Neural Architecture for Temporal Relation Extraction: A Bi-LSTM Approach for Detecting Narrative Containers

Julien Tourille
LIMSI, CNRS
Univ. Paris-Sud
Université Paris-Saclay
julien.tourille@limsi.fr

Olivier Ferret
CEA, LIST,
Gif-sur-Yvette,
F-91191 France.
olivier.ferret@cea.fr

Xavier Tannier
LIMSI, CNRS
Univ. Paris-Sud
Université Paris-Saclay
xavier.tannier@limsi.fr

Aurélie Névéol
LIMSI, CNRS
Université Paris-Saclay
aurelie.neveol@limsi.fr

Abstract

We present a neural architecture for containment relation identification between medical events and/or temporal expressions. We experiment on a corpus of de-identified clinical notes in English from the Mayo Clinic, namely the THYME corpus. Our model achieves an F-measure of 0.613 and outperforms the best result reported on this corpus to date.

1 Introduction

Temporal information extraction from clinical health records allows for a fine-grained analysis of patient health history. Providing medical staff with patient timelines could lead to improved diagnostic and care. Important temporal information (such as when a patient started a treatment, or when they started experiencing side effects from a treatment) can be found only within the narrative portion of records and needs the development of new Natural Language Processing methods in order to be accessed.

In this paper, we present a neural architecture for narrative container identification between medical events (EVENT) and/or temporal expressions (TIMEX3). We experiment on the THYME corpus (Styler IV et al., 2014), a corpus of de-identified clinical notes in English from the Mayo Clinic. We use the Gold Standard annotations for EVENT and TIMEX3 entities and we focus on containment relation extraction where the objective is to identify temporal relations between pairs of entities formalized as narrative container relations.

2 Related Work

SemEval has been offering a shared task related to temporal relation extraction from clinical narratives over the past two years (Bethard et al., 2015, 2016). Relying on the THYME corpus, the task challenged participants to extract EVENT and TIMEX3 entities and then to extract narrative container relations and document creation time relations. Herein, we focus on the second part of the challenge, temporal relation extraction and more specifically the narrative container relations. Different approaches have been implemented by the participants, including Support Vector Machine (SVM) classifiers (AAl Abdulsalam et al., 2016; Cohan et al., 2016; Lee et al., 2016; Tourille et al., 2016), Conditional Random Fields (CRF) and convolutional neural networks (CNNs) (Chikka, 2016). Beyond the challenges, Leeuwenberg and Moens (2017) propose a model based on a structured perceptron to jointly predict both types of temporal relations. Lin et al. (2016) performs training instance augmentation to increase the number of training examples and implement a SVM based model for containment relation extraction. Dligach et al. (2017) implement models based on CNNs and Long Short-Term Memory Networks (LSTMs) (Hochreiter and Schmidhuber, 1997) to extract containment relations from the THYME corpus.

From a more general perspective, relation extraction and classification is a task explored by many approaches, from fully unsupervised to fully supervised. Recent years have seen an increasing interest for the use of neural approaches.

Recursive neural networks (Socher et al., 2011, 2013) have proved useful for tasks involving long-distance relations, such as semantic relation extraction (Hashimoto et al., 2013; Li et al., 2015). Convolutional networks have also been used (dos Santos et al., 2015; Zeng et al., 2014) and more recently, recurrent networks such as LSTM showed to be more robust for learning long-distance semantic information (Miwa and Bansal, 2016; Xu et al., 2015; Zhou et al., 2016).

3 Data

3.1 Corpus Presentation

The THYME corpus is a collection of clinical texts written in English from a cancer department that have been released during the Clinical TempEval campaigns (Bethard et al., 2015, 2016). This corpus contains documents annotated with medical events and temporal expressions as well as narrative container relations.

According to the annotation guidelines of the THYME corpus, a medical event is anything that could be of interest on the patient's clinical timeline. It could be for instance a medical procedure, a disease or a diagnosis. There are five attributes given to each event: *Contextual Modality*, *Degree*, *Polarity*, *Type* and *DocTimeRel*.

Temporal expressions are assigned a *Class* attribute. Possible values for these attributes are presented in Table 2.

Narrative containers can be apprehended as temporal buckets in which several events may be included. These containers are anchored by temporal expressions, medical events or other concepts. Styler IV et al. (2014) argue that the use of narrative containers instead of classical temporal relations (Allen, 1983) yields better annotation while keeping most of the useful temporal information intact. The concept of narrative container is illustrated in Figure 1 and described further in Pustejovsky and Stubbs (2011).

