Original Paper

Extraction of Information Related to Drug Safety Surveillance From Electronic Health Record Notes: Joint Modeling of Entities and Relations Using Knowledge-Aware Neural Attentive Models

Bharath Dandala, BTECH, MS, PhD; Venkata Joopudi, BTECH, MS; Ching-Huei Tsou, BS, MEng, PhD; Jennifer J Liang, SB, MD; Parthasarathy Suryanarayanan, BSc, BTECH
IBM Research, Yorktown Heights, NY, United States

Corresponding Author:
Bharath Dandala, BTECH, MS, PhD
IBM Research
1101 Kitchawan Rd
Yorktown Heights, NY, 10598
United States
Phone: 1 9403673972
Email: bdand@us.ibm.com

Abstract

Background: An adverse drug event (ADE) is commonly defined as “an injury resulting from medical intervention related to a drug.” Providing information related to ADEs and alerting caregivers at the point of care can reduce the risk of prescription and diagnostic errors and improve health outcomes. ADEs captured in structured data in electronic health records (EHRs) as either coded problems or allergies are often incomplete, leading to underreporting. Therefore, it is important to develop capabilities to process unstructured EHR data in the form of clinical notes, which contain a richer documentation of a patient’s ADE. Several natural language processing (NLP) systems have been proposed to automatically extract information related to ADEs. However, the results from these systems showed that significant improvement is still required for the automatic extraction of ADEs from clinical notes.

Objective: This study aims to improve the automatic extraction of ADEs and related information such as drugs, their attributes, and reason for administration from the clinical notes of patients.

Methods: This research was conducted using discharge summaries from the Medical Information Mart for Intensive Care III (MIMIC-III) database obtained through the 2018 National NLP Clinical Challenges (n2c2) annotated with drugs, drug attributes (ie, strength, form, frequency, route, dosage, duration), ADEs, reasons, and relations between drugs and other entities. We developed a deep learning–based system for extracting these drug-centric concepts and relations simultaneously using a joint method enhanced with contextualized embeddings, a position-attention mechanism, and knowledge representations. The joint method generated different sentence representations for each drug, which were then used to extract related concepts and relations simultaneously. Contextualized representations trained on the MIMIC-III database were used to capture context-sensitive meanings of words. The position-attention mechanism amplified the benefits of the joint method by generating sentence representations that capture long-distance relations. Knowledge representations were obtained from graph embeddings created using the US Food and Drug Administration Adverse Event Reporting System database to improve relation extraction, especially when contextual clues were insufficient.

Results: Our system achieved new state-of-the-art results on the n2c2 data set, with significant improvements in recognizing crucial drug–reason (F1=0.650 versus F1=0.579) and drug–ADE (F1=0.490 versus F1=0.476) relations.

Conclusions: This study presents a system for extracting drug-centric concepts and relations that outperformed current state-of-the-art results and shows that contextualized embeddings, position-attention mechanisms, and knowledge graph embeddings effectively improve deep learning–based concepts and relation extraction. This study demonstrates the potential for deep learning–based methods to help extract real-world evidence from unstructured patient data for drug safety surveillance.

(JMIR Med Inform 2020;8(7):e18417) doi: 10.2196/18417

https://medinform.jmir.org/2020/7/e18417
KEYWORDS
electronic health records; adverse drug events; natural language processing; deep learning; information extraction; adverse drug reaction reporting systems; named entity recognition; relation extraction

Introduction

Background
An electronic health record (EHR) is the systematized collection of electronically stored health information of patients and the general population in a digital format [1]. Clinical notes in EHRs summarize interactions that occur between patients and health care providers [2]. These notes include observations, impressions, treatments, drug use, adverse drug events (ADEs), and other activities arising from each interaction between the patient and the health care system. Extracting useful information such as ADEs from these notes and alerting caregivers at the point of care has the potential to improve patient health outcomes.

An ADE is commonly defined as “an injury resulting from medical intervention related to a drug” [3]. ADEs are a major public health concern and one of the leading causes of morbidity and mortality [4]. Studies have shown the substantial economic burden of these undesired effects [5,6]. Although drug safety and efficacy are tested during premarketing randomized clinical trials, these trials may not detect all ADEs because such studies are often small, short, and biased by the exclusion of patients with comorbid diseases. With the limited information available when a drug is marketed, postmarketing surveillance has become increasingly important. Spontaneous reporting systems, such as the US Food and Drug Administration Adverse Event Reporting System (FAERS) [7], are monitoring mechanisms for postmarketing surveillance that enable both physicians and patients to report ADEs. However, previous studies [8-10] have exposed various inadequacies with such systems, including underreporting, reporting biases, and incomplete information, prompting researchers to explore additional sources to detect ADEs from real-world data.

Several efforts have been made to extract ADEs automatically from disparate information sources, including EHRs [11-13], spontaneous reporting systems [14-16], social media [17-20], search queries on the web via search engine logs [21,22], and biology and chemistry knowledge bases [23-25]. Furthermore, the clinical natural language processing (NLP) community has organized several open challenges such as the 2010 Informatics for Integrating Biology & the Bedside/Veterans Affairs NLP Challenge [26], Text Analysis Conference 2017 Adverse Drug Reactions Track [27], and BioCreative V Chemical Disease Relation task [28]. Recently, 2 such challenges, Medication and Adverse Drug Events from Electronic Health Records (MADE) [29] and the 2018 National NLP Clinical Challenges (n2c2) Shared Task Track 2 [30], were organized to extract drugs, drug attributes, ADEs, reasons for prescribing drugs, and their relations from clinical notes. The results from these 2 challenges showed that deep learning techniques outperform traditional machine learning techniques for this task, and significant improvement is still required for drug–{ADE, reason} relation extraction. Specifically, the organizers of these challenges hypothesized that models that can effectively incorporate the larger context to capture long-distance relations or leverage knowledge to capture implicit relations will likely improve the performance of future systems.

Considering these conclusions, we developed a joint deep learning–based relation extraction system that helps in extracting long-distance relations through a position-attention mechanism and implicit relations through external knowledge from the FAERS. To the best of our knowledge, no previous research has been conducted on using the position-attention mechanism and domain-specific knowledge graph embeddings in ADE detection.

Relevant Literature

Adverse Drug Event Detection

From the viewpoint of NLP, effective techniques for entity and relation extraction are fundamental requirements in automatic ADE extraction. Entity and relation extraction from text has traditionally been treated as a pipeline of 2 separate subtasks: named entity recognition (NER) and relation classification. Previous studies employed traditional machine learning techniques [31-34], such as conditional random fields (CRF) [35] for NER and support vector machines [36] for relation classification. Several recent approaches [37-44], developed on MADE 1.0 [29] and 2018 n2c2 Shared Task Track 2 [30] data sets, employed deep learning techniques, such as bidirectional, long short-term memory–conditional random fields (BiLSTM-CRFs) [45], for NER and convolutional neural network (CNN) [46] for relation classification, and showed numerous advantages resulting in better performance and less feature engineering. However, there is an inevitable error propagation issue with pipeline-based methods because of the following:

1. NER relying on sequence-labeling techniques suffers from lossy representation when there are overlapping annotations on entities. For example, in “she was on furosemide and became hypotensive requiring norepinephrine,” hypotensive is an ADE with respect to furosemide but a reason with respect to norepinephrine.

2. NER approaches usually take an input context window that may not contain the necessary information to determine the appropriate label (ie, ADE, reason, no label). For example, in “Patient reports nausea. Started on ondansetron,” the identification of nausea as a reason requires information from both sentences.

3. Signs or symptoms are only labeled as ADE or reason if they are related to a drug (ie, not all signs or symptoms in the clinical note are annotated). This makes the corpus less suitable to train an effective relation classification model as it misses negative candidate pairs for drug–{ADE, reason} relations.

