Concept, Assertion and Relation Extraction from clinical narratives using contextual word embedding

Manas Ranjan Mohanty
mohanty.m@northeastern.edu

Rohit Thakur
thakur.ro@northeastern.edu

Nikhil Jyoti
jyoti.n@northeastern.edu

1 Summary

Moving away from the traditional paper and pen methods, clinical workers have now started to record patient data such as discharge summaries, clinical narratives, lab reports, etc electronically leading to the availability of voluminous and useful Electronic Health Records. However, such data is still not annotated or structured but has the potential to provide rich and revealing information. Natural Language Processing (NLP) is effective and useful in extracting and discovering information from such clinical data/narratives. NLP methods have been used in clinical domain for quite some time now and been successful in their operation.

The problem we are addressing is extracting concepts, assertions, and relations from clinical tests. Firstly, concept extraction refers to extracting entities of interest such as medical problems, tests, treatments from clinical text. It can be achieved by tagging words or tokens in the text with classes representing these entities. Thus, with concept extraction we are trying to perform Named Entity Recognition(NER)[2]. We are also interested to know the state of the extracted problems; i.e. it is important to identify if the problem is present, absent, possible, conditionally present, hypothetically present in the patient or associated with some other patient. This is what we are trying to achieve as part of assertion extraction. Similar to concept extraction it is also a token classification problem. Finally, the relation between the problems and remaining entities should be established. It's a relation extraction task. Relation extraction is about building triplets (Entity 1, Entity 2, Relation).

For addressing the above problems, we explored using contextual word embedding from pre-trained state of the arts language models available in clinical domain. Recent developments in this field such as BioBERT, Clinical BERT and NCBI BERT (presently known as BlueBERT) were considered for this. We used NCBI BERT in all our experiments as it is trained on more clinical literature than its competitive models.

1.1 Related Work

For more than a decade now, researchers have been using NLP for extracting all kinds of information from clinical data. For Named Entity Recognition and for other Token classification tasks pre-trained language models such as BERT has been used recently in clinical domain[3]. Relation extraction has become quite useful for clinical research and application. Previous systems for

Type	Description	Number
Concept	Problem	19665
	Treatment	14188
	Test	13833
Assertion	Present	13246
	Absent	4190
	Possible	961
	Hypothetical	827
	Conditional	221
	Associated with someone else	220
Relation	TeRP - Treatment reveals medical problem	3053
	TrAP - Treatment is administered for medical problem	2617
	PIP - Medical problem indicates medical problem	2203
	TrCP - Treatment causes medical problem	526
	TeCP - Test conducted to investigate medical problem	504
	TrIP - Treatment improves medical problem	203
	TrNAP - Treatment is not administered because of medical problem	174
	TrWP - Treatment worsens medical problem	133

Table 1: Details of the dataset

relation extraction can be basically classified as rule-based methods and machine learning based methods. Rule based methods such as dependency trees[4] were implemented earlier for relation extraction tasks. Machine learning methods involving Support Vector Machines[5] and Conditional Random Field[6] were initially used but required huge pre-processing which introduced some errors in the RE task and also effected the accuracy of the system. More recently, deep learning sequential models such as recurrent neural networks (RNN), convolution neural networks(CNN) are being used in the field of relation extraction. While all these works focus on using deep neural networks, limited or very minimal work has been done on using pre-trained language models such as BERT for relation extraction in clinical domain.

1.2 Dataset

The data we are using for our project is the I2B2 2010 challenge data[1]. This data consists of 426 discharge summaries and has been provided by Partners Healthcare and Beth Israel Center. The data is being currently managed by Harvard Medical School. The dataset has 3 types of concepts, 6 types of assertions and 8 types of relations which are shown in table 1. All these types are self explanatory. Though there is data imbalance in case of assertion and relations for some types, we have not implemented any data imbalance handling technique in this project and are aiming to achieve that in future. For all the experiments of this project we used a training to test data ratio as 4:1, which is achieved by using 80 percent of the files (341 files) for training and remaining for testing.

