2018-04-16

A hybrid Neural Network Model for Joint Prediction of Presence and Period Assertions of Medical Events in Clinical Notes

Li Rumeng
University of Massachusetts Amherst

Et al.

A hybrid Neural Network Model for Joint Prediction of Presence and Period Assertions of Medical Events in Clinical Notes

Rumeng Li[1], Abhyuday N Jagannatha[1], Hong Yu, PhD[2,3]
[1]University of Massachusetts, Amherst, MA, USA; [2]Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA;[3]Bedford VAMC and CHOIR, Bedford, MA,USA

Abstract

In this paper, we propose a novel neural network architecture for clinical text mining. We formulate this hybrid neural network model (HNN), composed of recurrent neural network and deep residual network, to jointly predict the presence and period assertion values associated with medical events in clinical texts. We evaluate the effectiveness of our model on a corpus of expert-annotated longitudinal Electronic Health Records (EHR) notes from Cancer patients. Our experiments show that HNN improves the joint assertion classification accuracy as compared to conventional baselines.

Introduction

In recent years natural language processing techniques have demonstrated increasing effectiveness in clinical text mining. Electronic health record (EHR) narratives, e.g., discharge summaries and progress notes contain a wealth of medically relevant information like diagnosis information, adverse drug events etc. Automatic extraction of such information and representation of clinical knowledge in standardized formats could be used for a variety of purposes like clinical event surveillance and decision support, pharmacovigilance and drug efficacy studies etc. Extracting information from EHR narratives presents challenges unique to this domain. Unlike open-domain data, clinical text normally contain a substantial amount of in-domain terminology and domain-specific knowledge. Therefore, accurately recognizing and understanding the medical entities in clinical text become essential for useful information extraction.

This work addresses the problem of identifying *assertion* in EHRs. Assertion is an important attribute to any event in information extraction. In EHRs, assertation can be understood as a physician's belief status with regards to a particular patient's medical problem. Specifically, as shown in Table 1, a medical problem could be current or happened in the past. The problem could be present, absent, or hypothetical and conditional. Knowing the assertion status of a clinical event (e.g., bleeding) is important for physicians to make clinical decisions (e.g., prescribing anticoagulants). Therefore assertion identification of clinical events is critical for information extraction and data mining from EHRs. Assertion identification was also one of the 2010 i2b2/VA challenge task[1], which classified assertion as *present*, *absent*, *possible*, *conditionally present*, *hypothetically present* and *not associated with patients*. In our task, we classify the presence status of a clinical event conditioned on the current time period. We have the same six categories for presence assertion as the 2010 i2b2/VA challenge task, and we have extra four classes of period assertion for the same clinical event as *current*, *history*, *future*, and *unknown* to capture the temporal information about statements. To better describe the task and our model, we present an exemplary explanation in Table 1.

Table 1: Presence and Period Assertions.

Adverse Drug Event	Period	Presence
He has fever (caused by the drug)	Current	Present
He had fever (due to the drug)	History	Present
He has no fever (from the drug)	Current	Absent
His fever (caused by the drug is) resolved	History or current	Present or Absent
He has a fever, (possibly caused by the drug)	Current	Possible
He might have a fever	Current	Possible
If he is infected/(takes the drug), he will run a fever	Future	Conditional
He may develop a fever (with this drug)	Future	Hypothetical

Table 1 shows the *period* and *presence* assertion categories and corresponding representative texts to the clinical event "fever" (an adverse drug event). Our goal is to develop natural language processing (NLP) approaches to automatically identify both the belief status of a clinical event and its period status. It is important to identify both types of assertaion as the task represents a more accurate scheme for reasoning about the physician's belief status of the patient's medical problem. For example, to identify the assertion in the sentence "His fever caused by the drug is resolved", one needs to consider the dependency between period and presence. In this particular scenario, "fever" is present if period is history. In contrast, "fever" is absent if period is current. Also as the presence and period assertions are related to each other or even conditionally dependent on each other, a joint learning model can be of advantage. We incorporate such relations in our model and propose a neural network based framework to jointly predict the two types assertions for a given medical entity.

