CDE-IIITH at SemEval-2016 Task 12: Extraction of Temporal Information from Clinical documents using Machine Learning techniques

Veera Raghavendra Chikka
Center For Data Engineering
International Institute of Information Technology
Hyderabad, India
raghavendra.ch@research.iiit.ac.in

Abstract

In this paper, we demonstrate our approach for identification of events, time expressions and temporal relations among them. This work was carried out as part of SemEval-2016 Challenge Task 12: Clinical TempEval. The task comprises six sub-tasks: identification of event spans, time spans and their attributes, document time relation and the narrative container relations among events and time expressions. We have participated in all six sub-tasks. We have provided with a manually annotated dataset which comprises of training dataset (293 documents), development dataset (147 documents) and 151 documents as test dataset. We have submitted our work as two systems for the challenge. One system is developed using machine learning techniques, Conditional Random Fields (CRF) and Support Vector machines (SVM) and the other system is developed using deep neural network (DNN) techniques. The results show that both systems have given relatively same performance on these tasks.

1 Introduction

The interest on extracting temporal information is well versed from the creation of time bank corpus (Pustejovsky et al., 2003b) in 2003. A specification language has been developed, TimeML (Markup Language for Temporal and Event Expressions) (Pustejovsky et al., 2003a) to conceptualize the events, time expressions and temporal relations using tags (EVENT, TIMEX, TLINKS, ALINKS, SLINKS). Various algorithms have been developed to tag events and time expressions on time bank corpus. Initial works used machine learning algorithms with manually extracted features (Mani et al., 2006), syntax and clausal features on inter-sentential events. Later, automated feature selection were used for extracting events and finding the temporal relation(Chambers et al., 2007; Lapata and Lascarides, 2006). TARSQI (Verhagen and Pustejovsky, 2008) is a project employed by the creators of TimeML to develop algorithms for tagging these tags in text.

A series of challenges have been organized on TempEval comprising evaluation tasks on events, time expressions and temporal relations on news data. I2b2 2012 (Sun et al., 2013) is the first conference to study the temporal information extraction in clinical domain using THYME corpus. It is followed by Clinical TempEval tasks in semEval2015. Most of the participants of these challenges used CRF and SVM for event extraction with features including the information gathered from different resources like UMLS (Unified Medical Language System), output of TARSQI toolkit, Brown Clustering, Wikipedia and Metamap (Aronson and Lang, 2010). For time expression extraction, various existing tools like SUTIME (Chang and Manning, 2012), HeidelTime (Strötgen and Gertz, 2010) and GUTIME were used. And for temporal relation extraction various machine learning methods ranging from MaxEnt, Bayesian and SVM to CRF were used incorporating the heuristics and rule based components. Out of many participants of these challenges/workshops, the top performing systems used hybrid approaches with machine learning techniques and rule based tools.

1128

2 Methods

The SemEval 2016 Clinical TempEval challenge (Bethard et al., 2016) is on identification of event spans (ES), time spans (TS) and their attirbutes (EA and TA), document time relation (DR) and narrative container relations (CR) among events and time expressions. In this paper, we describe two approaches using machine learning techniques for these tasks. First, we broadly classify the tasks into three tasks:

1. Sequence labeling tasks: These tasks involves tagging the sequence of words with the output tags. For example, tasks like part of speech tagging (POS), Named Entity recognition (NER) are sequence labeling tasks.

2. Classification tasks: Classification tasks focuses on classifying the entities into one of the output classes.

3. Relation Extraction Tasks : These tasks involves extracting relation of the entities with the other entities. That is, identifying temporal relations among event/time expressions.

In the following sections, we group the SemEval tasks into one of the above three tasks and provide the methodology dealt with each of these tasks.

