Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning

Fei Li  
*University of Massachusetts Medical School*

Weisong Liu  
*University of Massachusetts Medical School*

Hong Yu  
*University of Massachusetts Medical School*

Follow this and additional works at: [https://escholarship.umassmed.edu/oapubs](https://escholarship.umassmed.edu/oapubs)

Part of the [Artificial Intelligence and Robotics Commons](https://escholarship.umassmed.edu/oapubs/685), [Databases and Information Systems Commons](https://escholarship.umassmed.edu/oapubs/686), [Health Information Technology Commons](https://escholarship.umassmed.edu/oapubs/687), and the [Library and Information Science Commons](https://escholarship.umassmed.edu/oapubs/688)

This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

Repository Citation

Li, Fei; Liu, Weisong; and Yu, Hong, "Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning" (2018). *Open Access Articles*. 3672.  
[https://escholarship.umassmed.edu/oapubs/3672](https://escholarship.umassmed.edu/oapubs/3672)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Open Access Articles by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Original Paper

Extraction of Information Related to Adverse Drug Events from Electronic Health Record Notes: Design of an End-to-End Model Based on Deep Learning

Fei Li¹,²,³, PhD; Weisong Liu¹,²,³, PhD; Hong Yu¹,²,³,⁴, PhD

¹Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States
²Center for Healthcare Organization and Implementation Research, Bedford Veterans Affairs Medical Center, Bedford, MA, United States
³Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States
⁴School of Computer Science, University of Massachusetts, Amherst, MA, United States

Corresponding Author:
Hong Yu, PhD
Department of Computer Science
University of Massachusetts Lowell
1 University Avenue
Lowell, MA,
United States
Phone: 1 508 612 7292
Email: Hong_Yu@uml.edu

Abstract

Background: Pharmacovigilance and drug-safety surveillance are crucial for monitoring adverse drug events (ADEs), but the main ADE-reporting systems such as Food and Drug Administration Adverse Event Reporting System face challenges such as underreporting. Therefore, as complementary surveillance, data on ADEs are extracted from electronic health record (EHR) notes via natural language processing (NLP). As NLP develops, many up-to-date machine-learning techniques are introduced in this field, such as deep learning and multi-task learning (MTL). However, only a few studies have focused on employing such techniques to extract ADEs.

Objective: We aimed to design a deep learning model for extracting ADEs and related information such as medications and indications. Since extraction of ADE-related information includes two steps—named entity recognition and relation extraction—our second objective was to improve the deep learning model using multi-task learning between the two steps.

Methods: We employed the dataset from the Medication, Indication and Adverse Drug Events (MADE) 1.0 challenge to train and test our models. This dataset consists of 1089 EHR notes of cancer patients and includes 9 entity types such as Medication, Indication, and ADE and 7 types of relations between these entities. To extract information from the dataset, we proposed a deep-learning model that uses a bidirectional long short-term memory (BiLSTM) conditional random field network to recognize entities and a BiLSTM-Attention network to extract relations. To further improve the deep-learning model, we employed three typical MTL methods, namely, hard parameter sharing, parameter regularization, and task relation learning, to build three MTL models, called HardMTL, RegMTL, and LearnMTL, respectively.

Results: Since extraction of ADE-related information is a two-step task, the result of the second step (ie, relation extraction) was used to compare all models. We used microaveraged precision, recall, and F1 as evaluation metrics. Our deep learning model achieved state-of-the-art results (F1=65.9%), which is significantly higher than that (F1=61.7%) of the best system in the MADE1.0 challenge. HardMTL further improved the F1 by 0.8%, boosting the F1 to 66.7%, whereas RegMTL and LearnMTL failed to boost the performance.