Previous efforts for the assertion identification (or classification) task include rule-based and machine-learning-based methods. Rule-based approaches required hand-crafted rules, which limited their performance. Therefore, it is no surprise that in the 2010 i2b2/VA challenge task, eight of the top 10 participating systems employed machine-learning approaches (e.g.,SVM-based classifier, sometimes employing millions of features). With the recent advance in deep learning, neural network models have shown in automatically capturing semantics and syntax as compared to traditional SVM based models, which require significant feature engineering. Additionally, neural network models also have the added advantage of capturing long-distance dependency in text.

In this study, we explored deep learning models. Specifically, we used recurrent neural network with gated recurrent unit(GRU) to represent a clinical event using the left and right context of the event in its sentence. For each generated hidden unit, residual neural network[12] was used for better representation generation. After combining the entity and context representations with additional attention weights, the framework outputs two labels for presence and period assertion. In this architecture, the two tasks leveraged a common feature-set generated by the recurrent neural networks and then used different attention weights for each assertion task. We also used two extra parameters to add mutual influence for the final prediction of the two types of assertion. Experiments on an expert-annotated EHR narratives show the effectiveness of our deep learning model. We provide more details about our model in Section 3.

The main contributions of this paper include a novel neural network architecture that not only leverages recurrent residual network for assertion classification task, but also jointly predicts both the presence and period assertions in one framework. Our method obtains good results on both types of assertion classification tasks. The rest of the paper is organized as follows: we present the related work in Section 2 and introduce our proposed model in Section 3. In Section 4 we report the experiment results and our analysis. And we conclude the paper in Section 5.

Related Work

Medical Assertion Classification

Determination of the assertion status of clinical events is an important area of clinical NLP research. Previous efforts mainly include rule-based methods and machine learning approaches. Popular rule-based methods include the NegEx algorithm[13] and ConText algorithm[14,30]. The NegEx algorithm is a simple regular expressions algorithm to determine whether a medical entity is present or absent in a patient. The ConText algorithm extends the NegEx algorithm to detect four assertion categories: absent, hypothetical,historical, and not associated with the patient. Uzuner et al.[15] studied the rule-based Extended NegEx system and a SVM-based Statistical Assertion Classifier (StAC) and showed that a machine learning approach achieved competitive results for assertion classification. Four assertion classes as present, absent, uncertain in the patient, or not associated with the patient were used in their system. Wu et al.[13] conducted a multi-corpus analysis of negation detection and concluded that it was easy to optimize for a single corpus but not to generalize to arbitrary clinical text.

The 2010 i2b2/VA Challenge designed a specific assertion classification task[1]. For each "problem" concept mentioned in a clinical text, systems were built to classify the concept's status associated with the patients as "present", "absent", "possible in the patient", "conditionally present", and "hypothetically present", or mentioned in the patient report but "associated with someone other than the patient" based on the context that describes it. The task as a multi-class categorization problem allows the use of machine learning classification methods. SVMs were still the common theme for the task[2–7]. For some SVM models, millions of features were employed from lexical, syntactic to contextual level.

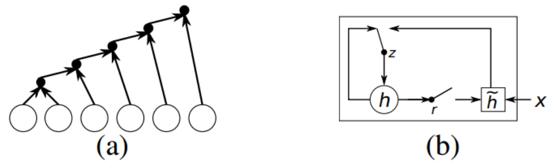

Figure 1: The graphical illustration[22] of (a) the recurrent neural network and (b) the hidden unit of GRU that adaptively forgets and remembers.

In addition to those text features, concept-mapping features derived from existing annotation tools like cTAKEs[8], MetaMap(UMLS)[9] etc. are also used. The best system[10] addressed the task in two stages: in stage 1, assertion class predictions were generated for every word that was part of a "problem" concept by using three parallel different svm classifiers. In stage 2, a secondary classifier predicted a class for the complete concept, based on the separate per-word predictions from the ensemble of stage-1 classifiers, by using SVM-multi-class with a linear kernel. Later Kim et al.[11] revised their participating system and added specific features to improve the performance on minority classes (e.g., the conditional class) and obtained better results.

Our task is more challenging than the 2010 i2b2/VA Challenge assertion task in that we not only classify the presence assertion, we also jointly classify the period assertion as being "Current", "History", "Future", or "Unknown" in a joint model architecture.

