Ghent University at TAC KBP

Distributed Representations of Relation Paths to Bootstrap Relation Extractors

Lucas Sterckx

Ghent University - iMinds

September 27, 2017

Overview

TAC Knowledge Base Population

UNIVERSITEIT
GENT

Goals : Organized by NIST to promote research in he extraction of binary relations between named and numeric entities from free text.

- ▶ **English Slot Filling** (2009-2014): extract values of specified attributes for a given entity from a large collection of natural language texts,
  e.g., (Teresa Heinz, LTW_ENG_20070425, `per:spouse`, ?).
- ▶ **Cold Start KBP** - **Slot Filling Variant** (2015-...): Predict multiple hops of slot fillers

Relation Extractors

UNIVERSITEIT
GENT
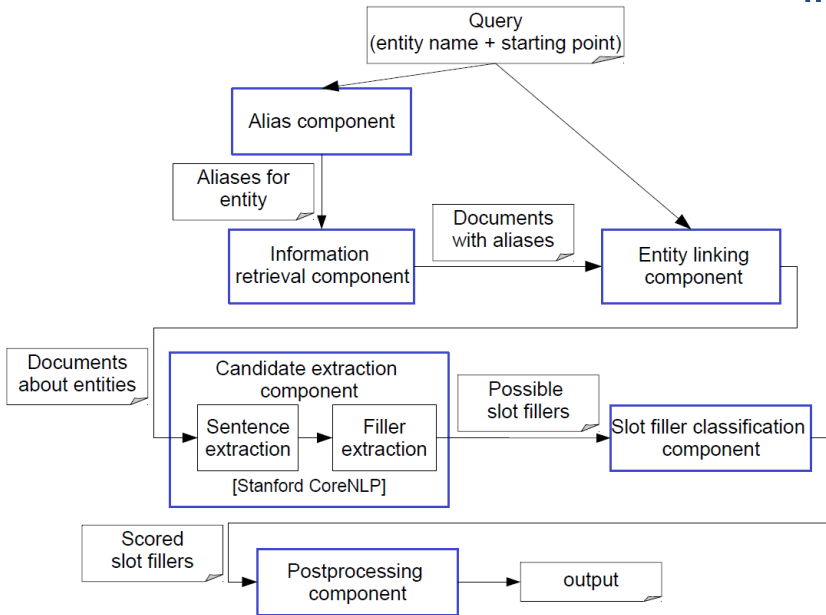
- Patterns
  - Manual Engineering (Semgrex, Tokensregex, . . .)
  - OpenIE
  - Universal Schema
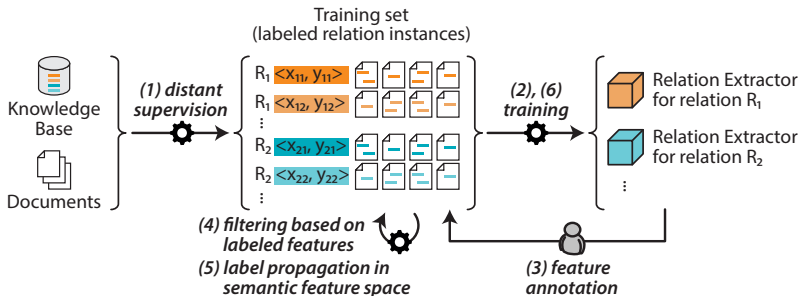 - Coreferences (*-year-old, *-based, . . .)

Relation Classifiers

- Linear classifier + sparse feature vectors
- Neural network sentence classifiers (LSTM, CNN, . . .)

Training:

- Distant Supervision
- Semi-Supervision, Bootstrapping
- Supervision

Ensembles!

Guiding Bootstrapped Relation Extractors

UNIVERSITEIT
GENT

Training set
(labeled relation instances)

(1) distant supervision

R_1 $<x_{11}, y_{11}>$
R_1 $<x_{12}, y_{12}>$
⋮
R_2 $<x_{21}, y_{21}>$
R_2 $<x_{22}, y_{22}>$
⋮

(2), (6) training

Relation Extractor
for relation R_1

Relation Extractor
for relation R_2

⋮

Knowledge Base

Documents

(4) filtering based on labeled features
(5) label propagation in semantic feature space
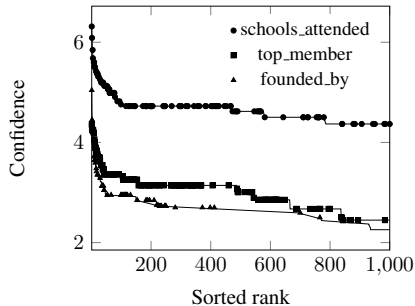
(3) feature annotation
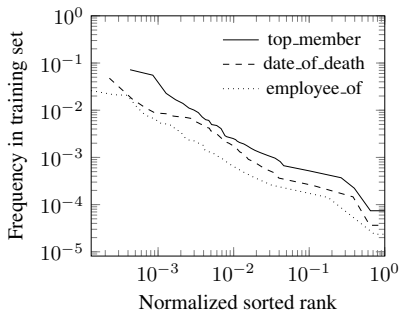
Guiding Bootstrapped Relation Extractors

