

Annotating Inter-Sentence Temporal Relations in Clinical Notes

Jennifer D'Souza and Vincent Ng
Human Language Technology Research Institute
The University of Texas at Dallas

Temporal Relation Extraction

- ⦿ Given two entities (i.e. **events** or **time expressions**) in a text document,
 - Determine whether a temporal relation exists between them (**temporal relation identification**)
 - If so, classify the pair as one of a set of predefined temporal relations (**temporal relation classification**)

Temporal Relation Extraction

- Given two entities (i.e. **events** or **time expressions**) in a text document,
 - Determine whether a temporal relation exists between them (**temporal relation identification**)
 - If so, classify the pair as one of a set of predefined temporal relations (**temporal relation classification**)

She had **a normal pancreas** at **that time** , however , **hyperdense kidneys** .

Temporal Relation Extraction

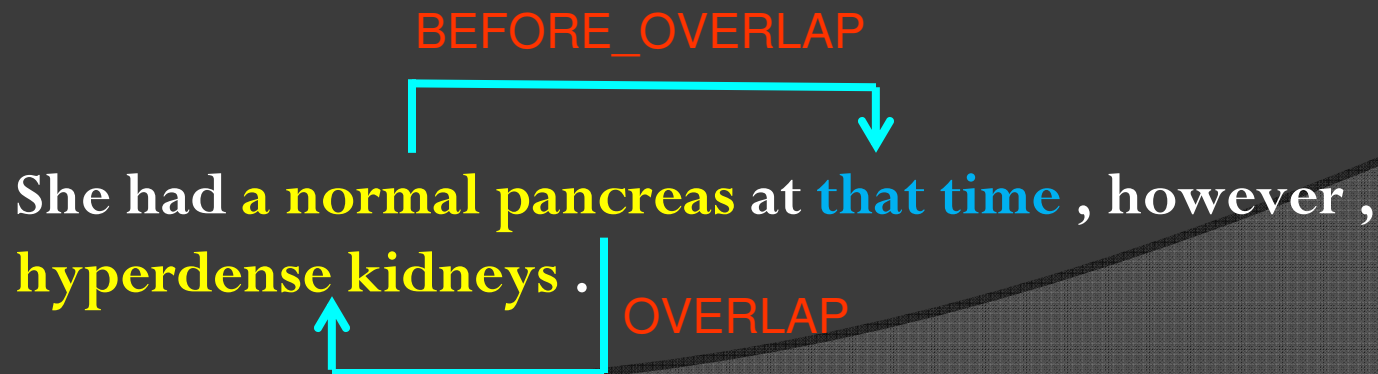
- Given two entities (i.e. **events** or **time expressions**) in a text document,
 - Determine whether a temporal relation exists between them (**temporal relation identification**)
 - If so, classify the pair as one of a set of predefined temporal relations (**temporal relation classification**)

She had **a normal pancreas** at **that time** , however ,
hyperdense kidneys .

The diagram illustrates the process of identifying a temporal relation between two entities in a sentence. Two blue arrows, each labeled 'YES' in green, point from the entities 'a normal pancreas' and 'hyperdense kidneys' to each other, indicating that a temporal relation exists between them.

Temporal Relation Extraction

- Given two entities (i.e. **events** or **time expressions**) in a text document,
 - Determine whether a temporal relation exists between them (**temporal relation identification**)
 - If so, classify the pair as one of a set of predefined temporal relations (**temporal relation classification**)



The Missing Annotation Problem

- ◎ Many temporal relations that exist between entities in a document are not being annotated
 - due to **annotator fatigue** (Verhagen 2005; Verhagen et al., 2006; UzZaman et al., 2012)

The Missing Annotation Problem

- Many temporal relations that exist between entities in a document are not being annotated
 - due to **annotator fatigue** (Verhagen 2005; Verhagen et al., 2006; UzZaman et al., 2012)
 - If there are N entities in a document, $O(N^2)$ pairs of entities need to be annotated

The Missing Annotation Problem

- The presence of missing links presents problems to both the training and evaluation of temporal relation extraction systems