Related Neural Networks

Our deep learning model is based on neural network models to learn feature representations for clinical events and their context for classification. Our model is related to learning representations for long text (sentence/paragraph/document), an important task which draws much efforts. Recurrent neural networks (RNN)[17], and their variants are widely used. Closely related work to our model is the recurrent neural network with gated recurrent unit (GRU)[18] and deep residual network[12].

RNN can be used effectively to learn distributed representations over a variable-length sequence. At each time-step, it takes both the output of the previous step and the current token as input, convolutes the inputs, and forwards it to the next step. A gated recurrent unit (GRU), a variant of RNN, was proposed by Cho et al.[18] to make each recurrent unit to adaptively capture dependencies of different time scales. It has similar unit as the long-short term memory unit (LSTM)[19] with two gating units named reset and update gates modulating the flow of information inside the unit. However, without having a separate memory cells as LSTM. GRU has been proved comparable results and faster training with less parameters than LSTM[20]. Figure 1 provides an illustration of RNN and GRU.

The deep residual network has two significant characters compared with RNN. The first one is the residual learning. In a neural network model, normally the data is passed from one layer to the adjacent layer. In the residual network, an additional layer is used to connect layers that are far away. During the back propagation, errors can be passed from a higher layer to a lower layer directly. This character is of advantage as it may capture the long-distance context determining the assertion. The second character is the depth of such models. A typical residual network has hundreds of layers, which is much deeper than most existing models. Thus the training of such a model becomes a challenge. In addition, given the number of the layers, the number of parameters also exceeds most networks. When trained on a small dataset, a deep residual network may suffer from over-fitting[12]. The residual neural network has been proved useful in capturing information from images for classification. A number of variants have been introduced for a series of tasks[23,24], we do not go in to the details due to the space limitation.

Attention mechanism is also adopted in our system. Attention mechanisms in neural networks are inspired by the presence of attention in human visual system[25,26]. Human beings' visual system is able to focus on the most salient part of an image and adjust the focal point over time. The concept of "attention" has gained popularity in training neural networks and have been applied to various computer vision and NLP tasks[27–29] and we don't enumerate here.

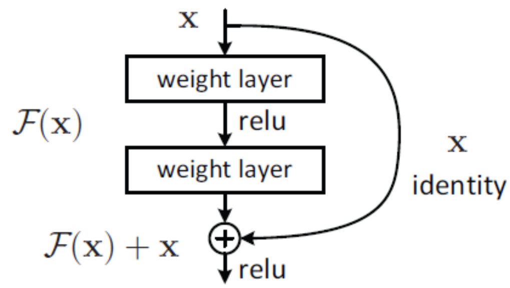

Figure 2: A Residual Learning Block[12]. Compared with a traditional multi-layer perceptron, the change is that some layers that used to be non-adjacent are connected.

In the following Section, we will show how to employ the recurrent neural network with gated recurrent unit and residual network in the targeted task.

Methods

We address the assertion classification task as a supervised learning problem and build one framework to jointly predict both presence and period assertion labels. Given a sentence with an annotated clinical event, the model predicts both the presence tag and the period tag for the event. The presence tag is chosen from six categories and the period tag has four categories.

Given a clinical event and its sentence, our HNN model use *three* recurrent neural networks with gated recurrent unit (GRU) and residual networks to generate representations respectively for the clinical event, sentence tokens to the left of the clinical event entity, and the right. These representations are further passed to following layers conducting interactions and predicting the tags.

For better illustration, we provide Figure 3 and Figure 4 to show how the proposed HNN model works for the medical entity "fever" in the sentence "He has fever caused by the drug". [1] The sentence is splitted into three parts, the entity, its left neighbor and its right neighbor. We firstly have a representation for the entity "fever", as this is a single word, RNN GRU is not needed to generate the representation, we can just use the embedding of "fever" instead. But for an entity with longer sequence, for example like "lung cancer" or any entity longer, an RNN model is needed for generating the representation. Two RNNs are also employed for "He has" and "caused by the drug" which are the left and right neighbors of the entity. The three representations we get from the three RNNs are fed to the following part of the network as shown in Equation 1. In Equation 1, R_{Block} stands for residual learning block. L stands for linear function.

