DeviceBERT: Transfer Learning for Medical Device Entity Recognition

Miriam Farrington [1]

[1]Computer Science, Stanford

Introduction & Problem

The rapid search and identification of impacted medical devices in **FDA medical device recall summaries** is crucial for timely action and ensuring patient safety. However, the process of **quickly searching** and **identifying** affected device terms from recall action text summaries is tedious and time-consuming. Existing **Named Entity Recognition (NER) models**, even those trained on biomedical corpora, fail to identify medical device terminology due to the unique and domain-specific vocabulary and acronyms. To overcome these challenges, we propose **DeviceBERT**, a **transfer learning pipeline** which makes use of **targeted annotations**, **vocabulary enrichment**, and **regularization techniques** to more accurately identify medical device trade names, part numbers, and component terms in recall summaries. This targeted approach aims to streamline recall communications and aid downstream systems to quickly and accurately identify all device component parts included in a recall action, utilizing a small dataset for training.

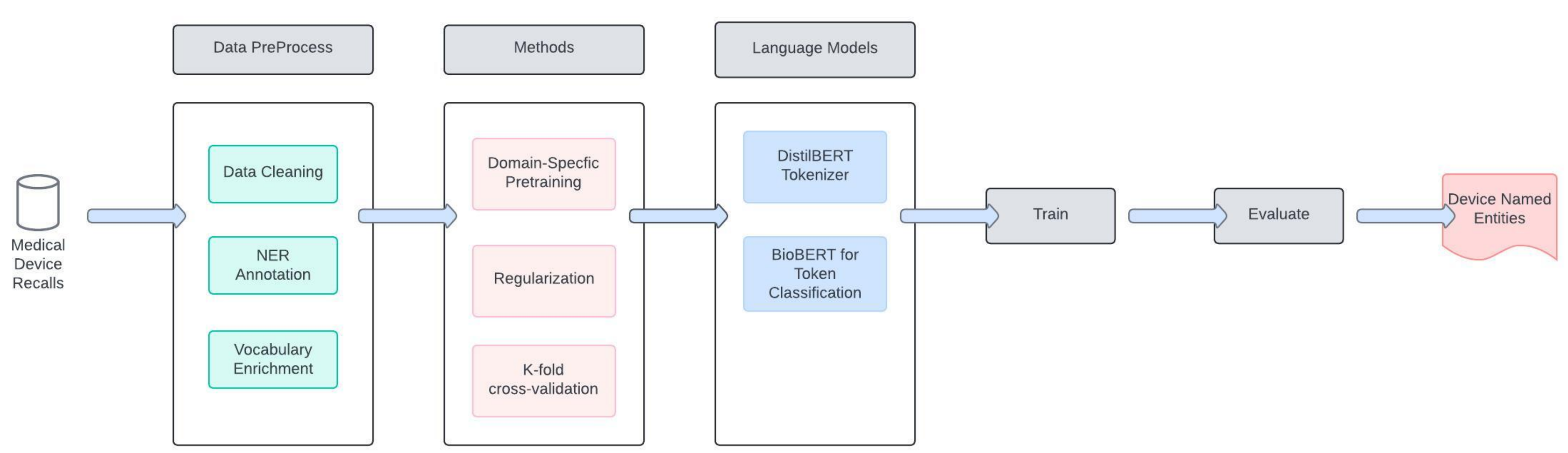

Figure 1. DeviceBERT Process Overview

Background: Transfer Learning for Domain-Specific Entity Recognition

Obtaining satisfactory performance utilizing pre-trained models to perform biomedical entity recognition is a known challenge in the NLP space. Prior work has revealed inherent challenges when attempting to perform NER tasks in the biomedical domain, such as **limited availability** of training data, **ambiguity** in medical entity terminology, and **heavy reliance on acronyms** in training corpora [2].

Transfer learning is a widely adopted technique in machine learning that utilizes pre-trained model weights from models trained on large-scale datasets to fine-tune the model on smaller, downstream tasks [3]. This approach takes advantage of the knowledge captured by the pre-trained model on the larger dataset and adapts it to the specific requirements of the target task, thereby reducing the need for extensive retraining and improving performance.To accomplish this, we initialized a pretrained BioBERT model with weights as a starting point for this task [1].

Pre-Processing & Annotations

To train DeviceBERT using the pre-trained weights of BioBERT, we needed to extract, pre-process and annotate a **custom NER dataset** of 2000 recall actions. We extracted the data from source into an annotations database and utilizing the following methodology to apply **BIO (Beginning, Inside, Outside) labels** as follows:

- **The beginning word** of a device name/part/component is assigned the label **B-DEVICE**
- **Subsequent words** in the device name which should be given attention are assigned **I-DEVICE**
- **Outside words** part of the device name ('and', 'the', etc) are assigned the label **O-DEVICE**
- **Remaining words** are assigned the value of **'O'**, indicating they are not Device terms.

Pre-Processing & Annotations Contd.

Additionally, the following rules were applied when labeling device terms:

- Software components of a medical device are excluded from NER labeling, which includes software trade names, version numbers, and operating systems.
- Different model numbers/names are treated as separate devices.
- Context of words is considered; if a word has multiple meanings in different contexts, the NER label is only applied where the term refers to a medical device.
- Special characters in/around the device words were excluded.

Sammons Preston Roylan, the U.S. agent for Ito Co., Ltd. issued recall letters dated 7/13/05 to all of their

customers who purchased the Performa TM-300 traction unit, informing them that the manufacturer, Ito

B-DEVICE I-DEVICE I-DEVICE

I-DEVICE

Figure 2. Example of an annotated recall action with BIO tagging applied to a device trade name

The NER label tags are created and mapped to each token at the positions defined in the original label span and converts tokens to numeric (0-3, with a special tag -100 designated for unlabeled tokens). This effectively masks the unlabeled tokens to ensure they do not influence the gradients during back-propagation.

