Technology developed at CLiPS

Simon Šuster, Madhumita Sushil, Pieter Fivez and Walter Daelemans (with cooperation from UZA)

Accumulate Industrial Meeting

6 September 2018

Overview of our work

WP2

- Normalization
  - Spelling correction for EN and NL
  - Synonymy discovery and KB completion
- Terminology extraction
  - Concept extraction and disambiguation

New annotated data, trained model and results for supervised concept labelling

WP3

- Event extraction
  - Machine reading comprehension and QA
  - Negation and modality detection
  - Relation extraction

Supervised relation extraction for English

Overview of our work

~WP6

- Model interpretability and document-level representations
  - Patient vectors for clinical prediction tasks
  - Explaining model decisions through salient features and rules
- Cohort selection for clinical trials
  - Participation at the n2c2-2018 shared task (with Radboud University)
  - Identifying patients fitting selected criteria
  - Hybrid approach using ML, rules and embedding similarity

Supervised concept labelling for Dutch

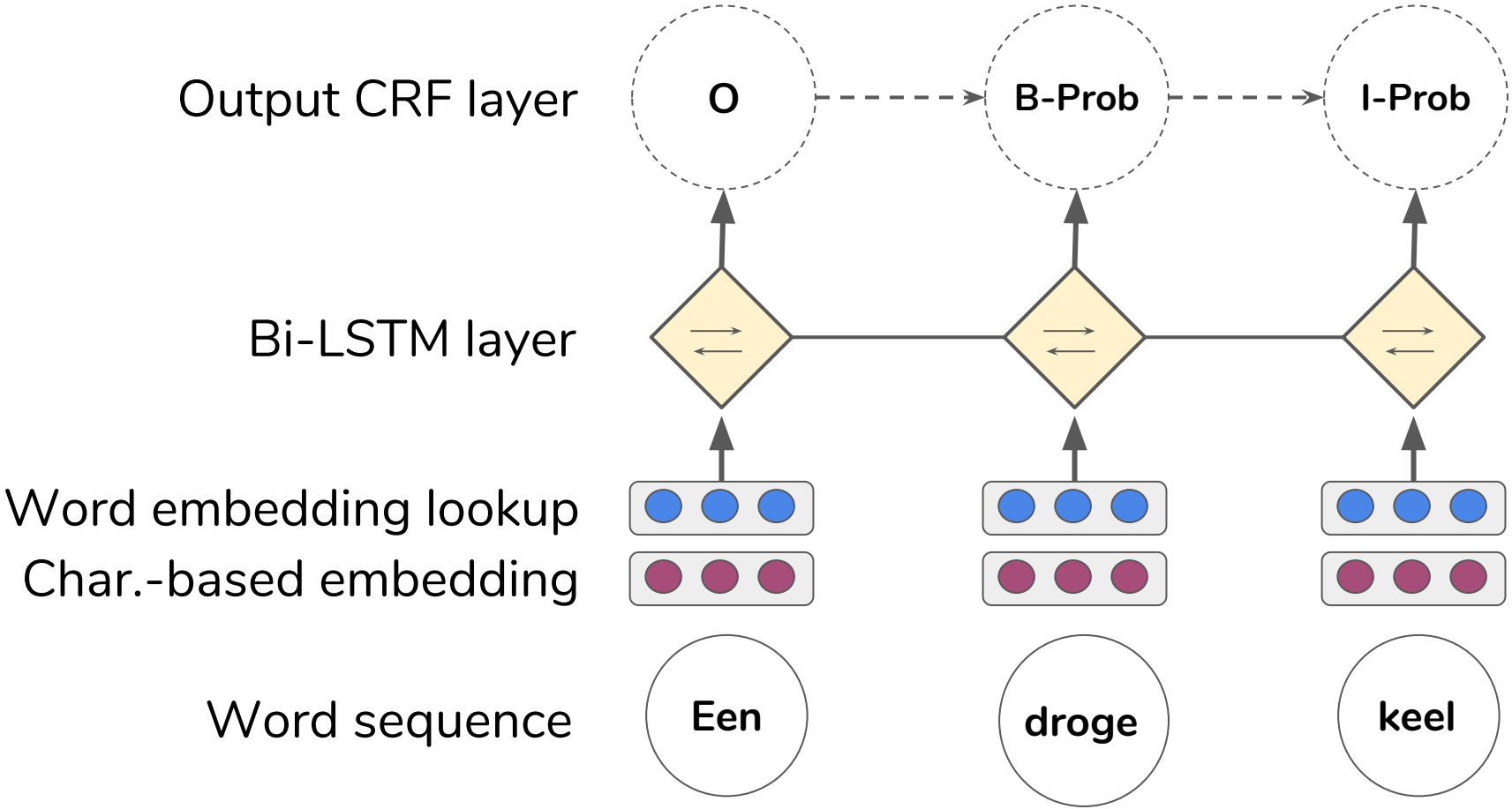
Concept-annotated data at UZA

15,900 × **test**
15,500 × **problem**
8,100 × **treatment**
8,100 × **anatomical site**
1,500 × **observation**
130 × **behavior**