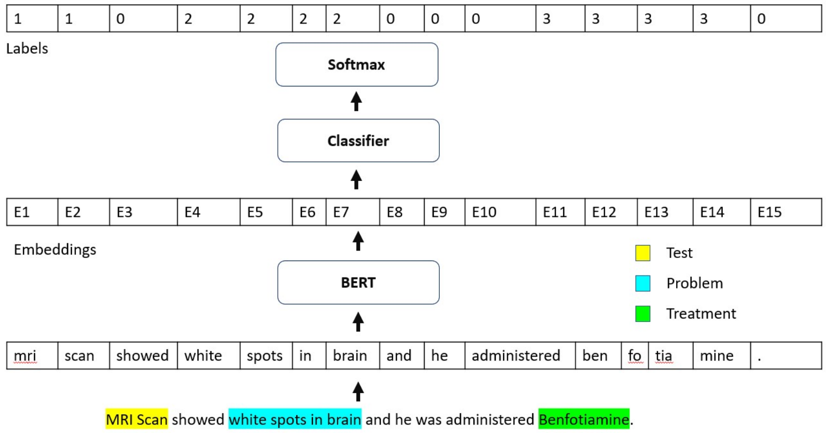

Figure 1: Fine-tuning BERT for concept extraction

2 Methods

2.1 Data and Instance Preparation

We collected data from all the files and validated that all the relations are present in the same sentence. In addition to that, we also validated that the relations are only between a problem and another concept. We discarded invalid relations from data, All the code for this project were on google colab, where BERT finetuning was done on a Tesla V100 GPU(16 GB).

2.2 Fine-tuning BERT for Concept Extraction

Clinical concept extraction is a task of identifying medical concepts such as Problem, Test and Treatment from clinical notes. This task is usually considered as a Named Entity Recognition(NER) which can be solved by machine learning based models such as conditional random fields. But after the breakthrough achieved through BERT by google, similar tasks have shown promising results with word embeddings as input.

We decided to perform both transfer learning and finetuning using NCBI BERT. Input to BERT are text sentences represented in word encodings which are integer sequences of input sentences as well as positional encodings which refer to the token sequence of input. Padding for each sentence can be of length of the largest sentence but since for any BERT model the maximum input sequence length after tokenization is 512, we kept the padding of size 512 as well. In case of transfer learning, we simply used the word embeddings for tokens generated by NCBI BERT to train a classifier. Fine-tuning BERT clinical models on the downstream tasks require some adjustments. Main idea was to train the base NCBI BERT model on our relatively small dataset, feeding the output to softmax layer and backpropagating the error through the entire architecture to update the weights.

Figure 1 shows the end to end process of using BERT for concept extraction task with a given example. As shown in the example, the tokens which do not represent any class can be treated as null class or class 0. As shown, one word can be divided in to multiple tokens by BERT tokenizer. In prediction time, all consecutive tokens with same class type which is not 0, are considered as a single concept.

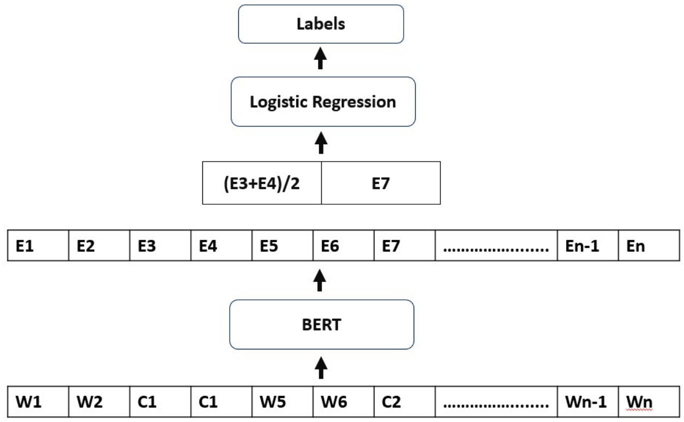

Figure 2: Using word embedding from concepts for relation extraction

2.3 Assertion Extraction

Assertion Extraction is considered to be a significant and challenging task in clinical NLP. There are two approaches that could be taken for this task. First detecting context of the sentence and then identifying entities affected by it. Second, focuses on detecting entity first and then identifying assertion type from the context. We followed the second approach, as it is in continuation of the concept extraction task. Thus, we formulated the problem statement as 'Given an entity in a discharge summary, identify the asserted class from text'. We decided to use the transfer learning approach where we used BERT for generating word embedding of the tokens and using it to train a classifier for assertion extraction. A simple classification problem is described as a mapping function which expects an already defined input,**X** in one or more dimensional space and maps it to a certain category from a predefined set of possible values,**Y**. In our approach we used a classification layer followed by softmax layer similar to concept extraction.