We identify two types of narrative container relations: relations that stay within sentence boundaries ($\approx 75\%$ of the instances) and relations that spread over several sentences. In the rest of the paper, we will refer to them as intra- and inter-sentence relations. Descriptive statistics on the corpus are presented in Table 1.
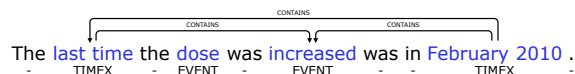

Figure 1: Examples of intra-sentence narrative container relations.

	Train	Test
EVENT	49,147	15,503
TIMEX3	5,791	1,917
Intra-sentence relations	12,855	4,365
Inter-sentence relations	4,582	1,565

Table 1: Descriptive statistics about the train and test parts of the THYME corpus.

3.2 Preprocessing

We preprocessed the corpus using cTAKES (Savova et al., 2010), an open-source natural language processing system for the extraction of information from electronic health records. We extracted sentence and token boundaries, as well as token types and semantic types of the entities that have a span overlap with a least one gold standard EVENT entity of the THYME corpus. Semantic types are a set of subject categories (organized as a tree) that are used to categorize concepts in the UMLS® (Unified Medical Language System) Metathesaurus. There are currently 135 types (Bodenreider, 2004). This information was added to the set of Gold Standard attributes available for EVENT entities in the corpus. An overview of the attributes available for each token is presented in Table 2.

4 Task Description

The container relation extraction task can be cast as a 3-class classification problem. For each combination of EVENT and/or TIMEX3 from left to right, three cases are possible:

- the first entity temporally *contains* the second entity,
- the first entity *is* temporally *contained* by the second entity,
- there is no temporal containment relation between the entities.

Intra- and inter-sentence relation detection can be seen as two different tasks with specific features. Intra-sentence relations can benefit from intra-sentential clues such as adverbs (e.g. *during*) or pronouns (e.g. *which*) which are not available at the inter-sentence level. Furthermore, past work on the topic seems to indicate that this differentia-

Source	Attribute	Values
Corpus	Contextual Modality	Actual, Hypothetical, Hedged, Generic or no-value
	Degree	Most, Little, N/A or no-value
	Polarity	Pos, Neg or no-value
	Type	Aspectual, Evidential, N/A or no-value
	DocTimeRel	Before, Before-Overlap, Overlap, After or no-value
	Entity	EVENT, TIMEX3 or no-entity
cTAKES	Entity Type[a]	DiseaseDisorderMention, LabMention, MedicationEventMention, MedicationMention, ProcedureMention, SignSymptomMention or no-value
	Semantic Type[a]	list of semantic types extracted from the training corpus or no-value

[a] If the token is not part of an EVENT entity, the value is automatically *no-value*.

Table 2: Attributes available for each token of the corpus after preprocessing.

tion improves overall performance (Tourille et al., 2016). We have adopted this approach by building two separate classifiers, one for intra-sentence relations and one for inter-sentence relations.

If we were to consider all combinations of entities within documents for inter-sentence relations, it would result in a very large training corpus with very few positive examples. In order to cope with this issue, we limit our experiments to inter-sentence relations that do not span over more than three sentences. By doing so, we obtain a manageable training corpus size with less unbalanced classes while keeping a good coverage. It results in 2,085 inter-sentences relations for the training corpus and 743 for the test corpus.

5 Neural Network Model

First, we present the main component of our model in Section 5.1. Then, we describe the word embeddings that we use as input for the model in Section 5.2. Finally, we present the parameters used for network training in Section 5.3.

5.1 Temporal Relation Extraction

Our approach relies on Long Short-Term Memory Networks (LSTMs) (Hochreiter and Schmidhuber, 1997). The architecture of our model is presented in Figure 2. For a given sequence of tokens separating two entities (EVENT and/or TIMEX3), represented as vectors, we compute a representation by going from left to right in the sequence (forward LSTM in figure 2).

As LSTMs tend to be biased toward the most recent inputs, this implementation would be biased toward the second entity of each pair processed by the network. To counteract this effect, we compute the reverse representation with an LSTM reading the sequence backwards, from right to left (backward LSTM in figure 2). By doing so, we keep as

much information as possible about the two entities.