~30,000 sentences in **training set**
~4,000 in **development set**
~4,000 in **test set**

\+ PHI identified
\+ concept negation and modality

Concept labeller (Lample et al. 2016)



Concept extraction results

Label	Precision	Recall	**F-1**
Test	.71	.70	**.71**
Problem	.75	.74	**.74**
Treatment	.78	.72	**.75**
Anatomical	.62	.60	**.61**
Observation	.45	.37	**.41**
Behavior	.75	.69	**.72**

These results are for overarching concepts, ignoring potential embedded concepts

Practical

Training takes a couple of days, but inference is fast

Implementation licensed under Apache-2.0

Trained model available at UZA

Further work

Add identification of embedded concepts

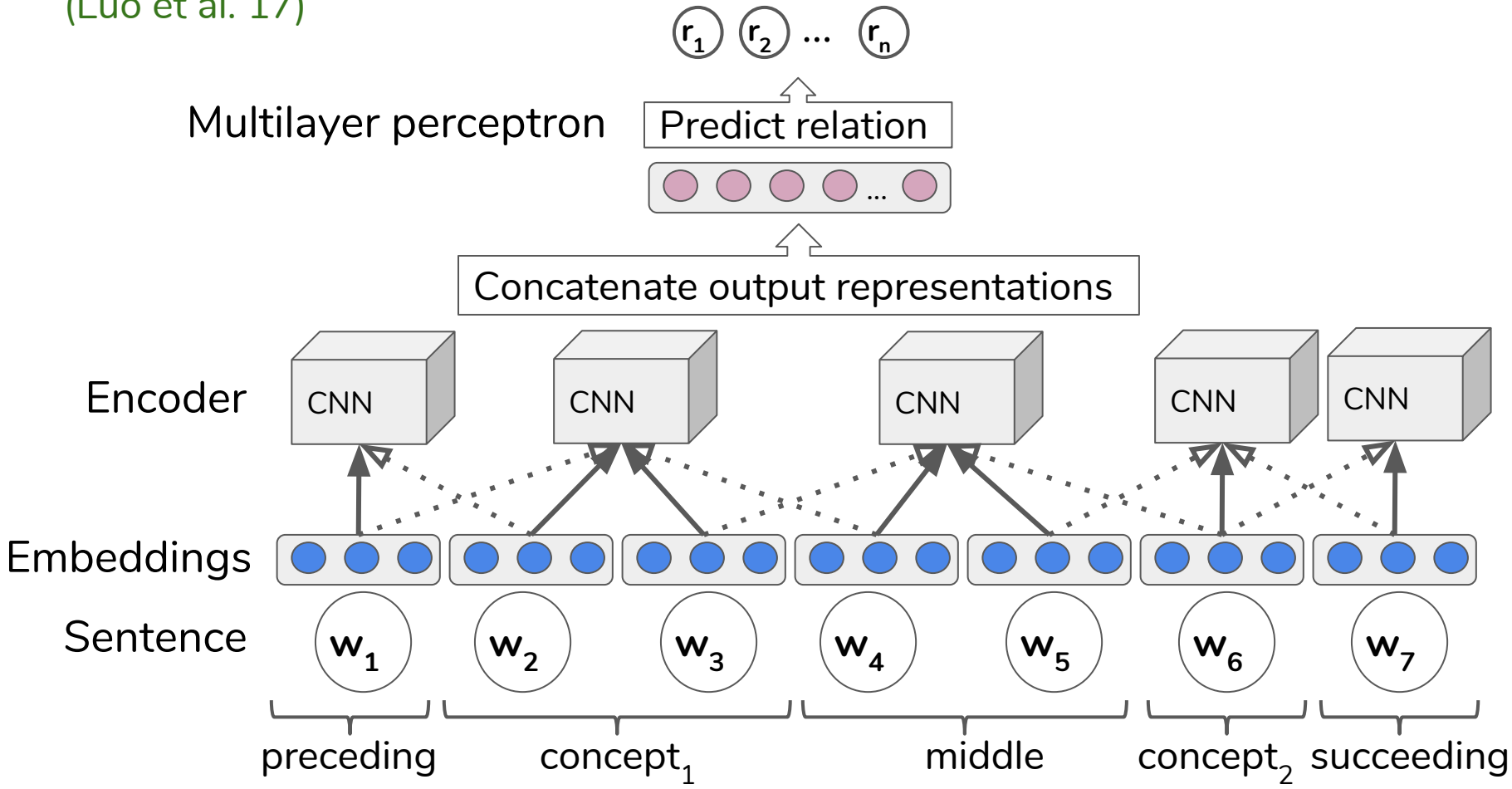
Compare to classifiers with hand-crafted features