To address the first 2 issues, we previously proposed a joint method that outperformed the pipeline method for concept and...
relation extraction on a similar data set (MADE 1.0) [37]. In a separate study, Li et al [47] proposed a joint method using multitask learning [48] and made similar observations. To address the third issue, which was introduced with the n2c2 data set, Wei et al [38] proposed a novel label-encoding scheme to jointly extract ADE, reason, drug attributes, and their relations.

Attention-Based Relation Extraction
The attention mechanism allows neural networks to selectively focus on specific information [49-51]. This has proven to be effective for NLP problems with long-distance dependencies such as NER and relation extraction. Zhou et al [52] proposed an attention-based BiLSTM network and demonstrated its effectiveness in selectively focusing on words that have decisive effects on relation classification. Next, Zhang et al [53] extended the attention mechanism to help networks not only focus on words based on the semantic information of the sentence but also the global positions of entities within the sentence. Recently Dai et al [54] introduced a position-attention mechanism for joint extraction of entities and overlapping relations. The position-attention mechanism builds on self-attention by focusing on both the global dependencies of the input and tokens of the target entities of interest for relation extraction. Recent research [37,55] on ADE extraction showed the benefits of self-attention mechanisms in pipeline-based methods, specifically for relation classification. However, to the best of our knowledge, no previous work has focused on using self-attention or position-attention mechanisms for joint extraction of entities and relations for ADE extraction.

Knowledge-Aware Relation Extraction
Several approaches [56-59] in the open domain have shown that incorporating embeddings learned from knowledge bases benefit deep learning–based relation classification. These embeddings are typically learned using translation-based methods such as TransE [60], TransH [61], and TransR [62]; walk-based methods such as DeepWalk [63] and node2vec [64]; or neural network–based methods such as large-scale information network embedding (LINE) [65] and bipartite network embedding [66].

Clinical notes are typically written for medical professionals. Hence, a certain degree of medical knowledge is assumed by the authors, which is not explicitly expressed in the text. This is especially true for relations between clinical findings and drugs, where a drug could either cause (ADE) or treat (reason) a clinical finding. In our previous study [37], we showed that augmenting knowledge base features such as proportional report ratio and reporting odds ratio calculated from the FAERS into deep learning models can benefit relation classification. Recently, Chen et al [67] proposed a hybrid clinical NLP system by combining a general knowledge-based system using the Unified Medical Language System (UMLS) and BiLSTM-CRF for concept extraction and attention-BiLSTM for relation classification. However, to the best of our knowledge, no previous work has focused on using knowledge graph embeddings generated from the FAERS for joint extraction of entities and relations for ADE extraction.

Methods

Data Set
The n2c2 data set consists of 505 deidentified clinical narratives, of which 303 and 202 narratives were released as train and test data sets, respectively. Each narrative was manually annotated with drug-centric entities, including drugs, their attributes (strength, form, frequency, route, dosage, and duration), ADEs, reasons, and relations between drugs and other entities (drug−{attributes, ADE, reason}). Drug−{attributes} represent 6 different types of relations: drug−{strength, form, frequency, route, dosage, duration}. Figure 1 presents an example with annotations. Tables 1 and 2 present the statistical overview of the annotated entities and relations.

Figure 1. An illustration with annotations for entities and relations. ADE: adverse drug event; HTN: hypertension; QHS: every night at bedtime.
Table 1. Entities in the data set.

| Entity type | Number of annotations | Example | Description |
|-------------|-----------------------|---------|-------------|
| Drug        | 16,225 (31.84)        | Coumadin | Name of the drug |
| Strength    | 6691 (13.13)          | 5 mg    | Strength of the drug |
| Form        | 6651 (13.05)          | Tablet  | Form of the drug |
| Frequency   | 6281 (12.32)          | Daily   | Frequency of the drug |
| Route       | 5476 (10.75)          | By mouth | Route in which the drug is administered |
| Dosage      | 4221 (8.28)           | 1       | Dosage of the drug |
| Duration    | 592 (1.16)            | For 5 days | Duration of the drug |
| ADEа        | 959 (1.88)            | Rash    | Adverse reaction of the drug |
| Reason      | 3855 (7.57)           | Constipation | Indication if it is an affliction that a physician is actively treating with a drug |
| Total       | 50,951 (100.00)       | N/Aб  | N/A |

аADE: adverse drug event.
бNot applicable.

Table 2. Relations in the data set.

| Relation type | Relations | Inter sentential relations | Exampleа |
|---------------|-----------|----------------------------|---------|
|               | Train, n (%) | Test, n (%) | Train, n (%) | Test, n (%) |
| Drug–strength | 6702 (18.44) | 4244 (18.09) | 80 (1.19) | 59 (1.39) | Lisinopril 1×5 mg tablet orally daily for 7 days |
| Drug–form     | 6654 (18.31) | 4374 (18.64) | 259 (3.89) | 144 (3.29) | Lisinopril 1×5 mg tablet orally daily for 7 days |
| Drug–frequency| 6310 (17.36) | 4034 (17.19) | 372 (5.90) | 238 (5.90) | Lisinopril 1×5 mg tablet orally daily for 7 days |
| Drug–route    | 5538 (15.24) | 3546 (15.11) | 199 (3.59) | 149 (4.20) | Lisinopril 1×5 mg tablet orally daily for 7 days |
| Drug–dosage   | 4225 (11.62) | 2695 (11.49) | 135 (3.20) | 102 (3.78) | Lisinopril1×5 mg tablet orally daily for 7 days |
| Drug–duration | 643 (1.80)   | 426 (1.80)   | 34 (5.4)   | 43 (10.0)  | Lisinopril 1×5 mg tablet orally daily for 7 days |
| Drug–ADEб     | 1107 (3.05)  | 733 (3.10)   | 254 (22.94)| 139 (18.9) | Patient is experiencing muscle pain, secondary to statin therapy for coronary artery disease |
| Drug–reason   | 5169 (14.22) | 3410 (14.53) | 1638 (31.69)| 1088 (31.91)| Patient is experiencing muscle pain, secondary to statin therapy for coronary artery disease |
| Total         | 36,348 (100.00) | 23,462 (100.00) | 2971 (8.17) | 1947 (8.30) | N/Aс |

аItalics indicate entities participating in the specified relation type.
бADE: adverse drug event.
сNot applicable.

Preprocessing

Sentence boundary detection (SBD) and tokenization are often treated as solved problems in NLP and carried out using off-the-shelf toolkits such as Apache Natural Language Toolkit [68], Explosion AI spaCy [69] or the Stanford CoreNLP toolkit [70]. However, these are still difficult and critical problems [71] in the clinical domain because (1) sentence ends are frequently indicated by layout and not by punctuation and (2) white space is not always present to indicate token boundaries (eg, 50 mg). To address these issues, we incorporated domain-specific rules sensitive to low-level features such as capitalization, text-wrap properties, indentation, and punctuation into the spaCy tokenizer and SBD models. These custom rules are provided in Multimedia Appendix 1.

Representation Learning

Static Word Representations

Word embedding is a text vectorization technique that transforms words or subwords into vectors of real numbers. Pretrained word embeddings created using Word2Vec [72], Glove [73], and fastText [74] have been broadly used to initialize deep learning architectures for NLP tasks and have shown substantial improvement over random initialization. Recent research [75] showed that NER performance is significantly affected by the overlap between the pretrained

https://medinform.jmir.org/2020/7/e18417 JMIR Med Inform 2020 | vol. 8 | iss. 7 | e18417 | p. 4 (page number not for citation purposes)
word embedding vocabulary and the vocabulary of the target NER data set. Thus, we used Word2Vec with skip-gram to pretrain word embeddings over the Medical Information Mart for Intensive Care III (MIMIC-III) [76] with the default parameters provided in a study by Mikolov et al [72].