Conclusions: Deep learning models can significantly improve the performance of ADE-related information extraction. MTL may be effective for named entity recognition and relation extraction, but it depends on the methods, data, and other factors. Our results can facilitate research on ADE detection, NLP, and machine learning.
KEYWORDS
adverse drug event; deep learning; multi-task learning; named entity recognition; natural language processing; relation extraction

Introduction

Background
An adverse drug event (ADE) is an injury resulting from a medical drug intervention [1]. Previous studies reported that ADEs could account for up to 41% of all hospital admissions [2,3]. An ADE may cause a prolonged length of stay in the hospital and increase the economic burden [4]. The annual cost of ADEs for a 700-bed hospital is approximately $5.6 million [5]. Moreover, the total number of iatrogenic deaths can reach nearly 800,000 per year, which is higher than the death rate of heart disease or cancer [6]. In 2013, medical error, including ADEs, is the third most-common cause of death in the United States [7]. Therefore, ADE detection and report are crucial for pharmacovigilance and drug-safety surveillance [8,9].

Two methods are usually used to detect and report ADE. In premarketing surveillance, ADEs can be discovered during phase III clinical trials for drug development. In postmarketing surveillance, ADEs are discovered by patients and physicians using the Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS). These traditional methods are limited by the number of participants [10], underreporting [11], and missing patterns of drug exposure [12]; for example, underreporting is a known issue in FAERS and may occur due to several reasons. First, the objective and content of the report in FAERS change over time, which may confuse physicians and the general public. Second, patients may choose not to mention some reactions, due to which practitioners fail to report them. Third, ADEs with long latency or producing unusual symptoms may be unrecognized. Other reasons may include payments from pharmaceutical companies to doctors [13] and ineffective communication between patients, physicians, and pharmacists. Recently, the FDA made the FAERS data available through a public dashboard [14]. Since anyone can view ADE reports online, this venture may help the FDA receive feedback to improve the FAERS.

Many researchers have used other resources to identify ADEs, such as biomedical publications [15,16], social media [17,18], and electronic health record (EHR) notes [19-21]. The ADEs extracted from these resources are an important complement to traditional ADE-surveillance systems. However, manual collection of ADEs from these data is laborious [22]. As such, the use of computer systems is a good choice to automatically detect ADEs, but may fail since these data are often unstructured text. Therefore, natural language processing (NLP) techniques are employed for this significant task [15,20,21,23].

From the viewpoint of NLP, ADE detection is covered under the task of information extraction, which includes ADE extraction as well as information related to ADE, such as medications and indications. Extraction of ADE-related information can be casted as a two-step pipeline. The first step is named entity recognition (NER) [24], which recognizes a string of text as an entity (eg, medication or ADE) that is predefined by dataset annotators. The second step is relation extraction (RE) [15], which is a model that determines whether two entities have a specific relation (eg, medication and ADE).

Previous studies employed traditional machine-learning techniques [15,16,23,24] such as condition random field (CRF) [25] or support vector machine (SVM) [26]. Recently, deep learning attracted much attention in NLP due its numerous advantages such as better performances and less feature engineering compared to other systems [27,28]. However, only a few studies have addressed extraction of ADE-related information via deep learning. Since ADE detection can be divided into two tasks (ie, NER and RE), it is logical to incorporate multi-task learning (MTL) [29] into ADE detection. However, only limited prior work has investigated the impact of MTL on ADE detection.

Relevant Literature

Adverse Drug Event Detection
Since ADEs play an important role in drug-safety surveillance, ADE detection receives increasing attention from both the federal regulation and the research community. Besides the official reporting system FAERS, there are other databases that collect data on known ADEs, such as the Comparative Toxicogenomics Database [30] and SIDER [31]. Various resources have been used to detect ADEs, such as biomedical publications [15,16], social media [17,18], and electronic health record (EHR) notes [19-21]. For example, Gurulingappa et al [16] leveraged medical case reports to build a corpus in order to support drug-related adverse effects. Wei et al [15] organized a challenge task to extract chemical-induced disease relations from the literature and created an annotated corpus from 1500 articles. With respect to the methods, both supervised and unsupervised methods were used. Ramesh et al [32] developed a supervised machine-learning model to extract adverse event entities from FAERS narratives. Xu and Wang [33] used a semisupervised bootstrapped method to construct a knowledge base for the drug-side-effect association. Liu et al [34] proposed a causality-analysis model based on structure learning for identifying factors that contribute to adverse drug reactions. Yildirim et al [35] applied the k-mean algorithm to identify adverse reactions. Xu et al [23] used SVM to extract ADEs between sentence-level and document-level drug-disease pairs. Recently, Munkhdalai et al [21] attempted to use deep learning to address ADE extraction, but their model was not end-to-end and relied on the entities.