The Missing Annotation Problem

- ◎ The presence of missing links presents problems to both the training and evaluation of temporal relation extraction systems
 - Inaccurate estimation of system performance
 - Systems trained on instances with incorrect labels
 - In temporal relation identification, we will have to posit an **unannotated pair** of entities as having **no temporal relation** even if a relation exists

Goals

- Manually annotate certain types of missing temporal links in the i2b2 Clinical Temporal Relations Challenge Corpus
- Empirically determine their impact on the performance of a temporal relation extraction system

Plan for the Talk

- ⦿ Corpus and its annotations
- ⦿ Annotation procedure and results
- ⦿ Temporal relation extraction results

Plan for the Talk

- ⦿ Corpus and its annotations
- ⦿ Annotation procedure and results
- ⦿ Temporal relation extraction results

Corpus

- ⦿ Used in the 2012 i2b2 Challenge (shared task on clinical temporal relation extraction)
- ⦿ 310 de-identified discharge summaries
 - 190 summaries for training; 120 for testing
 - Each summary is annotated with
 - **event** expressions
 - **time** expressions
 - temporal relations between an **event-event** pair and an **event-time** pair

Our work: focus on **event-event** pairs

Events

- ⦿ An event
 - refers to a clinically relevant patient-related action
 - belongs to one of six event types

Event Types	Examples
TEST	CT scan
PROBLEM	the tumor
TREATMENT	operation
CLINICAL DEPARTMENTS	ICU
OCCURRENCE	discharge
EVIDENTIAL information	complained

Events

- An event
 - refers to a clinically relevant patient-related action
 - belongs to one of six event types

Event Types	Examples
TEST	CT scan
PROBLEM	the tumor
TREATMENT	operation
CLINICAL DEPARTMENTS	ICU
OCCURRENCE	discharge
EVIDENTIAL information	complained

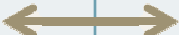
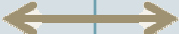
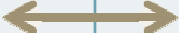
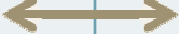
Events

- ⦿ An event
 - refers to a clinically relevant patient-related action
 - belongs to one of six event types

Event Types	Examples
TEST	CT scan
PROBLEM	the tumor
TREATMENT	operation
CLINICAL DEPARTMENTS	ICU
OCCURRENCE	discharge
EVIDENTIAL information	complained

Temporal Relation Types

- 12 types of temporal relations

Simultaneous (32.5%)		Overlap (40.2%)
Before (11.1%)		After (4.1%)
Before_Overlap (3.6%)		Overlap_After (11.6%)
During (2.7%)		During_Inv (4.5%)
Begins (5.5%)		Begun_By (1.4%)
Ends (2.3%)		Ended_By (3.1%)

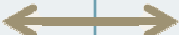
Temporal Relation Types

- 12 types of temporal relations

Simultaneous (32.5%)	↔	Overlap (40.2%)
Before (11.1%)	↔	After (4.1%)
Before_Overlap (3.6%)	↔	Overlap_After (11.6%)
During (2.7%)	↔	During_Inv (4.5%)
Begins (5.5%)	↔	Begun_By (1.4%)
Ends (2.3%)		Ended_By (3.1%)

Temporal Relation Types

- 12 types of temporal relations

Simultaneous (32.5%)		Overlap (40.2%)
Before (11.1%)		After (4.1%)
Before_Overlap (3.6%)		Overlap_After (11.6%)
During (2.7%)		During_Inv (4.5%)
Begins (5.5%)		Begun_By (1.4%)
Ends (2.3%)		Ended_By (3.1%)

Plan for the Talk

- ⦿ Corpus and existing annotations
- ⦿ **Annotation procedure and results**
- ⦿ Temporal relation extraction results