Contextualized Word Representations

A well-known limitation of word embedding methods is that they produce a single representation of all possible meanings of a word. To tackle these deficiencies, advanced approaches have attempted to model the word’s context into a vector representation. Embeddings from Language Models (ELMo) [77] is a prominent model that generates contextualized word representations by combining the internal states of different layers in a neural language model. Bidirectional Encoder Representations from Transformers (BERT) [78] furthered this idea by training bidirectional transformers [50] using subwords. Contextualized embeddings are particularly useful for clinical NER as entities (eg, cold as low temperature versus infection) have different meanings in different contexts. Recent research [79] showed that deep learning architectures with contextualized embeddings pretrained on a large clinical corpus achieve state-of-the-art performance on several clinical NER data sets. Inspired by these, we trained contextualized representations using ELMo on MIMIC-III. Detailed explanations of ELMo and training parameters are provided in Multimedia Appendix 2.

Knowledge Representations

To introduce medical knowledge, we built knowledge representations on the FAERS, a database for postmarketing drug safety monitoring. Specifically, we used 2 tables from Adverse Event Open Learning through Universal Standardization (AEOLUS) [14], a curated and standardized FAERS resource, to generate 2 separate graph embeddings. As shown in Figure 2, standard drug outcome count contains case frequencies for drug outcomes, including ADEs, and standard drug indication count contains case frequencies for drug indications (ie, reasons).

Let \( G=(D,O,E) \) be a weighted bipartite network, where \( D \) and \( O \) denote the set of drug concept id and outcome concept id in standard drug outcome count, and \( E \subseteq D \times O \) defines the iset edges. \( D_i \) and \( O_j \) denote the \( i^{th} \) and \( j^{th} \) vertex in \( D \) and \( O \) respectively, where \( i=\{1,2,\ldots,|D|\} \) and \( j=\{1,2,\ldots,|O|\} \). Each edge \( E_{ij} \in E \) carries a frequency \( f_{ij} \) provided by the drug outcome pair count field in standard drug outcome count, indicating the strength between the connected vertices \( D_i \) and \( O_j \); if \( D_i \) and \( O_j \) are not connected, \( f_{ij} \) is set to zero. To integrate this knowledge into our proposed architecture, we computed token-level embeddings by transforming \( G \) to \( G' \) as follows:

Given a drug concept id (RxNorm) or outcome concept id (Medical Dictionary for Regulatory Activities) from AEOLUS, we mapped it to its concept unique identifiers (CUIs) in UMLS [80] and obtained a set of tokens from all CUI variants. Let \( d=d_1, d_2, \ldots, d_L \) and \( o=o_1, o_2, \ldots, o_M \) represent all unique drug and outcome tokens obtained from mapping all \( D_i \in D \) and \( O_j \in O \). Let \( T_1:D \to d \) and \( T_2:O \to o \) represent 2 multivalued functions that associate each element in the set of drug concept id and outcome concept id to a set of tokens. Let \( G'=(d,o,e) \) be a weighted bipartite graph and each edge \( e_{lm} \in E \) of \( G' \) is associated with a nonnegative weight \( w_{lm} \) indicating the strength between the drug token \( d_l \) and the outcome token \( o_m \). We calculated \( w_{lm} \) as token-level co-occurrence between \( d_l \) and \( o_m \) normalized for the drug token \( d_l \):

\[
    w_{lm} = \frac{\sum_{(i,j) \in E} E_{ij} \cdot f_{ij}}{\sum_{(i,j) \in E} f_{ij}}
\]

In \( w_{lm} \), the numerator represents the sum of frequencies of all drug concept id and outcome concept id pairs that contain drug token \( d_l \) and outcome token \( o_m \), and the denominator represents the sum of frequencies of all pairs whose drug concept id contains the drug token \( d_l \).

From the generated bipartite weighted graph \( G'=(d,o,e) \), we used the LINE approach to generate drug-adverse knowledge embeddings. We used LINE because (1) relations between drugs and other concepts in the FAERS form a weighted bipartite graph with a long-tail distribution of vertex degrees and (2) it helps in embedding implicit connectivity relations between vertices of the same type. Similarly, we generated drug-reason knowledge embeddings from the standard drug indication count table. Detailed explanations of LINE and training parameters are provided in Multimedia Appendix 2.
**Figure 2.** Excerpts from the standard drug outcome count and standard drug indication count tables from adverse event open learning through universal standardization.

| standard_drug_outcome_count | standard_drug_indication_count |
|-----------------------------|--------------------------------|
| **drug_concept_id** | **Drug name** | **outcome_concept_id** | **outcome name** | **drug_outcome_pair_count** | **drug_concept_id** | **Drug name** | **indication_concept_id** | **indication name** | **drug_indication_pair_count** |
| 29046 | Lisinopril | 10011224 | Cough | 103 | 29046 | Lisinopril | 10020772 | Hypertension | 9003 |
| 6809 | Metformin | 10047700 | Vomiting | 399 |

**Architecture**

In the following sections, we present our system, illustrated in **Figure 3**, in an incremental fashion: *joint method*, contextual-joint, positional-joint, and knowledge-joint. A detailed explanation of the deep learning architecture, BiLSTM-CRF [81], and input embeddings used in this system is included in the Multimedia Appendix 3.

**Figure 3.** Canonical architecture of the proposed system. ADE: adverse drug event; BReason: beginning of reason annotation; CRF: conditional random field; ELMo: Embeddings from Language Models; KB: knowledge base; LSTM: long short-term memory; POS: part-of-speech.

**Joint Method**

We developed a *drug recognition model* followed by 2 joint drug-centric relation extraction models (*drug−{attributes} and *drug−{ADE, reason})*, as explained in the following sections.

**Drug Recognition Model**

We modeled drug recognition as a sequence-labeling task using BiLSTM-CRF and a beginning, inside, and outside of a drug mention (BIO) tagging scheme. The input layer of the
BiLSTM-CRF takes word, character, and part-of-speech embeddings. The word embeddings were obtained using Word2Vec representations generated using MIMIC-III. The character and part-of-speech embeddings were initialized randomly. We used CNNs [46] to encode a character-level representation for a word.

Drug-Centric Relation Extraction Models

To extract entities and relations jointly, we used the encoding scheme proposed in [38], which takes annotated sentences and produces drug-centric sequences for a specified target-drug. For sentences containing multiple identified drugs, 1 drug-centric sequence was generated for each target-drug. For example, for the sentence in Figure 4, the encoding scheme produced 2 labeled sequences: one with lisinopril as the target-drug and the other with mirtazapine. In each sequence, associated entities with the target-drug were labeled using a BIO scheme enhanced with their types. Hence, for the sequence generated with lisinopril as the target-drug, only 30 mg and the first QHS were labeled using B and I tags, and other entities (eg, 15 mg, PO, and the second QHS) were labeled as O.

We trained 2 separate models with the BiLSTM-CRF to jointly recognize (1) drug attributes and drug−{attributes} relations and (2) ADE, reason, and their corresponding relations (drug−{ADE, reason}). Similar to the drug recognition model, the input layer of these models takes word, character, and part-of-speech representations, with additional positional and semantic-tag embeddings. We used the positional embedding technique introduced in [82] to represent the positional distance from target-drug to each word in the input context. We used 3 different semantic tags, target-drug, duplicate-target-drug, and nontarget-drugs, to represent tokens of the current target-drug, other mentions of the same target-drug, and other drugs in the input context, respectively.

To handle intersentential relations, we provided adjacent sentences as an input context to the sentence containing the target-drug. We used training data to determine the optimal input context for the 2 models empirically. For the drug−{attributes} model, we determined the optimal context as the current sentence with the target-drug and the sentences preceding and following it. For the drug−{ADE, reason} model, the optimal context was the current sentence and the 4 sentences preceding and following it.

Figure 4. Label-encoding scheme used in drug-centric relation extraction models. B: beginning; I: inside; PO: orally; QHS: every night at bedtime.

Contextual-Joint Model

We obtained domain-specific contextualized representations for input contexts by pretraining ELMo on MIMIC-III. These contextualized representations were used to augment the representations used in the input layers of the models in the joint method. With the augmented input representations, we trained (1) a drug recognition model and (2) 2 drug-centric relation extraction models (drug−{attributes} and drug−{ADE, reason}).

