A Comparative Study on Feature Selection in Relation Extraction from Electronic Health Records

Ilseyar Alimova[0000−0003−4528−6631] and Elena Tutubalina[0000−0001−7936−0284]

Kazan (Volga Region) Federal University, Kazan, Russia
{alimovailseyar,evtutubalina}@gmail.com

Abstract. In this paper, we focus on clinical relation extraction; namely, given a medical record with mentions of drugs and their attributes, we identify relations between these entities. We propose a machine learning model with a novel set of knowledge and context embedding features. We systematically investigate the impact of these features with popular distance and word-based features. Experiments are conducted on a benchmark dataset of clinical texts from the MADE 2018 shared task. We compare the developed feature-based model with BERT and several state-of-the-art models. The obtained results show that distance and word features are significantly beneficial to the classifier. The knowledge-based features increase classification results on particular types of relations only. The context embedding feature gives the highest increase in results among the other explored features. The classifier obtains state-of-the-art performance in clinical relation extraction with 92.6% of F-measure improving F-measure by 3.5% on the MADE corpus.

Keywords: relation extraction, electronic health records, natural language processing, machine learning, clinical data, hand-crafted features

1 Introduction

Electronic health records (EHRs) contain rich information that can be applied to different research purposes in the field of medicine such as adverse drug reaction (ADR) detection, revealing unknown disease correlations, design and execution of clinical trials for new drugs, clinical decision supports and evidence-based medicine [16, 1, 14, 9, 12, 2]. Despite the enormous potential contained in the clinical notes, there are a lot of technical challenges devoted to the extraction of necessary information from EHRs [16]. EHRs describing the treatment of patients represents a massive volume of an underused text data source. Natural language processing (NLP) can be a solution to provide fast, accurate, and automated information extraction methods that can yield high cost and logistical advantages.

The relation extraction, which identifies important links between entities is one of the crucial steps of natural language processing (NLP). In this paper, we consider the relation extraction task as a binary classification. The classifier

takes as an input pre-annotated pairs of entities and have to identify the relation between them. Let us consider the sentence: "The patient has received *4 cycles* of *Ruxience* plus *Cyclophosphamide* in the last day". In this sentence the entities *Ruxience* and *4 cycles* are related to each other, while *Cyclophosphamide* and *4 cycles* are not related.

Considerable efforts have been devoted to relation extraction research in biomedical domain, including MADE shared-task challenge [15], i2b2 competition [31] and BioCreative V chemical-disease relation extraction task [33]. The aim of the MADE competition was unlocking ADR related information, which can be further used by pharmacovigilance and drug safety surveillance. The organizers provided EHRs texts annotated with medications and their relations to corresponding attributes, indications, and adverse events. All participants of competition developed system based on the machine learning approaches [4, 7, 20, 34]. The winning system obtained 86.8%, while other participants achieved comparable results [15]. However, for the real-world application of extracting drug-related information, the results need to be further improved. Moreover, the contribution of different feature types has not been extensively investigated yet.

To fill this gap, we systematically evaluate four types of features on drug-related information extraction from EHRs: distance, word-based, knowledge, and embedding. In addition to popular features, we propose novel features: (i) number of sentences and punctuation characters between entities, (ii) the previous co-occurrence of entities in biomedical documents from different sources, (iii) semantic types from Medical Subject Headings (MeSH), and (iv) context embedding feature obtained with sent2vec model [6]. We apply a random forest model and perform experiments on the MADE corpus. For comparison, we evaluate a classifier based on Bidirectional Encoder Representations from Transformers (BERT) and approaches of teams participated in the MADE shared task.

The classifier with a combination of baseline and context embedding feature obtains the best results of 92.6% of F-measure and outperforms the previous state-of-the-art results [22] on 3.5%. BERT achieves 90.5% of F-measure. The obtained results show that distance and word features are significantly beneficial to the machine learning classifier. The knowledge features can increase results only on particular types of relations. We also found out that the context embedding feature gives the highest increase in results among the other explored features.

The rest of the paper is structured as follows. We discuss related work in Section 2. Section 3 devoted to corpus description. We describe our set of features in Section 4. Section 5 provides experimental evaluation and discussion. Section 6 concludes this paper.