Annotation Procedure

- Two steps
 1. Automatic
 2. Manual

Annotation Procedure

- Two steps
 1. Automatic
 2. Manual

Automatic Annotation Procedure

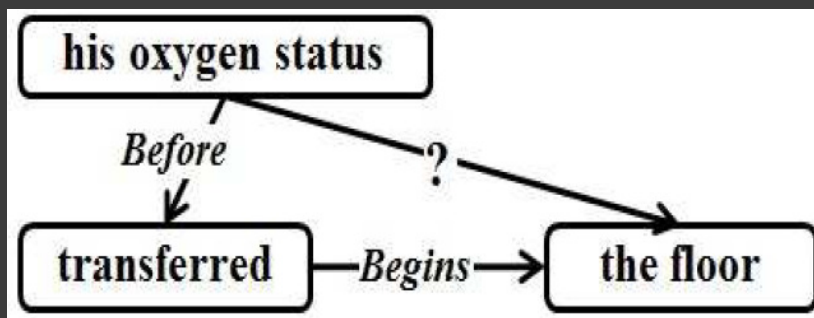
⦿ Motivation

- some missing temporal links can be recovered **automatically** by computing the **transitive closure** over existing links

Automatic Annotation Procedure

⦿ Motivation

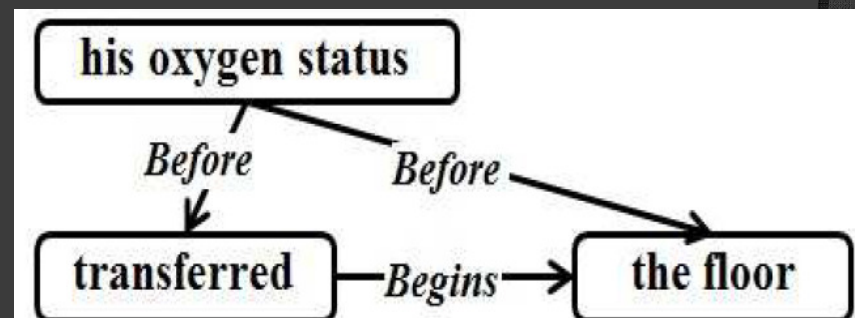
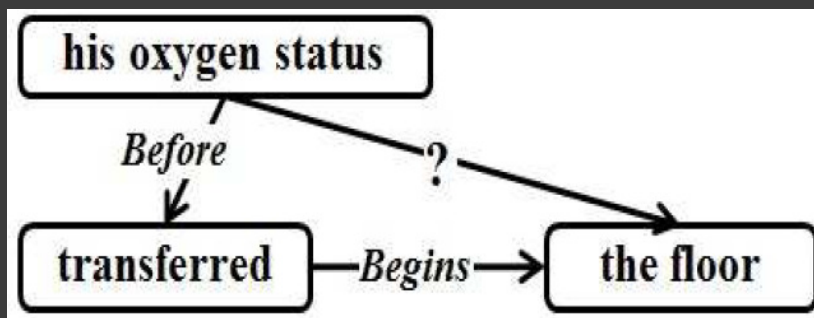
- some missing temporal links can be recovered **automatically** by computing the **transitive closure** over existing links



Automatic Annotation Procedure

⦿ Motivation

- some missing temporal links can be recovered **automatically** by computing the **transitive closure** over existing links



Automatic Annotation Procedure

- ⦿ Transitive closure can be computed using transitivity rules
 - $A \text{ Before } B \wedge B \text{ Begins } C \rightarrow A \text{ Before } C$
 - $A \text{ Before } B \wedge B \text{ Before } C \rightarrow A \text{ Before } C$
- ⦿ Transitivity rules can be derived from Allen's (1983) interval algebra