Positional-Joint Model

As the task involves extraction of drug-centric entities and relations, we used the position-attention mechanism to extract entities and relations jointly with respect to an entity of interest (target-drug).

Let $H = \{ h_l \}_{l=1}^L$ represent the hidden representations of an input sequence obtained from the BiLSTM layer of the contextual-joint model. Positional representations $P = \{ p_l \}_{l=1}^L$ were generated as follows:

$$p_l = \sum_{i=1}^{N} a_i h_i$$
were
denote representations of the input sequence
tokens obtained from the embeddings created using the FAERS database.
As introduced earlier, background knowledge and hidden
knowledge embeddings, were computed by
feeding these concatenated
representations to a CRF layer. During the test phase, we used the
drug recognition model from the contextual-joint model for predicting drugs and the trained drug-centric
relation extraction models for predicting drug−{attributes} and
drug−{ADE, reason} relations.

Knowledge-Joint Model
As introduced earlier, background knowledge and hidden
relations beyond the contextual and positional information play
a crucial role in extracting drug−{ADE, reason} relations. To
address this, we propose the knowledge-joint model by
enhancing the positional-joint model with knowledge
embeddings created using the FAERS database.

Let \( \{r_j\}_{j=1}^{n}, \{a_j\}_{j=1}^{n} \) denote representations of the input sequence
tokens obtained from the drug-reason and drug-adverse
knowledge embeddings, respectively. Let \( l \) and \( m \) be
the beginning and end indices of target-drug in the input
sequence. The target-drug \( D_r \) and \( D_a \), corresponding to
drug-reason and drug-adverse knowledge embeddings, were computed by
averaging the representations of target-drug tokens:
\[
D_r = \frac{\sum_{j=l}^{m} r_j}{m - l + 1}
\]
\[
D_a = \frac{\sum_{j=l}^{m} a_j}{m - l + 1}
\]
The target-drug–centric representations \( \{r'_j\}_{j=1}^{n} \) and \( \{a'_j\}_{j=1}^{n} \) were obtained by computing similarities between input sequence tokens and the target-drug:
\[
r'_j = w_r(D_r \cdot r_j)
\]
\[
a'_j = w_a(D_a \cdot a_j)
\]
where \( w_r \) and \( w_a \) represent the scalar weights corresponding to
drug-reason, and drug-adverse knowledge embeddings learned
during training. Finally, for a token at position \( t \), we
concatenated its target-drug–centric similarities \( \{r'_j; a'_j\} \)
with positional and hidden representations \( u_t \) to produce \( k_t \):
\[
k_t = [u_t; r'_t; a'_t]
\]

Evaluation Metrics and Significance Tests
We evaluated the proposed system using the evaluation script
released by the organizers of the n2c2 challenge to measure the lenient precision, recall, and \( F_1 \) scores, explained as follows. For NER, a predicted entity is considered as a true-positive if
its span overlaps with a gold annotation and is the correct entity type. For relation extraction, a predicted relation is considered as
a true-positive if both entities in the relation are true-positives
and the relation type matches the gold annotation. We also report
statistical significance on these results with 50,000 shuffles and
a significance level set to .05 by using a test script released by
the n2c2 organizers based on the approximate randomization test [83].

In the following sections, we present the results of our system.
The experimental settings used to achieve these results are
provided in Multimedia Appendix 4.

Results

Named Entity Recognition
Table 3 presents the results for each proposed incremental
approach for NER. Compared with the joint method,
incorporating contextualized embeddings (contextual-joint
model) improved the overall microaveraged \( F_1 \) score by 0.3
percentage points. The improvement was mainly observed in
recognizing drugs (0.6 points), with some improvements in
recognizing strength and reason. Compared with the
contextual-joint model, the positional-joint model improved the
overall micro-F1 score by 0.2 points, with significant
improvements observed in identifying reason (2.1 points) and
ADE (6.8 points). Compared with the positional-joint model,
the knowledge-joint model further improved the overall micro-F1
score by 0.1 points, with significant improvements observed in
accurately determining reason (1.9 points) and ADE (1.7 points).
Note that the overall improvement between the positional-joint
and knowledge-joint models is relatively small due to the biased
distribution of annotations, as ADE and reason together
constitute less than 10% of the entities.

Note that the overall improvement between the positional-joint
and knowledge-joint models is relatively small due to the biased
distribution of annotations, as ADE and reason together
constitute less than 10% of the entities.
Significance tests showed that the improvements in micro-$F_1$ score observed with each incremental approach are statistically significant with $P$ values of .001, <.001, and <.001 for the contextual-joint, positional-joint, and knowledge-joint models, respectively. As the contextual-joint and positional-joint models share the same drug recognition model, we ignored drug predictions when performing significance tests. Similarly, the positional-joint and knowledge-joint models share the same drug recognition model and drug−{attributes} model; therefore, we considered only ADE and reason predictions when performing significance tests.

Table 3. Lenient precision, recall, and $F_1$ score of the proposed approaches for named entity recognition.

| Entity type | Joint | Contextual-joint | Positional-joint | Knowledge-joint |
|-------------|-------|------------------|------------------|------------------|
|             | Precision | Recall | $F_1$ score | Precision | Recall | $F_1$ score | Precision | Recall | $F_1$ score | Precision | Recall | $F_1$ score |
| Drug        | 0.956 | 0.954 | 0.956 | 0.964 | 0.960 | 0.964 | 0.960 | 0.964 | 0.960 | 0.964 | 0.960 |
| Strength    | 0.980 | 0.974 | 0.982 | 0.971 | 0.976 | 0.985 | 0.976 | 0.980 | 0.985 | 0.976 | 0.980 |
| Form        | 0.974 | 0.958 | 0.975 | 0.939 | 0.957 | 0.972 | 0.943 | 0.958 | 0.972 | 0.943 | 0.958 |
| Frequency   | 0.981 | 0.958 | 0.970 | 0.981 | 0.958 | 0.969 | 0.979 | 0.964 | 0.971 | 0.964 | 0.971 |
| Route       | 0.964 | 0.942 | 0.953 | 0.962 | 0.943 | 0.952 | 0.950 | 0.949 | 0.949 | 0.950 | 0.949 |
| Dosage      | 0.943 | 0.938 | 0.941 | 0.941 | 0.937 | 0.939 | 0.936 | 0.957 | 0.946 | 0.936 | 0.957 | 0.946 |
| Duration    | 0.887 | 0.788 | 0.835 | 0.914 | 0.791 | 0.848 | 0.880 | 0.815 | 0.846 | 0.880 | 0.815 | 0.846 |
| ADE$^a$     | 0.649 | 0.358 | 0.462 | 0.643 | 0.346 | 0.450 | 0.660 | 0.426 | 0.518 | 0.589 | 0.490 | 0.535 |
| Reason      | 0.757 | 0.611 | 0.676 | 0.747 | 0.636 | 0.687 | 0.747 | 0.672 | 0.708 | 0.753 | 0.702 | 0.727 |
| Overall (micro) | 0.948 | 0.912 | 0.929 | 0.947 | 0.917 | 0.932 | 0.943 | 0.926 | 0.934 | 0.941 | 0.930 | 0.935 |

$^a$ADE: adverse drug event.

Relation Extraction

Table 4 presents the results for each proposed incremental approach for relation extraction. Compared with the joint method, the contextual-joint model improved the overall micro-$F_1$ score by 0.5 percentage points, with the majority of improvements observed in accurately recognizing drug−strength, drug−frequency, drug−reason, and drug−dosage relations. Compared with the contextual-joint model, the positional-joint model improved the $F_1$ score by 0.4 points with significant improvements observed in determining drug−ADE (5.6 points) and drug−reason (2.9 points) relations. The knowledge-joint model further improved the overall $F_1$ score by 0.1 points, with specific improvements in drug−ADE by 3.0 points and drug−reason by 1.7 points when compared with the positional-joint model. Similar to the NER significance results, significance testing for relation extraction showed that the improvements observed with each incremental approach are statistically significant with $P$ values of <.001, <.001, and <.001 for the contextual-joint, positional-joint, and knowledge-joint models, respectively.