$$
\begin{cases}
h_i^{tmp} = GRU(h_{i-1}, x_i) \\
R\_Block(h_{tmp}^i) = \\
\quad Dropout(L(Dropout(\underbrace{L(tanh}_{\text{repeat twice}}(h_i^{tmp}))))) \\
ResidualNet(h_{i-1}, x_i) = \underbrace{R\_Block}_{\text{repeat 7 times}}(h_i^{tmp}) \\
h_i = ResidualNet(h_{i-1}, x_i)
\end{cases}
\tag{1}
$$

Equation 1 and Figure 4 explain the residual recurrent neural networks with gated recurrent unit (GRU) we used in this work.

Their representations are then fed to the linear transformation node which outputs the final score. The linear function

[1] Please note that in this example sentence, "drug" is also a medical entity, but we just address the entity "fever" as an example here.

takes as input the final left representation $out_{L_i} = h^L_{final}$, final right representation $out_{R_j} = h^R_{final}$ and the final representation for the medical entity $out_{E_k} = h^E_{final}$, and outputs the comprehensive representation for this entity R_e as in Equation 2.

$$\begin{aligned} &R_e(out_{L_i}, out_{R_j}, out_{E_k}) = \\ &f(out_{L_i}W_0 + out_{R_j}W_1 + out_{E_k}W_2 + b) \end{aligned} \tag{2}$$

Then we get to consider features that are specific to each sub-task. We assume that for each sub-task, the model should attend to different parts of the original sentence. In the mean time since the two sub-tasks are related as we have stated, the knowledge what sub-task A is attending to will help improve the performance of sub-task B, we formulate the representations of the two sub-tasks to interact with each other. Equation 3 defines the details.

$$\begin{aligned} &Step\ 1 \\ &R_e^1 = f_1(R_e) + \sum_{i=1,n} att_i^1 w_i; w_i \in S \\ &R_e^2 = f_2(R_e) + \sum_{i=1,n} att_i^2 w_i; w_i \in S \\ &Step\ 2 \\ &R_e^1 = R_e^1 + \alpha R_e^2 \\ &R_e^2 = R_e^2 + \beta R_e^1 \end{aligned} \tag{3}$$

We use S to represent the sentence which contains the entity. In Equation 3 n is the number of its containing words. w_i is the embedding of the i_{th} word in S. We use R_e to distribute attentions over the n words in S. The weights are calculated using a softmax over the sum of all w in S. $\vec{att^j} = Softmax(Linear^j([R_e : \sum_{w_i \in S} w_i]$. $Step2$ defines our way to interact the two sub-tasks with weights α and β. α and β are learned according to $\alpha = WR_e^2 + b; \beta = W'R_e^1 + b'$.

To boost performance, we add some basic features here including the entity position in the sentence, entity length, entity bag-of-words features, sentence length, number of nouns in the sentence, number of verbs in the sentence, the verb tense in the sentence, and part-of-speech tags of words in the sentence. They are concatenated with R_e^1 and R_e^2.

The score for each label was fed into the softmax classifier using R_e^1 and R_e^2 to make predication of the presence and period assertion.

Equation 4 defines our loss function:

$$\begin{aligned} &loss = hinge_{loss}(pred_{lable1}, gold_{label1}) + hinge_{loss}(pred_{lable2}, gold_{label2}) \\ &hinge_{loss} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}[\max(0, 1 - \delta\{l_n = k\}t_{nk})]^p \end{aligned} \tag{4}$$

Experiments

Datasets

We used an annotated corpus of 1089 EHR notes. The corpus comprises of a three-year longitudinal provider notes of 21 cancer patients, which include progress notes and discharge summaries – essentially all note types in the longitudinal EHRs of the cancer patients. Each note was annotated by two clinical professionals. For each annotated medical entity, including four different types respectively as Drug (Medication information of drug and its attributes), Indication (reason for prescribing medication), ADE (Adverse Drug Event as an injury resulting from the normal use

