LIMSI-COT at SemEval-2017 Task 12
Neural Architecture for Temporal Information Extraction from Clinical Narratives

Julien Tourille[1,2,3], Olivier Ferret[4], Xavier Tannier[1,2,3], Aurélie Névéol[1,3]

[1]LIMSI, CNRS
[2]Univ. Paris-Sud
[3]Université Paris-Saclay
[4]CEA, LIST, Gif-sur-Yvette, F-91191 France

SemEval Workshop 2017 - Vancouver - August 4, 2017

Overview

1. Task Description
2. Medical Event and Temporal Expression Extraction
3. Attribute and DCT Relation Classification
4. Containment Relation Extraction
5. Input Embedding
6. Domain Adaptation
7. Results

1. Task Description

2. Medical Event and Temporal Expression Extraction

3. Attribute and DCT Relation Classification

4. Containment Relation Extraction

5. Input Embedding

6. Domain Adaptation

7. Results

Subtasks and Phases

Subtasks

1. **Medical event (EVENT) and temporal expression (TIMEX3) extraction**
   - spans + attributes (*type*, *modality*, *degree*, *polarity* for EVENTs and *class* for TIMEX3s)
2. **Temporal relation extraction**
   - **Document Creation Time** (DCT) **relations** between EVENTs and documents (*Before*, *Before-Overlap*, *Overlap*, *After*)
   - **Containment relation** extraction between EVENTs and/or TIMEX3s

Phases (domain adaptation)

1. Unsupervised: **source and target domain are different**
2. Semi-supervised: **a little training data** is given for the target domain

Subtasks and Phases

Subtasks

1. **Medical event (EVENT) and temporal expression (TIMEX3) extraction**
   - spans + attributes (*type*, *modality*, *degree*, *polarity* for EVENTs and *class* for TIMEX3s)
2. **Temporal relation extraction**
   - **Document Creation Time** (DCT) **relations** between EVENTs and documents (*Before*, *Before-Overlap*, *Overlap*, *After*)
   - **Containment relation** extraction between EVENTs and/or TIMEX3s

Phases (domain adaptation)

1. Unsupervised: **source and target domain are different**
2. Semi-supervised: **a little training data** is given for the target domain

1. Task Description

2. Medical Event and Temporal Expression Extraction

3. Attribute and DCT Relation Classification

4. Containment Relation Extraction

5. Input Embedding

6. Domain Adaptation

7. Results

Methodology

- **Named Entity Recognition** task (EVENT and TIMEX3 entities)
- **Sequence labeling problem** where each token of a given sentence is assigned a label.
- **IOB format** (**I**nside, **O**utside, **B**eginning)
  - *Type* attribute for medical events (3 types → 7 IOB labels)
  - *Class* attribute for temporal expressions (6 classes → 13 IOB labels)
- **Two separate models** for EVENT and TIMEX3 entities

The	**last**	**time**	the	**dose**	was	**increased**	was	in	**February**	**2010**	.
O	B-Date	I-Date	O	B-N/A	O	B-N/A	O	O	B-Date	I-Date	O

