A Framework for Cross-Domain Clinical Temporal Information Extraction

Sean MacAvaney, Arman Cohan, Nazli Goharian

July 15, 2018

(Appearing at SemEval '17)

Motivation

There is a significant amount of unstructured data in the form of clinical notes.

"The intraoperative as well as the postoperative course were uncomplicated. The patient was transferred from the Vision Care Center to the Laurel Beach Medical Center where she was observed overnight."

This information could be the basis of a patient ”timeline”, which could be used to track disease status, monitor treatment outcomes, discover side-effects, etc.

Clinical TempEval

SemEval-2017 Task 12

Follows Clinical TempEval 2015 and 2016 and the i2b2 2013 temporal challenge

Clinical TempEval

SemEval-2017 Task 12

Follows Clinical TempEval 2015 and 2016 and the i2b2 2013 temporal challenge

Focus: Domain adaptation

Source domain: colon cancer notes

Target domain: brain cancer notes

Clinical TempEval

SemEval-2017 Task 12

Follows Clinical TempEval 2015 and 2016 and the i2b2 2013 temporal challenge

Focus: Domain adaptation

Source domain: colon cancer notes

Target domain: brain cancer notes

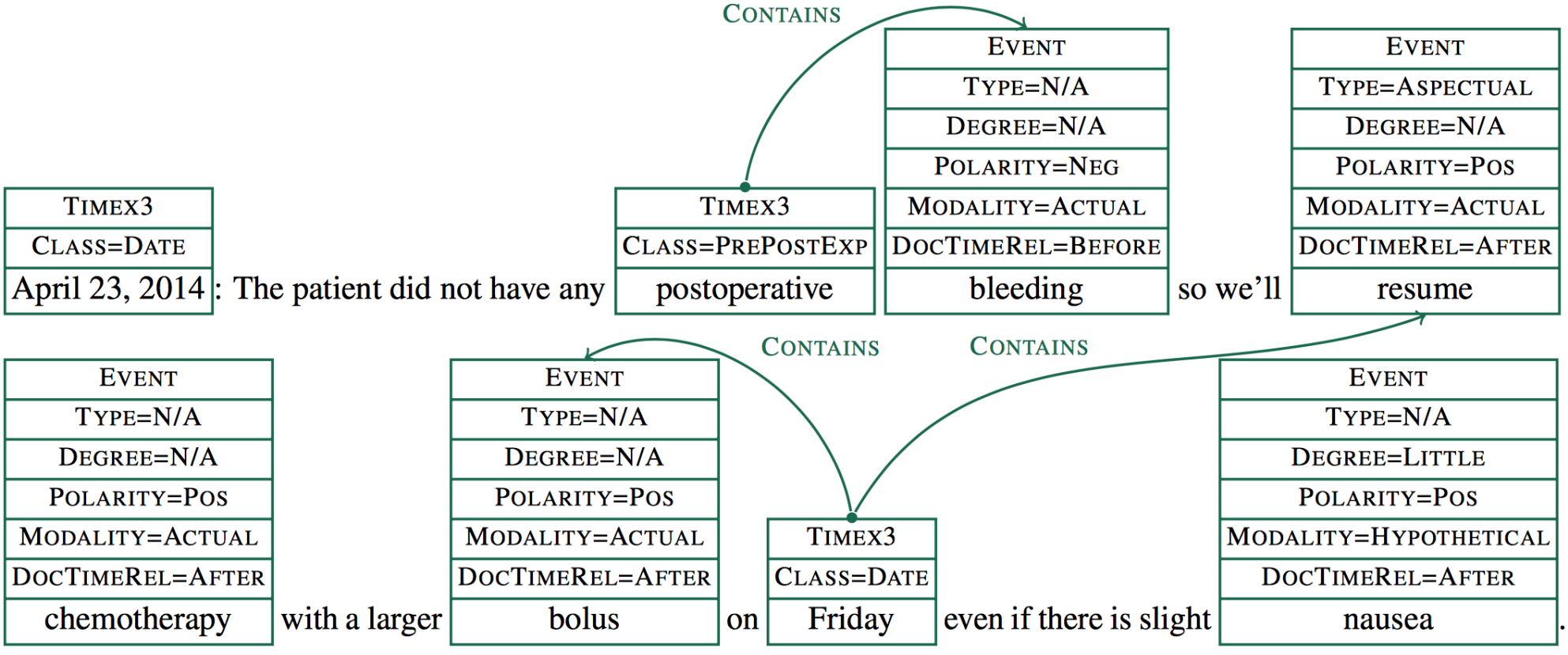
Unsupervised and supervised domain adaptation