Table 4. Lenient precision, recall, and $F_1$ score of the proposed approaches for relation extraction.

| Relation type | Joint | Contextual-joint | Positional-joint | Knowledge-joint |
|---------------|-------|------------------|------------------|------------------|
|               | Precision | Recall | $F_1$ score | Precision | Recall | $F_1$ score | Precision | Recall | $F_1$ score | Precision | Recall | $F_1$ score |
| Drug−strength | 0.966 | 0.962 | 0.964 | 0.977 | 0.964 | 0.971 | 0.978 | 0.971 | 0.975 | 0.978 | 0.971 | 0.975 |
| Drug−form     | 0.963 | 0.936 | 0.949 | 0.972 | 0.936 | 0.953 | 0.969 | 0.939 | 0.954 | 0.969 | 0.939 | 0.954 |
| Drug−frequency| 0.961 | 0.949 | 0.955 | 0.972 | 0.950 | 0.961 | 0.969 | 0.955 | 0.962 | 0.969 | 0.955 | 0.962 |
| Drug−route    | 0.943 | 0.931 | 0.937 | 0.954 | 0.933 | 0.943 | 0.936 | 0.939 | 0.937 | 0.936 | 0.939 | 0.937 |
| Drug−dosage   | 0.921 | 0.928 | 0.924 | 0.933 | 0.931 | 0.932 | 0.925 | 0.950 | 0.937 | 0.925 | 0.950 | 0.937 |
| Drug−duration | 0.814 | 0.718 | 0.763 | 0.880 | 0.723 | 0.794 | 0.823 | 0.739 | 0.779 | 0.823 | 0.739 | 0.779 |
| Drug−ADE$^a$  | 0.590 | 0.322 | 0.417 | 0.592 | 0.307 | 0.404 | 0.590 | 0.377 | 0.460 | 0.544 | 0.446 | 0.490 |
| Drug−reason   | 0.682 | 0.526 | 0.594 | 0.676 | 0.546 | 0.604 | 0.680 | 0.593 | 0.633 | 0.673 | 0.628 | 0.650 |
| Overall (micro) | 0.912 | 0.859 | 0.885 | 0.920 | 0.862 | 0.890 | 0.912 | 0.877 | 0.894 | 0.906 | 0.884 | 0.895 |

$^a$ADE: adverse drug event.
Discussion

Principal Findings

Contextualized representations (contextual-joint) are effective in differentiating between words and abbreviations that could have multiple meanings. For example, ensure and contrast can be understood as either a drug (“Ensure: 1 can PO three times daily” and “contrast-induced nephropathy”) or a verb, and terms such as blood could either refer to a drug (“transfused 1 unit of blood”), that is, substance given to a patient, a test for the drug (“blood alcohol concentration”), or a natural occurring substance in the body (“blood pressure”). Additionally, abbreviations such as PE (physical examination versus pulmonary embolism) and pcp (primary care physician versus pneumocystis pneumonia) can have multiple expansions. In all the examples above, the contextual-joint correctly identifies these entities.

One prevailing challenge in ADE extraction is the presence of long-distance or intersentential relations. As shown in Table 2, a significant portion of drug−{ADE, reason} in the data set is intersentential (23% of drug−ADE and 31.7% of drug−reason). These relations typically span long distances, making them more difficult to capture. To study the effectiveness of the proposed approaches over long-distance relations, we calculated the F1 scores on drug−{ADE, reason} with an increasing number of tokens between entities. As shown in Figure 5, we find that the positional-joint model performs significantly better than the contextual-joint model with increasing distance between entities, suggesting that the positional-joint can effectively model long-distance relations.

Incorporating knowledge embeddings learned on the FAERS improved drug−{ADE, reason} relation extraction, especially in the case of long-distance relations or when contextual clues are insufficient. As shown in Figure 5, the knowledge-joint model further improved on the positional-joint model at all distances. The knowledge-joint model was also useful in cases of insufficient or ambiguous context in extracting the correct relation. For example, in the phrase “Wellbutrin - nausea and vomiting,” the relation is indicated only by an uninformative hyphen, with no contextual clues to indicate the type of relation. Similarly, in “Patient had history of depression and was on elavil previously;” it is unclear whether the history of depression was previously treated by drug−reason or caused by drug−ADE of the drug elavil. Furthermore, the knowledge-joint also helped to extract correct relations when multiple drugs and candidate ADEs and reasons are discussed in a given context. For example, in “Upon arrival, she was hypertensive and had a fever. She was given Tylenol,” based on sentence construction, 2 candidate reasons (hypertensive and fever) may be associated with the drug Tylenol. Knowledge is required to infer that of the two, only fever is related to Tylenol.

Figure 5. F1 scores of approaches with increasing distance between entities for relation extraction. ADE: adverse drug event.

Error Analysis

We investigated the most common error categories by entity and relation type and present these in Table 5. Most of the errors in recognizing drugs were due to abbreviations, misspellings, generic terms, or linguistic shorthand. For strength and dosage, these entities were often mislabeled as each other—both are often numeric quantities and used in similar contexts. For duration and frequency, most of the errors resulted from these entities being expressed in colloquial language.

Intersentential relations remain a major category of false-negative errors for all relations despite improvements from the position-attention mechanism. For drug−{attributes}, these errors were likely due to an insufficient number of such examples in the training data (approximately 4%). In addition...
to errors from intersentential relations, other important categories for false-negative drug–{ADE, reason} include (1) ADE or reasons expressed in generic terms, (2) reasons such as procedures and activities (eg, angioplasty/stenting) that occur infrequently in the training set, and (3) ADE or reasons expressed as abbreviations that are nonstandard or ambiguous.

False-positive errors in drug–{ADE, reason} mainly fall into 2 categories. In the first, one of the entities participating in the relation is negated, hypothetical, or conditional, such as when a drug is withheld to avoid an anticipated ADE (eg, contraindications). In the second, the same concept (drug, ADE, or reason) is mentioned multiple times in the same context, and the system associated the relation to one mention whereas the ground truth to the other. To add further complexity, these mentions may be synonyms, for example, “the pain medications (morphine, vicodin, codeine) worsened your mental status and made you delirious.” With multiple possible drug–ADE relations, some combinations were not captured in the ground truth, resulting in false-positives that may not be true errors.

Table 5. Error analysis on our best-performing model (knowledge-joint).

| Entity/relation, Error category | Texta | Explanation |
|--------------------------------|-------|-------------|
| Drug                           |       |             |
| Abbreviation                   | Hyponatremia due to HCTZb | HCTZ—abbreviated drug |
| Misspelled words               | 30 units of Lantus in addition to humalong | Humalog is incorrectly written as humalong |
| Short forms                    | She was given vanco | Vancomycin is expressed in shorthand |
| Generic phrase                 | He was advised to not take any of his blood pressure medications | Antihypertensives are expressed through generic terms |
| **Strength**                   |       |             |
| Contextual ambiguity           | Patient received 1 unit of blood | Strength (1 unit) wrongly predicted as dosage; usually, the unit token is associated with dosage |
| **Duration**                   |       |             |
| Colloquial language            | Only take Hydroxyzine as long as your rash is itching | Duration is expressed colloquially |
| **Drug – strength**            |       |             |
| Intersentential                | Continued Carvedilol. INRc initially slightly supratherapeutic, but then his home regimen of 4 mg alternating with 2 mg daily was started | Intersentential relation between carvedilol and 4 mg |
| **Drug – ADEd; Drug – reason** |       |             |
| Intersentential                | He underwent coronary artery bypass x5, please see operative report for further details. He was transferred to the CSRUe on Neo with IABPf | Intersentential relation between neo and coronary artery bypass graft |
| Generic terms                  | Start a baby aspirin every day to protect the heart | Reason is expressed in generic terms |
| Abbreviation                   | Detro was discontinued on suspicion that it might contribute to AMS | AMS has multiple possible expansions |
| Procedure                      | Angioplasty of the left tibial artery; had been on Plavix prior to NSTEMIf | Procedure angioplasty is annotated as reason |
| Contraindication               | Avoiding NSAIDsb to prevent gastrointestinal bleed | Drug was not given to this patient |
| Negated                        | Heparin-induced thrombocytopenia negative | ADE thrombocytopenia is negated |

aItalics indicate text that contributes to the specified error category.
bHCTZ: hydrochlorothiazide.
cINR: international normalized ratio.
dADE: adverse drug event.
eCSRU: cardiac surgery recovery unit.
fIABP: intra-aortic balloon pump.
fNSTEMI: non–ST-elevation myocardial infarction.
bNSAIDs: nonsteroidal anti-inflammatory drugs.
Document-Level Analysis