JMIR Med Inform 2018 | vol. 6 | iss. 4 | e12159 | p.2 (page number not for citation purposes)

**Named Entity Recognition**

NER is a standard task that has been studied for many years in NLP [25]. Many researchers made important contributions to dataset construction including the GENIA corpus [36], the NCBI disease corpus [37], and the ShARe/CLEF eHealth evaluation [38]. Early studies addressed NER by diverse approaches such as rule-based [39] and machine-learning approaches [40-42], among which CRF-based approaches predominated. For example, Campos et al [40] presented a CRF model to recognize biomedical names, which achieved state-of-the-art performance at the time by incorporating rich features. Tang et al [43] modified the label scheme of CRF to make it be capable of recognizing disjoint clinical concepts. However, such approaches need many efforts for feature engineering. Recently, a bidirectional LSTM (BiLSTM) model [44,45] was proposed and became a popular method for NER. In the biomedical domain, Jagannatha and Yu [20] used such a model to detect medical events from EHR notes.

**Relation Extraction**

RE has been widely studied, and some typical RE corpora in the biomedical domain include the 2010 i2b2/VA challenge [46] and BioCreative V CDR task [15]. Early work used rules and NLP structures such as dependency trees [47] and coreference chains [48] to help extract relations. Others usually leveraged machine-learning approaches such as SVM [49,50] and structured learning [51]. As deep learning developed, researchers proposed a number of neural network models to handle RE [52,53]. Our study is related to the joint or end-to-end entity and RE, which allows performance of NER and RE simultaneously. Miwa and Bansal [54] proposed an end-to-end model based on the sequence and tree LSTM. Similarly, Mehrkry et al [55] proposed an end-to-end system to extract information about bacteria and their habitats.

**Multi-Task Learning**

MTL [29] refers to training the model for multiple related tasks. It is widely used in artificial intelligence research such as computer vision [56] and NLP [57]. Learning these tasks simultaneously may improve the performance as compared to learning the tasks individually. Prior MTL studies mostly focused on homogeneous MTL that consists of tasks with only one type such as classification or regression [58]. Some of their tasks are closely related, such as cross-lingual [59] and synthetic data [60]. Based on a previous study [58], MTL can be roughly divided into two categories—hard and soft parameter sharing. For hard parameter sharing, the lower layers are shared among multiple tasks and each task has its own higher layer [54]. For soft parameter sharing, each task has its own model with its own parameters. There are some representative methods for soft parameter sharing such as regularization [59] or learning task relations [56].

**Objective**

Since only a few prior studies have addressed end-to-end detection of ADE via deep learning, we aimed to design a two-step pipeline model that consists of two submodels: a BiLSTM [61] CRF [25] network for NER and a BiLSTM-Attention [62] network for RE. Since extraction of ADE-related information includes two steps, it is possible to study the impact of MTL on NER and RE. However, only limited prior work has focused on MTL with such heterogeneous and loosely related tasks. Therefore, our second objective was to fill this gap by proposing three MTL models and comparing them with the pipeline model. An overview of this study is shown in Figure 1.

