasks, respectively. The model can be deployed to other, similar problems in medication-related clinical texts.

Keywords. Transfer learning, multi-task learning, rule-based NLP, BERT, medication errors, incident reports, patient safety

1. Introduction

Patient safety incident reporting and learning systems were specified as one of the 35 items comprising the World Health Organization's Global Patient Safety Action Plan [1]. The application of computational techniques to the analysis and synthesis of natural language within free-text incident reports is regarded as a promising direction for the elimination of avoidable harm in healthcare. As of today, clinical narratives remain an essential component of electronic health records, such as discharge summaries, procedure and progress notes, laboratory reports and physician orders. This diversity of

[1] Corresponding Author: Zoie S. Y. Wong, email: zoiesywong@gmail.com.

clinical texts illustrates the need for natural language processing (NLP) approaches that can effectively identify medication incidents from a range of clinical sources. For instance, a drug mentioned in one encounter could be linked to an event mentioned later in a different document. Furthermore, the ability to effectively identify a drug-related concept in one type of document could later be applied to other document types.

Common techniques to detect medication errors using NLP include keyword and trigger phrase-based methods, symbolic methods and custom-built ontologies [2]. Integrating the strengths of these methods could enhance the performance of NLP in clinical contexts. For instance, custom-built ontologies explore the ontological structure to better characterize semantic similarities between concepts in a specific domain, e.g., medication-related incident reports [3, 4] or discharge summaries relating to adverse drug events. Hybrid models that incorporate both rule-based and machine learning approaches have also demonstrated potential in enhancing deep neural model performance [5, 6].

Transfer learning involves machine-learning techniques that investigate how to improve a predictive model's performance for one problem using the knowledge from a different but relevant domain. Improved learning can reduce computation cost, accelerate training and improve performance for novel but relevant problems, even if there are little data for the new problem [7]. Model-based transfer learning assumes the source task and the target task share some common knowledge on the model level, allowing weights in one or more deep learning layers from a pre-trained network to be reused and fine-tuned in a new model.

Transfer learning can be effective for medication extraction from clinical texts [8]. In this paper, we address how named entity recognition and relation extraction can be performed for medication-related incident reports, even when few gold-standard data are available [4]. We examine how rule-based approaches can be used to enhance the performance of deep-learning models using transfer learning via fine-tuning techniques.

2. Methods

Rules for organizing drug incident concepts have been presented in annotation guidelines for medication-related incident reports elsewhere [9, 10, 4]. This set of concepts is widely agreed upon, as are the rules relating to how these concepts can be arranged in meaningful ways. The associated NLP tasks essentially involve identifying medication error concepts and identifying relationships between the concepts, including intention and factuality analysis, which is presented in [4]. The medication error concepts (i.e., named entities), regarded as the 'things of interest' within incident reports and the key units of information that need to be extracted, include 'drug', 'form', 'strength', 'duration', 'timing', 'frequency', 'date', 'dosage' and 'route.' The inputs are named entity labels, registered in IOB (inside-outside-beginning) format, on every single token. The rules we designed to extract each of these named entities are presented below.

- Drug: We first extracted the full list of generic drug names from the 2022 'Standard Collection of Medicines' from the Japan Ministry of Health, Labor and Welfare website [11]. Then, we strictly cleaned the drug list so that only the name of the drug remained. This involved removing manufacturer names and other named entities that might be contained within the drug name, such as strength-amount, form-mode, form-form and route. As a result, we have a list of 6,406 unique drug names.

- Form, Mode, Route: We extracted unique lists of form, mode and route that were identified by the gold-standard IFMIR corpus data [4]. Ultimately, 41 distinctive forms, 16 modes and 42 routes were recorded.
- Strength-amount, Strength-rate, Strength-concentration, Frequency, Date, Dosage, Timing and Duration: We employed 'RegEx' to create custom combinations of special character operators. These character sequences are used to specify patterns that are associated with target named entities. For specific details, one may refer to the script.

As shown in Figure 1, we underwent two stages of rule-based pipeline evaluation: internal validation and external validation. In the internal validation phase, we split the gold-standard IFMIR data into training and testing groups, 75% and 25%, respectively. We used the training group set to develop the rule-based pipeline, as described above, and tested the model using the testing group. In the external validation exercise, we first randomly selected and manually annotated 50 medication-related incident reports from the Project to Collect Medical Near-Miss/Adverse Event Information by the Japan Council for Quality Health Care (JQ) [12] (The full corpus of incident reports of medication errors contains 58,658 free-text reports). Then, we employed the developed rule-based pipeline and evaluated the performance of using rules to extract the targeted named entities. Entity level performance was evaluated using precision, recall and F1-score; the model's overall performance was evaluated using macro-F-1 score.