2.4 Relation Extraction

Relation extraction is the task of extracting semantic relations from text, usually occurring between two or more entities. Once concepts are extracted from text, it is crucial to also establish the relation between the entities of interest. For example if there is a pair of concepts in a sentence, one of which represents a medical problem and the other one is a treatment, it is necessary to know the relation between the two entities. Similarly, a test and a problem can be related with each other. Two problems can also be interrelated. To keep the problem simple, the authors of 2010 i2b2 challenge only considered the relations which exist between concepts of the same sentence. In order to perform relation extraction, we considered two approaches which are mentioned below.

2.4.1 Using word embedding from concepts

In this approach, we can simply train a relation extraction model by passing the individual concept embedding as input to the model, while treating the relation type between the concepts as output to the model. We decided to train a logistic regression model for this purpose. If a concept has more than one token, than the average of all the tokens are taken to prepare the embedding of the concept. Figure 1 shows the end to end process for this approach.

Relation	Sentence
TrIP	{0} improves {1}.
TrWP	{0} worsens {1}.
TrCP	{0} causes {1}.
TrAP	{0} is administered for {1}.
TrNAP	{0} is not administered due to {1}.
TeRP	{0} reveals {1}.
TeCP	{0} is conducted to investigate {1}.
PIP	{1} indicates {0}.
No Relation	{0} and {1} are not related.

Table 2: Sentences used for relations

2.4.2 Next Sentence Prediction

BERT was originally pretrained on two self supervised activities. Those are masked language modeling and next sentence prediction. In next sentence prediction task, the model receives pairs of sentences as input and is trained to learn if the second sentence in the pair is the subsequent sentence of the first sentence or not. To help the model distinguish between the sentences, special tokens '[CLS]' and '[SEP]' are added. '[CLS]' token represents the entire sequence which follows it and thus it's embedding is used for sequence classification related tasks. '[SEP]' token is used to separate the two sentences and also mark the end of the sequence. '[CLS]' token embedding is passed to a classification layer, followed by a softmax layer which computes the probabilities of individual classes associated with it.

In order to use next sentence prediction for extracting relation between concepts, we prepared sentences associated with relations with place holders for words/phrases representing the concepts. Table 2 shows sentences for the relations. First placeholder is for the associated concept (test/treatment/problem), where as the second placeholder is for the problem in context. We use true as the classification label, if the sentence following the second sentence is true, or else the label is false. Thus, the classification in this case is a binary classification problem. Figure 2 shows an example of performing relation extraction using next sentence prediction task with BERT. As shown in example, the second sentence represents the sentence associated with relation TeCP(Treatment causes problem). In this sentence the placeholders for Treatment and problem have been replaced by their corresponding phrases from the sentence.

2.4.3 Focus on remaining words for Relation Extraction

In concept extraction task, the actual word/phrase in a sentence plays an important role in understanding the concept type. Similarly, for understanding the relation between the extracted concepts in a sentence, other words in the sentence play an important role. Keeping this in mind, we decided to replace the words representing concepts by their types in the sentences. With this approach, the model will learn to pay attention into the other words in the sentence apart from the concepts to find out relations. Figure 3 represents the end to end process of performing relation extraction using this approach.