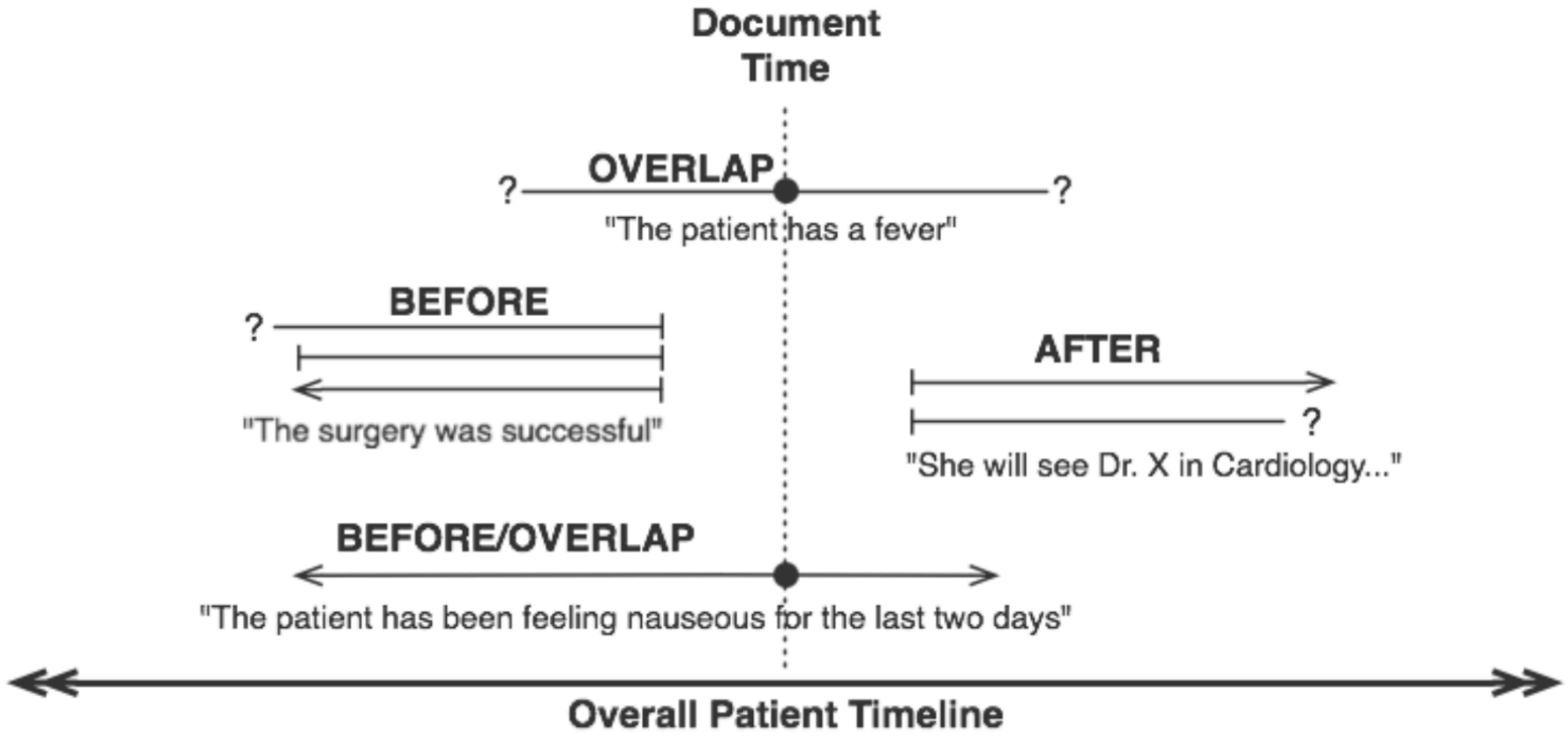
THYME Corpus

1,200 notes, two domains, three types (clinical, pathology, radiology)

Annotations of events (EVENT), and time expressions (TIMEX3)

Relations between them (TLINK)


CONTAINS
TIMEX3
CLASS=DATE
April 23, 2014
: The patient did not have any
TIMEX3
CLASS=PREPOSTEXP
postoperative
EVENT
TYPE=N/A
DEGREE=N/A
POLARITY=NEG
MODALITY=ACTUAL
DOCTIMEREL=BEFORE
bleeding
so we'll
EVENT
TYPE=ASPECTUAL
DEGREE=N/A
POLARITY=POS
MODALITY=ACTUAL
DOCTIMEREL=AFTER
resume
CONTAINS
CONTAINS
EVENT
TYPE=N/A
DEGREE=N/A
POLARITY=POS
MODALITY=ACTUAL
DOCTIMEREL=AFTER
chemotherapy
with a larger
EVENT
TYPE=N/A
DEGREE=N/A
POLARITY=POS
MODALITY=ACTUAL
DOCTIMEREL=AFTER
bolus
on
TIMEX3
CLASS=DATE
Friday
even if there is slight
EVENT
TYPE=N/A
DEGREE=LITTLE
POLARITY=POS
MODALITY=HYPOTHETICAL
DOCTIMEREL=AFTER
nausea
.


Document Time
OVERLAP
?
?
"The patient has a fever"
BEFORE
?
"The surgery was successful"
AFTER
?
"She will see Dr. X in Cardiology..."
BEFORE/OVERLAP
"The patient has been feeling nauseous for the last two days"
Overall Patient Timeline