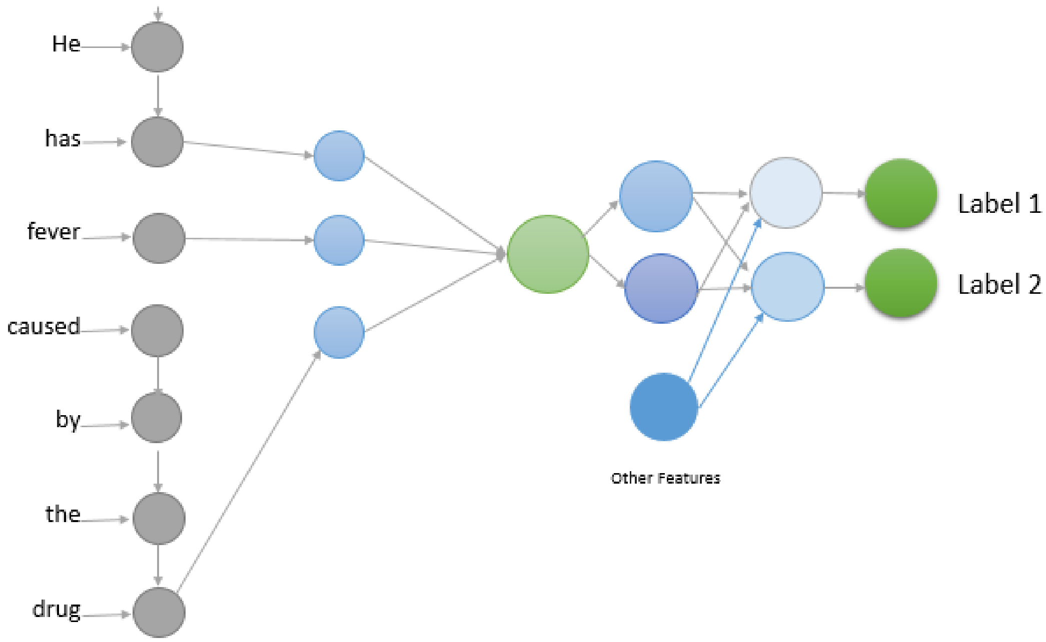

Figure 3: The architecture of the proposed model.

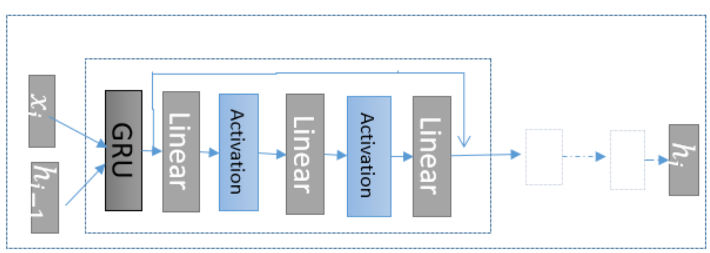

Figure 4: How the residual network functions in the proposed model. Note that it contains dropout which is not shown in this figure explicitly.

of a drug at a normal dose), and SSLIF (other signs which the physician can observe and symptoms which the patient reports) in the corpus, presence and period assertions are used together to assign property to it. *Presence assertions* include six categories which are respectively *present*, *absent*, *possible*, *conditional*, *hypothetical*, and *not associated with patient*. The *period assertions* include four categories as *current*, *history*, *future*, and *unknown*. *Presence* and *period* assertion annotation statistics in the corpus are provided in the Table 2 and Table 3.

Table 2: Distribution of Presence Assertions.

	Training No.	Training Proportion %	Testing No.	Testing Proportion %	Total No.
Present	13960	52.09	6978	52.07	20938
Absent	10680	39.85	5391	40.23	16071
Hypothetical	582	2.17	307	2.29	889
Possible	679	2.53	315	2.35	994
Conditional	68	0.25	36	0.27	104
Not Patient	830	3.10	373	2.78	1203

Table 3: Distribution of Period Assertions.

	Training No.	Training Proportion %	Testing No.	Testing Proportion %	Total No.
Current	20881	77.92	10485	78.25	31366
History	5156	19.24	2503	18.68	7659
Future	696	2.60	369	2.75	1065
Unknown	66	0.25	43	0.32	109

Experiments Setup

Our experiments used word embeddings of 100 dimensions learnt on a combined corpus of PubMed open access articles, English Wikipedia and an unlabeled corpus of around hundred thousand Electronic Health records which are not the EHR used in our dataset[16]. We used Chainer, a flexible framework for neural networks for the implementation of the proposed model [2]. We report the micro-averaged and macro-averaged accuracy results and also the performance on each class.