Table 1: Event and Time attributes and their values

Attributes	Values
EVENT:Modality	ACTUAL, HEDGED, HYPOTHETICAL or GENERIC
EVENT:Degree	N/A, MOST or LITTLE
EVENT:Polarity	POS or NEG
EVENT:Type	N/A, ASPECTUAL, EVIDENTIAL
TIMEX:Class	DATE, TIME, DURATION, QUANTIFIER, PREPOSTEXP or SET

Identification of events spans (ES) and Time Spans (TS)

Identifying events spans and time spans comes under sequence labeling. We use Conditional Random Field (CRF) which is a popularly used probabilistic graphical model for sequence labelling to extract event spans and time expressions. We use CRF++ suite (CRFPP,) tool for training conditional

random field model with the features:

Term feature: The word itself and its stem are used as features.

POS and Chunk tags: Parts of speech and chunk tags of the word. OpenNLP tagger is used for POS and Chunk tagging (Baldridge, 2005).

Orthographic features: Orthographic features like AlphaNumeric, IsNumeral, isUpperCase, startsWithUpperCase, etc.

Stopword: We use our custom English and Medical stopword list to tag this binary feature.

Train Events Dictionary: Dictionary of events build from the training dataset is used as a feature to check an event has already occurred in training dataset.

In addition to the above features, HeidelTime and HeidelTime Class are used as features for the identification of time expression.

Identifying Events Attributes (EA) and Time Attributes (TA)

Table 1 shows the Event and Time attributes and their classes. Assigning these attributes to one of its values is an classification task. We train a separate Support Vector Machine (SVM) (Chang and Lin, 2011) for each of the attributes to classify in to their respective classes. We use word representations or word embeddings as the features for training the SVM classifier. The word representations are generated based on the co-occurrence count modeling using Stanford Glove tool (Pennington et al., 2014). We trained this count-based model on a text window size of 5, to obtain words vector representations of dimension 25. These word representations are used to train separate SVM classifier for each attribute.

Identify Document Time Relation (DR) and Narrative Container Relation (CR)

Document Time Relation (DR) is the temporal relation of the entities with respect to the document create time. The document relations can take BEFORE, OVERLAP, BEFORE-OVERLAP or AFTER temporal relations.

Temporal links are used to identify the temporal ordering of the events in the timeline. These links are only provided for the events that happen within a temporal bucket, called narrative container, to avoid

Table 2: Phase 1 Evaluation on events, times and temporal relations

Task	Approach 1 (CRF and SVM)			Approach 2 (DNN)		
	Precision	Recall	F-Score	Precision	Recall	F-Score
EVENT:<span> (ES)	0.835	0.797	0.815	0.838	0.786	0.811
EVENT:Modality	0.764	0.729	0.746	0.779	0.731	0.754
EVENT:Degree	0.830	0.793	0.811	0.834	0.783	0.807
EVENT:Polarity	0.750	0.716	0.733	0.813	0.764	0.788
EVENT:Type	0.806	0.769	0.787	0.814	0.765	0.789
TIMEX3:<span> (TS)	0.752	0.515	0.612	0.614	0.560	0.586
TIMEX3:Class	0.644	0.439	0.522	0.468	0.426	0.446
EVENT:DocTimeRel (DR)	0.481	0.460	0.470	0.643	0.604	0.623
TLINK:Type (CR)	0.431	0.167	0.241	0.285	0.225	0.252

Table 3: Phase 2 Evaluation on temporal relations

Task	Approach1 (CRF and SVM)			Approach2 (DNN)		
	Precision	Recall	F-Score	Precision	Recall	F-Score
EVENT:<span>	0.935	0.912	0.923	0.994	0.994	0.994
EVENT:DocTimeRel (DR)	0.724	0.705	0.714	0.588	0.588	0.588
TIMEX3:<span>	0.965	0.794	0.871	0.999	0.999	0.999
TLINK:Type (CR)	0.348	0.284	0.313	0.493	0.185	0.269

the heap of links between all possible events in the document (Styler IV et al., 2014). In an example sentence, "When *compared* with *ECG* of yesterday no significant change is found.". Here, temporal link "BEFORE" is used to notify the event "*ECG*" is occurred BEFORE the event "*compared*". The temporal relations BEFORE, OVERLAP, BEGINS-ON and CONTAINS are used for Narrative Container Relation (CR). We train CRF model similar to that of event span (ES) model to identify DR and CR.