Manual Annotation

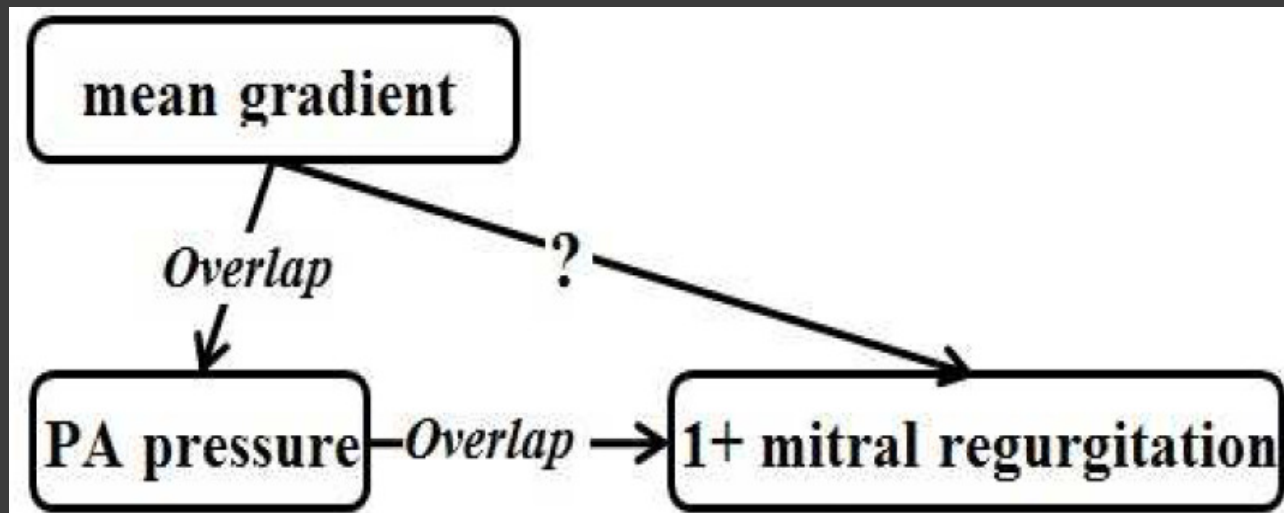
- ④ Motivation

- Not all missing links can be recovered by transitive closure

Manual Annotation

⦿ Motivation

- Not all missing links can be recovered by transitive closure



Manual Annotation

- Focus on annotating missing temporal links between events in **adjacent sentences**

Manual Annotation

- ◎ Focus on annotating missing temporal links between events in **adjacent sentences**
 - They account for 41% of the inter-sentence temporal relations in the i2b2 corpus
 - Classifying inter-sentence relations is harder than classifying intra-sentence relations

Manual Annotation

⦿ Challenge

- Determining whether a temporal relation exists between two events
 - original annotation guideline is vague

Manual Annotation

⦿ Challenge

- Determining whether a temporal relation exists between two events
 - original annotation guideline is vague
 - Two events have a temporal relation when they have an explicit temporal influence on each other in a patient's medical timeline

Manual Annotation

⦿ Challenge

- Determining whether a temporal relation exists between two events
 - original annotation guideline is vague
 - Two events have a temporal relation when they have an explicit temporal influence on each other in a patient's medical timeline
 - Hypothesis: can lead to low inter-annotator agreement

Manual Annotation

- ⦿ Goal

- Develop annotation guidelines that clarify when a temporal relation exists between two events
- 10 rules

Rule 1

- ⦿ Two events have a temporal relation if
 - they are **coreferent**

Rule 1

- ⦿ Two events have a temporal relations if
 - they are **coreferent**
- ⦿ ... his usual monthly change of his suprapubic tube and felt discomfort
- ⦿ ... he was evaluated by Urology and had his tube changed

Rule 2

- ⦿ A TEST event and a PROBLEM event have a temporal relation if
 - the test event reveals a patient's problem

Rule 2

- ⦿ A TEST event and a PROBLEM event have a temporal relation if
 - the test event reveals a patient's problem
- ⦿ A follow-up CT scan was done which did not show any evidence for splenomegaly ...
- ⦿ The 1cm cyst which was seen in 10/92 was still present.

Rule 3

- ⦿ Two events have a temporal relation if
 - one is a TREATMENT/TEST/OCCURRENCE event that happens in a CLINICAL DEPARTMENT event

Rule 3

- ⦿ Two events have a temporal relation if
 - one is a TREATMENT/TEST/OCCURRENCE event that happens in a CLINICAL DEPARTMENT event
- ⦿ He stayed in the unit for about one day.
- ⦿ During that period of time he received nebulizer treatments.