Experiment Results

For comparison, we used a standard SVM and a standard LSTM model as baseline. To be specific, the Linear SVM classifier[1] with bag-of-words features is used. The LSTM is implemented using Chainer. It runs through the whole sentence, without splitting the sentence according to entities as what we do in the proposed model. Bag-of-words features is also employed. In addition, the presence assertion and period assertion classification task are conducted separately for the baseline systems.

Experiment results for presence assertion is shown in Table 4 and results for period assertion is shown in Table 5. We use HNN as abbreviation for our model. As can be seen, the proposed model by jointly training the two task in one unified framework obtain good results in both two tasks.

Analysis

As mentioned in the previous sections, HNN addresses the combined classification tasks of presence and period assertion in one joint framework. Assertion prediction is a complex task from the NLP perspective. The difficulty

[2] https://chainer.org/

[1] Trained using LIBSVM[21]

Table 4: Presence Assertion Accuracy Results Comparison with Baseline.

	SVM %	LSTM %	HNN %
Present	93.81	93.01	91.27
Absent	83.29	90.41	**91.02**
Not Patient	0	61.13	**81.50**
Conditional	0	0	2.78
Possible	0	2.22	**20.95**
Hypothetical	0	0	0
Micro Avg.	82.3	86.56	**86.92**
Macro Avg.	29.52	41.13	**47.92**

Table 5: Period Assertion Accuracy Results Comparison with Baseline.

	SVM %	LSTM %	HNN %
History	0	31.76	**55.41**
Current	100	95.32	93.14
Future	0	5.69	**27.64**
Unknown	0	4.65	**34.88**
Micro Avg.	78.25	80.69	**84.10**
Macro Avg.	25.00	34.36	**52.77**

of this task is further exacerbated by the fact that we predict assertions on four different underlying medical entities (Drug, ADE, Indication and SSLIF). The difference among these entity types induces a domain difference in assertion prediction which increases the complexity of our task.

The HNN model in general achieves competitive results on both tasks. The overall result of all methods on the period task is lower than that on the presence task. One possible factor behind this discrepancy is that an EHR document is largely written in past or present perfect tense, irrespective of the relative timeline of events in that document. Due to this ambiguous use of tense in EHRs, learning algorithms are often forced to rely on contextual clues, which are more difficult to recognize and extract. Presence assertions on the other hand have more direct linguistic clues such as negation. However, even though the period assertion task is a more difficult one, our system is able to outperform the baseline by a large margin. In fact, the performance gap between HNN and the next best baseline (around 3.4% micro accuracy and 18.4% macro accuracy) for period task is significantly higher than that in presence task. This indicates that HNN can learn to recognize a more diverse set of contextual patterns, and is less dependent on tense based indicators.

The SVM model is ineffective for presence and period assertion in minority classes. LSTM model also only shows small improvements in minority class prediction. In contrast, our HNN model leads to significant improvement on several minority classes such as “Not Patient”, “Possible”, “Future” and “Unknown”. This performance indicates that HNN can better capture the semantics of minority class and subsequently learn to generalize better from fewer examples.

All the three models get zero accuracy on the conditional and hypothetical classes (except a very small accuracy of our model on conditional class) because these two are the rarest classes in our dataset. As a result, our data-driven models do not have enough corresponding positive samples for meaningful training.

Conclusion

We have proposed a novel hybrid neural network framework for prediction of assertions in Electronic Health Records. Our model can jointly predict both the presence and period assertion values ascribed to the medical entities associated with patients in clinical texts. Our experimental results show that the HNN model leads to significant improvement in both minority and majority class for period and presence assertion.