Method

1. Span extraction

2. Attribute extraction

3. Relation extraction

Span Extraction

Supervised learning: CRF

Classifying tokens in text

Labels: BIO

Features: lexical, syntactic, distributional, semantic, rule-based

Window: ± 1 token

Span Extraction » Labels

"Patient taken to the operating theater on January 13, 2006 where he underwent a left-sided suboccipital craniectomy for two hours."

Tokenization

`Patient␣taken␣to␣the␣operating␣theater␣on␣January␣13␣,␣2006␣where␣he␣underwent␣a␣left␣-␣sided␣suboccipital␣craniectomy␣for␣two␣hours␣.`

B-I-O Labeling

$\texttt{Patient}_O\ \texttt{taken}_O\ \texttt{to}_O\ \texttt{the}_O\ \texttt{operating}_O\ \texttt{theater}_O\ \texttt{on}_O\ \texttt{January}_B\ \texttt{13}_I\ \texttt{,}_I\ \texttt{2006}_I\ \texttt{where}_O\ \texttt{he}_O\ \texttt{underwent}_O\ \texttt{a}_O\ \texttt{left}_O\ \texttt{-}_O\ \texttt{sided}_O\ \texttt{suboccipital}_O\ \texttt{craniectomy}_O\ \texttt{for}_O\ \texttt{two}_B\ \texttt{hours}_I\ \texttt{.}_O$

Span Extraction » Features

Lexical: word, shape, lemma, casing flags, etc.