Network Architecture

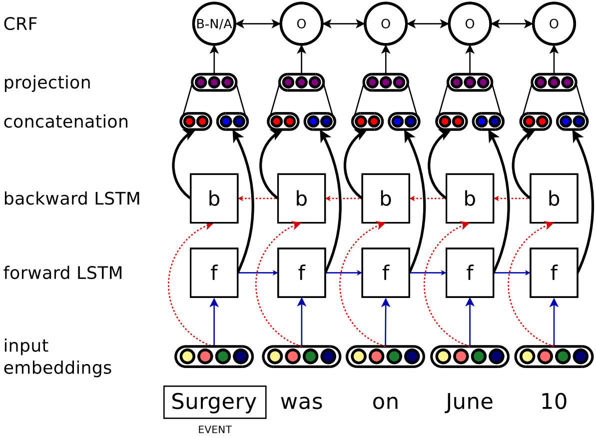

Figure: EVENT and TIMEX3 extraction - BiLSTM-CRF architecture

1. Task Description

2. Medical Event and Temporal Expression Extraction

3. Attribute and DCT Relation Classification

4. Containment Relation Extraction

5. Input Embedding

6. Domain Adaptation

7. Results

Attribute Classification - Pipeline Structure

- **Attribute** (*modality*, *degree*, *polarity*, ~~*type*~~, ~~*class*~~) and **DCT relation** extraction seen as **supervised classification problems**
- **Common pipeline structure**: optimization via cross-validation of *C*, *number of features to keep* and *windows around entities* via a Tree-structured Parzen Estimator approach (Bergstra et al. 2011)
- Lexical, contextual and structural features

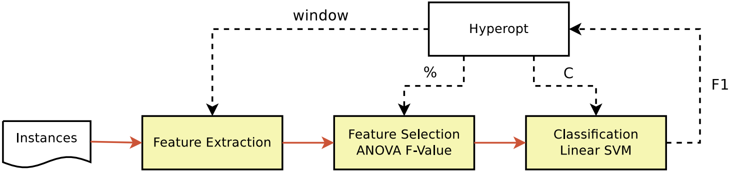

Figure: Attribute classification - Optimization Pipeline

1. Task Description

2. Medical Event and Temporal Expression Extraction

3. Attribute and DCT Relation Classification

4. Containment Relation Extraction

5. Input Embedding

6. Domain Adaptation

7. Results

Methodology

- **Relation extraction between two entities** e1 and e2 (left to right)
  - e1 temporally *contains* e2
  - e1 *is* temporally *contained* by e2
  - there is *no temporal containment relation* between e1 and e2
- **One model** for EVENT-EVENT, EVENT-TIMEX3 and TIMEX3-TIMEX3 relations
- **Two models** for **within**- and **cross-sentence relations**

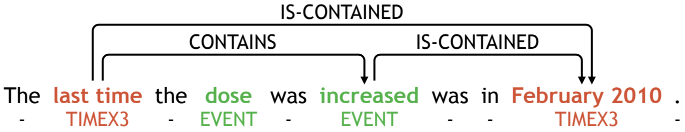

Figure: Within-sentence relation example

Network Architecture (Tourille et al. 2017)

The **[last time the dose was increased]** *was in February 2010*

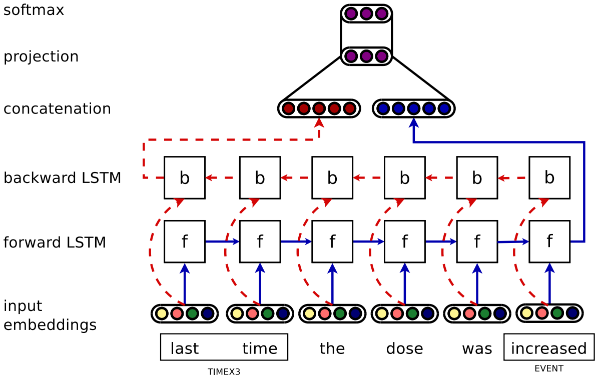

Figure: Containment relation extraction - System architecture

1. Task Description

2. Medical Event and Temporal Expression Extraction

3. Attribute and DCT Relation Classification

4. Containment Relation Extraction

5. Input Embedding

6. Domain Adaptation

7. Results

Feeding the Networks

Preprocessing

- **cTAKES** (Savova et al. 2010): Sentence and token boundaries, token type, semantic type
- **HeidelTime** (Strötgen et al. 2015)

Concatenation

- a **character-based embedding**
- a pre-computed **word embedding** on MIMIC-III corpus (Johnson et al. 2016)
- **one embedding per gold standard attribute** (DCT and CONTAINS subtasks)
- **one embedding per cTAKES and HeidelTime attribute** (EVENT and TIMEX3 subtasks)