**Methods**

**Deep Learning Pipeline Model**

**BiLSTM-CRF Submodel for NER**

Our NER submodel is presented in Figure 2. We extended the state-of-the-art BiLSTM-CRF model [44,45] by enriching its features. For a sentence, we extracted four kinds of features for each token, namely, its word, whether the initial character is capital, its part-of-speech (POS) tag, and its character representation. We employed a convolutional neural network to obtain character representation. After the token representations are obtained by concatenating the features, we fed them into a bidirectional LSTM layer to learn the hidden representations. Subsequently, the hidden representations were input into the CRF layer to determine the optimal labels for all
the tokens in the sentence. For labels, we use the BMES (Begin, Middle, End, Singular) label scheme [45] plus entity types. For example, the label of the token “Renal” is “B_Disease.” The details of the BiLSTM-CRF submodel for NER are provided in Multimedia Appendix 1.

**BiLSTM-Attention Submodel for RE**

Our RE submodel is presented in Figure 3. A relation instance can be considered as a token sequence and two target entities. Here, the token sequence did not necessarily have to be one sentence, as we could also extract intersentence relations. For each token, we extracted four kinds of features, namely, its word, its POS tag, and the position embeddings [63]. Here, the character representation was not used, because it reduced the performance in our preliminary experiments. Similar to the case for NER, we employed a BiLSTM layer to generate the hidden representations. Subsequently, the attention method [62] was used to obtain context features.

Because only context features may not be enough to capture the semantic relation, we also employed other features for concision, which are not shown in Figure 3. Considering previous work [21], these features included words of two target entities, types of two target entities, the token number between two target entities, and the entity number between two target entities. Like the word or POS embeddings, these features can be represented as vectors. Therefore, the output layer considers the concatenation of all these features to determine the relation of target entities. The details of the BiLSTM-Attention submodel for RE are provided in Multimedia Appendix 2.

**Multi-Task Learning Models**

In this section, we propose three MTL models: one model used hard parameter sharing [54] and two models used soft parameter sharing, namely, regularization [59] and task relation learning [56].

**HardMTL**

Our MTL model for hard parameter sharing is presented in Figure 4. We employed the shared-private architecture [64] to make each submodel of each task retain its private parts and share some parts for multi-task learning. The NER and RE submodels had their own BiLSTM layers, namely, \( \text{LSTM}^{\text{ner}} \) and \( \text{LSTM}^{\text{re}} \), and shared a BiLSTM layer, \( \text{LSTM}^{\text{share}} \). During training, the shared BiLSTM layer \( \text{LSTM}^{\text{share}} \) was used by both the NER and RE submodels, so that it was tuned during the back-propagation by both submodels. Therefore, the model was able to learn useful knowledge from both tasks. The details of the HardMTL model are provided in Multimedia Appendix 3.

**RegMTL**

Our first MTL model for soft parameter sharing was based on regularization, and its architecture is presented in Figure 5. With reference to previous studies [59,60], we employed the L2 regularization in order to encourage the parameters of the NER and RE submodels to be similar instead of sharing some parts in the networks. Two BiLSTM layers were considered because different inputs of the NER and RE submodels lead to different dimensions of the first BiLSTM layer; therefore, L2 regularization of the parameters of the first BiLSTM layer was computationally intractable. We resolved this issue by performing L2 regularization in the second BiLSTM layer. The details of the RegMTL model are provided in Multimedia Appendix 3.

**LearnMTL**

Our second MTL model for soft parameter sharing was based on task relation learning [56], and its architecture is illustrated in Figure 6. After generating hidden representations from the BiLSTM and attention layers, we used a linear layer, \( W_s \), to exchange information between the NER and RE submodels. To utilize task-specific and shared information, the concatenation of hidden representations of the BiLSTM and information exchange layers was fed into the upper decoders \( D^{\text{ner}} \) and \( D^{\text{re}} \). The details of the LearnMTL model are provided in Multimedia Appendix 3.

