



Structured Learning for Temporal Relation Extraction from Clinical Records

Artuur Leeuwenberg and Marie-Francine Moens
Department of Computer Science, KU Leuven

7th of April, 2017 (EACL)



Temporal Information



- Find all patients that:
 - Used to smoke, but now stopped...
 - Used a particular medication in the past
 - Showed certain symptoms for more than a year

Task Description



We will set-up an appointment with Medical Oncology in four to six-weeks.

Task Description



Given: **Events**, **Temporal Expressions**

We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.

Task Description



Given: **Events**, **Temporal Expressions**

Predict: **Document-time Relations**

We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.
AFTER **AFTER**

Task Description



Given: **Events**, **Temporal Expressions**

Predict: **Document-time Relations**, Temporal Links

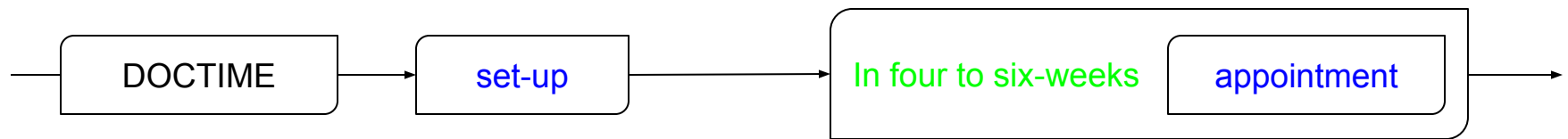


Task Description



Given: **Events**, **Temporal Expressions**

Predict: **Document-time Relations**, Temporal Links



Classical Approach



We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.

Classical Approach



We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.
?

Classical Approach



We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.
AFTER **?**

Classical Approach



We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.
AFTER **AFTER**

Classical Approach



We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.
AFTER **AFTER**

↑
?

Classical Approach



?

We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.

AFTER **AFTER**

Classical Approach



before
?

We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.
AFTER **AFTER**

Classical Approach



before
?

We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.

AFTER **AFTER**

The diagram illustrates a process flow. A curved arrow labeled "before" points from the start of the sentence to the word "appointment". A second curved arrow labeled "?" points from the word "appointment" to the end of the sentence, "in four to six-weeks". Below the words "set-up" and "appointment" are the red words "AFTER".

Classical Approach

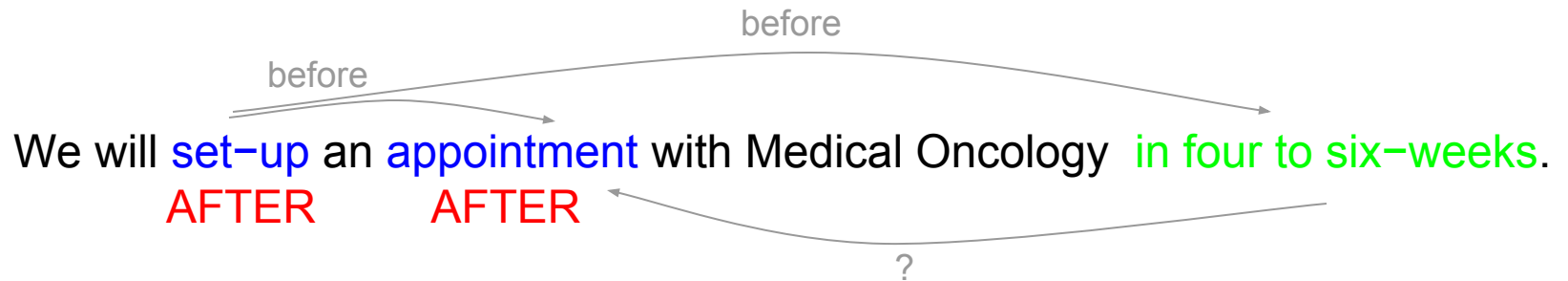


We will **set-up** an **appointment** with Medical Oncology **in four to six-weeks**.