2 Related Work

The first attempts to relation extraction from EHRs were made in 2010. One of the challenges of i2b2 competition was devoted to assigning relation types that hold between medical problems, tests, and treatments in clinical health records

[31]. This challenge aimed to classify relations of pairs of given reference standard concepts from a sentence. The system based on maximum entropy with a set of features from [25], semantic features from Medline abstracts and parsing trees feature performed the best results among challenge participants [27]. The described system obtained 73.7% of F-measure. The model developed by the team from NRC Canada achieved 73.1% of F-measure [3]. This model is also based on the maximum entropy classification algorithm with the following set of features: based on parsing trees, Pointwise Mutual Information between two entities calculated on Medline abstracts, word surface, concept mapping and context, section, sentence, document-level features. Besides, category balancing and semi-supervised training were applied. The third-place system is based on a hybrid approach that combines machine-learning techniques and constructed linguistic patterns matching [13]. The authors trained SVM with three types of features: surface, lexical, and syntactic. The system obtained 70.9% of F-measure. The rest of the participants applied supervised machine-learning approaches and achieved the results varying from 70.2% to 65.6% of F-measure [24, 17, 11, 28, 8]. One of the main problems faced by participants was varying number of examples for each relation types. The developed classifiers could capture the larger classes accurately by using basic textual features. However, to recognize less relevant relation types, hand-built rules have to be developed.

Natural Language Processing Challenge for Extracting Medication, Indication, and Adverse Drug Events from Electronic Health Record Notes was organized in 2018 [15]. The aim of the competition was extracting ADRs and detecting relations between drugs, their attributes, and diseases. In contrast to i2b2 competition, in this case, only entities are annotated in the corpus. Thus, it is necessary to make candidate pairs and then determine if there is a relation between them. The first place obtained system based on a random forest model with following a set of features, including, candidate entity types and forms, the number of entities between and their types, tokens and part of speech tags between and neighboring the candidate entities [4]. According to the competition resulting table, the described system obtained 86.8% of micro-averaged F_1. Dandala et al. applied the combination of Bidirectional LSTM and attention network and achieved the second place results with 84% of micro-averaged F_1 score [7]. The third place was taken by the system based on the support vector machine model [34]. The classifiers use four types of features: position, distance, a bag of words, and a bag of entities and obtained 83.1% of micro-averaged F_1 measure. Magge et al. employed random forest with entity types, number of the word in entities, number of words between entities, averaged word embeddings of each entity and indicator of presence in the same sentence as a feature [20]. This approach obtained 81.6% of micro-averaged F_1. As can be seen, the most participant teams applied machine learning models, and the only one utilized neural networks while the results were on par.

Munkhdalai et al. conducted additional experiments on MADE corpus and explored three supervised machine learning systems for relation identification: (1) a support vector machines (SVM) model, (2) an end-to-end deep neural net-

work system, and (3) a supervised descriptive rule induction baseline system [22]. For the SVM system entity types, a number of clinical entities, tokens between entities, n-grams between two entities and of surrounding tokens, character n-grams of named entities were applied as features. The combination of BiLSTM and attention was utilized as a neural network model. The maximum averaged F-measure of 89.1% was obtained by the SVM based approach, while the neural network achieved only 65.72% of F-measure.

According to the reviewed studies, the machine learning approaches have a high potential for clinical relation extraction task. However, for real-world biomedical applications, the results need to be improved [15]. The error analysis of systems shows that the most common errors: (i) related entities more than two sentences away from each other, (ii), not related entities occur together in a small distance marks as related (iii) there is more than one entity related to the same entity and only the closest relation is detected. We suppose that these errors can be eliminated if the context is taken into account. Also, most of the previously proposed studies devoted to relation extraction from EHRs largely ignore valuable supportive information, such as the context and knowledge sources. Therefore, the machine learning approach proposed in this paper can be viewed as an extension of the previous work on extracting relations from clinical notes.

3 Corpus

We evaluated our model on the MADE competition corpus [15]. MADE corpus consists of de-identified electronic health records (EHRs) from 21 cancer patients. The EHRs include discharge summaries, consultation reports, and other clinic notes. The overall number of records is 1089, where 876 records were selected for the training split, and the remaining 213 notes formed the testing split. Several annotators participated in the annotation process, including physicians, biologists, linguists, and biomedical database curators. Each document was annotated with two annotators, one of which carried out the initial annotation, the second reviewed the annotations and modified them to produce the final version.