The two final states are then concatenated and linearly transformed to a 3-dimensional vector representing the number of categories (concatenation and projection in figure 2). Finally, a softmax function is applied.
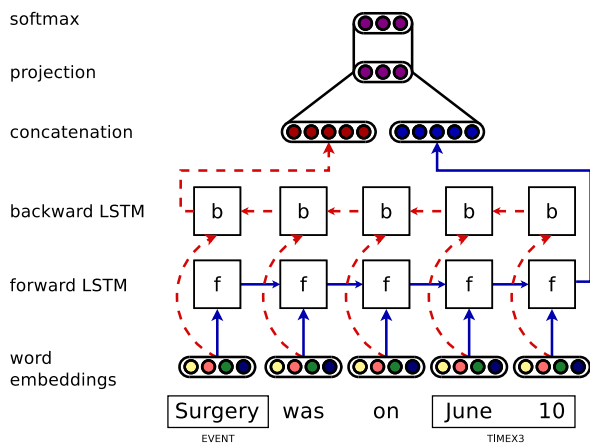

Figure 2: Neural architecture for containment relation extraction.

5.2 Input Embeddings

Vectors representing tokens are built by concatenating a character-based embedding, a word embedding, one embedding per Gold Standard attribute and one embedding per cTAKES attribute. While the word embedding is a classical option in the context of neural models, the embeddings for Gold Standard and cTAKES attributes are a way of integrating in such model features that have been demonstrated as useful in previous work. Finally, temporal clues such as verbs, and more particularly their tense, which are important in assessing if one entity temporally contains another, are taken into account by our character-based representation

226

	Intra-sentence classifier						Inter-sentence classifier					
	ref	pred	corr	P	R	F1	ref	pred	corr	P	R	F1
No Features	4365	4529	3035	0.670	0.681	0.675	743	895	377	0.421	**0.498**	0.456
+ GS	4365	4253	2980	**0.701**	0.661	0.680	743	692	349	**0.504**	0.462	**0.482**
+ cTAKES	4365	4780	3170	0.663	**0.704**	**0.683**	743	628	305	0.486	0.408	0.443

Table 3: Results obtained by the intra-sentence and inter-sentence classifiers for each model of this paper. We report the number of Gold Standard relations (ref), the number of relations predicted by our system (pred), the number of true positives (corr), the precision (P), the recall (R) and the F1-measure (F1).

of tokens.

An overview of the embedding computation is presented in Figure 3. Following Lample et al. (2016), the character-based representation is constructed with a Bi-LSTM. First, a random embedding is generated for every character present in the training corpus. Token characters are then processed with a forward and backward LSTM similar to the one we use in our general architecture. The final character-based representation is the result of the concatenation of the forward and backward representations.
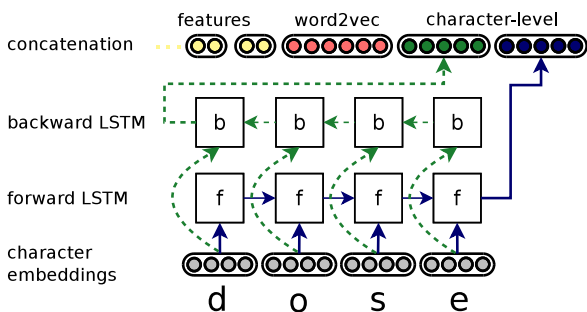

Figure 3: Neural architecture for input embeddings.

We use a character embedding size of 8 and hidden dimensions of 25 for the forward and backward LSTMs, resulting in a final representation size of 50 after concatenation. This representation is randomly initialized and incrementally defined by the training of the whole network.

For word embeddings, our vectors are pretrained by applying *word2vec* (Mikolov et al., 2013) on the Mimic 3 corpus (Johnson et al., 2016)[1]. In order to account for unknown tokens during the test phase, we train a special embedding UNK by replacing randomly some singletons with the UNK token (probability of replacement = 0.5).

In the inter-sentence relation classifier, we introduce a specific token for identifying sentence breaks. This token is composed of one distinctive character and it is associated to a specific word embedding.

Similarly to the character embeddings, we randomly initialize one embedding per token attribute value, with an embedding size of 4. All these embeddings are then concatenated in a final representation.