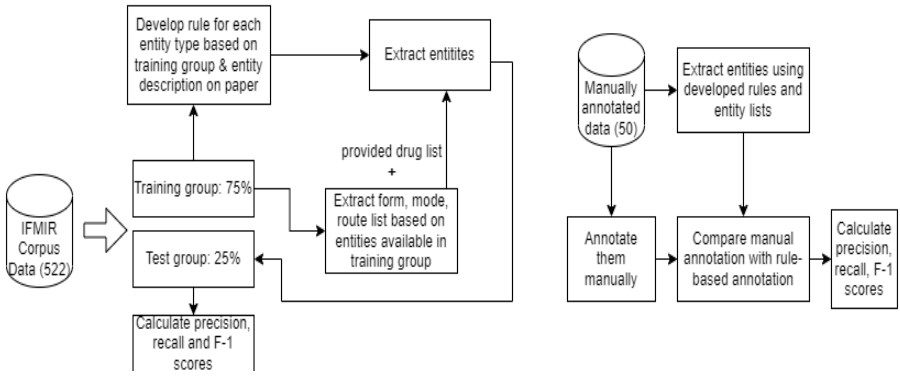

Figure 1. Design and evaluation of the rule-based approach. The left chart shows internal validation, the right shows external validation.

A BERT model with the SentencePiece tokenizer, pre-trained on Japanese Wikipedia and Twitter corpora [13], was adopted and was then pre-trained with the JQ incident report corpus of 121,244 unannotated free-text documents of all incident types [12] before undergoing two phases of fine-tuning. The first phase of fine-tuning used JQ medication-related reports with ruled-based annotations of named entities (58,658 reports). The second phase of fine-tuning employed the gold-standard data (522 reports) with annotations for both named entities and relations. As a result, a three-layer multi-task BERT model for named entity recognition and relation extraction was developed. Model improvement was assessed by comparing the model's performance using transfer learning with the standalone model, which was considered baseline performance. The standalone model in this study is the abovementioned BERT model [13] fine-tuned with only the gold-standard data (522 reports).

3. Results

Table 1 presents drug-specific performance using the gold-standard data. Overall, the rule-based model achieved an F1-score of 0.81 (internal validation) and 0.73 (external validation).

Table 1. Drug-named entity specific performance (internal validation using gold standard data).

Entity_type	TP	FP	FN	Precision	Recall	F1-score
Strength_rate	68	0	3	1	0.96	0.98
Date	165	26	8	0.86	0.95	0.91
Frequency	71	5	9	0.93	0.89	0.91
Duration	87	15	6	0.85	0.94	0.89
Strength_concentration	28	6	1	0.82	0.97	0.89
Form_form	237	58	15	0.8	0.94	0.87
Timing	269	56	45	0.83	0.86	0.84
Form_mode	16	3	3	0.84	0.84	0.84
Drug	852	109	260	0.89	0.77	0.82
Strength_amount	454	42	196	0.92	0.7	0.79
Route	132	104	9	0.56	0.94	0.7
Dosage	204	190	26	0.52	0.89	0.65

The model fine-tuned with rule-based annotated data achieved an F1-score of 0.97 (validation set) and 0.84 (hold-out testing set) for the named entity recognition task. For the relation extraction task (intention and factuality analysis), the model achieved an F1-score of 0.75. For the standalone model, we recorded a macro-F1-score of 0.15 and 0.61 for drug named-entity recognition and relation extraction, respectively. This indicates our method accomplished a 460% and 23% improvement over the standalone model.

Table 2. Performance of multi-task BERT model fine-tuned with rule-based annotated data.