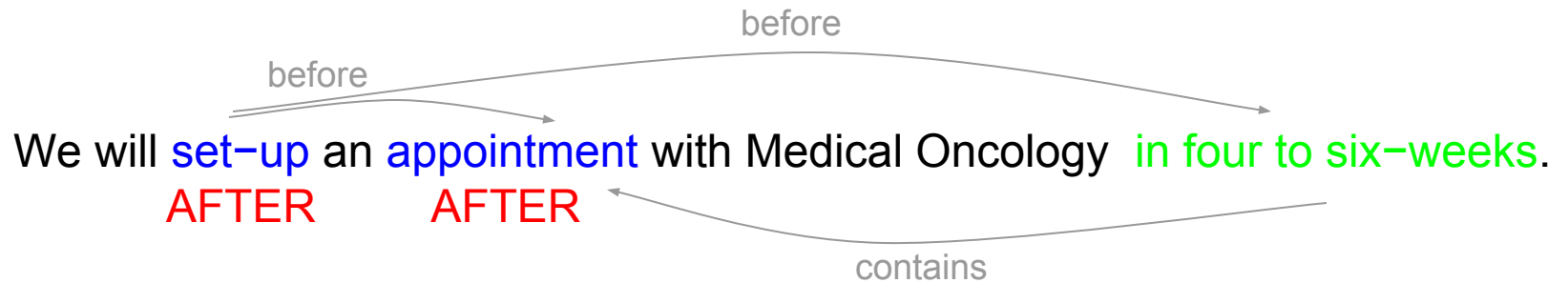
AFTER **AFTER**

before before

Classical Approach



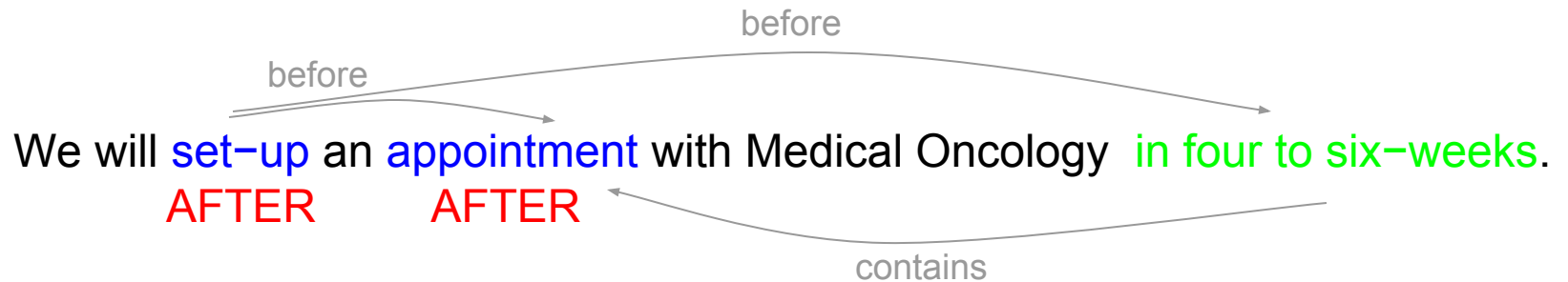
Classical Approach



Classical Approach



- Local Learning
- Independent Predictions



Classical Approach



- Local Learning
- Independent Predictions

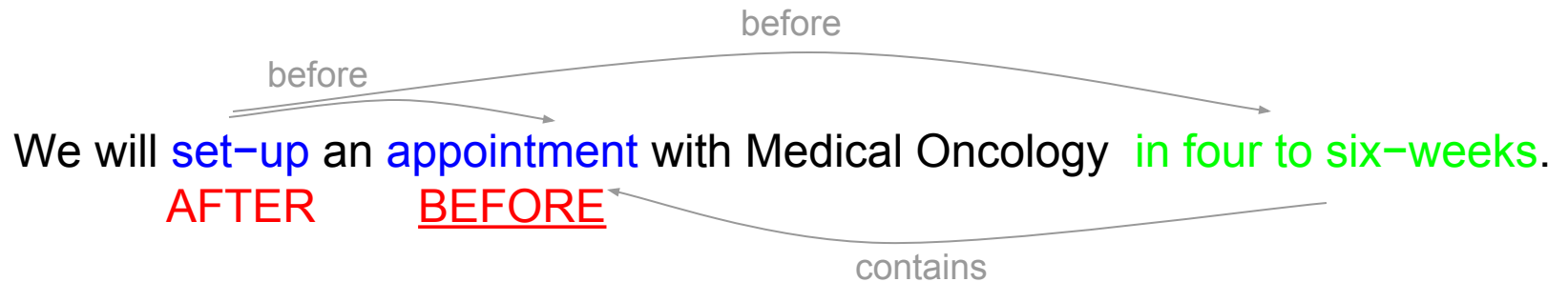


What if we make one wrong prediction?

Classical Approach



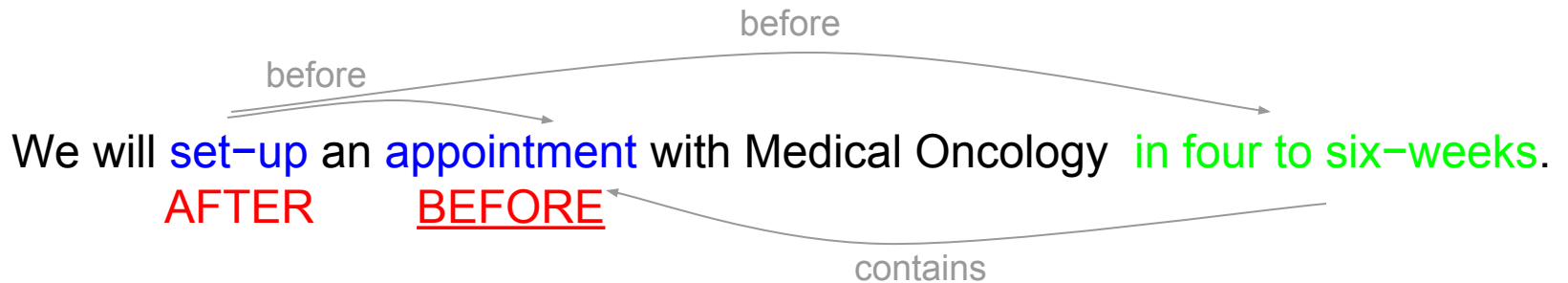
- Local Learning
- Independent Predictions



Classical Approach



- Local Learning
- Independent Predictions



Problem: Timeline construction from inconsistent relations

Proposed Model



Proposed Model



1. Document Level Learning
2. Global Constraints on Predictions
3. Learning Output Structure through Global Features

Document-Level Learning



- Scoring Label Assignments

$$S(X, Y) = \lambda\Phi(X, Y)$$

Document-Level Learning



- Scoring Label Assignments

$$S(X, Y) = \lambda\Phi(X, Y)$$

- Structured Perceptron Training [Collins, 2002]
 - Averaging [Freund and Schapire, 1999]
 - Loss augmented sub-sampling