5.3 Network Training

We implemented the network using Tensor-Flow (Abadi et al., 2015). We trained our network with mini-batch Stochastic Gradient Descent using Adam (Kingma and Ba, 2014) with a batch-size of 256. The learning rate was set to 0.001. The hidden layers of our forward and backward LSTMs have a size of 512. We kept 10% of the training corpus for a development corpus and we implemented early stopping with a patience of 10 epochs without performance improvement. Finally, we used dropout training to avoid overfitting. We applied dropout on input embeddings with a rate of 0.5.

6 Experiments and Discussion

We experimented with three configurations. In the first one, we used only word embeddings and character embeddings. In the second one, we added the feature embeddings related to the Gold Standard (GS) attributes. Finally, in a third experiment, we added the feature embeddings related to cTAKES. For each experiment, we report precision (P), recall (R) and F1-measure (F1) computed with the official evaluation script[2] provided during the Clinical TempEval challenges. Results of the experiments are presented in Table 4. For comparison, we report the baseline provided as reference during the Clinical TempEval shared tasks,

---

[1]Parameters used during computation: algorithm = CBOW; min-count = 4; vector size = 100; window = 8.

[2]https://github.com/bethard/anaforatools

	P	R	F1
baseline (closest)	0.459	0.154	0.231
Lee et al. (2016)	0.588	0.559	0.573
Lin et al. (2016)	0.669	0.534	0.594
No features	0.646	0.568	0.605
+ GS features	**0.687**	0.549	0.610
+ cTAKES features	0.657	**0.575**	**0.613**

Table 4: Experimentation results. We report precision (P), recall (R) and F1-measure (F1) for each configuration of our model, for the best system of the Clinical TempEval 2016 challenge (Lee et al., 2016) and for the best result obtained so far on the corpus (Lin et al., 2016).

the results of the best system of the Clinical TempEval 2016 challenge (Lee et al., 2016) and the best scores obtained after the challenge (Lin et al., 2016) on the test portion of the corpus. Both Lee et al. (2016) and Lin et al. (2016) rely on SVM classifiers using hand-engineered linguistic features.

All three of our models perform better in terms of F1-measure than Lee et al. (2016) and Lin et al. (2016). Our two best models also outperform Leeuwenberg and Moens (2017), who report an F-measure of .608 using a structured perceptron. Interestingly, their model did not distinguish between intra- and inter- sentence relations, but instead considered that related entities had to occur within a window of 30 tokens. We see that the addition of attribute embeddings slightly improves the overall performance of our system (+0.008). Adding the embeddings of GS features contributes to the major part of this improvement but tends to increase the imbalance between recall and precision. On the contrary, while the attribute embeddings related to cTAKES seem to have little impact on the overall performance, they tend to restore more balanced precision and recall.

The results for respectively intra- and inter-sentence relations are presented in Table 3. Similarly to our global results, the intra-sentence classifier benefits from the addition of feature embeddings with a small increase for GS features and only a very little improvement for cTAKES features.

The inter-sentence classifier exhibits the same trend: GS features do improve the performance. However, adding cTAKES features degrades it slightly (-0.013).

The closest work compared to ours is clearly Dligach et al. (2017) as it also heavily relies on neural models for extracting temporal containment relations between medical events. Dligach et al. (2017) tested both CNN and LSTM models and found CNN superior to LSTM. However, this work addressed intra-sentence relations only. Moreover, its LSTM model was not a Bi-LSTM model as ours and it did not include character-based or attribute embeddings. Finally, it distinguished EVENT-TIMEX3 and EVENT-EVENT relations while we have only one model for the two types of relations.

7   Conclusion and Perspectives

From a global perspective, the work we have presented in this article shows that in accordance with a more general trend, our neural model for extracting containment relations clearly outperforms classical approaches based on feature engineering. However, it also shows that incorporating classical features in such a model is a way to improve it, even if all kinds of features do not contribute equally to such improvement. A more fine-grained study has now to be performed to determine the most meaningful features in this perspective and to measure the contribution of each feature to the overall performance, with a specific emphasis on character-based embeddings.

Beyond a further analysis of the characteristics of our model, we are interested in two main extensions. The first one will investigate whether training two models, one for EVENT-TIMEX3 relations and one for EVENT-EVENT relations, as done by Dligach et al. (2017), is a better option than training one model for all types of containment relations as presented herein. The second extension consists in transposing the model we have defined in this work for English to French, as done by Tourille et al. (2017) for a more traditional approach based on a feature engineering approach.

