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 Li[1,2,3], PhD; Weisong Liu[1,2,3], PhD; Hong Yu[1,2,3,4], PhD

[1]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States

[2]Center for Healthcare Organization and Implementation Research, Bedford Veterans Affairs Medical Center, Bedford, MA, United States

[3]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[4]School 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.

XSL•FO
RenderX

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 inefficient 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.

Figure 1. Study overview. NER: named entity recognition. RE: relation extraction. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. MTL: multi-task learning. MADE: Medication, Indication, and Adverse Drug Events. HardMTL: multi-task learning model for hard parameter sharing. RegMTL: multi-task learning model for soft parameter sharing based on regularization. LearnMTL: multi-task learning model for soft parameter sharing based on task-relation learning.
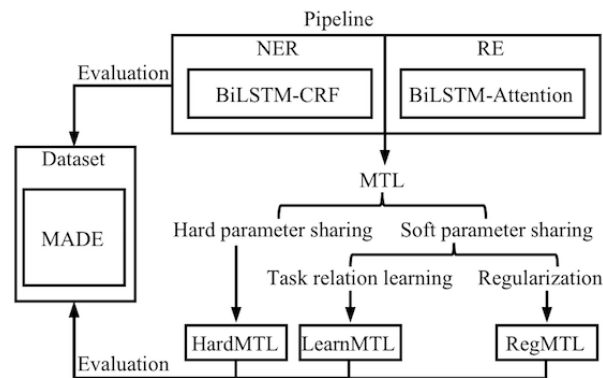
Named Entity Recognition

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, Mehryary 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, $LSTM^{ner}$ and $LSTM^{re}$, and shared a BiLSTM layer, $LSTM^{share}$. During training, the shared BiLSTM layer $LSTM^{share}$ 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_5, 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^{ner} and D^{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.
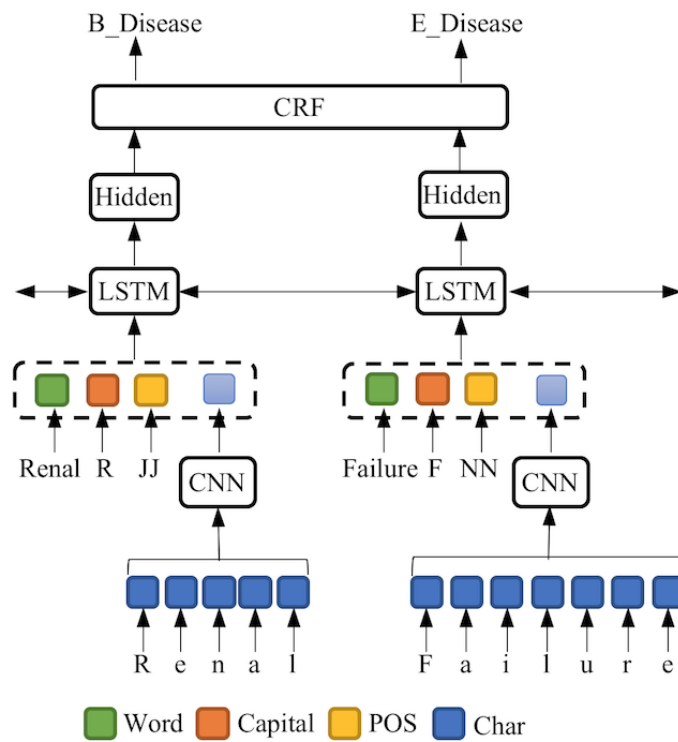