Manual Annotation Rules

- ⦿ 10 rules for determining whether a temporal relation exists between two events
- ⦿ They follow from the (vague) original annotation guideline
- ⦿ But... they are not necessarily complete
 - they don't cover all the event-event pairs that have a temporal relation under the original definition

Annotation Statistics: Training Set

Relation	Original	Closure	Manual	% New Links
Simutaneous	197	210	744	254
Overlap	211	211	3237	1434
Before	60	576	2975	416
After	9	130	562	332
Before_Overlap	69	69	1696	2358
Overlap_After	0	6	1202	19333
During	11	15	157	947
During_Inv	0	0	601	Inf
Begins	0	63	127	102
Begun_By	8	50	66	32
Ends	9	18	58	222
Ended_By	42	83	143	72

Annotation Statistics: Training Set

Relation	Original	Closure	Manual	% New Links
Simutaneous	197	210	744	254
Overlap	211	211	3237	1434
Before	60	576	2975	416
After	9	130	562	332
Before_Overlap	69	69	1696	2358
Overlap_After	0	6	1202	19333
During	11	15	157	947
During_Inv	0	0	601	Inf
Begins	0	63	127	102
Begun_By	8	50	66	32
Ends	9	18	58	222
Ended_By	42	83	143	72

Annotation Statistics: Training Set

Relation	Original	Closure	Manual	% New Links
Simutaneous	197	210	744	254
Overlap	211	211	3237	1434
Before	60	576	2975	416
After	9	130	562	332
Before_Overlap	69	69	1696	2358
Overlap_After	0	6	1202	19333
During	11	15	157	947
During_Inv	0	0	601	Inf
Begins	0	63	127	102
Begun_By	8	50	66	32
Ends	9	18	58	222
Ended_By	42	83	143	72

Annotation Statistics: Training Set

Relation	Original	Closure	Manual	% New Links
Simutaneous	197	210	744	254
Overlap	211	211	3237	1434
Before	60	576	2975	416
After	9	130	562	332
Before_Overlap	69	69	1696	2358
Overlap_After	0	6	1202	19333
During	11	15	157	947
During_Inv	0	0	601	Inf
Begins	0	63	127	102
Begun_By	8	50	66	32
Ends	9	18	58	222
Ended_By	42	83	143	72

Annotation Statistics: Training Set

Relation	Original	Closure	Manual	% New Links
Simutaneous	197	210	744	254
Overlap	211	211	3237	1434
Before	60	576	2975	416
After	9	130	562	332
Before_Overlap	69	69	1696	2358
Overlap_After	0	6	1202	19333
During	11	15	157	947
During_Inv	0	0	601	Inf
Begins	0	63	127	102
Begun_By	8	50	66	32
Ends	9	18	58	222
Ended_By	42	83	143	72

Annotation Statistics: Training Set

Relation	Original	Closure	Manual	% New Links
Simutaneous	197	210	744	254
Overlap	211	211	3237	1434
Before	60	576	2975	416
After	9	130	562	332
Before_Overlap	69	69	1696	2358
Overlap_After	0	6	1202	19333
During	11	15	157	947
During_Inv	0	0	601	Inf
Begins	0	63	127	102
Begun_By	8	50	66	32
Ends	9	18	58	222
Ended_By	42	83	143	72

Annotation Statistics: Test Set

Relation	Original	Closure	Manual	% New Links
Simutaneous	172	174	676	289
Overlap	130	130	2138	1545
Before	42	429	2192	411
After	5	100	445	345
Before_Overlap	68	68	1348	1882
Overlap_After	0	5	866	17220
During	3	3	146	4767
During_Inv	0	0	414	Inf
Begins	0	52	109	110
Begun_By	4	52	52	0
Ends	0	21	73	248
Ended_By	42	96	167	74

Manual Annotation

- ◎ The missing inter-sentence temporal links were
 - identified using 10 rules we developed
 - classified using the i2b2 guidelinesby one human annotator.