Vocabulary Enrichment

To compile the data to enrich the tokenizer vocabulary, we extract the device_name and product_description fields from the **Device Recalls** dataset, and combine with the FDA **Device Registration** database. The data is cleaned, shuffled, tokenized into words and de-duplicated. We identified 172,821 new tokens, increasing the size from 28,996 to 191,049 when 100% of the identified tokens were added to the vocabulary.

Before Vocabulary Enrichment	After Vocabulary Enrichment
[CLS], An, Advisory, Letter, was, sent, to, the, customers, via, certified, mail, ., To, ##shi, ##ba, issued, a, Field, Mo, ##di, ##fica-tion, In, ##struction, (, FM, ##I, X, ##RA, ##29, -, 90, ##8, ##28,), to, correct, that, software, bug, and, bring, the, D, ##FP, -, 800, ##0, ##D, into, compliance, ., The, FM, ##I, is, provided, to, the, customers, at, no, charge, ., [SEP]	[CLS], An, Advisory, Letter, was, sent, to, the, customers, via, certified, mail, ., Toshiba, issued, a, Field, Modification, Instruction, (, FMI, XR, A2, 9, -, 90, ##8, ##28,), to, correct, that, software, bug, and, bring, the, DF, P, -, , 800, 0D, into, compliance, ., The, FMI, is, provided, to, the, customers, at, no, charge, ., [SEP]

Table 1. Comparison of a Tokenized Recall Action Before and After Vocabulary Enrichment

After vocabulary enrichment, the tokenizer displayed significantly reduced frequency of sub-word tokenization of device terms, indicating an increased semantic understanding of device related terms.

Experiments

We conducted several experiments to evaluate the performance of DeviceBERT:

- DeviceBERT utilizing combined regularization techniques on the base DistilBERT tokenizer
- DeviceBERT utilizing DistilBERT tokenizer with 100% device term vocabulary enrichment
- DeviceBERT utilizing DistilBERT tokenizer with 50% device term vocabulary enrichment
- DeviceBERT utilizing DistilBERT tokenizer with 25% device term vocabulary enrichment
- DeviceBERT with combined regularization techniques and vocabulary enrichment (best % split)

Results

Results for DeviceBERT across all experiments show that the combination of knowledge distillation pre-training steps, coupled with vocabulary enrichment achieves the highest overall F1 score. One observation is the relative high performance of DeviceBERT utilizing only regularization techniques, which we feel warrants further investigation, as the use of regularization techniques alone appear to correlate with a significant impact on overall model accuracy across all 3 metrics.

Table 2. Performance comparison of language models on NER dataset of devices

Model	Precision (%)	Recall (%)	F1 Score (%)
BERT	72.96	73.99	73.47
BioBERT	73.42	73.29	73.35
DeviceBERT			
(+Reg only)	82.37	78.52	80.37
(+Vocab 100%)	75.59	73.46	74.51
(+Vocab 50%)	81.56	80.11	80.83
(+Vocab 25%)	80.14	77.87	78.91
(+Reg+Vocab)	85.14	82.07	83.56

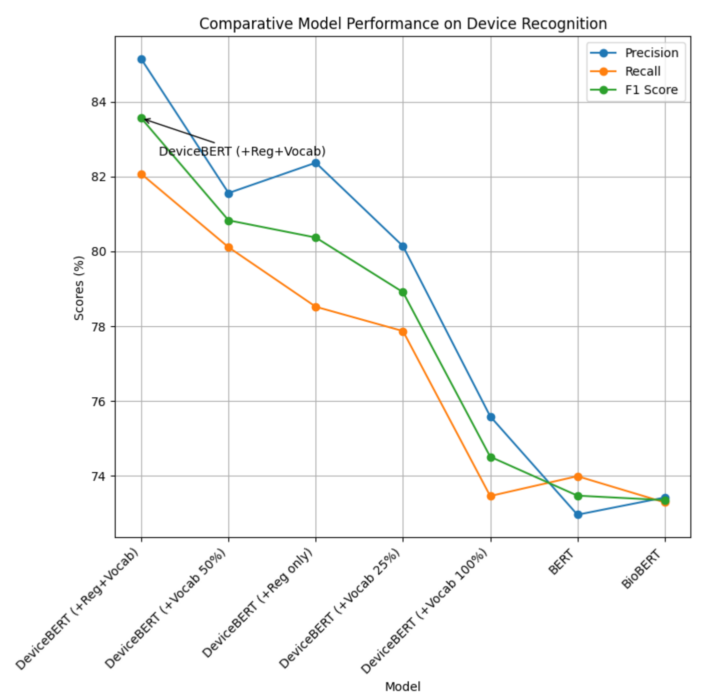

Figure 3. Comparative score of all models on device entity recognition task.

Conclusion & Future Work

DeviceBERT is proposed to maximize the use of a pre-trained domain-specific model (BioBERT) for performing domain-speciifc NER tasks. DeviceBERT adds a new domain-specific medical device vocabulary, while using an ensemble of annotation, cross-validation and regularization to avoid common pitfalls when working with limited training data. DeviceBERT also paves the way for potential new applications of the model to downstream Named Entity Linking tasks.

References

[1] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, September 2019.

[2] Usman Naseem, Matloob Khushi, Vinay Reddy, Sakthivel Rajendran, Imran Razzak, and Jinman Kim. Bioalbert: A simple and effective pre-trained language model for biomedical named entity recognition, 2020.

[3] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021.