**Dataset**

We used the MADE dataset from the MADE1.0 challenge for detecting medications and ADEs from EHR notes [65]. It consists of 1089 EHR notes of patients with cancer, from which data for 18 common Protected Health Information aspects were removed according to the Health Insurance Portability and Accountability Act. The dataset was separated into 876 notes for training and 213 notes for testing. In this dataset, the annotators annotated not only ADEs, but also other ADE-related information. They predefined 9 entity types, namely, Medication, Indication, Frequency, Severity, Dosage, Duration, Route, ADE, and SSLIF (any sign, symptom, and disease that is not an ADE or Indication). In addition, they predefined 7 relation types between these entity types, namely, Dosage-Medication, Route-Medication, Frequency-Medication, Duration-Medication, Medication-Indication, Medication-ADE, and Severity-ADE.
Figure 2. NER submodel. For simplicity, here we use “Renal Failure” to illustrate the architecture. For “Renal,” the word feature is “Renal,” the capital feature of the initial character is “R,” the POS feature is “JJ,” and the character representation is generated from CNN. NER: named entity recognition. CNN: convolutional neural network. CRF: condition random field. LSTM: long short-term memory. CNN: convolutional neural network. POS: part of speech.

Figure 3. RE submodel. The target entities are “renal failure” (e₁) and “antibiotics” (e₂). Positions represent token distances to the target entities. RE: relation extraction. LSTM: long short-term memory. POS: part of speech.
**Figure 4.** The high-level view of HardMTL. For conciseness, “LSTM” indicates a BiLSTM layer, and the layers above the BiLSTM layer are denoted as $D_{\text{new}}$ and $D_{\text{re}}$. The forward procedures for an NER instance and an RE instance are indicated by blue and green arrow lines, respectively. HardMTL: multi-task learning model for hard parameter sharing. LSTM: long short-term-memory. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. NER: named entity recognition. RE: relation extraction.

**Figure 5.** The high-level view of RegMTL. $LSTM_1^{\text{ner}}$ and $LSTM_2^{\text{ner}}$ indicate the first and second BiLSTM layers of the NER model. $LSTM_1^{\text{re}}$ and $LSTM_2^{\text{re}}$ indicate the first and second BiLSTM layers of the RE model. NER: named entity recognition. RE: relation extraction. RegMTL: multi-task learning model for soft parameter sharing based on regularization. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. LSTM: long short-term-memory.
Figure 6. The high-level view of LearnMTL. LearnMTL: multi-task learning model for soft parameter sharing based on task-relation learning. CRF: conditional random field. LSTM: long short-term-memory.

Results

The experimental settings used to obtain these results are provided in Multimedia Appendix 4.

Comparison Between Our Best Model and Existing Systems

We compared our models with the top three systems in the MADE1.0 challenge. Chapman et al [66] used CRF for NER and random forest for RE. Specifically, two random forest models were used—one for detecting whether relations exist between entities and the other for classifying what specific relation type exists. Xu et al [67] used BiLSTM-CRF for NER with word, prefix, suffix, and character features. For RE, they used SVM and designed features such as positions, distances, bag of words, and bag of entities. Dandala et al [68] also used BiLSTM-CRF for NER, but they input different features into the model such as words, POS tags, and characters. For RE, they employed the BiLSTM-Attention model that takes tokens, entity types, and positions as input.

Full neural systems ([68] and our study) achieve better performances with the MADE dataset than with other systems (Table 1). Although the main methods between the study of Dandala et al [68] and our study are similar, our model is much better, as it significantly improved the F1 for RE by 5%. The reasons for this superiority may be that we used more features than previous work, such as capital information and entity words, and our model attained approximately 0.8% improvement in F1 from MTL.

Comparison Between the Pipeline and MTL Models

The HardMTL model outperforms other models, achieving an F1 of 84.5% in NER and 66.7% in RE (Table 2); the pipeline model ranks second, with F1 values of 84.1% and 65.9%, respectively. The RegMTL model obtains the best recall in both NER (84.5%) and RE (63.6%). Surprisingly, the most-complex MTL model LearnMTL ranked last.

In our experiments, HardMTL successfully boosted the NER F1 by 0.4% ($P=0.003$) and the RE F1 by 0.8% ($P=0.01$), but RegMTL and LearnMTL failed to boost the performances. Thus, the effectiveness of different MTL methods depends on the selected tasks and data. For heterogenous and loosely related tasks such as NER and RE, it is more difficult for MTL to be effective.