Inter-Annotator Agreement

- ⦿ Second human annotator annotate 50 documents
- ⦿ Measured inter-annotator agreement on
 - Identification of temporal links
 - Classification of only those temporal links that both annotators identified
 - Identification & classification of temporal links

Inter-Annotator Agreement

- Second human annotator annotate 50 documents
- Measured inter-annotator agreement on
 - Identification of temporal links
 - Classification of only those temporal links that both annotators identified
 - Identification & classification of temporal links
- Cohen's kappa
 - Identification: 0.82
 - Classification: 0.78
 - Identification & classification: 0.65

Observation

- ⦿ Agreement on classification is worst for Overlap, Before_Overlap, and Overlap_After
 - due to different perception of overlapping events

Observation

- ⦿ Agreement on classification is worst for Overlap, Before_Overlap, and Overlap_After
 - due to different perception of overlapping events
 - E.g., for a treatment administered to a patient before his admission to a hospital, some may label the treatment event as having Before_Overlap relation with the admission event while others may label it as Overlap

Observation

- ⦿ Agreement on classification is worst for Overlap, Before_Overlap, and Overlap_After
 - due to different perception of overlapping events
 - E.g., for a treatment administered to a patient before his admission to a hospital, some may label the treatment event as having Before_Overlap relation with the admission event while others may label it as Overlap
 - Merge these classes into one?

Plan for the Talk

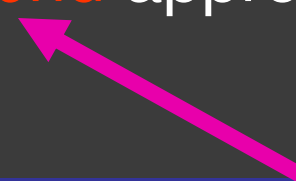
- ⦿ Corpus and existing annotations
- ⦿ Annotation procedure and results
- ⦿ **Temporal relation extraction results**
 - **Goal:** Determine how the performance of a temporal relation extraction system changes when it is trained and tested on the corpus with additional links

Temporal Relation Extraction System

- ◉ D'Souza and Ng (2013)

- First system evaluated on the 12 temporal relation types defined in the i2b2 corpus
- Adopts a knowledge-rich, hybrid approach

- 
- lexical
 - dependency relations
 - lexical semantic relations (WordNet, Webster)
 - predicate-argument relations
 - discourse relations



rule-based and learning-based

Evaluation Settings

- ◎ Evaluate system on temporal relation **identification** & **classification** on 3 versions of the corpus
 - Original
 - Original + Closure
 - Original + Closure + Manual

Results on Identification + Classification

Corpus	All event pairs		
	R	P	F
Original	52.9	21.0	30.0
Closed	56.4	32.7	41.4
Manual	33.5	57.3	42.3

- Results not directly comparable because they were not evaluated on same set of annotations

Results on Identification + Classification

Corpus	All event pairs		
	R	P	F
Original	52.9	21.0	30.0
Closed	56.4	32.7	41.4
Manual	33.5	57.3	42.3

- But.. results on bottom row are a more accurate reflection of the system's performance

Results on Identification + Classification

Corpus	All event pairs		
	R	P	F
Original	52.9	21.0	30.0
Closed	56.4	32.7	41.4
Manual	33.5	57.3	42.3

- But.. results on bottom row are a more accurate reflection of the system's performance
 - trained and tested on fewer missing links
 - performance is better than on the Original corpus

Results on Identification + Classification

Corpus	All event pairs			Event pairs in adjacent sentences		
	R	P	F	R	P	F
Original	52.9	21.0	30.0	52.7	6.7	11.9
Closed	56.4	32.7	41.4	58.4	7.2	12.7
Manual	33.5	57.3	42.3	21.3	31.9	25.5

Results on Identification + Classification

Corpus	All event pairs			Event pairs in adjacent sentences		
	R	P	F	R	P	F
Original	52.9	21.0	30.0	52.7	6.7	11.9
Closed	56.4	32.7	41.4	58.4	7.2	12.7
Manual	33.5	57.3	42.3	21.3	31.9	25.5

- Same performance trend when evaluated only on relations between entities in adjacent sentences

Summary

- Manually annotated missing temporal links in the i2b2 Clinical Temporal Relations Challenge Corpus
- Empirically determined the impact of the new annotations on the performance of a state-of-the-art temporal relation extraction system