From an end user perspective, the core information needed for patient care purposes is a patient-level summary of these relations, which is a unique set of extracted relations after normalization. To evaluate our system for this purpose, we measured drug–ADE and drug–reason F1 scores by considering unique pairs of relation mentions at the document level, presented in Table 6. We observed scores at the document level to be 1 to 2 percentage points higher than the instance level.

Table 6. Document-level analysis for drug–reason and drug–adverse drug event relations.

| Model          | Drug–reason Instance level |            | Document level | Drug–ADEa Instance level |            | Document level |
|----------------|----------------------------|------------|----------------|-------------------------|------------|----------------|
|                | Precision | Recall | F1 score | Precision | Recall | F1 score | Precision | Recall | F1 score | Precision | Recall | F1 score |
| Joint          | 0.682     | 0.526  | 0.594   | 0.691     | 0.542  | 0.607   | 0.590     | 0.322  | 0.417   | 0.631     | 0.322  | 0.426   |
| Contextual-joint| 0.675     | 0.546  | 0.604   | 0.685     | 0.560  | 0.616   | 0.592     | 0.307  | 0.404   | 0.630     | 0.308  | 0.414   |
| Position-joint | 0.680     | 0.593  | 0.633   | 0.692     | 0.611  | 0.649   | 0.590     | 0.376  | 0.460   | 0.647     | 0.384  | 0.482   |
| Knowledge-joint| 0.673     | 0.628  | 0.650   | 0.687     | 0.647  | 0.666   | 0.544     | 0.446  | 0.490   | 0.579     | 0.444  | 0.503   |

aADE: adverse drug event.

Comparison With Previous Work

For NER, the state-of-the-art system [38] used an ensemble (committee) of 3 different methods: CRF, BiLSTM-CRF, and joint approach. They showed that the BiLSTM-CRF is the best among the single models. Thus, we compare our best model (knowledge-joint) with their best-performing single model and committee approach, as shown in Table 7. Overall, the knowledge-joint model outperformed the single model by 0.2 percentage points and achieved similar micro-F1 to the committee approach. Notably, the knowledge-joint model significantly outperformed the committee approach in recognizing the crucial ADE (0.5 points) and reason (5.2 points) entities.

For relation extraction, the state-of-the-art system used the committee approach for NER, convolutional neural network – recurrent neural network (CNN-RNN) for relation classification, and postprocessing rules. Although postprocessing rules are commonly used in competitions, they often do not generalize across data sets and therefore are of limited interest in this research. As shown in Table 7, the knowledge-joint model outperformed the state-of-the-art approach, both with (0.4 points) and without rules (1.6 points). Notably, the knowledge-joint model achieved the best results and outperformed the state-of-the-art in recognizing the most crucial and difficult to extract relations: drug–reason (7.1 points) and drug–ADE (1.4 points).
Table 7. The lenient F1 scores for named entity recognition of single and state-of-the-art ensemble models compared with our best model. The lenient F1 scores for relation extraction of state-of-the-art ensemble models with and without rules, compared with our best model.

| NER^a | Entity type | BiLSTM-CRFb | Committee [38] | Knowledge-joint | Relation extraction | Committee + CNN-RNNc | Committee + CNN-RNN + Rules [38] | Knowledge-joint |
|-------|-------------|-------------|----------------|-----------------|--------------------|-----------------------|-----------------------------------|----------------|
|       | Drug        | 0.955       | 0.956          | 0.960           | N/A^d              | N/A                   | N/A                               | N/A            |
|       | Strength    | 0.982       | 0.983          | 0.980           | Drug–strength      | 0.964                 | 0.972                             | 0.975           |
|       | Form        | 0.958       | 0.958          | 0.958           | Drug–form          | 0.940                 | 0.952                             | 0.954           |
|       | Frequency   | 0.974       | 0.975          | 0.971           | Drug–frequency     | 0.941                 | 0.958                             | 0.962           |
|       | Route       | 0.956       | 0.956          | 0.949           | Drug–route         | 0.930                 | 0.942                             | 0.937           |
|       | Dosage      | 0.943       | 0.948          | 0.946           | Drug–dosage        | 0.923                 | 0.935                             | 0.937           |
|       | Duration    | 0.856       | 0.862          | 0.846           | Drug–duration      | 0.740                 | 0.786                             | 0.779           |
|       | ADE^e       | 0.422       | 0.530          | 0.535           | Drug–ADE           | 0.475                 | 0.476                             | 0.490           |
|       | Reason      | 0.680       | 0.675          | 0.727           | Drug–reason        | 0.572                 | 0.579                             | 0.650           |
|       | Overall (mi- | 0.933       | 0.935          | 0.935           | Overall (micro)    | 0.879                 | 0.891                             | 0.895           |

aNER: named entity recognition.
bBiLSTM-CRF: bidirectional long short-term memory–conditional random field.
cCNN-RNN: convolutional neural network–recurrent neural network.
dNot applicable.
eADE: adverse drug event.

Limitations and Future Work

We acknowledge several limitations of this study. First, these results are specific to the n2c2 data set, which contains only intensive care unit (ICU) discharge summaries from a single health care organization. Ground truth generation and evaluation on a more diverse data set is needed to better understand the effectiveness of these proposed approaches. Second, we observed some annotation errors in the ground truth, likely due to the complex nature of the task. Further investigation is needed to quantify the prevalence of such errors and their impact on the results.

Despite achieving state-of-the-art results, the proposed system still has room for improvement, specifically in recognizing intersentential drug–{ADE, reason} relations. To further improve ADE extraction, we plan to explore the following research areas:

1. Although we incorporated knowledge graph embeddings, other advanced methods that use higher-order proximity and role-preserving network embedding techniques have shown promising results in the general domain. We plan to explore methods such as Edge Label Aware Network Embedding [84] rather than training separate graph embeddings for drug–{ADE, reason} relations.

2. The field of contextual embeddings has evolved quickly along with the release of newer language representation models trained on clinical text. We plan to explore BERT [78,85], which utilizes a transformer network to pretrain a language model for extracting better contextual word embeddings.

3. To address some of the findings from the error analysis, we plan to leverage our clinical abbreviation expansion components [86] to help resolve ambiguous mentions and also incorporate assertion recognition [26] to capture the belief state of the physician on a concept (negated, hypothetical, conditional).

4. As mentioned earlier, the proposed models performed poorly on intersentential relation extraction. To address this, we plan to explore N-ary relation extraction for cross-sentence relation extraction using graph long short-term memory networks [87].

Conclusions

We presented a system for extracting drug-centric concepts and relations that outperformed current state-of-the-art results. Experimental results showed that contextualized embeddings, position-attention mechanisms, and knowledge embeddings effectively improve deep learning-based concepts and relation extraction. Specifically, we showed the effectiveness of a position-attention mechanism in extracting long-distance relations and knowledge embeddings from the FAERS in recognizing relations where contextual clues are insufficient.