Performance of Each Entity Type

Table 3 shows the performance of each entity type. Medication and Route (both F1>90%) were easier to recognize than other types. In contrast, ADE is the most-difficult type to recognize (F1=55%). Other entity types with lower performances included Indication and Duration.

Performance of Each Relation Type

Table 4 shows the performance of each relation type. Medication-ADE relations are the most-difficult type to extract (F1=45.5%). Severity-ADE ranks second (F1=54.1%), followed by Duration-Medication (F1=59.5%). In contrast, Route-Medication and Dosage-Medication extraction are relatively easier, with F1>80%.

Comparison Between the Pipeline Model and MedEx System

MedEx [69] is an end-to-end system used to identify medications and their attributes such as routes and dosages. Therefore, the final results of MedEx correspond to our results for extracting 4 kinds of relations: Route-Medication, Dosage-Medication, Duration-Medication, Frequency-Medication. Table 5 compares MedEx with our model. Our model significantly outperformed MedEx, which demonstrates that our model is a competitive system in this domain.
Table 1. Comparison of our model with the existing systems in the Medication, Indication, and Adverse Drug Events dataset. The microaveraged F1s of relation extraction are shown according to the official evaluation report.

<table>
<thead>
<tr>
<th>System</th>
<th>Named entity recognition</th>
<th>Relation extraction</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapman et al [66]</td>
<td>CRF&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Random forest</td>
<td>59.2</td>
</tr>
<tr>
<td>Xu et al [67]</td>
<td>BiLSTM&lt;sup&gt;b&lt;/sup&gt;-CRF</td>
<td>Support vector machine</td>
<td>59.9</td>
</tr>
<tr>
<td>Dandala et al [68]</td>
<td>BiLSTM-CRF</td>
<td>BiLSTM-Attention</td>
<td>61.7</td>
</tr>
<tr>
<td>Our Best (HardMTL&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>BiLSTM-CRF</td>
<td>BiLSTM-Attention</td>
<td>66.7</td>
</tr>
</tbody>
</table>

<sup>a</sup>CRF: conditional random field  
<sup>b</sup>BiLSTM: bidirectional long short-term memory  
<sup>c</sup>HardMTL: multi-task learning model for hard parameter sharing

Table 2. Performances (%) of the pipeline and multi-task learning models. The values presented are the means of 5 runs of each model. The microaveraged P, R, and F1s of all entity or relation types are shown.

<table>
<thead>
<tr>
<th>Method</th>
<th>Entity recognition</th>
<th>Relation extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Pipeline</td>
<td>85.0</td>
<td>83.2</td>
</tr>
<tr>
<td>HardMTL&lt;sup&gt;a&lt;/sup&gt;</td>
<td>85.0</td>
<td>84.1</td>
</tr>
<tr>
<td>RegMTL&lt;sup&gt;b&lt;/sup&gt;</td>
<td>84.5</td>
<td>84.5</td>
</tr>
<tr>
<td>LearnMTL&lt;sup&gt;c&lt;/sup&gt;</td>
<td>84.5</td>
<td>82.8</td>
</tr>
</tbody>
</table>

<sup>a</sup>HardMTL: multi-task learning model for hard parameter sharing  
<sup>b</sup>RegMTL: multi-task learning model for soft parameter sharing based on regularization  
<sup>c</sup>LearnMTL: multi-task learning model for soft parameter sharing based on task relation learning

Table 3. Performance (%) of each entity type.