Document-Level Features

- Joint Local Features
 - E.g. Tokens, POS, Dependency Path
- Global Features
 - E.g. DCTR label n-grams

$$\Phi(X, Y) = \Phi_{joint}(X, Y) \oplus \Phi_{global}(X, Y)$$

Prediction with Constraints



- Prediction

$$\hat{Y}_k = \arg \max_Y S(X, Y)$$

- Integer Linear Programming (Gurobi, 2015)

$$O = \arg \max_W \sum_{x_{i,d} \in C_{dr}(X)} \sum_{l \in L_{dr}} w_{i,d}^l \cdot S(x_{i,d}, y_{i,d}^l) \\ + \sum_{x_{i,j} \in C_{tl}(X)} \sum_{l \in L_{tl}} w_{i,j}^l \cdot S(x_{i,j}, y_{i,j}^l)$$

Temporal Constraints



- Label dependencies as Constraints on decision variables

Abbrev.	Label Dependencies	Constraints
\mathcal{C}_{Ctrans}	$\text{CONTAINS}_{i,j} \wedge \text{CONTAINS}_{j,k} \rightarrow \text{CONTAINS}_{i,k}$	$\forall i,j,k : w_{i,k}^{\text{contains}} - w_{i,j}^{\text{contains}} - w_{j,k}^{\text{contains}} \geq -1$
\mathcal{C}_{Btrans}	$\text{BEFORE}_{i,j} \wedge \text{BEFORE}_{j,k} \rightarrow \text{BEFORE}_{i,k}$	$\forall i,j,k : w_{i,k}^{\text{before}} - w_{i,j}^{\text{before}} - w_{j,k}^{\text{before}} \geq -1$
\mathcal{C}_{CBB}	$\text{CONTAINS}_{i,j} \wedge \text{BEFORE}_{i,d} \rightarrow \text{BEFORE}_{j,d}$	$\forall i,j : w_{j,d}^{\text{before}} - w_{i,j}^{\text{contains}} - w_{i,d}^{\text{before}} \geq -1$
\mathcal{C}_{CAA}	$\text{CONTAINS}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$	$\forall i,j : w_{j,d}^{\text{after}} - w_{i,j}^{\text{contains}} - w_{i,d}^{\text{after}} \geq -1$
\mathcal{C}_{BBB}	$\text{BEFORE}_{i,j} \wedge \text{BEFORE}_{j,d} \rightarrow \text{BEFORE}_{i,d}$	$\forall i,j : w_{i,d}^{\text{before}} - w_{i,j}^{\text{before}} - w_{j,d}^{\text{before}} \geq -1$
\mathcal{C}_{BAA}	$\text{BEFORE}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$	$\forall i,j : w_{j,d}^{\text{after}} - w_{i,j}^{\text{before}} - w_{i,d}^{\text{after}} \geq -1$

Temporal Constraints



- Label dependencies as Constraints on decision variables
 - **E.g. Transitivity of BEFORE^{TLINK}**

Abbrev.	Label Dependencies	Constraints
\mathcal{C}_{trans}	$\text{CONTAINS}_{i,j} \wedge \text{CONTAINS}_{i,k} \rightarrow \text{CONTAINS}_{i,k}$	$\forall_{i,j,k} : w_{i,j}^{\text{contains}} - w_{i,k}^{\text{contains}} - w_{j,k}^{\text{contains}} > -1$
\mathcal{C}_{Btrans}	$\text{BEFORE}_{i,j} \wedge \text{BEFORE}_{j,k} \rightarrow \text{BEFORE}_{i,k}$	$\forall_{i,j,k} : w_{i,k}^{\text{before}} - w_{i,j}^{\text{before}} - w_{j,k}^{\text{before}} \geq -1$
\mathcal{C}_{CBB}	$\text{CONTAINS}_{i,j} \wedge \text{BEFORE}_{i,d} \rightarrow \text{BEFORE}_{j,d}$	$\forall_{i,j} : w_{j,d}^{\text{before}} - w_{i,j}^{\text{contains}} - w_{i,d}^{\text{before}} \geq -1$
\mathcal{C}_{CAA}	$\text{CONTAINS}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$	$\forall_{i,j} : w_{j,d}^{\text{after}} - w_{i,j}^{\text{contains}} - w_{i,d}^{\text{after}} \geq -1$
\mathcal{C}_{BBB}	$\text{BEFORE}_{i,j} \wedge \text{BEFORE}_{j,d} \rightarrow \text{BEFORE}_{i,d}$	$\forall_{i,j} : w_{i,d}^{\text{before}} - w_{i,j}^{\text{before}} - w_{j,d}^{\text{before}} \geq -1$
\mathcal{C}_{BAA}	$\text{BEFORE}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$	$\forall_{i,j} : w_{j,d}^{\text{after}} - w_{i,j}^{\text{before}} - w_{i,d}^{\text{after}} \geq -1$

Temporal Constraints



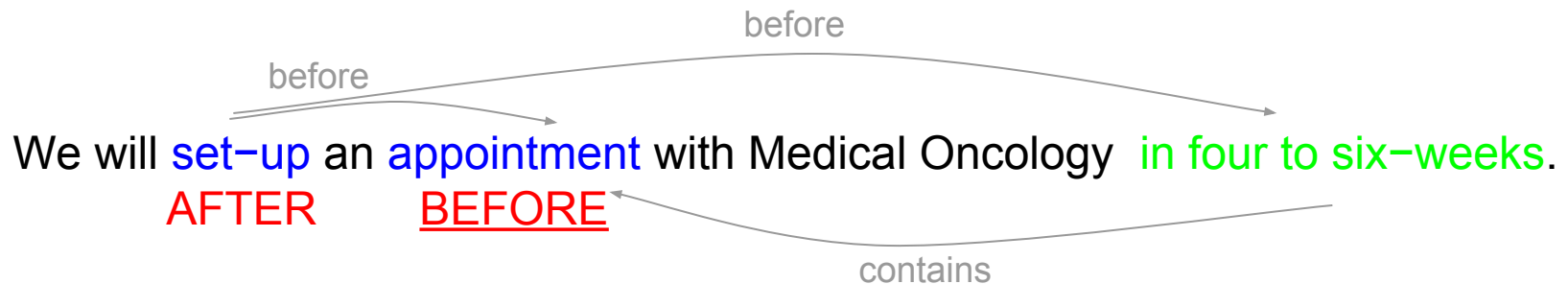
- Label dependencies as Constraints on decision variables
 - E.g. Transitivity of $\text{BEFORE}^{\text{TLINK}}$
 - E.g. Connecting $\text{BEFORE}^{\text{TLINK}}$ and $\text{AFTER}^{\text{DCTR}}$