Network Architecture

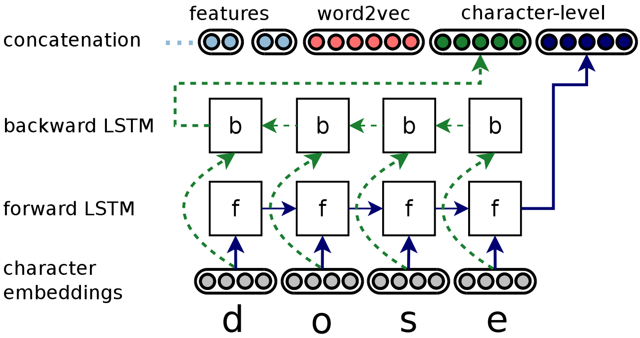

Figure: Input embeddings - System architecture

1. Task Description

2. Medical Event and Temporal Expression Extraction

3. Attribute and DCT Relation Classification

4. Containment Relation Extraction

5. Input Embedding

6. Domain Adaptation

7. Results

Strategies

Phase 1 (no target domain data)

1. **STATIC** - **Block further tuning** of the word embeddings during network training
2. **REPLACE** - Randomly replace tokens from event entities with the 'unknown' token → force the network to use context

→ No strategy applied for TIMEX3

Phase 2 (a little target domain data)

1. **ALL** - **Mix both domain and target data** with stratified randomization
2. **30-30** - Match the number of target domain documents with an equal number of source domain documents → Balanced training corpus

Strategies

Phase 1 (no target domain data)

1. **STATIC - Block further tuning** of the word embeddings during network training
2. **REPLACE** - Randomly replace tokens from event entities with the 'unknown' token → force the network to use context

→ No strategy applied for TIMEX3

Phase 2 (a little target domain data)

1. **ALL - Mix both domain and target data** with stratified randomization
2. **30-30** - Match the number of target domain documents with an equal number of source domain documents → Balanced training corpus

1. Task Description

2. Medical Event and Temporal Expression Extraction

3. Attribute and DCT Relation Classification

4. Containment Relation Extraction

5. Input Embedding

6. Domain Adaptation

7. Results

Performance on Test Set - Phase 1

	STATIC			REPLACE			'17[a]	'16[b]
	P	R	F1	P	R	F1	F1	F1
EVENT Span	.62*	.84*	.72*	.61	.84	.71	.72	.90
EVENT Modality	.55*	.75*	.64*	.54	.75	.62	.64	.86
EVENT Degree	.62*	.83*	.71*	.60	.83	.70	.71	.90
EVENT Polarity	.60*	.82*	.69*	.59	.82	.68	.69	.89
EVENT Type	.61*	.82*	.70*	.59	.82	.69	.70	.88
TIMEX3 Span	.42	.66	.51	-	-	-	.57	.80
TIMEX3 Class	.40	.63	.49	-	-	-	.53	.77
DCT Relation	.44*	.60	.51*	.44	.60*	.51	.51	.76
CONTAINS	.28*	.40	.33*	.26	.41*	.32	.34	.48

[a] Best performance for the phase (SemEval 2017)
[b] Best performance during Clinical TempEval 2016 (same domain for train and test)
* Best performance between STATIC and REPLACE strategies