Each record annotated with the following types of entities: drug, adverse drug reaction (ADR), indication, dose, frequency, duration, route, severity, and SSLIF (other signs/symptoms/illnesses). There are 7 types of relations: drug−ade (adverse), sslif−severity (severity), drug−route (route), drug−dosage (do), drug−duration (du), drug−frequency (fr), drug−indication (reason). The detailed statistic of annotated relations is presented in Table 1. According to statistics, the most common relation types are drug-dose, drug-indication, and frequency. Two types of relationships (reason and adverse) have the maximum distance between entities more than 900 characters, which complicates the identification of relations between them.

Table 1. The overall statistic for MADE corpus

	Number			Avg. distance			Max. dist		
Relation type	**train**	**test**	**all**	**train**	**test**	**all**	**train**	**test**	**all**
do	5176	866	6042	8.4	7.7	8.3	215	143	215
reason	4523	870	5393	89.3	63.8	85.2	981	868	981
fr	4417	729	5146	17.7	18.6	17.8	201	178	201
severity_type	3475	557	4032	2.6	1.8	2.5	259	188	259
adverse	1989	481	2470	59.4	45.6	56.7	937	718	937
manner/route	2550	455	3005	13.5	12.9	13.4	191	137	191
du	906	147	1053	18.5	15.0	18.0	272	121	272
all	23036	4109	27145	30.6	26.0	29.9	981	868	981

4 Features

We have divided features into four categories: distance, word, embedding, and knowledge. Distance features are based on counting different metrics between entities. Word features were derived using various properties of context and entity words. Embedding features were received from word embedding models pre-trained on a large number of biomedical texts. Knowledge features were obtained from biomedical resources. The description of each type of feature set out below.

1. Distance features:
   - *word distance* (word_dist): the number of words between entities;
   - *char distance* (char_dist): the number of characters between entities;
   - *sentence distance* (sent_dist): the number of sentences between entities;
   - *punctuation* (punc_dist): the number of punctuation characters between entities;
   - *position* (position): the position of the entity candidate (drug or SSLIF type entity) with respect to the attribute among the entire entity candidates of the attribute, where the position of medical attribute is set to 0.
2. Word features:
   - *bag of words* (bow): all words within a 10-word window before and after the entities plus the entities text. We utilized as features only words that appeared in such context windows with frequencies $\geq$500 across the dataset. Thus, for each entity pair we generated 847 features;
   - *bag of entities* (boe): the counts of all annotation types between the entities;
   - *entity types* (type): binary vector with the number of entities length and units at positions of entity types.
3. Embedding features:
   - *entities embeddings* (ent_emb): the vectors obtained from pre-trained word embedding models for each entity. We explored two word embedding models, including trained on the concatenation of Wikipedia and

PubMed, PMC abstracts [21], and BioWordVec created using PubMed and the clinical notes from MIMIC-III Clinical Database [6]. For entities represented by several words the averaged vector value was applied;
  - *context embedding* (cont_emb): vector obtained from pre-trained BioSentVec model for words between two entities [6]. BioSentVec was obtained using sent2vec library and consists of 700-dimensional sentence embeddings;
  - *similarity* (sim): similarity measure between entities embedding vectors. Four types of similarity measures were employed: taxicab, Euclidean, cosine, coordinate. The vectors were obtained from BioWordVec model [6].
4. Knowledge features:
  - *UMLS concept types* (umls): UMLS[1] (Unified Medical Language System) semantic types of entities represented with binary vector;
  - *MeSH concept types* (mesh): MeSH[2] (Medical Subject Headings) categories of entities represented with a binary vector;
  - *fda clinical trials occurrence* (fda): the number of co-occurrence of both entities in approval document received from FDA[3] for each drug of dataset;
  - *biomedical texts co-occurrence* (bio_texts): the number of entities co-occurrence in biomedical texts. The detailed description of this feature is provided below.