Abbrev.	Label Dependencies	Constraints
\mathcal{C}_{Ctrans}	$\text{CONTAINS}_{i,j} \wedge \text{CONTAINS}_{j,k} \rightarrow \text{CONTAINS}_{i,k}$	$\forall_{i,j,k} : w_{i,k}^{\text{contains}} - w_{i,j}^{\text{contains}} - w_{j,k}^{\text{contains}} \geq -1$
\mathcal{C}_{Btrans}	$\text{BEFORE}_{i,j} \wedge \text{BEFORE}_{j,k} \rightarrow \text{BEFORE}_{i,k}$	$\forall_{i,j,k} : w_{i,k}^{\text{before}} - w_{i,j}^{\text{before}} - w_{j,k}^{\text{before}} \geq -1$
\mathcal{C}_{CBB}	$\text{CONTAINS}_{i,j} \wedge \text{BEFORE}_{i,d} \rightarrow \text{BEFORE}_{j,d}$	$\forall_{i,j} : w_{j,d}^{\text{before}} - w_{i,j}^{\text{contains}} - w_{i,d}^{\text{before}} \geq -1$
\mathcal{C}_{CAA}	$\text{CONTAINS}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$	$\forall_{i,j} : w_{j,d}^{\text{after}} - w_{i,j}^{\text{contains}} - w_{i,d}^{\text{after}} \geq -1$
\mathcal{C}_{BBB}	$\text{BEFORE}_{i,i} \wedge \text{BEFORE}_{i,d} \rightarrow \text{BEFORE}_{i,d}$	$\forall_{i,i} : w_{i,i}^{\text{before}} - w_{i,i}^{\text{before}} - w_{i,i}^{\text{before}} > -1$
\mathcal{C}_{BAA}	$\text{BEFORE}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$	$\forall_{i,j} : w_{j,d}^{\text{after}} - w_{i,j}^{\text{before}} - w_{i,d}^{\text{after}} \geq -1$



Example Revisited

- Global Constraints on Predictions
- Learning Output Structure through Global Features



DCTR-label bigram feature

$$\text{BEFORE}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$$



Example Revisited

- **Global Constraints on Predictions**
- Learning Output Structure through Global Features



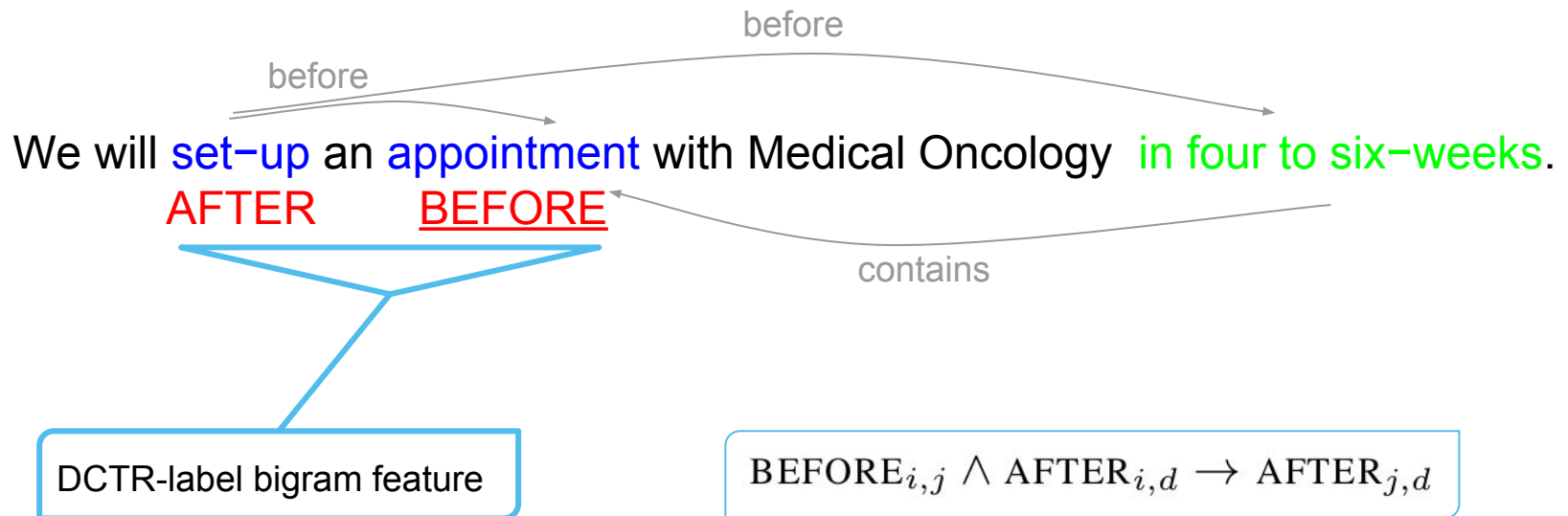
DCTR-label bigram feature

$$\text{BEFORE}_{i,j} \wedge \text{AFTER}_{i,d} \rightarrow \text{AFTER}_{j,d}$$



Example Revisited

- Global Constraints on Predictions
- **Learning Output Structure through Global Features**