Table: Results obtained by our system for phase 1

Performance on Test Set - Phase 2

	ALL			30-30			'17[a]	'16[b]	Φ_1[c]
	P	R	F1	P	R	F1	F1	F1	F1
EVENT Span	.69*	.85	.76*	.66	.87*	.75	.76	.90	.72
EVENT Modality	.63*	.78	.70*	.60	.78*	.68	.69	.86	.64
EVENT Degree	.68*	.84	.75*	.65	.85*	.74	.75	.90	.71
EVENT Polarity	.68*	.84	.75*	.64	.84*	.73	.75	.89	.69
EVENT Type	.68*	.83	.75*	.64	.84*	.73	.75	.88	.70
TIMEX3 Span	.51*	.67*	.58*	.45	.62	.52	.59	.80	.49
TIMEX3 Class	.49*	.64*	.55*	.43	.59	.50	.56	.77	.51
DCT Relation	.54*	.66	.59*	.51	.67*	.58	.59	.76	.51
CONTAINS	.24*	.44*	.32*	.21	.42	.28	.32	.48	.33

[a] Best performance for the phase
[b] Best performance during Clinical TempEval 2016 (same domain for train and test)
[c] Best score obtained by our system during phase 1
* Best performance between ALL and 30-30 strategies

Table: Results obtained by our system for phase 2

Performance on Test Set - Phase 2

	ALL			30-30			'17[a]	'16[b]	Φ_1[c]
	P	R	F1	P	R	F1	F1	F1	F1
EVENT Span	.69*	.85	.76*	.66	.87*	.75	.76	.90	.72
EVENT Modality	.63*	.78	.70*	.60	.78*	.68	.69	.86	.64
EVENT Degree	.68*	.84	.75*	.65	.85*	.74	.75	.90	.71
EVENT Polarity	.68*	.84	.75*	.64	.84*	.73	.75	.89	.69
EVENT Type	.68*	.83	.75*	.64	.84*	.73	.75	.88	.70
TIMEX3 Span	.51*	.67*	.58*	.45	.62	.52	.59	.80	.49
TIMEX3 Class	.49*	.64*	.55*	.43	.59	.50	.56	.77	.51
DCT Relation	.54*	.66	.59*	.51	.67*	.58	.59	.76	.51
CONTAINS	.24*	.44*	.32*	.21	.42	.28	.32	.48	.33

[a] Best performance for the phase
[b] Best performance during Clinical TempEval 2016 (same domain for train and test)
[c] Best score obtained by our system during phase 1
* Best performance between ALL and 30-30 strategies

Table: Results obtained by our system for phase 2

Conclusion

Present work

- Our models perform well for most tasks except for TIMEX3 extraction
- Training with no target domain data is hard
- Little target domain data significantly improves performance
- More data seem to work better than having a balanced dataset

Future work

- Take into account left and right contexts $\rightarrow$ Done !
- Joint learning of entities and relations
- Predict coherent temporal graphs
- Switch to neural architecture for attribute classification

Thank You

LIMSI-COT at SemEval-2017 Task 12
Neural Architecture for Temporal Information Extraction from Clinical Narratives

Julien Tourille[1,2,3], Olivier Ferret[4], Xavier Tannier[1,2,3], Aurélie Névéol[1,3]

[1] LIMSI, CNRS
[2] Univ. Paris-Sud
[3] Université Paris-Saclay
[4] CEA, LIST, Gif-sur-Yvette, F-91191 France

James S. Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. "Algorithms for Hyper-Parameter Optimization". In: *Advances in Neural Information Processing Systems 24*. Ed. by J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger. Curran Associates, Inc., 2011, pp. 2546–2554.

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

Guergana K. Savova, James J Masanz, Philip V. Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C. Kipper-Schuler, and Christopher G. Chute. "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications". In: *Journal of the American Medical Informatics Association* 17.5 (2010), pp. 507–513.

Jannik Strötgen and Michael Gertz. "A Baseline Temporal Tagger for all Languages". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sept. 2015, pp. 541–547.

Julien Tourille, Olivier Ferret, Xavier Tannier, and Aurélie Névéol. "Neural Architecture for Temporal Relation Extraction: A Bi-LSTM Approach for Detecting Narrative Containers". In: *Proceedings of the 55th Conference of the Association for Computational Linguistics*. Vancouver, Canada: Association for Computational Linguistics, Aug. 2017.