Figure 3. RE submodel. The target entities are "renal failure" (e_1) and "antibiotics" (e_2). Positions represent token distances to the target entities. RE: relation extraction. LSTM: long short-term memory. POS: part of speech.
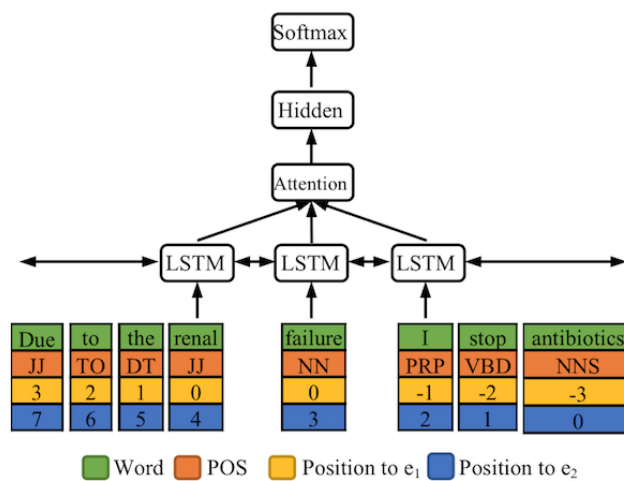
Figure 4. The high-level view of HardMTL. For conciseness, "LSTM" indicates a BiLSTM layer, and the layers above the BiLSTM layer are denoted as D^{new} and D^{re}. The forward procedures for an NER instance and an RE instance are indicated by blue and green arrow lines, respectively. HardMTL: multi-task 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.
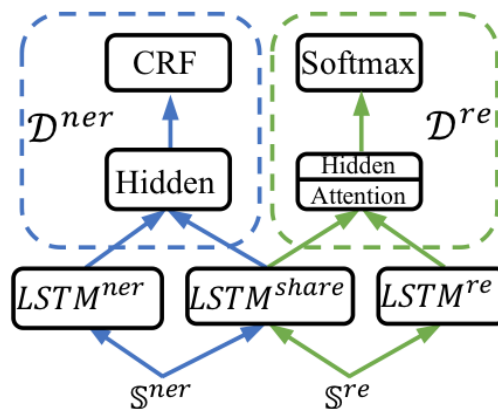

Figure 5. The high-level view of RegMTL. $LSTM_1^{ner}$ and $LSTM_2^{ner}$ indicate the first and second BiLSTM layers of the NER model. $LSTM_1^{re}$ and $LSTM_2^{re}$ indicate the first and second BiLSTM layers of the RE model. NER: named entity recognition. RE: relation extraction. RegMTL: multi-task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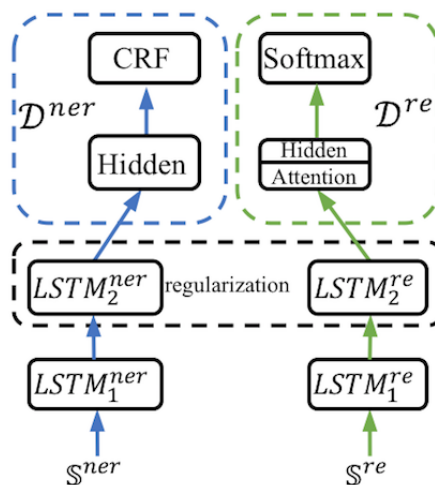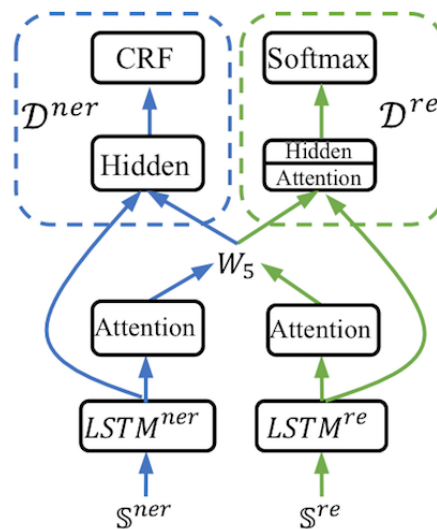
Figure 6. The high-level view of LearnMTL. LearnMTL: multi-task learning model for soft parameter sharing based on task-relation learning. CRF: conditional random field. LSTM: long short-term-memory.

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=.003) and the RE F1 by 0.8% (P=.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.