Prior knowledge retrieved from available sources is essential for today's health specialists to keep up with and incorporate new health information into their practices [23]. This process of retrieving relevant information is usually carried out by querying and checking medical articles. We propose a set of features based on primary sources of information to analyze the influence of this process on clinical decision making. In particular, we utilize statistics from various resources using *Pharmacognitive*[4]. This system provides access to databases of grants, publications, patents, clinical trials, and others.

For our experiments, we focus on three sources: (i) scientific abstracts from MEDLINE, (ii) USPTO patents, and (iii) projects from the grant-making Agencies of USA, Canada, EU, and Australia. The *Pharmacognitive* system allows retrieving statistics such as the number of documents or overall funding per year matching a query. The queries are generated using terms from entities of three types: Medication, Indication, and ADR. We extend all queries with terms' synonyms provided by the Pharmacognitive tools. We consider the following features for a individual query *Medication*, *Condition*, *ADR*:

- the number of publications/patents/projects published in the particular year (3 features for each year from 1952 to 2018);

[1] https://www.nlm.nih.gov/research/umls/
[2] https://www.nlm.nih.gov/mesh/meshhome.html
[3] https://www.fda.gov/
[4] https://pharmacognitive.com

– the number of publications/patents/projects published before the particular year (3 features for each year from 1953 to 2018);
– the total number of publications/patents/projects published for all time (3 features);
– the average and sum of projects' funding published in the particular year (2 features for each year from 1974 to 2018);
– the average and sum of projects' funding published before the particular year (3 features for each year from 1975 to 2018);
– the average and sum of projects' funding published for all time (2 features).

We also generate features based on statistics of publications and projects for joint queries of two terms: *Drug* and a disease-related entity (*ADR* or *Indication*).

5 Experiments

In this section, we describe our classifier model, entity pair generation, experiments, and results.

5.1 Classifier

We build a system to resolve the task as a set of independent Random Forest classifiers, one for each relation type. The Random Forest model was implemented with the Scikit-learn library [26]. We tuned the parameters on 5-fold cross-validation and set the number of estimators equal to 100 and the weight balance: 0.7 for positive and 0.3 for negative classes to mitigate the imbalanced class issues.

5.2 Bidirectional Encoder Representations from Transformers (BERT)

BERT (Bidirectional Encoder Representations from Transformers) is a recent neural network model for NLP presented by Google [10]. The model obtained state-of-the-art results in various NLP tasks, including question answering, dialog systems, text classification and sentiment analysis [18, 35, 30, 5, 29]. BERT neural network based on bidirectional attention-based transformer architecture [32]. One of the main model advantages is the ability to give it a row text as the input. In our experiments, we utilized the entity texts combined with a context between them as an input.

5.3 Entities Pair Generation

For each entity we obtained a set of candidate entities following the rules from [34]: the number of characters between the entities is smaller than 1000, and the number of other entities that may participate in relations and locate between the candidate entities is not more than 3. These restrictions allow to reduce infrequent negative pairs and mitigate the imbalanced class issues, while more than 97% of the positive pairs remain in the dataset.

Table 2. The results of F-measure for each relation type and averaged micro F-measure of all relation types for MADE corpus. The distance and word features are applied as a baseline.

Features	**severity**	**route**	**reason**	**do**	**du**	**fr**	**adverse**	**all**
baseline: distance & word feat-s	.933	.918	.806	.906	.905	.896	.729	.866
Munkhdalai et al. [22]	.950	.960	.750	.880	.910	**.950**	.850	.891
Li et al. [19]	-	-	-	-	-	-	-	.872
baseline-word_dist	.923	.922	.812	.900	.860	.909	.716	.864
baseline-char_dist	.929	.916	.810	.908	.869	.890	.731	.864
baseline-sent_dist	.933	.919	.807	.910	.880	.906	.719	.866
baseline-punc_dist	.926	.912	.798	.907	.836	.906	.735	.863
baseline-position	.931	.917	.803	.897	.865	.883	.723	.858
distance	.918	.843	.683	.859	.713	.780	.525	.766
baseline-boe	.932	.897	.775	.888	.861	.868	.715	.845
baseline-bow	.918	.906	.726	.895	.810	.843	.712	.828
baseline-type	.934	.906	.779	.899	.891	.891	.562	.839
word	.542	.777	.645	.662	.718	.846	.511	.672
baseline+emb_pubmed_pmc_wiki	.927	.898	.730	.887	.684	.900	.605	.827
baseline+emb_bio	.920	.903	.772	.893	.602	.908	.613	.833
baseline+cont_emb	.936	.954	**.937**	.929	.854	.938	**.869**	**.926**
cont_emb	.932	.935	.909	.915	.854	.835	.782	.884
baseline+sim	.920	.908	.796	.905	.880	.902	.737	.862
baseline+umls	.936	.915	.815	.922	.883	.891	.734	.870
baseline+mesh	.938	.918	.812	.910	.856	.904	.730	.868
baseline+fda	.936	.912	.808	.906	.895	.909	.730	.868
baseline+bio_text	.934	.918	.805	.906	.905	.896	.749	.866
baseline+knowledge	.936	.914	.806	.916	.889	.896	.736	.848
BERT	**.951**	**.976**	.845	**.934**	**.946**	**.950**	.767	.905