Experiments & Results



Experiments



- Dataset
 - THYME corpus [Styler IV et al., 2014]
 - 440 documents for training, and 151 for testing
- Preprocessing
 - Ground-truth **events** and **temporal expressions**
 - Tokenization, POS Tagging (cTAKES), Dependency Parsing (cTAKES)
- Local Candidate Generation
 - **DCTR**: all events
 - **TLINKS**: pairs within each paragraph
- Parameter tuning on development set

Baselines

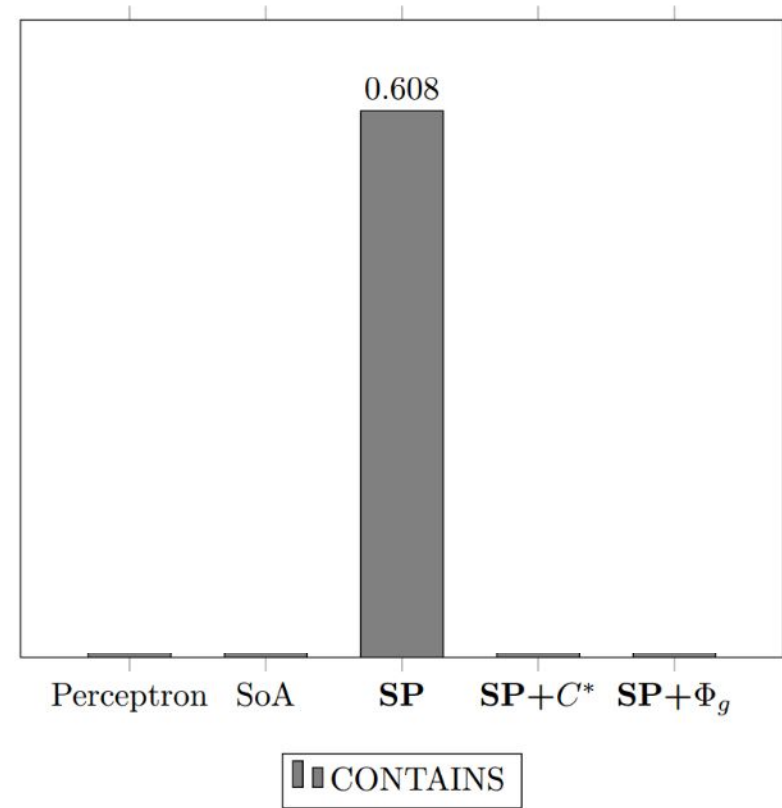
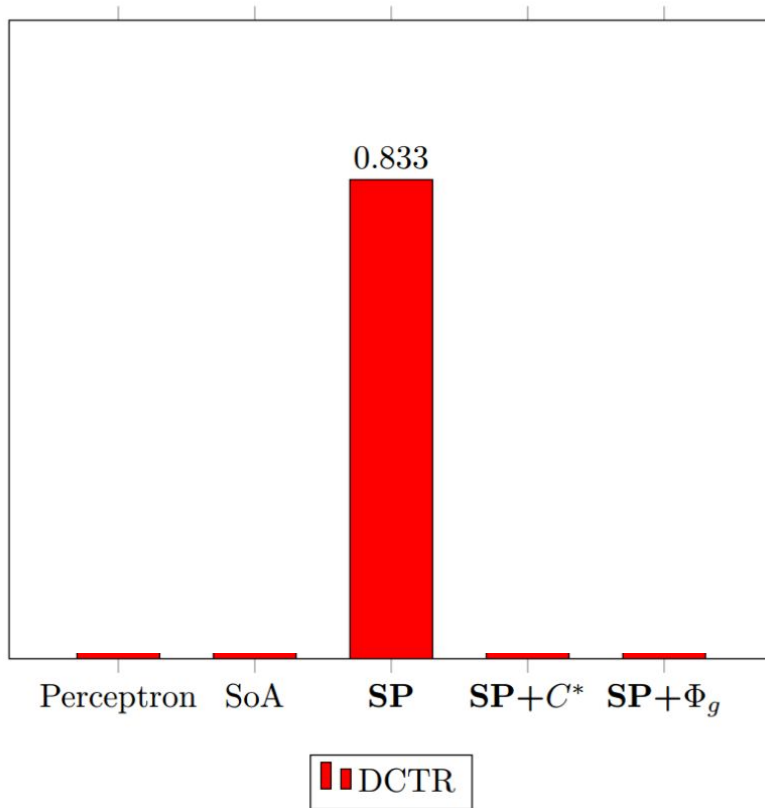


- **DCTR:**
 - Local Perceptron
 - CRF (Khalifa et al., 2016)
- **TLINKS:**
 - Local Perceptron
 - SVM (Lin et al., 2016)