System	Named entity recognition	Relation extraction	F1
Chapman et al [66]	CRF[a]	Random forest	59.2
Xu et al [67]	BiLSTM[b]-CRF	Support vector machine	59.9
Dandala et al [68]	BiLSTM-CRF	BiLSTM-Attention	61.7
Our Best (HardMTL[c])	BiLSTM-CRF	BiLSTM-Attention	66.7

[a]CRF: conditional random field

[b]BiLSTM: bidirectional long short-term memory

[c]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.

Method	Entity recognition			Relation extraction		
	P	R	F1	P	R	F1
Pipeline	85.0	83.2	84.1	69.8	62.4	65.9
HardMTL[a]	85.0	84.1	84.5	70.2	63.6	66.7
RegMTL[b]	84.5	84.5	84.5	66.7	63.6	65.1
LearnMTL[c]	84.5	82.8	83.6	67.2	61.5	64.2

[a]HardMTL: multi-task learning model for hard parameter sharing

[b]RegMTL: multi-task learning model for soft parameter sharing based on regularization

[c]LearnMTL: multi-task learning model for soft parameter sharing based on task relation learning

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

Entity type	P	R	F1
Medication	91.1	92.0	91.3
Indication	65.4	64.8	64.8
Frequency	87.1	86.5	86.3
Severity	84.6	84.7	84.7
Dosage	87.9	86.4	88.0
Duration	75.3	76.6	77.6
Route	91.6	91.9	91.9
Adverse drug events	59.5	57.6	55.4
SSLIF[a]	83.9	84.8	84.9

[a]SSLIF: any sign, symptom, and disease that is not an ADE or Indication

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

Relation type	P	R	F1
Severity-Adverse drug events	55.0	54.4	54.1
Route-Medication	81.0	82.5	82.1
Medication-Indication	53.9	52.5	52.9
Dosage-Medication	80.9	79.8	81.0
Duration-Medication	60.3	63.7	59.5
Frequency-Medication	77.7	78.6	78.4
Medication-Adverse drug events	50.4	47.6	45.5

Table 5. Results (%) of comparisons between our pipeline model and the MedEx system.

Entity type	MedEx system			Pipeline model		
	P	R	F1	P	R	F1
Route-Medication	71.9	47.9	57.5	81.0	82.5	82.1
Dosage-Medication	29.7	3.5	6.2	80.9	79.8	81.0
Duration-Medication	25.5	15.6	19.4	60.3	63.7	59.5
Frequency-Medication	52.5	36.2	42.8	77.7	78.6	78.4

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

XSL•FO

RenderX

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

Conflicts of Interest

None declared.

Multimedia Appendix 1

BiLSTM-CRF submodel for NER. BiLSTM: bidirectional long short-term-memory. CRF: conditional random field. NER: named entity recognition.

[PDF File (Adobe PDF File), 75KB-Multimedia Appendix 1]

Multimedia Appendix 2

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

[PDF File (Adobe PDF File), 69KB-Multimedia Appendix 2]

Multimedia Appendix 3

MTL models. MTL: multi-task learning.

[PDF File (Adobe PDF File), 181KB-Multimedia Appendix 3]

Multimedia Appendix 4

Experimental settings.

[PDF File (Adobe PDF File), 73KB-Multimedia Appendix 4]

Multimedia Appendix 5

Examples.

[PDF File (Adobe PDF File), 22KB-Multimedia Appendix 5]

References

1. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA 1995 Jul 05;274(1):29-34. [Medline: 7791255]

2. Nebeker JR, Hoffman JM, Weir CR, Bennett CL, Hurdle JF. High rates of adverse drug events in a highly computerized hospital. Arch Intern Med 2005 May 23;165(10):1111-1116. [doi: 10.1001/archinte.165.10.1111] [Medline: 15911723]

3. Fattinger K, Roos M, Vergères P, Holenstein C, Kind B, Masche U, et al. Epidemiology of drug exposure and adverse drug reactions in two swiss departments of internal medicine. Br J Clin Pharmacol 2000 Feb;49(2):158-167 [FREE Full text] [Medline: 10671911]

4. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. JAMA 1997;277(4):301-306. [Medline: 9002492]

5. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA 1997;277(4):307-311. [Medline: 9002493]

6.   Null G, Dean C, Feldman M, Rasio D. Death by Medicine. Journal of Orthomolecular Medicine 2005;20(1):34 [FREE Full text]

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

8.   Handler SM, Altman RL, Perera S, Hanlon JT, Studenski SA, Bost JE, et al. A systematic review of the performance characteristics of clinical event monitor signals used to detect adverse drug events in the hospital setting. J Am Med Inform Assoc 2007;14(4):451-458 [FREE Full text] [doi: 10.1197/jamia.M2369] [Medline: 17460130]

9.   Kaushal R, Jha AK, Franz C, Glaser J, Shetty KD, Jaggi T, BrighamWomen's Hospital CPOE Working Group. Return on investment for a computerized physician order entry system. J Am Med Inform Assoc 2006 May;13(3):261-266 [FREE Full text] [doi: 10.1197/jamia.M1984] [Medline: 16501178]

10. Haas JS, Iyer A, Orav EJ, Schiff GD, Bates DW. Participation in an ambulatory e-pharmacovigilance system. Pharmacoepidemiol Drug Saf 2010 Sep;19(9):961-969. [doi: 10.1002/pds.2006] [Medline: 20623512]

11. Edlavitch SA. Adverse Drug Event Reporting. Arch Intern Med 1988 Jul 01;148(7):1499. [doi: 10.1001/archinte.1988.00380070017005]

12. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? British Journal of Clinical Pharmacology 2012 Jul 05;38(5):401-404. [doi: 10.1111/j.1365-2125.1994.tb04373.x]

13. ProPublica. Dollars for Docs URL: https://projects.propublica.org/docdollars/ [accessed 2018-10-28] [WebCite Cache ID 73UddA54z]

14. Kumar A. The Newly Available FAERS Public Dashboard: Implications for Health Care Professionals. Hospital Pharmacy 2018 Aug 23. [doi: 10.1177/0018578718795271]

15. Wei C, Peng Y, Leaman R, Davis AP, Mattingly CJ, Li J, et al. Assessing the state of the art in biomedical relation extraction: overview of the BioCreative V chemical-disease relation (CDR) task. Database (Oxford) 2016 [FREE Full text] [doi: 10.1093/database/baw032] [Medline: 26994911]

16. Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. J Biomed Inform 2012 Oct;45(5):885-892 [FREE Full text] [doi: 10.1016/j.jbi.2012.04.008] [Medline: 22554702]

17. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]

18. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]

19. Gurwitz JH. Incidence and Preventability of Adverse Drug Events Among Older Persons in the Ambulatory Setting. JAMA 2003 Mar 05;289(9):1107. [doi: 10.1001/jama.289.9.1107]

20. Jagannatha AN, Yu H. Bidirectional RNN for Medical Event Detection in Electronic Health Records. Proc Conf 2016 Jun;2016:473-482 [FREE Full text] [Medline: 27885364]

21. Munkhdalai T, Liu F, Yu H. Clinical Relation Extraction Toward Drug Safety Surveillance Using Electronic Health Record Narratives: Classical Learning Versus Deep Learning. JMIR Public Health Surveill 2018 Apr 25;4(2):e29 [FREE Full text] [doi: 10.2196/publichealth.9361] [Medline: 29695376]

22. Hurdle JF, Weir CR, Roth B, Hoffman J, Nebeker JR. Critical gaps in the world's largest electronic medical record: Ad Hoc nursing narratives and invisible adverse drug events. AMIA Annu Symp Proc 2003:309-312 [FREE Full text] [Medline: 14728184]

23. Xu J, Wu Y, Zhang Y, Wang J, Lee H, Xu H. CD-REST: a system for extracting chemical-induced disease relation in literature. Database (Oxford) 2016 May 25. [doi: 10.1093/database/baw036] [Medline: 27016700]