Another Approach using Deep Neural Networks (DNN)

Recent advances in deep learning architecture made us to try an another approach for this challenge. We have used deep neural networks (DNN) for all of the six sub-tasks of the challenge. Given a input sentence and output tags, the neural network learns the weights of nodes of each word of the sentence. The input words are represented as word embeddings which are same as that of word representations used for SVM classifier. A separate neural network is trained for each of the tasks using *deepnl* (Attardi, 2015) library. This neural network architecture follows the convolution method used for natural language processing (Collobert et al., 2011).

3 Results

The dataset used for the challenge comprises of de-identified cancer patient records from the Mayo Clinic with train dataset (195 clinical notes and 98 pathology reports), development dataset (98 clinical notes and 49 pathology reports) and test dataset (100 clinical notes and 51 pathology reports). All of the above described machine learning models are trained on the train and development dataset (340 documents) of THYME-corpus. The evaluation of the task is carried out in two phases. In the first phase, only plain clinical documents are provided. In the second phase, events and time expressions of clinical document are provided and the task is to find the document creation relation and narrative container relation. The results on extracting event, attributes and temporal relations on the test dataset (151 documents) are given in Table 2. Even though the values of DocTime Relation and Container relation are relatively less, they are comparable with the top performing system of SemEval challenge (Bethard et al., 2015). As the phase 2 of evaluation, we are provided with events, time expressions and their attributes, the results on extracting temporal relations are shown in Table 3.

4   Conclusion

In this paper, we present our work on Clinical TemEval task of SemEval 2016 challenge. We have used two approaches, first approach is based on CRF and SVM, and the second approach uses deep neural network to solve the tasks of the challenge. The results show that both approaches relatively same performance on the provided train and test datasets of the challenge.

References

Alan R Aronson and François-Michel Lang. 2010. An overview of metamap: historical perspective and recent advances. *Journal of the American Medical Informatics Association*, 17(3):229–236.

Giuseppe Attardi. 2015. Deepnl: a deep learning nlp pipeline. In *Proceedings of NAACL-HLT*.

Jason Baldridge. 2005. The opennlp project. *URL: http://opennlp. apache. org/index. html,(accessed 2 February 2012)*.

Steven Bethard, Leon Derczynski, 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*.

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)*, San Diego, California, June. Association for Computational Linguistics.

Nathanael Chambers, Shan Wang, and Dan Jurafsky. 2007. Classifying temporal relations between events. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 173–176. Association for Computational Linguistics.

Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*.

Angel X Chang and Christopher D Manning. 2012. Sutime: A library for recognizing and normalizing time expressions. In *LREC*, pages 3735–3740.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*.

CRFPP. Crf++: Yet another crf toolkit.

Maria Lapata and Alex Lascarides. 2006. Learning sentence-internal temporal relations. *J. Artif. Intell. Res.(JAIR)*, 27:85–117.

I. Mani, M. Verhagen, B. Wellner, C.M. Lee, and J. Pustejovsky. 2006. Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics*, page 760. ACL.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*, 12:1532–1543.

James Pustejovsky, José M Castano, Robert Ingria, Roser Sauri, Robert J Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R Radev. 2003a. Timeml: Robust specification of event and temporal expressions in text. *New directions in question answering*, 3:28–34.

James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. 2003b. The timebank corpus. In *Corpus linguistics*, volume 2003, page 40.

Jannik Strötgen and Michael Gertz. 2010. Heideltime: High quality rule-based extraction and normalization of temporal expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 321–324. Association for Computational Linguistics.

William F Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, et al. 2014. Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2:143–154.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *Journal of the American Medical Informatics Association*, 20(5):806–813.

Marc Verhagen and James Pustejovsky. 2008. Temporal processing with the tarsqi toolkit. In *22nd International Conference on on Computational Linguistics: Demonstration Papers*, pages 189–192. Association for Computational Linguistics.