Acknowledgments

The authors wish to thank Dr Kenneth J Barker for his assistance in providing valuable feedback on the manuscript.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Sentence segmentation and tokenization.
[PDF File (Adobe PDF File), 392 KB-Multimedia Appendix 1]

Multimedia Appendix 2
Embeddings from Language Models contextualized embeddings and large-scale information network embedding graph embeddings.
[PDF File (Adobe PDF File), 142 KB-Multimedia Appendix 2]

Multimedia Appendix 3
Detailed explanation of bidirectional long short-term memory–conditional random fields and input embeddings.
[PDF File (Adobe PDF File), 369 KB-Multimedia Appendix 3]

Multimedia Appendix 4
Experimental settings used in the proposed system.
[PDF File (Adobe PDF File), 106 KB-Multimedia Appendix 4]

References
1. Gunter TD, Terry NP. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. J Med Internet Res 2005 Mar 14;7(1):e3 [FREE Full text] [doi: 10.2196/jmir.7.1.e3] [Medline: 15829475]
2. Rosenbloom S, Stead W, Denny J, Giuse D, Lorenzi N, Brown S, et al. Generating clinical notes for electronic health record systems. Appl Clin Inform 2010 Jan 1;1(3):232-243 [FREE Full text] [doi: 10.4338/ACI-2010-03-RA-0019] [Medline: 2103148]
3. 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. J Am Med Assoc 1995 Jul 5;274(1):29-34. [Medline: 7791255]
4. Johnson JA, Bootman JL. Drug-related morbidity and mortality. A cost-of-illness model. Arch Intern Med 1995 Oct 9;155(18):1949-1956. [Medline: 7575048]
5. 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. J Am Med Assoc 1997;277(4):301-306. [Medline: 9002492]
6. Chiatti C, Bustacchini S, Furneri G, Mantovani L, Cristiani M, Misuraca C, et al. The economic burden of inappropriate drug prescribing, lack of adherence and compliance, adverse drug events in older people: a systematic review. Drug Saf 2012 Jan;35(Suppl 1):73-87. [doi: 10.1007/BF03319105] [Medline: 23446788]
7. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. J Gen Intern Med 2003 Jan;18(1):57-60 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.20130.x] [Medline: 12534765]
8. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. Drug Saf 2006;29(5):385-396. [doi: 10.2165/00002018-200629050-00003] [Medline: 16689555]
9. Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. J Am Med Inform Assoc 2012;19(1):79-85 [FREE Full text] [doi: 10.1136/amiajnl-2011-000214] [Medline: 21676938]
10. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. Br Med J 2004 Jul 3;329(7456):15-19 [FREE Full text] [doi: 10.1136/bmj.329.7456.15] [Medline: 15231615]
11. Haerian K, Varn D, Vaidya S, Ena L, Chase HS, Friedman C. Detection of pharmaco vigilance-related adverse events using electronic health records and automated methods. Clin Pharmacol Ther 2012 Aug;92(2):228-234 [FREE Full text] [doi: 10.1038/clpt.2012.54] [Medline: 22713699]
12. Lependu P, Iyer SV, Fainon C, Shah NH. Annotation analysis for testing drug safety signals using unstructured clinical notes. J Biomed Semantics 2012 Apr 24;3(Suppl 1):S5 [FREE Full text] [doi: 10.1186/2041-1480-3-S1-S5] [Medline: 22541596]
13. LePendu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchiyska T, et al. Pharmacovigilance using clinical notes. Clin Pharmacol Ther 2013 Jun;93(6):547-555 [FREE Full text] [doi: 10.1038/clpt.2013.47] [Medline: 23571773]
14. Banda JM, Evans L, VanGuri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. Sci Data 2016 May 10;3:160026 [FREE Full text] [doi: 10.1038/sdata.2016.26] [Medline: 27193236]

15. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther 2012 Jun;91(6):1010-1021 [FREE Full text] [doi: 10.1038/clpt.2012.50] [Medline: 22549283]

16. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the observational medical outcomes partnership. Ann Intern Med 2010 Nov 2;153(9):600-606. [doi: 10.7326/0003-4819-153-9.201011010-00010] [Medline: 21041580]

17. Sarkar A, Ginn R, Nikfarjam A, O’Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]

18. Nikfarjam A, Sarkar A, O’Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]

19. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. Drug Saf 2014 May;37(5):343-350 [FREE Full text] [doi: 10.1007/s40264-014-0155-x] [Medline: 24777653]

20. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. Drug Saf 2014 Oct;37(10):777-790 [FREE Full text] [doi: 10.1007/s40264-014-0218-z] [Medline: 25151493]

21. Odgers D, Harpaz R, Callahan A, Stiglic G, Shah N. Analyzing search behavior of healthcare professionals for drug safety surveillance. Biocomputing 2014(2014):306-317 [FREE Full text] [doi: 10.1142/9789814644730_0030]

22. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. Clin Pharmacol Ther 2014 Aug;96(2):239-246 [FREE Full text] [doi: 10.1038/clpt.2014.77] [Medline: 24713590]

23. Abernethy DR, Woodcock J, Lesko LJ. Pharmacological mechanism-based drug safety assessment and prediction. Clin Pharmacol Ther 2011 Jun;89(6):793-797. [doi: 10.1038/clpt.2011.55] [Medline: 21490594]

24. Chiang A, Butte A. Data-driven methods to discover molecular determinants of serious adverse drug events. Clin Pharmacol Ther 2009 Mar;85(3):259-268 [FREE Full text] [doi: 10.1038/clpt.2008.274] [Medline: 19177064]

25. Vilar S, Harpaz R, Chase H, Costanzi S, Rabadan R, Friedman C. Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i73-i80 [FREE Full text] [doi: 10.1136/amiajnl-2011-000417] [Medline: 21946238]

26. Uzuner O, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i73-i80 [FREE Full text] [doi: 10.1136/amiajnl-2011-000417] [Medline: 21946238]

27. Roberts K, Denmer-Fushman D, Tonning JM. Overview of the TAC 2017 Adverse Reaction Extraction from Drug Labels Track. Semantic Scholar. 2017. URL: https://pdfs.semanticscholar.org/5b8a/7b11b987ddeb865d5bf3aaa7b745a86ea5bf0.pdf [accessed 2020-06-22]

28. Li J, Sun Y, Johnson R, Sciaky D, Wei C, Leaman R, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction. Database (Oxford) 2016;2016 [FREE Full text] [doi: 10.1093/database/baw068] [Medline: 27161011]

29. 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 (MADE 1.0). Drug Saf 2019 Jan;42(1):99-111 [FREE Full text] [doi: 10.1007/s40264-018-0762-z] [Medline: 30649735]

30. Henry S, Buchan K, Filannino M, Stubbs A, Uzuner O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. J Am Med Inform Assoc 2020 Jan;27(1):3-12. [doi: 10.1093/jamia/ocz166]

31. Xu J, Wu Y, Zhang Y, Wang J, Lee HJ, Xu J. CD-REST: a system for extracting chemical-induced disease relation in literature. Database (Oxford) 2016;2016 [FREE Full text] [doi: 10.1093/database/baw036] [Medline: 27016700]

32. 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]

33. 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;2016 [FREE Full text] [doi: 10.1093/database/baw032] [Medline: 26994911]

34. 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]

35. Sutton C. An introduction to conditional random fields. FNT in Mach Learn 2012;4(4):267-373 [FREE Full text] [doi: 10.1561/2200000013]