	Precision	Recall	F1-score	
NER	0.97	0.97	0.97	validation
	0.82	0.89	0.84	test
I&F	0.75	0.76	0.75	

4. Discussion

In this study, we used a rule-based model can improve the performance of a deep-learning BERT model using transfer learning with a fine-tuning procedure. Our reasoning is that if a model is provided with a large amount of annotated data (even non-gold standard, if proven to have satisfactory performance), the model can effectively serve as a generic NLP model for medication-related concept detection. Modelers can take advantage of these learned-feature maps, avoiding the need to train a large model on a large dataset. As a result, the model we fine-tuned with rule-based annotation can be reused as the starting point for another model and task associated with medication-related named entities. For instance, the model can be applied to other datasets, such as

free-text pharmaceutical lists in Japan [14]. Furthermore, the model is suitable for some problems where medication concept identification is required but data is limited. Our outcomes can also be useful for standardizing medication-related clinical notes. Potential applications can be made in combination with existing data standards, such as the ISO Identification of Medicinal Products, openEHR or Fast Healthcare Interoperability Resources.

Acknowledging that rule-based models do not always produce perfect labels, further research can be carried out to explore fine-tuning with pseudo-labelling using labeled datasets and pseudo-labels from rule-based pipelines. In theory, transfer learning may exhibit a performance advantage through its 'higher start' (i.e., the source model having a higher initial skill, before model refinement, than it otherwise would), slope (i.e., the source model's skill improvement occurring at a greater rate during training than it otherwise would) and asymptote (i.e., the converged skill of the trained model being better than it otherwise would be). Future studies can explore these boundaries and examine the stability of improved performance by splitting, sizing and shuffling datasets and examining prediction distributions.

5. Conclusions

We highlight rule-based systems improving BERT model outcomes via transfer learning.

References

[1] Global Patient Safety Action Plan 2021-2030, in, WHO, 2022.
[2] Luo Y, Thompson WK, Herr TM, Zeng Z, Berendsen MA, Jonnalagadda SR, Carson MB, Starren J. Natural Language Processing for EHR-Based Pharmacovigilance: A Structured Review. Drug Saf. 2017 Nov;40(11):1075-89. doi: 10.1007/s40264-017-0558-6.
[3] Z.S. Wong, AI for Patient Safety, in, 2022.
[4] Zhang H, Sasano R, Takeda K, Wong ZS. Development of a Medical Incident Report Corpus with Intention and Factuality Annotation, in: LERC 2020, Marseille, 2020. p. 4580-6.
[5] Ray P, Chakrabarti A. A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. Appl Comput Inform. 2022 Mar;18(1/2):163-78, doi: 10.1016/j.aci.2019.02.002.
[6] Tan HX, Teo CHD, Ang PS, Loke WPC, Tham MY, Tan SH, Soh BLS, Foo PQB, Ling ZJ, Yip WLJ, Tang Y, Yang J, Tung KHA, Dorajoo SR. Combining Machine Learning with a Rule-Based Algorithm to Detect and Identify Related Entities of Documented Adverse Drug Reactions on Hospital Discharge Summaries. Drug Saf. 2022 Aug;45(8):853-862. doi: 10.1007/s40264-022-01196-x.
[7] Yang Q, Zhang Y, Dai W, Pan SJ. Transfer Learning, Cambridge University Press, Cambridge, 2020.
[8] Narayanan S, Mannam K, Rajan SP, Rangan PV. Evaluation of transfer learning for adverse drug event (ade) and medication entity extraction. InProceedings of the 3rd Clinical Natural Language Processing Workshop 2020 Nov. p. 55-64, doi:10.18653/v1/2020.clinicalnlp-1.6.
[9] AI for Patient Safety, Annotation Guidelines for Incident Reports of Medication Errors (English Reports) MARCH 31, 2022 (V.1.0.2), in, GitHub, 2022.
[10] Wong ZSY, Qiao Y, Sasano R, Zhang H, Taneda K, Ushiro S. Annotation Guidelines for Medication Errors in Incident Reports: Validation Through a Mixed Methods Approach. Stud Health Technol Inform. 2022 Jun 6;290:354-8. doi: 10.3233/SHTI220095.
[11] List of drug price standards and information on generic drugs (until March 31, 2022) [Yakka kijun shūsai hinmoku risuto oyobi kōhatsu iyakuhin ni kansuru jōhō ni tsuite (ryō wa 4-nen 3 tsuki 31-nichi made)], in, Ministry of Helath Labour and Welfare (MHLW), 2022.
[12] Japan Council for Quality Health Care. Project to collect medical near-miss/adverse event information 2013 annual report.
[13] Y. Kikuta, Bert-Japanese, in, 27th Jul, 2022, 2021.
[14] Project to Collect and Analyze Pharmaceutical Near-Miss Events, in, The Japan Council for Quality Health Care (JQ) 2017.