Train an attribute classifier for negation and modality

Supervised relation extraction with external features

Relation extraction with Segment convolutional NNs

(Luo et al. 17)



Set of relations from i2b2-2010 English dataset

TrP: Treatment { improves, worsens, causes, is administered for, is not administered for } Problem

TeP: Test { reveals, is carried out for } Problem

PP: Problem { is related to } Problem

Relation extraction example

“Acetaminophen 325 mg Tablet Sig : Two (2) Tablet PO Q6H (every 6 hours) as needed for fever or pain”

“Acetaminophen” : ”fever” → Treatment administered for a problem
“Acetaminophen” : ”pain” → Treatment administered for a problem

Improving the relation extractor: error analysis

Analyze confusion matrices, e.g. for test-problem relations

gold\system	None	TeCP	TeRP
None	**575**	17	294
TeCP	41	**52**	36
TeRP	89	9	**612**

- Poor sensitivity
  - Missing a relation: “pt. was started on **zosyntr**_TREATMENT for suspected biliary obstruction and **cholangitis**_PROBLEM”
- False alarm
  - “pt. was treated with **tylenol**_TREATMENT orally as well as ativan for **anxiety**_PROBLEM”
- Confusables

Improving the relation extractor: external features

Drug-problem association

Extract knowledge from Drugbank indications and ADRs

Concept-concept association

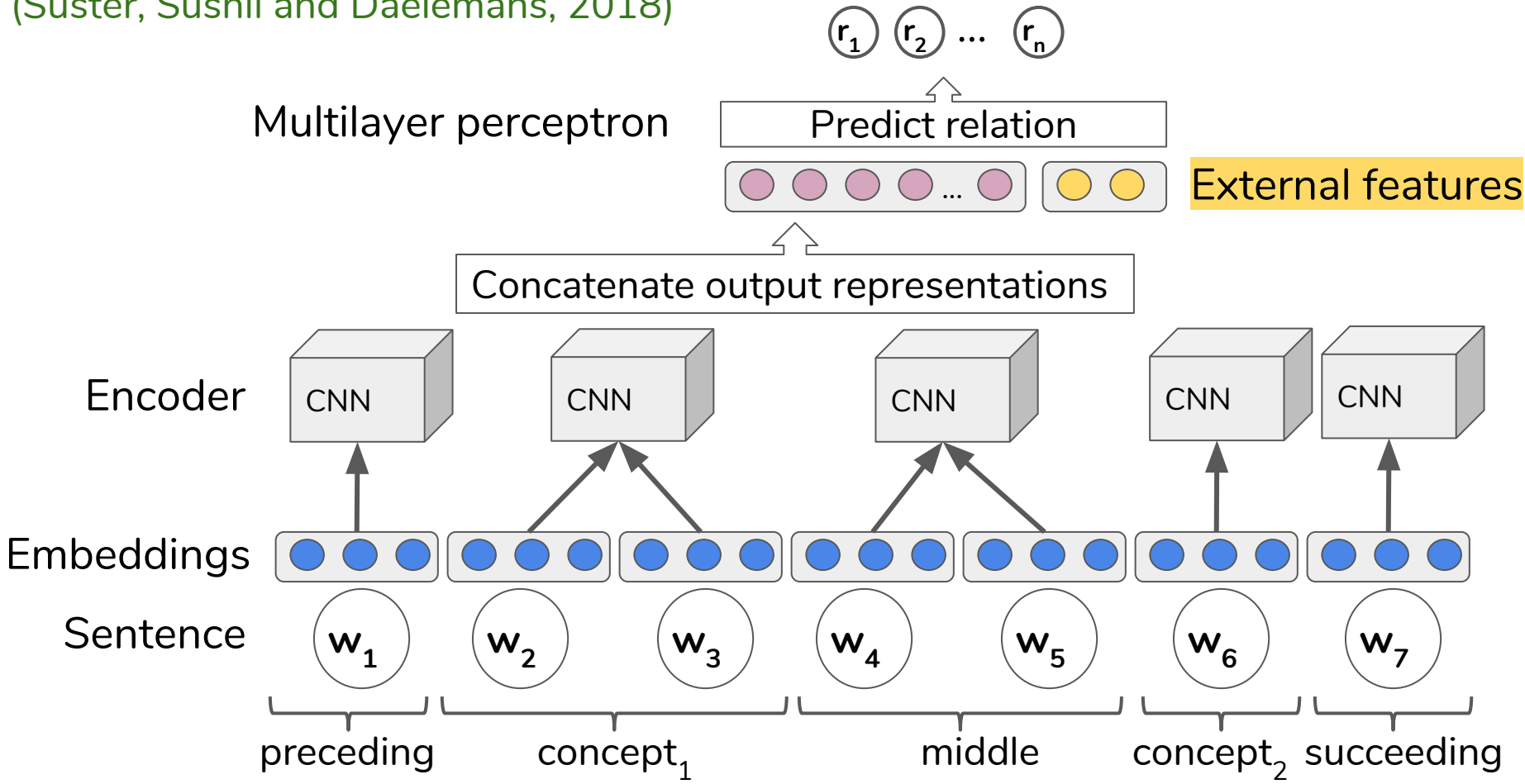
Estimate statistical associations from MIMIC-III using pointwise mutual information (PMI)