<table>
<thead>
<tr>
<th>Entity type</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>91.1</td>
<td>92.0</td>
<td>91.3</td>
</tr>
<tr>
<td>Indication</td>
<td>65.4</td>
<td>64.8</td>
<td>64.8</td>
</tr>
<tr>
<td>Frequency</td>
<td>87.1</td>
<td>86.5</td>
<td>86.3</td>
</tr>
<tr>
<td>Severity</td>
<td>84.6</td>
<td>84.7</td>
<td>84.7</td>
</tr>
<tr>
<td>Dosage</td>
<td>87.9</td>
<td>86.4</td>
<td>88.0</td>
</tr>
<tr>
<td>Duration</td>
<td>75.3</td>
<td>76.6</td>
<td>77.6</td>
</tr>
<tr>
<td>Route</td>
<td>91.6</td>
<td>91.9</td>
<td>91.9</td>
</tr>
<tr>
<td>Adverse drug events</td>
<td>59.5</td>
<td>57.6</td>
<td>55.4</td>
</tr>
<tr>
<td>SSLIF&lt;sup&gt;a&lt;/sup&gt;</td>
<td>83.9</td>
<td>84.8</td>
<td>84.9</td>
</tr>
</tbody>
</table>

<sup>a</sup>SSLIF: any sign, symptom, and disease that is not an ADE or Indication

Table 4. Performance (%) of each relation type.

<table>
<thead>
<tr>
<th>Relation type</th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity-Adverse drug events</td>
<td>55.0</td>
<td>54.4</td>
<td>54.1</td>
</tr>
<tr>
<td>Route-Medication</td>
<td>81.0</td>
<td>82.5</td>
<td>82.1</td>
</tr>
<tr>
<td>Medication-Indication</td>
<td>53.9</td>
<td>52.5</td>
<td>52.9</td>
</tr>
<tr>
<td>Dosage-Medication</td>
<td>80.9</td>
<td>79.8</td>
<td>81.0</td>
</tr>
<tr>
<td>Duration-Medication</td>
<td>60.3</td>
<td>63.7</td>
<td>59.5</td>
</tr>
<tr>
<td>Frequency-Medication</td>
<td>77.7</td>
<td>78.6</td>
<td>78.4</td>
</tr>
<tr>
<td>Medication-Adverse drug events</td>
<td>50.4</td>
<td>47.6</td>
<td>45.5</td>
</tr>
</tbody>
</table>
Table 5. Results (%) of comparisons between our pipeline model and the MedEx system.

<table>
<thead>
<tr>
<th>Entity type</th>
<th>MedEx system</th>
<th>Pipeline model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Route-Medication</td>
<td>71.9</td>
<td>47.9</td>
</tr>
<tr>
<td>Dosage-Medication</td>
<td>29.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Duration-Medication</td>
<td>25.5</td>
<td>15.6</td>
</tr>
<tr>
<td>Frequency-Medication</td>
<td>52.5</td>
<td>36.2</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

Existing systems usually selected a two-step pipeline to address ADE-related information extraction: recognizing entities and extracting relations. BiLSTM-CRF is the most-popular model for NER, whereas the selections of RE models are mixed. All our models outperformed the existing systems in the MADE1.0 challenge, which may be because of the following reasons: First, our models benefited from deep learning that is able to learn better from the data. Second, we enriched the features of deep learning models; therefore, our model outperformed the system [68] that used similar deep learning models as ours.

For MTL, we found that the model using hard parameter sharing (HardMTL) performed better than the other two models using soft parameter sharing (RegMTL and LearnMTL) and that the most complex MTL model, LearnMTL, performed the worst in our data. Our results are not surprising, as different MTL methods depend on different tasks and data [54,56,59]. Overall, MTL more difficult between heterogeneous and loosely related tasks such as NER and RE.

In our experiments, the entity type “ADE” and relation type “Medication-ADE” were the most difficult information to be extracted. Based on our analysis, this is not only due to a lack of training data, but also the intrinsic character of ADEs. ADEs are often implicit in the context without any obvious pattern, which negatively affects the model (Example 1 in Multimedia Appendix 5). In contrast, some entity or relation types with obvious patterns (eg, Medication-Dosage) are easier to identify (Example 2 in Multimedia Appendix 5).

Finally, we found that the performance improved when we used the pretrained word embeddings in the biomedical domain [70] rather than those in the general domain. Furthermore, if the pretrained word embeddings were not tuned, our models would perform better. One likely reason for this is that such a method can alleviate the overfitting problem.