5.4 Experiments and Results

We utilize the model with distance and word features as a baseline. In addition, we compare our results with two state-of-the-art approaches: proposed by Munkhdalai et al. [22] and by Li et al. [19]. Munkhdalai et al. applied SVM with following features: (i) token distance between the 2 entities, (ii) number of clinical entities between the 2 entities, (iii) n-grams between the 2 entities, (iv) n-grams of surrounding tokens of the 2 entities, (v) one-hot encoding of the left and right entities types, (vi) character n-grams of the named entities. Li et al. utilized modern capsule networks.

For distance and word features evaluation, we removed each of the features individually and in combination. To determine the most significant features from embedding and knowledge features sets, we add each of the features separately to the baseline model. The F_1-measure for each relation type and micro-averaged over all classes F_1 were used as evaluation metrics. The evaluation scripts provided by competition organizers were applied to compute these values. The results for each relation type and micro-averaged F-measure are shown in Table 2.

The combination of baseline selected features achieved 86.8% of micro F-measure. This result stays in pair with the best 86.84% F-measure achieved in the competition. The combination of baseline and context embedding features obtained the best results of 92.6% of micro-averaged F-measure. Thus our model outperformed the Munkhdalai et al. results on 3.5%, Li et al. approach on 5.4% and baseline approach on 6%. All reported improvements of the baseline model with context embedding feature over baseline and both state-of-the-art approaches are statistically significant with p-value < 0.01 based on the paired sample t-test. Further, we provided a more detailed analysis of the presented results.

According to Table 2, the classifier with distance features achieves 76.6% of micro-averaged F-measure. Different types of distance features seemed to be complementary to each other due to the absence of one of them leads to approximately the same loss of results. The baseline model without distance set of feature (see row 'word' in Table 2) decrease results on 19% of micro F-measure, which evidences the importance of these parameters for relation classification.

The word-based features also improved the performance of the relation extraction system. The most significant improvement of micro F-measure obtained with a bag of words feature (+3.8 %), which can be explained by a larger vector size compared to the rest of the word-based features. The entity type and a bag of entities feature increased the results of the baseline on 2.7% and 2.1% respectively (see rows 'baseline-type' and 'baseline-boe').

The results for embedding features show that entity embeddings and similarity feature decrease the results regardless of a word embedding model used. The context embedding feature achieved the most considerable improvement of baseline results and obtained 92.6% of micro F-measure. Moreover, the model trained only with the sent2vec feature, outperformed the baseline by 1.8%. This result leads to the conclusion that the context between candidate entities contains more useful information to make a conclusion about relations than candidate entities.

To evaluate knowledge features, it is better to consider the results for different relation types separately. The supplement of UMLS based feature to baseline model increased the results of baseline for severity, reason, and adverse relation types on 0.3%, 0.9% and 0.5% of F-measure respectively. The model with a combination of baseline and MeSH semantic types feature increased the results of baseline for severity and reason types on 0.5% and 0.6% of F-measure, respectively. The FDA co-occurrence feature increased the results for frequency type on 1.3%, while for the rest of the types results are in par. The number of co-occurrence in the biomedical texts feature improved the classifier performance for adverse relation type on 2%. Thus, the knowledge features improved model results for selected types of features.