Semantic classes

Group relation-triggering terms based on WordNet and thesauri

Relation extraction with external features

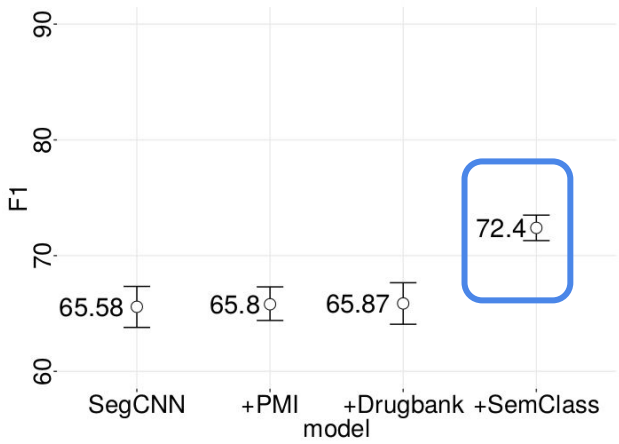
(Šuster, Sushil and Daelemans, 2018)



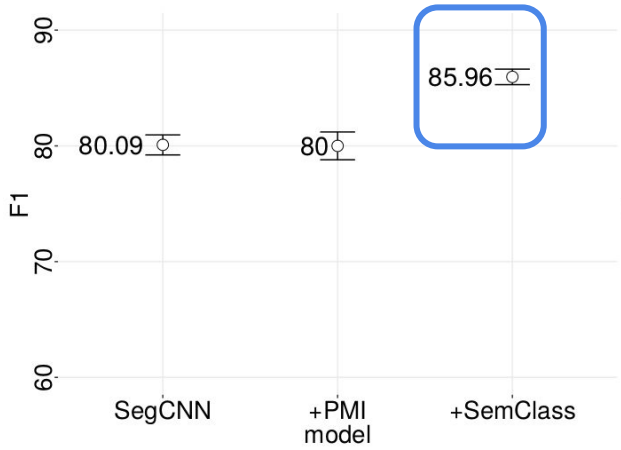
Results

1) Large gains from semantic classes
2) Only minor effect of drug-problem & concept-concept associations

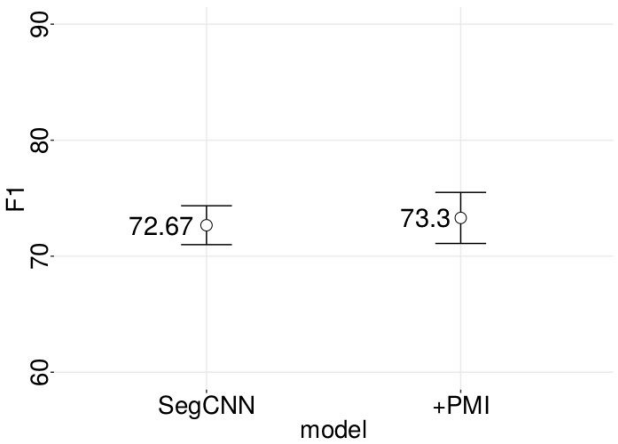
Treatment-Problem



Test-Problem



Problem-Problem



Software

https://github.com/clips/accumulate

https://github.com/SimonSuster/seg_cnn

https://github.com/glample/tagger