Error Analysis

We randomly sampled hundreds of error instances of NER and RE. Through the manual analyses, we found several sources of errors. For NER, the major false-negative errors are due to long expressions of entities (Examples 3 and 4 in Multimedia Appendix 5). These entities (eg, IgG kappa monoclonal protein) include multiple words; therefore, it is difficult to detect their boundaries. Moreover, the major false-positive errors for NER occur because some entity types are incorrectly recognized as SSLIF (Examples 5 and 6 in Multimedia Appendix 5). This may be because the training instances of SSLIF are ≥10 times those of other entity types such as ADE. Thus, imbalanced data distribution may lead to certain bias of our models.

With respect to RE, the major false-negative errors are due to long distances between target entities (Example 9 in Multimedia Appendix 5). The relation of two entities can be expressed through ≥6 sentences in EHRs; therefore, our model may miss such relations in a long context. In addition, the major false-positive errors for RE occur because relation expressions exist in the instance, but are not related to the target entities (Examples 7 and 8 in Multimedia Appendix 5). For instance, in Example 7 of Multimedia Appendix 5, “His current therapy includes [thalidomide]_{Entity1} 50 mg a day for 2 weeks out of the month. He had been on Velcade, which was stopped secondary to increasing [peripheral neuropathy]_{Entity2}.” “peripheral neuropathy,” and “thalidomide” have no Medication-ADE relation, but the model incorrectly predicts their relation due to the words “secondary to.”

Contributions

The main contributions of this work are as follows: (1) We proposed an up-to-date deep learning model to perform ADE-related information extraction in an end-to-end manner. Our model achieved new state-of-the-art performance, improving the F1 by 4.2% (absolute value). (2) To our knowledge, this is the first attempt to investigate the impact of MTL on two heterogeneous and loosely related tasks (ie, NER and RE). One of our MTL models further improved the F1 by 0.8% (absolute value). (3) Our manually annotated dataset—Medication, Indication, and Adverse Drug Events (MADE) [65]—will be publicly available to support the research on extraction of ADE-related information.

Conclusions

We proposed a deep learning model to detect ADEs and related information. We also investigated MTL on two ADE-related tasks, NER and RE. Our models achieved state-of-the-art performance in an ADE-detection dataset. MTL can improve performance, but it depends on the methods and data used. In the future, we plan to evaluate our models with more related datasets.
Acknowledgments

This work was supported by a grant (5R01HL125089) from the National Institutes of Health and the Investigator Initiated Research grant (1I01HX001457-01) from the Health Services Research & Development Program of the US Department of Veterans Affairs.

Conflicts of Interest

None declared.

Multimedia Appendix 1


[PDF File (Adobe PDF File), 75KB - medinform_v6i4e12159_app1.pdf]

Multimedia Appendix 2

BiLSTM-Attention submodel for RE. BiLSTM: bidirectional long short-term-memory. RE: relation extraction.

[PDF File (Adobe PDF File), 69KB - medinform_v6i4e12159_app2.pdf]

Multimedia Appendix 3

MTL models. MTL: multi-task learning.

[PDF File (Adobe PDF File), 181KB - medinform_v6i4e12159_app3.pdf]

Multimedia Appendix 4

Experimental settings.

[PDF File (Adobe PDF File), 73KB - medinform_v6i4e12159_app4.pdf]

Multimedia Appendix 5

Examples.

[PDF File (Adobe PDF File), 22KB - medinform_v6i4e12159_app5.pdf]

References

7. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ 2016 Dec 03. [Medline: 27143499]


Abbreviations

ADE: adverse drug event
BiLSTM: bidirectional long short-term-memory
CRF: conditional random field
EHR: electronic health record
FAERS: Food and Drug Administration Adverse Event Reporting System
LSTM: long short-term-memory
MADE: Medication, Indication, and Adverse Drug Events
MTL: multi-task learning
NER: named entity recognition
NLP: natural language processing
RE: relation extraction
POS: part of speech
SVM: support vector machine