24. Finkel J, Dingare S, Manning CD, Nissim M, Alex B, Grover C. Exploring the boundaries: gene and protein identification in biomedical text. BMC Bioinformatics 2005;6 Suppl 1:S5 [FREE Full text] [doi: 10.1186/1471-2105-6-S1-S5] [Medline: 15960839]

25. Lafferty J, McCallum A, Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. 2001 Presented at: The Eighteenth International Conference on Machine Learning; 2001; San Francisco, CA, USA p. 282-289 URL: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers

26. Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer-Verlag, LLC; 2009.

27. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]

28. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. Nature 2018. [doi: 10.1038/s41746-018-0029-1]

29. Caruana R. Multitask Learning. Machine Learning 1997 Jul;28(1):41-75. [doi: 10.1023/A:1007379606734]

30. Davis AP, Grondin CJ, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, et al. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. Nucleic Acids Res 2015 Jan;43(Database issue):D914-D920 [FREE Full text] [doi: 10.1093/nar/gku935] [Medline: 25326323]

XSL•FO

RenderX

31. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 2010;6:343 [FREE Full text] [doi: 10.1038/msb.2009.98] [Medline: 20087340]

32. Polepalli RB, Belknap SM, Li Z, Frid N, West DP, Yu H. Automatically Recognizing Medication and Adverse Event Information From Food and Drug Administration's Adverse Event Reporting System Narratives. JMIR Med Inform 2014 Jun 27;2(1):e10 [FREE Full text] [doi: 10.2196/medinform.3022] [Medline: 25600332]

33. Xu R, Wang Q. Automatic construction of a large-scale and accurate drug-side-effect association knowledge base from biomedical literature. J Biomed Inform 2014 Oct;51:191-199 [FREE Full text] [doi: 10.1016/j.jbi.2014.05.013] [Medline: 24928448]

34. Liu M, Cai R, Hu Y, Matheny ME, Sun J, Hu J, et al. Determining molecular predictors of adverse drug reactions with causality analysis based on structure learning. J Am Med Inform Assoc 2014 Mar;21(2):245-251 [FREE Full text] [doi: 10.1136/amiajnl-2013-002051] [Medline: 24334612]

35. Yildirim P, Majnarić L, Ekmekci O, Holzinger A. Knowledge discovery of drug data on the example of adverse reaction prediction. BMC Bioinformatics 2014;15 Suppl 6:S7 [FREE Full text] [Medline: 25079450]

36. Kim J, Ohta T, Tateisi Y, Tsujii J. GENIA corpus--semantically annotated corpus for bio-textmining. Bioinformatics 2003;19 Suppl 1:i180-i182. [Medline: 12855455]

37. Doğan RI, Leaman R, Lu Z. NCBI disease corpus: a resource for disease name recognition and concept normalization. J Biomed Inform 2014 Feb;47:1-10 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.006] [Medline: 24393765]

38. Suominen H, Salanterä S, Velupillai S, Chapman W, Savova G, Elhadad N, et al. Overview of the ShARe/CLEF eHealth Evaluation Lab. 2013 Presented at: International Conference of the Cross-Language Evaluation Forum for European Languages; 2013; Berlin. [doi: 10.1007/978-3-642-40802-1_24]

39. Hanisch D, Fundel K, Mevissen H, Zimmer R, Fluck J. ProMiner: rule-based protein and gene entity recognition. BMC Bioinformatics 2005;6 Suppl 1:S14 [FREE Full text] [doi: 10.1186/1471-2105-6-S1-S14] [Medline: 15960826]

40. Campos D, Matos S, Oliveira JL. Gimli: open source and high-performance biomedical name recognition. BMC Bioinformatics 2013 Feb 15;14:54 [FREE Full text] [doi: 10.1186/1471-2105-14-54] [Medline: 23413997]