Acknowledgement

We thank the UMassMed annotation team: Feifan Liu, Elaine Freund, Weisong Liu, Steve Belknap, Nadya Frid, Alex Granillo, Heather Keating, and Victoria Wang for creating the gold standard evaluation for this work. We also thank the anonymous reviewers for their comments and suggestions. This work was supported in part by the grant HL125089 from the National Institutes of Health. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

References

1. Uzuner , South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association. 2011 Jun 16;18(5):552-6.
2. Jiang M, Chen Y, Liu M, Rosenbloom T, Mani S, Denny JC, Xu H. Hybrid approaches to concept extraction and assertion classification-vanderbilts systems for 2010 I2B2 NLP Challenge. In Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 2010.
3. Roberts K, Rink B, Harabagiu S. Extraction of medical concepts, assertions, and relations from discharge summaries for the fourth i2b2/VA shared task. InProceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 2010.
4. Uzuner , South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association. 2011 Jun 16;18(5):552-6.
5. Demner-Fushman D, Apostolova E, Islamaj Dogan R. NLM's 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.
6. Grouin C, Abacha AB, Bernhard D, Cartoni B, Deleger L, Grau B, Ligozat AL, Minard AL, Rosset S, Zweigenbaum P. CARAMBA: concept, assertion, and relation annotation using machine-learning based approaches. Ini2b2 Medication Extraction Challenge Workshop 2010 Nov 12.
7. Chang E, Xu Y, Hong K, Dong J, Gu Z. A hybrid approach to extract structured information from narrative clinical discharge summaries. In Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data. Boston, MA, USA: i2b2 2010.
8. Savova G, Kipper-Schuler K, Buntrock J, Chute C. UIMA-based clinical information extraction system. Towards enhanced interoperability for large HLT systems: UIMA for NLP. 2008;39.
9. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. Journal of the American Medical Informatics Association. 2010 May 1;17(3):229-36.
10. de Bruijn B, Cherry C, Kiritchenko S, Martin J, Zhu X. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. Journal of the American Medical Informatics Association. 2011 May 12;18(5):557-62.
11. Kim Y, Riloff E, Meystre SM. Improving classification of medical assertions in clinical notes. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2 2011 Jun 19 (pp. 311-316). Association for Computational Linguistics.
12. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).
13. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of biomedical informatics. 2001 Oct 1;34(5):301-10.
14. Chapman WW, Chu D, Dowling JN. ConText: An algorithm for identifying contextual features from clinical text. In Proceedings of the workshop on BioNLP 2007: biological, translational, and clinical language processing 2007 Jun 29 (pp. 81-88). Association for Computational Linguistics.
15. Uzuner , Zhang X, Sibanda T. Machine learning and rule-based approaches to assertion classification. Journal of the American Medical Informatics Association. 2009 Jan 1;16(1):109-15.

16. Jagannatha AN, Yu H. Structured prediction models for RNN based sequence labeling in clinical text. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing 2016 Nov (Vol. 2016, p. 856). NIH Public Access.
17. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems 2013 (pp. 3111-3119).
18. Cho K, Van Merrinboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. 2014 Jun 3.
19. Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997 Nov 15;9(8):1735-80.
20. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555. 2014 Dec 11.
21. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST). 2011 Apr 1;2(3):27.
22. Cho K, Van Merrinboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259. 2014 Sep 3.
23. Wang Y, Tian F. Recurrent Residual Learning for Sequence Classification. In EMNLP 2016 (pp. 938-943).
24. Zhang Y, Chen G, Yu D, Yaco K, Khudanpur S, Glass J. Highway long short-term memory rnns for distant speech recognition. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on 2016 Mar 20 (pp. 5755-5759). IEEE.
25. Rensink RA. The dynamic representation of scenes. Visual cognition. 2000 Jan 1;7(1-3):17-42.
26. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nature reviews neuroscience. 2002 Mar 1;3(3):201-15.
27. Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025. 2015 Aug 17.
28. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y. Show, attend and tell: Neural image caption generation with visual attention. InInternational Conference on Machine Learning 2015 Jun 1 (pp. 2048-2057).
29. Rush AM, Chopra S, Weston J. A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685. 2015 Sep 2.
30. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experience, and temporal status from clinical reports. Journal of biomedical informatics. 2009 Oct 31;42(5):839-51.
31. Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, Clark C. Negations not solved: generalizability versus optimizability in clinical natural language processing. PloS one. 2014 Nov 13;9(11):e112774.