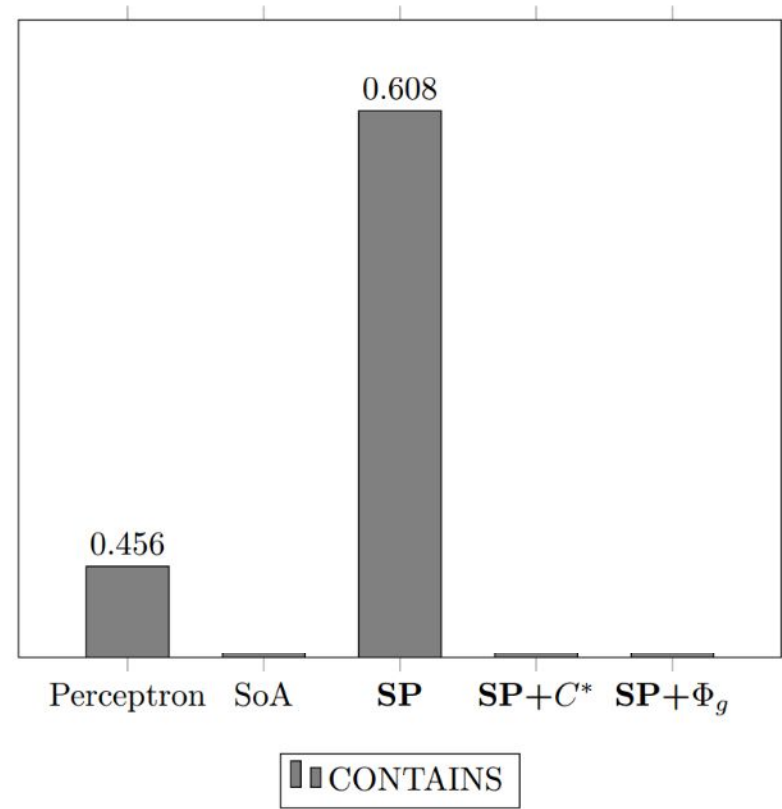
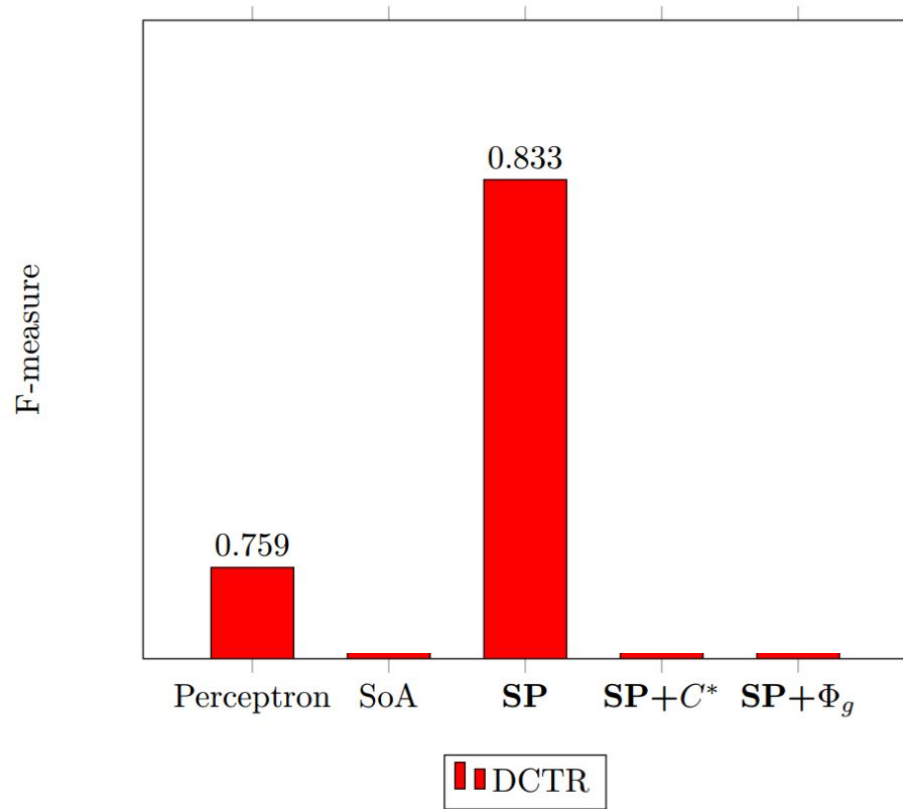
Structured Perceptron (SP)