Acknowledgements

References

Abdulrahman AAl Abdulsalam, Sumithra Velupillai, and Stephane Meystre. 2016. UtahBMI at SemEval-2016 Task 12: Extracting Temporal Information from Clinical Text. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 1256–1262.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org. http://tensorflow.org/.

James F. Allen. 1983. Maintaining Knowledge About Temporal Intervals. *Communications of the ACM* 26(11):832–843.

Steven Bethard, Leon Derczynski, Guergana Savova, James Pustejovsky, and Marc Verhagen. 2015. SemEval-2015 Task 6: Clinical TempEval. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Association for Computational Linguistics, Denver, USA, pages 806–814.

Steven Bethard, Guergana Savova, Wei-Te Chen, Leon Derczynski, James Pustejovsky, and Marc Verhagen. 2016. SemEval-2016 Task 12: Clinical TempEval. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 1052–1062.

Olivier Bodenreider. 2004. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research* 32:267–270.

Veera Raghavendra Chikka. 2016. CDE-IIITH at SemEval-2016 Task 12: Extraction of Temporal Information from Clinical documents using Machine Learning techniques. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 1237–1240.

Arman Cohan, Kevin Meurer, and Nazli Goharian. 2016. GUIR at SemEval-2016 task 12: Temporal Information Processing for Clinical Narratives. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 1248–1255.

Dmitriy Dligach, Timothy Miller, Chen Lin, Steven Bethard, and Guergana Savova. 2017. Neural temporal relation extraction. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, Valencia, Spain, pages 746–751.

Cicero dos Santos, Bing Xiang, and Bowen Zhou. 2015. Classifying Relations by Ranking with Convolutional Neural Networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Beijing, China, pages 626–634.

Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka, and Takashi Chikayama. 2013. Simple Customization of Recursive Neural Networks for Semantic Relation Classification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Seattle, Washington, USA, pages 1372–1376.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Liwei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo A. Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980. http://arxiv.org/abs/1412.6980.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural Architectures for Named Entity Recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, pages 260–270.

Hee-Jin Lee, Hua Xu, Jingqi Wang, Yaoyun Zhang, Sungrim Moon, Jun Xu, and Yonghui Wu. 2016. 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)*. Association for Computational Linguistics, San Diego, California, pages 1292–1297.

Artuur Leeuwenberg and Marie-Francine Moens. 2017. Structured Learning for Temporal Relation Extraction from Clinical Records. In *Proceedings*

of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. Association for Computational Linguistics, Valencia, Spain, pages 1150–1158.

Jiwei Li, Thang Luong, Dan Jurafsky, and Eduard Hovy. 2015. When Are Tree Structures Necessary for Deep Learning of Representations? In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 2304–2314.

Chen Lin, Timothy Miller, Dimitry Dligach, Steven Bethard, and Guergana Savova. 2016. Improving Temporal Relation Extraction with Training Instance Augmentation. In *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, Berlin, Germany, pages 108–113.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR* abs/1301.3781.

Makoto Miwa and Mohit Bansal. 2016. End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, pages 1105–1116.

James Pustejovsky and Amber Stubbs. 2011. Increasing Informativeness in Temporal Annotation. In *Proceedings of the 5th Linguistic Annotation Workshop*. Association for Computational Linguistics, Stroudsburg, PA, USA, LAW V '11, pages 152–160.

Guergana K. Savova, James J. Masanz, Philip V. Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C. Kipper-Schuler, and Christopher G. Chute. 2010. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association* 17(5):507–513.

Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. 2011. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Edinburgh, Scotland, UK., pages 151–161.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Seattle, Washington, USA, pages 1631–1642.

William Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, and James Pustejovsky. 2014. Temporal Annotation in the Clinical Domain. *Transactions of the Association for Computational Linguistics* 2:143–154.

Julien Tourille, Olivier Ferret, Aurélie Névéol, and Xavier Tannier. 2016. LIMSI-COT at SemEval-2016 Task 12: Temporal relation identification using a pipeline of classifiers. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 1136–1142.

Julien Tourille, Olivier Ferret, Xavier Tannier, and Aurélie Névéol. 2017. Temporal information extraction from clinical text. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017), short paper session*. Valencia, Spain, pages 739–745.

Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. 2015. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 1785–1794.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation Classification via Convolutional Deep Neural Network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, pages 2335–2344.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Berlin, Germany, pages 207–212.