BERT model achieved the best results for the severity, route, dose, duration, and frequency types of relation. However, for a reason and adverse types, this model obtained F-measure approximately lower on 10% than a random forest with baseline and context embedding features. Thus, BERT gained 90.5% of micro F-measure, and this is the second result among all evaluated models. We

suppose that the results reducing for adverse and reason types can be caused for two reasons: (i) the same disease in different cases could be an adverse drug reaction and a reason, (ii) the average length of the context for these relation types is too long to catch the relation between entities.

A comparison of results for different types of relation shows that the best result was achieved for route (97.6%). This result roughly stays on par with the best results for severity, reason, dose, duration, and frequency types, while the best results for adverse type lower on 10.7%. This difference in results could be due to the greater lexicon variety of adverse drug reaction entity type.

To sum up this section, three important conclusions can be drawn. First, the distance and word-based features are beneficial for the relation classifier. Second, the context embedding has more impact on entities relations than entities embeddings. Finally, the prior knowledge improves the results on particular relation types and the most improvement achieved on adverse relation type with biomedical text co-occurrence feature.

6 Conclusion

In this study, we have investigated the different types of features for drug-related information extraction tasks from EHRs. Our evaluation on MADE competition corpus shows that context embedding, distance, and word features bring the most beneficial to relation extraction task. The classifier with a combination of these sets of features outperformed state-of-the-art results. These facts lead to the conclusion that the context between entities plays a crucial role in relation detection. The detailed analysis of results showed that prior knowledge about entities co-occurrence improved the results for adverse relation type. Our future research will focus on the investigation of modern neural networks for relation extraction from EHRs. We also plan to analyze various context representation methods and extend experiments on other biomedical corpora.

Acknowledgments

This research was supported by the Russian Foundation for Basic Research grant no. 19-07-01115.

References

1. Bates, D.W., Cullen, D.J., Laird, N., Petersen, L.A., Small, S.D., Servi, D., Laffel, G., Sweitzer, B.J., Shea, B.F., Hallisey, R., et al.: Incidence of adverse drug events and potential adverse drug events: implications for prevention. Jama **274** (1), 29–34 (1995).
2. Batin, M., Turchin, A., Sergey, M., Zhila, A., Denkenberger, D.: Artificial intelligence in life extension: From deep learning to superintelligence. Informatica 41(4) (2017)
3. de Bruijn, B., Cherry, C., Kiritchenko, S., Martin, J., and Zhu, X.: Nrc at i2b2: one challenge, three practical tasks, nine statistical systems, hundreds of clinical

records, millions of useful features. In: Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 (2010).
4. Chapman, A.B., Peterson, K.S., Alba, P.R., DuVall, S.L., and Patterson, O.V.: Detecting adverse drug events with rapidly trained classification models. Drug safety, 1–10 (2019).
5. Chen, Q., Zhuo, Z., and Wang, W.: Bert for joint intent classification and slot filling. arXiv preprint arXiv:1902.10909 (2019).
6. Chen, Q., Peng, Y., and Lu, Z.: Biosentvec: creating sentence embeddings for biomedical texts. arXiv preprint arXiv:1810.09302 (2018).
7. Dandala, B., Joopudi, V., and Devarakonda, M.: Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks. Drug safety, 1–12 (2019).
8. Demner-Fushman, D., Apostolova, E., Islamaj Dogan, R., et al.: Nlms system description for the fourth i2b2/va challenge. In: Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 (2010).
9. Demner-Fushman, D., Chapman, W.W., and McDonald, C.J.: What can natural language processing do for clinical decision support? Journal of Biomedical Informatics **42** (5), 760–772 (2009).
10. Devlin, J., Chang, M.W., Lee, K., and Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).
11. Divita, G., Treitler, O., Kim, Y., et al.: Salt lake city vas challenge submissions. In: Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 (2010).
12. Frankovich, J., Longhurst, C.A., Sutherland, S.M.: Evidence-based medicine in the emr era. N Engl J Med 365(19), 1758–1759 (2011)
13. Grouin, C., Abacha, A.B., Bernhard, D., Cartoni, B., Deleger, L., Grau, B., Ligozat, A.L., Minard, A.L., Rosset, S., and Zweigenbaum, P.: Caramba: concept, assertion, and relation annotation using machine-learning based approaches. In: i2b2 Medication Extraction Challenge Workshop (2010).
14. Gurwitz, J.H., Field, T.S., Harrold, L.R., Rothschild, J., Debellis, K., Seger, A.C., Cadoret, C., Fish, L.S., Garber, L., Kelleher, M., et al.: Incidence and preventability of adverse drug events among older persons in the ambulatory setting. Jama **289** (9), 1107–1116 (2003).
15. Jagannatha, A., Liu, F., Liu, W., and 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 safety, 1–13 (2018).
16. Jensen, P.B., Jensen, L.J., and Brunak, S.: Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics **13** (6), 395 (2012).
17. Jonnalagadda, S. and Gonzalez, G.: Can distributional statistics aid clinical concept extraction. In: Proceedings of the 2010 i2b2/VA workshop on challenges in natural language processing for clinical data. Boston, MA, USA: i2b2 (2010).
18. Le, H., Hoi, S., Sahoo, D., and Chen, N.: End-to-end multimodal dialog systems with hierarchical multimodal attention on video features. In: DSTC7 at AAAI2019 Workshop (2019).
19. Li, F. and Yu, H.: An investigation of single-domain and multidomain medication and adverse drug event relation extraction from electronic health record notes