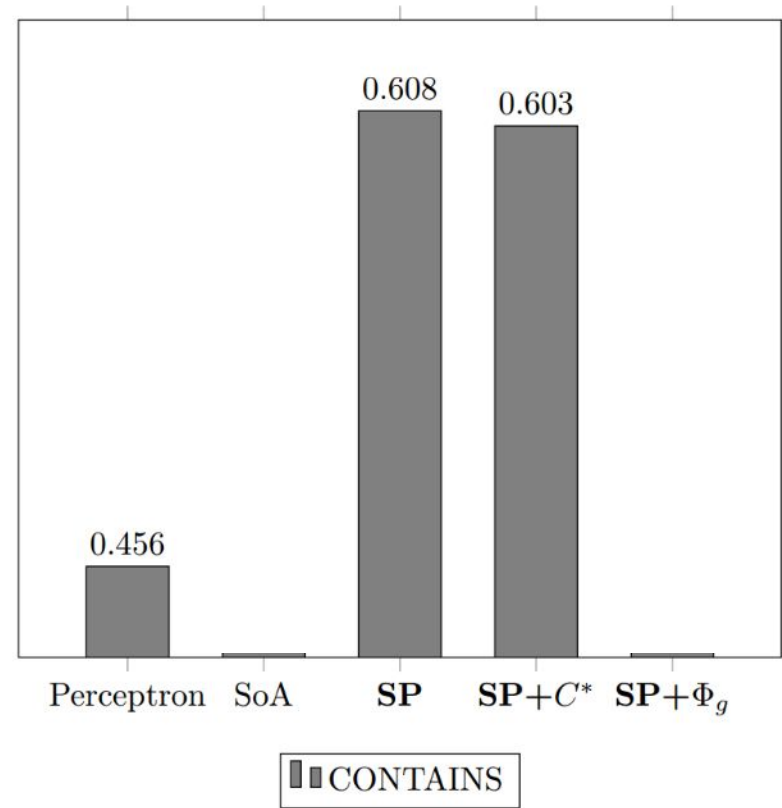
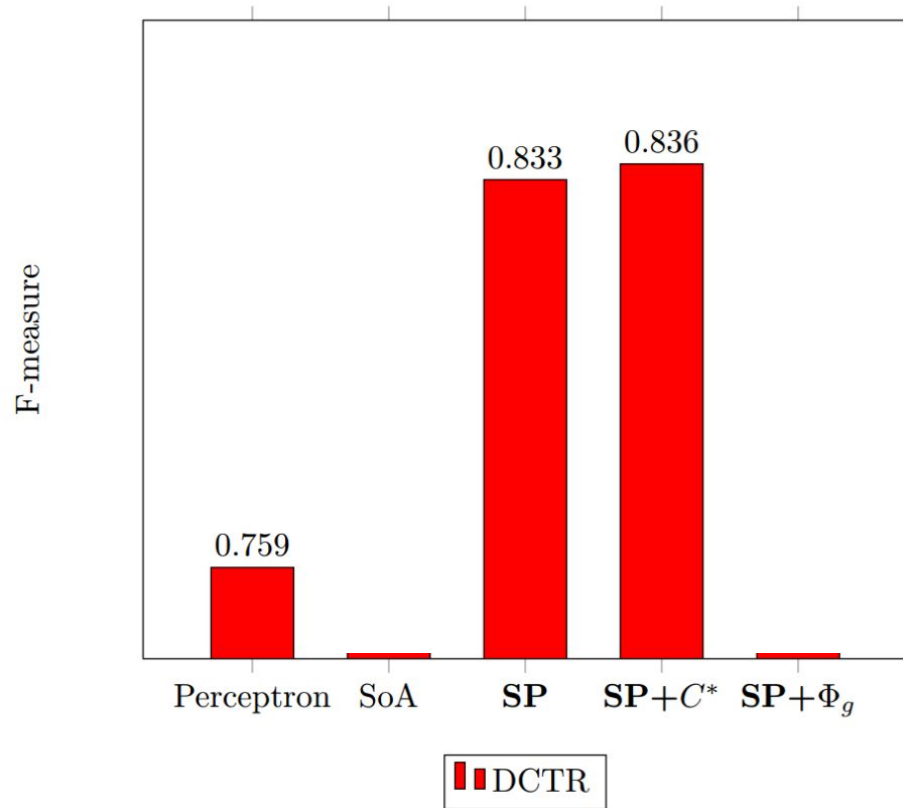
F-measure



(Local) Perceptron



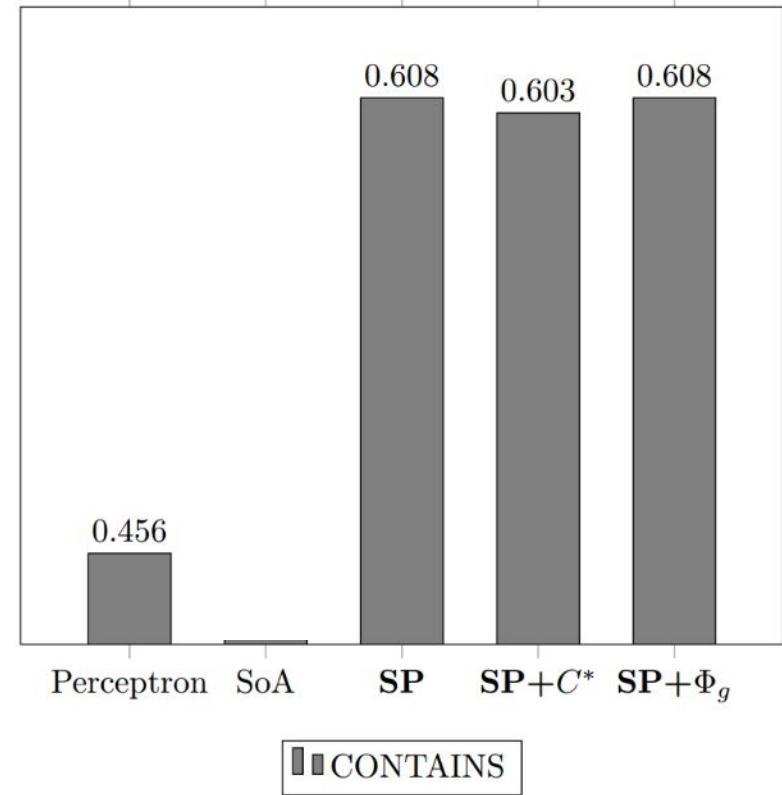
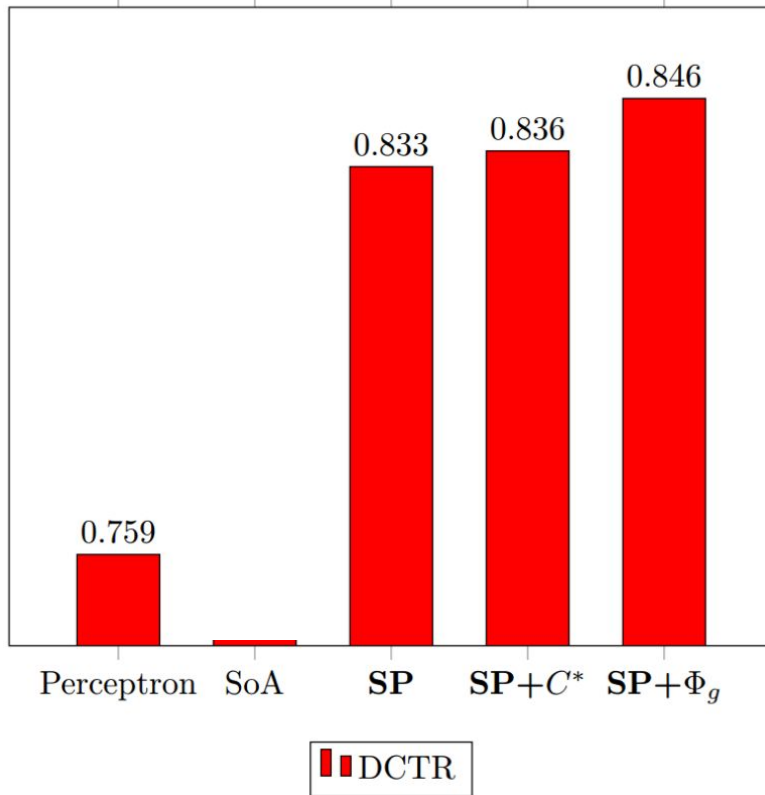
SP + Constraints (C^*)