36. Andrew AM. An introduction to support vector machines and other kernel - based learning methods. Kybernetes 2001 Feb;30(1):103-115. [doi: 10.1108/k.2001.30.1.103.6]
37. Dandala B, Joopudi V, Devarakonda M. Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks. Drug Saf 2019 Jan;42(1):135-146. [doi: 10.1007/s40264-018-0764-x] [Medline: 30649738]
38. Wei Q, Ji Z, Li Z, Du J, Wang J, Xu J, et al. A study of deep learning approaches for medication and adverse drug event extraction from clinical text. J Am Med Inform Assoc 2020 Jan 1;27(1):13-21 [FREE Full text] [doi: 10.1093/jamia/ocz063] [Medline: 31135882]
39. Li F, Liu W, Yu H. Extraction of information related to adverse drug events from electronic health record notes: design of an end-to-end model based on deep learning. JIMIR Med Inform 2018 Nov 26;6(4):e12159 [FREE Full text] [doi: 10.2196/12159] [Medline: 30478023]
40. Ju M, Nguyen N, Miwa M, Ananiadou S. An ensemble of neural models for nested adverse drug events and medication extraction with subwords. J Am Med Assoc 2020 Jan 1;27(1):22-30 [FREE Full text] [doi: 10.1093/jamia/ocz075] [Medline: 31197355]
41. Dai H, Su C, Wu C. Adverse drug event and medication extraction in electronic health records via a cascading architecture with different sequence labeling models and word embeddings. J Am Med Inform Assoc 2020 Jan 1;27(1):47-55. [doi: 10.1093/jamia/ocz120] [Medline: 31334805]
42. Wunnava S, Qin X, Kakar T, Sen C, Rundensteiner EA, Kong X. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. Drug Saf 2019 Jan;42(1):113-122. [doi: 10.1007/s40264-018-0765-9] [Medline: 30649736]
43. Chapman AB, Peterson KS, Alba PR, DuVall SL, Patterson OV. Detecting adverse drug events with rapidly trained classification models. Drug Saf 2019 Jan;42(1):147-156 [FREE Full text] [doi: 10.1007/s40264-018-0763-x] [Medline: 30649737]
44. Yang X, Bian J, Fang R, Bjarnadottir R, Hogan W, Wu Y. Identifying relations of medications with adverse drug events using recurrent convolutional neural networks and gradient boosting. J Am Med Inform Assoc 2020 Jan 1;27(1):65-72. [doi: 10.1093/jamia/ocz144] [Medline: 31504605]
45. Chalapathy R, Borzeshi E, Piccardi M. Bidirectional LSTM-CRF for Clinical Concept Extraction. arXiv 2016 epub ahead of print(1611.08373) [FREE Full text]
46. Kim Y, Jernite Y, Sontag D, Rush A. Character-Aware Neural Language Models. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016 Presented at: AAAI’16; February 12-17, 2016; Phoenix, Arizona, USA URL: https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/viewFile/12489/12017
47. Li F, Zhang M, Fu G, Ji D. A neural joint model for entity and relation extraction from biomedical text. BMC Bioinformatics 2017 Mar 31;18(1):198 [FREE Full text] [doi: 10.1186/s12859-017-1609-9] [Medline: 28559255]
48. Caruana R. Multitask learning. Mach Learn 1997;28(1):41-75. [doi: 10.1007/978-1-4615-5529-2_5]
49. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. In: Proceedings of the 3rd International Conference for Learning Representations. 2015 Presented at: ICLR’15; May 7-9, 2015; San Diego, CA, USA.
50. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;5998-6008 [FREE Full text]
51. Wang W, Yang N, Wei F, Chang B, Zhou M. Gated Self-Matching Networks for Reading Comprehension and Question Answering. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017 Presented at: ACL’17; July 30-August 4, 2017; Vancouver, Canada. [doi: 10.1093/jamia/ocz063]
52. Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, et al. Attention-based bidirectional long short-term memory networks for relation classification. 54th Annu Meet Assoc Comput Linguist ACL 2016:207-212 [FREE Full text] [doi: 10.1168/s12859-017-1606-8] [Medline: 28559234]
53. Zhang Y, Zhong V, Chen D, Angeli G, Manning C. Position-Aware Attention and Supervised Data Improve Slot Filling. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: EMNLP’17; September 7-11, 2017; Copenhagen, Denmark. [doi: 10.1093/jamia/ocz059]
54. Zhao Q, Reilly B, Zaidi R, Zhang J, Ramshaw L. An end-to-end model based on deep learning for Chinese relation classification. 54th Annu Meet Assoc Comput Linguist ACL 2016:207-212 [FREE Full text] [doi: 10.1168/s12859-017-1606-8] [Medline: 28559234]
58. Shen Y, Deng Y, Yang M, Li Y, Du N, Fan W, et al. Knowledge-Aware Attentive Neural Network for Ranking Question Answer Pairs. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018 Presented at: SIGIR'18; July 8-12, 2018; Ann Arbor, Michigan. [doi: 10.1145/3209978.3210081]

59. Li P, Mao K, Yang X, Li Q. Improving Relation Extraction with Knowledge-Attention. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019 Presented at: EMNLP-IJCNLP'19; November 3-7, 2019; Hong Kong, China. [doi: 10.18653/v1/D19-1022]

60. Bordes A, Usunier N, Garcia-Durán A, Weston J, YahSuggestions:...
78. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL’19; June 2-7, 2019; Minneapolis, Minnesota. [doi: 10.18653/v1/N19-1423]

79. Si Y, Wang J, Xu H, Roberts K. Enhancing clinical concept extraction with contextual embeddings. J Am Med Inform Assoc 2019 Nov 1;26(11):1297-1304. [doi: 10.1093/jamia/ocz096] [Medline: 31265066]

80. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [Medline: 14681409]

81. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv 2015 epub ahead of print(1508.01991) [FREE Full text]

82. Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation Classification via Convolutional Deep Neural Network. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014 Presented at: COLING’14; August 23-29, 2014; Dublin, Ireland URL: https://www.aclweb.org/anthology/C14-1220/

83. Edgington ES. Approximate randomization tests. J Psychol Interdiscip Appl 1969 Jul;72(2):143-149. [doi: 10.1080/00223980.1969.10543491]

84. Goyal P, Hosseinmardi H, Ferrara E, Galstyan A. Capturing edge attributes via network embedding. IEEE Trans Comput Soc Syst 2018 Dec;5(4):907-917. [doi: 10.1109/tcss.2018.2877083]

85. Lee J, Yoon W, Kim S, Kim D, Kim S, So C, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240. [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]

86. Joopudi V, Dandala B, Devarakonda M. A convolutional route to abbreviation disambiguation in clinical text. J Biomed Inform 2018 Oct;86:71-78 [FREE Full text] [doi: 10.1016/j.jbi.2018.07.025] [Medline: 30118854]

87. Peng N, Poon H, Quirk C, Toutanova K, Yih W. Cross-sentence n-ary relation extraction with graph LSTMs. Transact Assoc Comput Ling 2017:-101-115. [doi: 10.1162/tacl_a_00049]

Abbreviations

ADE: adverse drug event
AEOLUS: adverse event open learning through universal standardization
BERT: Bidirectional Encoder Representations from Transformers
BiLSTM-CRF: bidirectional, long short-term memory–conditional random fields
BIO: beginning, inside, and outside
CNN: convolutional neural network
CRF: conditional random field
CUI: concept unique identifier
EHR: electronic health record
ELMo: Embeddings from Language Models
FAERS: Food and Drug Administration Adverse Event Reporting System
LINE: large-scale information network embedding
MADE: Medication and Adverse Drug Events from Electronic Health Records
MIMIC: Medical Information Mart for Intensive Care III
n2c2: 2018 National NLP Clinical Challenges
NER: named entity recognition
NLP: natural language processing
SBT: sentence boundary detection
UMLS: Unified Medical Language System

Edited by C Lovis; submitted 25.02.20; peer-reviewed by T Mato, S Doan; comments to author 28.04.20; revised version received 12.05.20; accepted 13.05.20; published 10.07.20

Please cite as:
Dandala B, Joopudi V, Tsou CH, Liang JJ, Suryanarayanan P
Extraction of Information Related to Drug Safety Surveillance From Electronic Health Record Notes: Joint Modeling of Entities and Relations Using Knowledge-Aware Neural Attentive Models
JMIR Med Inform 2020;8(7):e18417
URL: https://medinform.jmir.org/2020/7/e18417
doi: 10.2196/18417
PMID: 32459650