Syntactic: part-of-speech

Distributional: Brown clusters, word embeddings

Semantic: UMLS semantic types (MetaMap)

Rules: Regex

Span Extraction » Features

Feature	Examples
Date	12/3/2010, 1965-01-21
Month	January, Aug
Day	1st, 31
Day-of-week	Monday, Wed
Season	summer, spring
Year	2013, 1990s
Time	8:42, a.m.
Time Unit	minute, sec
Number	4, seventeen
Temporal preposition	in, after
Temporal adverb	daily, lately
Temporal prefix	pre, post

Attribute Extraction

Same approach as span extraction

One model per attribute

Labels: attribute values

Extended window of ± 3 tokens

Relation Extraction

Supervised learning: Gradient-boosted trees

Classification of “candidate” relation pairs

(Intra-sentence heuristic)

Labels: relation type, or null

Features: entity features, relation features

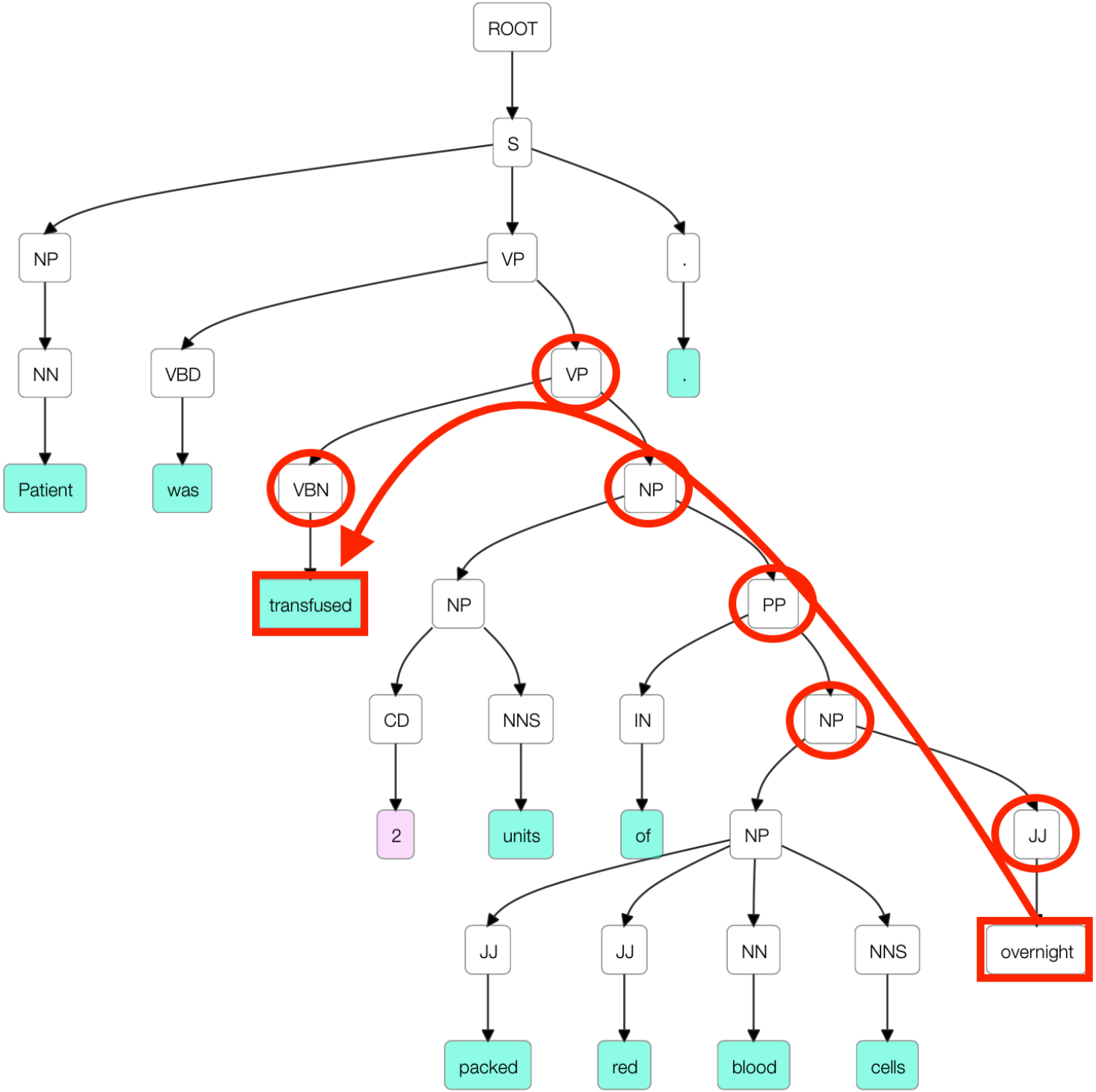
Relation Extraction » Features

Token features described previously

Annotation attributes as features

Token distance, character distance, etc.

Constituent/dependency tree path, path distance, etc.


ROOT
S
NP
VP
.
NN
VBD
VP
.
Patient
was
VBN
NP
transfused
NP
PP
CD
NNS
IN
NP
2
units
of
NP
JJ
JJ
JJ
NN
NNS
overnight
packed
red
blood
cells

Results » Spans

	TIMEX3 Spans			EVENT Spans		
	P	R	F1	P	R	F1
Phase 1 (Unsupervised)						
Our System	0.61	0.53	**†0.57**	0.64	**‡0.80**	**‡0.71**
Median	0.63	0.46	0.48	0.64	0.69	0.68
Phase 2 (Supervised)						
Our System	**‡0.57**	**‡0.62**	**†0.59**	**‡0.68**	0.82	**‡0.74**
Median	0.53	0.52	0.54	0.67	0.76	0.71