SP + Global Features (Φ_g)



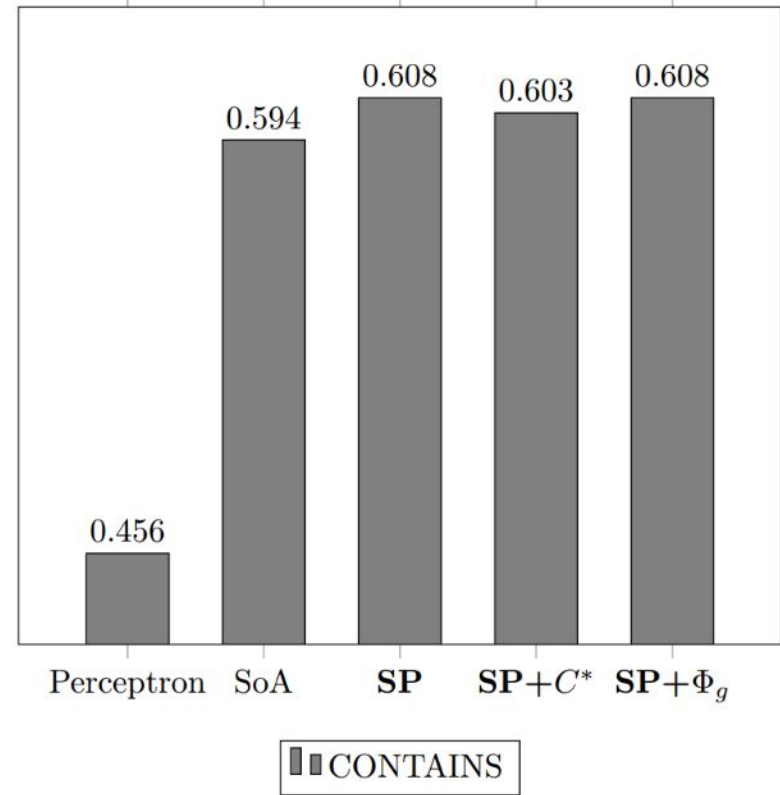
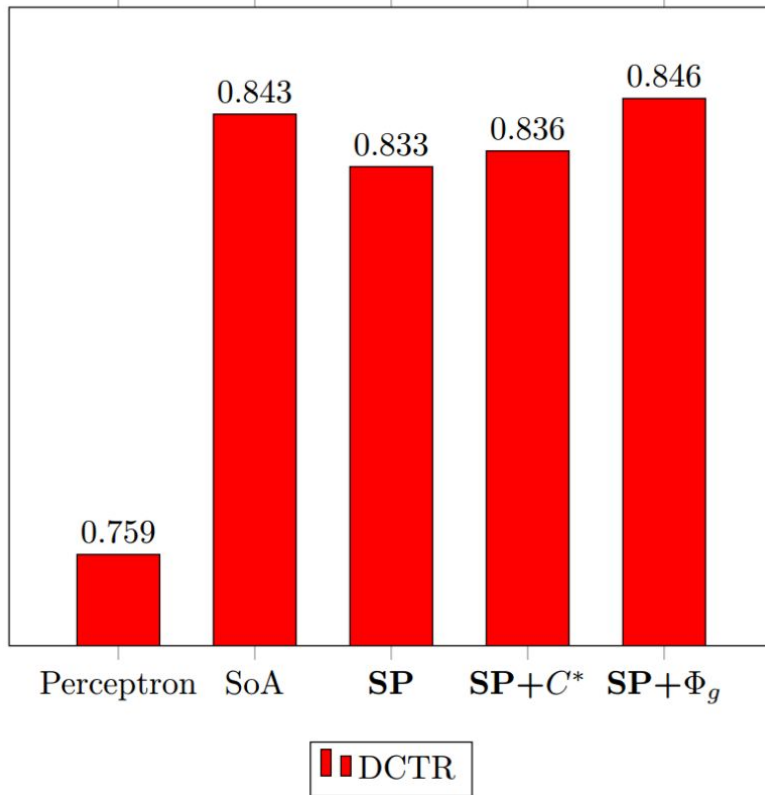
F-measure



Current State-of-the-Art (SoA)



F-measure



Recap



- Document level learning and prediction
- Modeling label dependencies / output structure
 - Hard Dependencies as Constraints
 - Soft Dependencies as Global Features
- Improve state-of-the-art results for temporal relations

Thank you for your attention!



Questions?

Code: <https://github.com/tuur/SPTempRels>

Project: www.accumulate.be