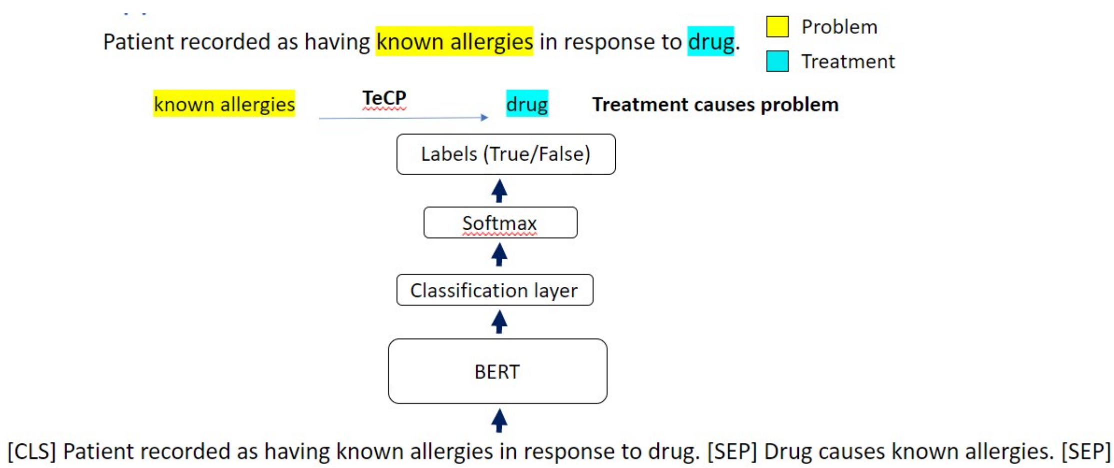

Figure 3: Using next sequence prediction for relation extraction

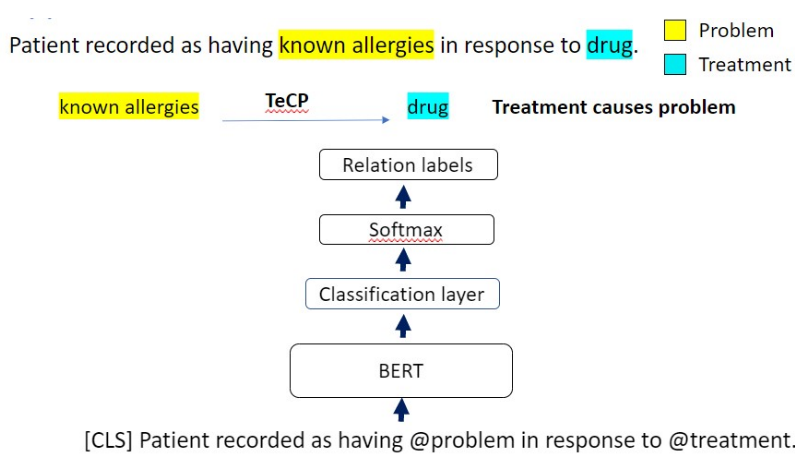

Figure 4: Using remaining words in sentence for relation extraction

Method	**Metric**	**Value**
Transfer learning	**Accuracy**	0.86
	Macro F1	0.78
Finetuning	**Accuracy**	0.96
	Macro F1	0.95

Concept	**Precision**	**Recall**	**F1**
Problem	0.93	0.94	0.94
Test	0.95	0.93	0.93
Treatment	0.94	0.92	0.93

Table 3: Concept extraction results

Assertion	**Precision**	**Recall**	**F1 score**
hypothetical	0.89	0.76	0.82
conditional	0.61	0.17	0.26
associated_with_someone_else	0.80	0.55	0.65
possible	0.70	0.54	0.61
present	0.94	0.97	0.96
absent	0.93	0.90	0.91

Table 4: Assertion extraction finetuning results

3 Results

3.1 Concept Extraction results

Table 3 shows the performance of concept extraction methods on test data. As shown on the left side of the table, finetuning leads to significant improvement in identifying concepts with higher f1 score. Also, from the right side it can be seen that individual concept f1 scores are good. This suggests that finetuning helps in improving the performance of concept extraction.

3.2 Assertion Extraction results

For assertion extraction we only used transfer learning. Table 4 shows the performance of assertion extraction method on test data. As shown in the table, class conditional, associated with someone else and possible have lower f1 scores. This may be because of their lower recall values, which might be due to the fact that definition of these assertions are very close to that of other assertions. So the model have a higher chances of misinterpreting these classes as any other class. This can be improved by finetuning BERT for this specific task.