MEMORIZE	0.64	0.22	0.33	0.61	0.51	0.56

† Top score; ‡ Second-best score

Results » Attributes

	TIMEX3 Class			EVENT Modality			EVENT Degree			EVENT Polarity			EVENT Type		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Phase 1 (Unsupervised)															
Ours	0.55	0.47	**‡0.51**	0.50	0.64	0.56	0.61	**‡0.77**	**‡0.68**	0.59	**‡0.74**	0.65	0.61	**‡0.76**	**‡0.68**
Median	0.56	0.45	0.46	0.55	0.63	0.59	0.62	0.71	0.68	0.60	0.70	0.66	0.61	0.70	0.66
Phase 2 (Supervised)															
Ours	**‡0.54**	**‡0.59**	**†0.56**	0.60	0.72	**‡0.66**	**‡0.67**	0.80	**‡0.73**	0.54	0.64	0.58	**‡0.66**	0.79	**‡0.72**
Median	0.49	0.48	0.48	0.57	0.68	0.63	0.66	0.77	0.71	0.62	0.70	0.66	0.65	0.76	0.70
MEMORIZE	0.49	0.17	0.25	0.29	0.24	0.26	0.47	0.40	0.43	0.56	0.47	0.51	0.50	0.42	0.46

† Top score; ‡ Second-best score

Results » Relations

	Contains			DocTimeRel		
	P	R	F1	P	R	F1
Phase 1 (Unsupervised)						
Our System	**†0.52**	0.25	**†0.34**	0.36	0.45	0.40
Median	0.33	0.25	0.32	0.39	0.45	0.41
Phase 2 (Supervised)						
Our System	**†0.59**	0.16	0.25	0.45	0.55	0.50
Median	0.20	0.16	0.16	0.42	0.51	0.46
Closest	0.33	0.08	0.12	-	-	-
Memorize	-	-	-	0.22	0.18	0.20

† Top score; ‡ Second-best score

Results » Errors

New date conventions (e.g. 12Jun2013)

Results » Errors

New date conventions (e.g. 12Jun2013)

“Cancer” as EVENT in “Cancer Research Hospital”

Results » Errors

New date conventions (e.g. 12Jun2013)

"Cancer" as EVENT in "Cancer Research Hospital"

Domain-specific overfitting – "temozolomide"

Questions?