41. Hsu C, Chang Y, Kuo C, Lin Y, Huang H, Chung I. Integrating high dimensional bi-directional parsing models for gene mention tagging. Bioinformatics 2008 Jul 01;24(13):i286-i294 [FREE Full text] [doi: 10.1093/bioinformatics/btn183] [Medline: 18586726]

42. Zhou G, Zhang J, Su J, Shen D, Tan C. Recognizing names in biomedical texts: a machine learning approach. Bioinformatics 2004 May 01;20(7):1178-1190. [doi: 10.1093/bioinformatics/bth060] [Medline: 14871877]

43. Tang B, Chen Q, Wang X, Wu Y, Zhang Y, Jiang M, et al. Recognizing Disjoint Clinical Concepts in Clinical Text Using Machine Learning-based Methods. AMIA Annu Symp Proc 2015;2015:1184-1193 [FREE Full text] [Medline: 26958258]

44. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition. : Association for Computational Linguistics; 2016 Presented at: the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2016; San Diego, California. [doi: 10.18653/v1/N16-1030]

45. Yang J, Liang S, Zhang Y. Design Challenges and Misconceptions in Neural Sequence Labeling. 2018 Presented at: Proceedings of the 27th International Conference on Computational Linguistics; 2018; Santa Fe, New Mexico, USA p. 3879-3889 URL: http://aclweb.org/anthology/C18-1327

46. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011;18(5):552-556 [FREE Full text] [doi: 10.1136/amiajnl-2011-000203] [Medline: 21685143]

47. Fundel K, Küffner R, Zimmer R. RelEx--relation extraction using dependency parse trees. Bioinformatics 2007 Feb 01;23(3):365-371. [doi: 10.1093/bioinformatics/btl616] [Medline: 17142812]

48. Kilicoglu H, Rosemblat G, Fiszman M, Rindflesch TC. Sortal anaphora resolution to enhance relation extraction from biomedical literature. BMC Bioinformatics 2016 Apr 14;17(1). [doi: 10.1186/s12859-016-1009-6] [Medline: 27080229]

49. Lavergne T, Grouin C, Zweigenbaum P. The contribution of co-reference resolution to supervised relation detection between bacteria and biotopes entities. BMC Bioinformatics 2015;16 Suppl 10:S6 [FREE Full text] [doi: 10.1186/1471-2105-16-S10-S6] [Medline: 26201352]

50. Airola A, Pyysalo S, Björne J, Pahikkala T, Ginter F, Salakoski T. All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. BMC Bioinformatics 2008 Nov 19;9 Suppl 11:S2 [FREE Full text] [doi: 10.1186/1471-2105-9-S11-S2] [Medline: 19025688]

51. Kordjamshidi P, Roth D, Moens M. Structured learning for spatial information extraction from biomedical text: bacteria biotopes. BMC Bioinformatics 2015 Apr 25;16:129 [FREE Full text] [doi: 10.1186/s12859-015-0542-z] [Medline: 25909637]

52. Luo Y, Cheng Y, Uzuner Ö, Szolovits P, Starren J. Segment convolutional neural networks (Seg-CNNs) for classifying relations in clinical notes. J Am Med Inform Assoc 2018 Jan 01;25(1):93-98. [doi: 10.1093/jamia/ocx090] [Medline: 29025149]

53. Verga P, Strubell E, McCallum A. Simultaneously Self-Attending to All Mentions for Full-Abstract Biological Relation Extraction. 2018 Presented at: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics; 2018; New Orleans, Louisiana p. 872-884 URL: http://aclweb.org/anthology/N18-1080

54. Miwa M, Bansal M. End-to-end Relation Extraction using LSTMs on Sequences and Tree Structures. 2016 Presented at: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics; 2016; Berlin, Germany p. 1105-1116. [doi: 10.18653/v1/P16-1105]

55. Mehryary F, Hakala K, Kaewphan S, Björne J, Salakoski T, Ginter F. End-to-End System for Bacteria Habitat Extraction. 2017 Presented at: BioNLP; 2017; Vancouver, Canada p. 80-90 URL: http://www.aclweb.org/anthology/W17-2310