3.3 Relation Extraction results

Table 5 shows the performance of all the methods implemented for relation extraction. As shown, we observed that the method in which concepts are replaced by their types performs the best for most of the relations. Next sentence prediction did not perform well in distinguishing the relations. This may be due to two facts. Firstly, the sentence representation for the relations are not sufficient for the model to differentiate the relation. Secondly, the next sentence prediction may not be truly used for relation extraction because of the fact that the all the second sentences representing relations may be in context with respect to the first sentence. To confirm either of these facts, we need to extend the experiment further. On the other hand, in the final method there is more focus on remaining words of the sentence apart from the words representing concepts,

Relation	Concept Embedding			NSP			Remaining words		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
PIP	0.36	0.28	0.32	0.12	0.28	0.17	0.47	0.38	**0.42**
TeCP	0.29	0.33	0.31	0.28	0.39	0.33	0.37	0.48	**0.42**
TeRP	0.55	0.74	0.63	0.31	0.46	0.37	0.67	0.82	**0.74**
TrCP	0.72	0.58	**0.64**	0.38	0.21	0.27	0.6	0.43	0.50
TrIP	0.28	0.12	0.17	0.6	0.32	0.42	1	0.43	**0.60**
TrAP	0.54	0.77	**0.63**	0.48	0.72	0.58	0.63	0.62	0.62
TrWP	0.42	0.29	0.34	0.27	0.69	0.39	0.35	0.43	**0.39**
TrNAP	0.29	0.47	**0.36**	0.21	0.05	0.08	0.28	0.37	0.32

Table 5: Relation extraction results

which is actually crucial in determining the relation. We conclude that's why the method performs the best.

4 Discussion and Conclusion

In this project, our focus was on using pretrained state of the arts language models to effectively extract the three primary clinical concepts which are problem, test and treatment from discharge notes, identify the assertions of the problems and finally establish relationship between all problems and all other concepts available in the same sentence. We finetuned BERT for performing concept extraction and could observe significant improvement in performance on test data. This suggests BERT performs well on Named Entity Recognition task. For assertion extraction, which is another token classification task, we used word embedding from BERT and trained Logistic regression model for the multi-class classification. Though the performance of the model was acceptable, we found that it did not perform well on classes which have similarity in definition with other classes. This can be improved by finetuning BERT similar to concept extraction. For relation extraction, we experimented with three different methods. We deduced that, when there is more focus on remaining words of the sentence, the model differentiates different relations better. We also tried to utilise next sentence prediction task for relation extraction, which did not perform well. We can extend the experiment further with different sentences, or adding negative sentences for relations in addition to positive sentence, or by focusing on relations which are difficult to differentiate.

5 Statement of Contributions

Manas Ranjan Mohanty: Report, Proposal, Training the models, Presentation, Project Research, Topic Research
Rohit Thakur: Report, Proposal, Training the models, Presentation, Project Research, Topic Research
Nikhil Jyoti: Report, Proposal, Training the models, Project Research, Topic Research

References

[1] I2B2 2010 challenge: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3168320/

[2] Named Entity Recognition(NER): https://en.wikipedia.org/wiki/Named-entity_recognition

[3] Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets: https://arxiv.org/abs/1906.05474v2

[4] Fundel K, Küffner R, Zimmer R. RelEx—Relation extraction using dependency parse trees. Bioinformatics. 2006 Dec 1;23(3):365–71: https://pubmed.ncbi.nlm.nih.gov/17142812/

[5] Lee HJ, Xu H, Wang J, Zhang Y, Moon S, Xu J, Wu Y. UTHealth at SemEval-2016 task 12: an end-to-end system for temporal information extraction from clinical notes. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) 2016:pp. 1292–1297

[6] Rink B, Harabagiu S, Roberts K. Automatic extraction of relations between medical concepts in clinical texts. Journal of the American Medical Informatics Association. 2011;18(5):594–600. :https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3168312/

Appendix

Github link: `https://github.com/ManasRMohanty/CS6120`