using advanced deep learning models. Journal of the American Medical Informatics Association **26** (7), 646–654 (2019).
20. Magge, A., Scotch, M., and Gonzalez-Hernandez, G.: Clinical ner and relation extraction using bi-char-lstms and random forest classifiers. In: International Workshop on Medication and Adverse Drug Event Detection, 25–30 (2018).
21. Moen, S. and Ananiadou, T.S.S.: Distributional semantics resources for biomedical text processing. Proceedings of LBM, 39–44 (2013).
22. Munkhdalai, T., Liu, F., and Yu, H.: Clinical relation extraction toward drug safety surveillance using electronic health record narratives: classical learning versus deep learning. JMIR public health and surveillance **4** (2), (2018).
23. Pao, M.L., Grefsheim, S.F., Barclay, M.L., Woolliscroft, J.O., McQuillan, M., and Shipman, B.L.: Factors affecting students' use of medline. Computers and Biomedical Research **26** (6), 541–555 (1993).
24. Patrick, J., Nguyen, D., Wang, Y., and Li, M.: I2b2 challenges in clinical natural language processing 2010. In: Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 (2010).
25. Patrick, J. and Li, M.: A cascade approach to extracting medication events. In: Proceedings of the Australasian Language Technology Association Workshop 2009, 99–103 (2009).
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011).
27. Roberts, K., Rink, B., and Harabagiu, S.: Extraction of medical concepts, assertions, and relations from discharge summaries for the fourth i2b2/va shared task. In: Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 (2010).
28. Solt, I., Szidarovszky, F.P., and Tikk, D.: Concept, assertion and relation extraction at the 2010 i2b2 relation extraction challenge using parsing information and dictionaries. Proc. of i2b2/VA Shared-Task. Washington, DC (2010).
29. Sun, C., Huang, L., and Qiu, X.: Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588 (2019).
30. Uglow, H., Zlocha, M., and Zmyślony, S.: Semeval 2019 task 6: An exploration of state-of-the-art methods for offensive language detection. arXiv preprint arXiv:1903.07445 (2019).
31. Uzuner, Ö., South, B.R., Shen, S., and DuVall, S.L.: 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association **18** (5), 552–556 (2011).
32. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, 5998–6008 (2017).
33. Wei, C.H., Peng, Y., Leaman, R., Davis, A.P., Mattingly, C.J., Li, J., Wiegers, T.C., and Lu, Z.: Assessing the state of the art in biomedical relation extraction: overview of the biocreative v chemical-disease relation (cdr) task. Database **2016** (2016).
34. Xu, D., Yadav, V., and Bethard, S.: Uarizona at the made1. 0 nlp challenge. Proceedings of machine learning research **90**, 57 (2018).
35. Zhu, C., Zeng, M., and Huang, X.: Sdnet: Contextualized attention-based deep network for conversational question answering. arXiv preprint arXiv:1812.03593 (2018).