56. Misra I, Shrivastava A, Gupta A, Hebert M. Cross-Stitch Networks for Multi-task Learning. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, Nevada p. 3994-4003 URL: http://openaccess.thecvf.com/content_cvpr_2016/papers/Misra_Cross-Stitch_Networks_for_CVPR_2016_paper.pdf [doi: 10.1109/CVPR.2016.433]

57. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. The Journal of Machine Learning Research 2011;12:2537 [FREE Full text]

58. Ruder S. arXiv. 2017. An Overview of Multi-Task Learning in Deep Neural Networks URL: https://arxiv.org/abs/1706.05098 [accessed 2018-11-15] [WebCite Cache ID 73wxJmB0m]

59. Duong L, Cohn T, Bird S, Cook P. Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser. 2015 Presented at: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics; 2015; Beijing, China p. 845-850 URL: http://anthology.aclweb.org/P/P15/P15-2139.pdf

60. Argyriou A, Evgeniou T, Pontil M. Multi-task Feature Learning. 2006 Presented at: Proceedings of the 19th International Conference on Neural Information Processing Systems; 2006; Cambridge, MA p. 41-48 URL: https://papers.nips.cc/paper/3143-multi-task-feature-learning.pdf

61. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation 1997 Nov;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735]

62. Luong T, Pham H, Manning C. Effective Approaches to Attention-based Neural Machine Translation. 2015 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2015; Lisbon, Portugal p. 1412-1421 URL: http://aclweb.org/anthology/D15-1166

63. Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation Classification via Convolutional Deep Neural Network. 2014 Presented at: Proceedings of the 25th International Conference on Computational Linguistics; 2014; Dublin, Ireland p. 2335-2344 URL: http://www.aclweb.org/anthology/C14-1220

64. Chen X, Cardie C. Multinomial Adversarial Networks for Multi-Domain Text Classification. 2018 Presented at: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics; 2018; New Orleans, Louisiana p. 1226-1240 URL: http://www.aclweb.org/anthology/N18-1111

65. Jagannatha A, Liu F, Liu W, Yu H. Overview of the First Natural Language Processing Challenge for Extracting Medication, Indication and Adverse Drug Events from Electronic Health Record Notes (MADE1.0). Drug Safety 2018 [FREE Full text]

66. Chapman A, Peterson K, Alba P, DuVall S, Patterson O. Hybrid system for adverse drug event detection. 2018 Presented at: International Workshop on Medication and Adverse Drug Event Detection; 2018; Worcester, MA, United States p. 16-24 URL: http://proceedings.mlr.press/v90/chapman18a/chapman18a.pdf

67. Xu D, Yadav V, Bethard S. UArizona at the MADE1.0 NLP Challenge. 2018 Presented at: International Workshop on Medication and Adverse Drug Event Detection; 2018; Worcester, MA, United States p. 57-65 URL: http://proceedings.mlr.press/v90/xu18a/xu18a.pdf

68. Dandala B, Joopudi V, Devarakonda M. IBM Research System at MADE 2018: Detecting Adverse Drug Events from Electronic Health Records. 2018 Presented at: International Workshop on Medication and Adverse Drug Event Detection; 2018; Worcester, MA, United States p. 39-47 URL: http://proceedings.mlr.press/v90/dandala18a/dandala18a.pdf

69. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc 2010;17(1):19-24 [FREE Full text] [doi: 10.1197/jamia.M3378] [Medline: 20064797]

70. Pyysalo S, Ginter F, Moen H, Salakoski T, Ananiadou S. Distributional semantics resources for biomedical text processing. 2013 Presented at: the 5th International Symposium on Languages in Biology and Medicine; 2013; Tokyo, Japan p. 39-43 URL: http://bio.nlplab.org/pdf/pyysalo13literature.pdf

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

XSL•FO
RenderX

NLP: natural language processing
RE: relation extraction
POS: part of speech
SVM: support vector machine

XSL•FO
RenderX