Table: Examples of top-ranked patterns

Relation	Top SDP	Assessment
top_members_employees	PER $\xleftarrow{appos}$ executive $\xrightarrow{prep\_of}$ ORG	✓
	PER $\xleftarrow{appos}$ chairman $\xrightarrow{appos}$ ORG	✓
	ORG $\xleftarrow{nn}$ founder $\xrightarrow{prep\_of}$ PER	✗
children	PER-2 $\xleftarrow{appos}$ son $\xrightarrow{prep\_of}$ PER-1	✓
	PER-1 $\xleftarrow{appos}$ father $\xrightarrow{prep\_of}$ PER-2	✓
	PER-2 $\xleftarrow{nn}$ grandson $\xrightarrow{prep\_of}$ PER-1	✗
city_of_birth	PER $\xleftarrow{rcmod}$ born $\xrightarrow{prep\_in}$ LOC	✓
	PER $\xleftarrow{nsubj}$ mayor $\xrightarrow{prep\_of}$ LOC	✗
	PER $\xleftarrow{appos}$ historian $\xrightarrow{prep\_from}$ LOC	✗
schools_attended	PER $\xleftarrow{nsubj}$ graduated $\xrightarrow{prep\_from}$ ORG	✓
	PER $\xleftarrow{dep}$ student $\xrightarrow{prep\_at}$ ORG	✓
	PER $\xleftarrow{appos}$ teacher $\xrightarrow{prep\_at}$ ORG	✗
(org:)parents	ORG-2 $\xleftarrow{appos}$ subsidiary $\xrightarrow{prep\_of}$ ORG-1	✓
	ORG-1 $\xleftarrow{appos}$ division $\xrightarrow{prep\_of}$ ORG-2	✓
	ORG-2 $\xleftarrow{prep\_to}$ shareholder $\xrightarrow{dep}$ ORG-1	✗

Research and Future Directions

Results
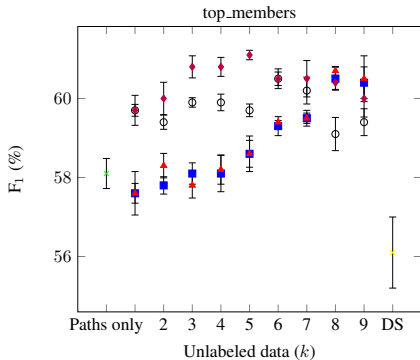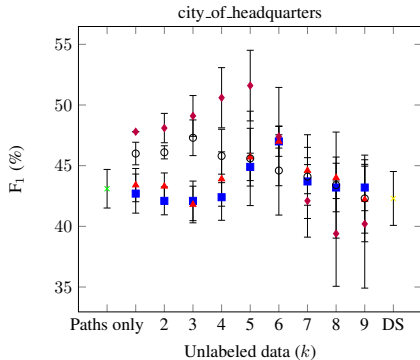
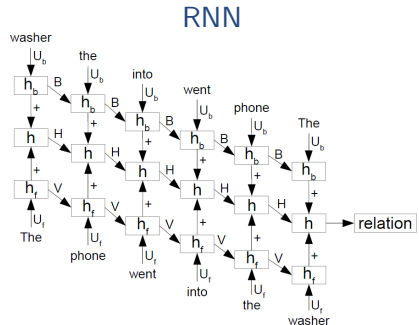| Relation	Distant Supervision (Mintz++)			SDP Filtered			SDP Guided Bootstrapping		
	P	R	F_1	P	R	F_1	P	R	F_1
title	22.3	58.8	32.3	36.1	39.1	37.5	37.3	41.2	39.2
top_members_employees	50.6	63.4	56.3	51.3	63.4	56.7	63.5	62.5	**63.0**
employee_or_member_of	31.4	34.0	32.6	33.8	51.0	**40.7**	27.9	51.0	36.1
age	71.6	72.5	72.0	75.6	70.0	72.7	68.8	82.5	**75.0**
origin	100.0	23.0	37.4	28.5	80.0	42.0	31.7	86.6	**46.4**
countries_of_residence	100.0	23.0	37.4	22.4	84.6	35.4	35.2	46.1	39.9
charges	45.0	52.9	48.6	40.9	52.9	46.1	44.3	68.1	53.7
cities_of_residence	22.9	45.8	30.5	31.5	25.0	27.9	34.4	41.6	**37.7**
cause_of_death	30.7	36.3	33.3	29.4	45.4	35.7	33.3	45.4	**38.4**
spouse	50.0	45.4	47.6	50.0	45.4	47.6	71.4	45.4	**55.5**
city_of_death	100.0	16.6	28.5	14.2	16.6	15.3	20.0	33.3	**25.0**
country_of_headquarters	22.7	41.6	29.4	62.5	41.6	**50.0**	100.0	33.3	**50.0**
date_of_death	66.6	50.0	**57.1**	66.6	50.0	**57.1**	66.6	50.0	**57.1**
TAC-scorer	29.3	28.1	28.7	35.5	33.7	34.7	36.9	35.9	**36.4**

Results
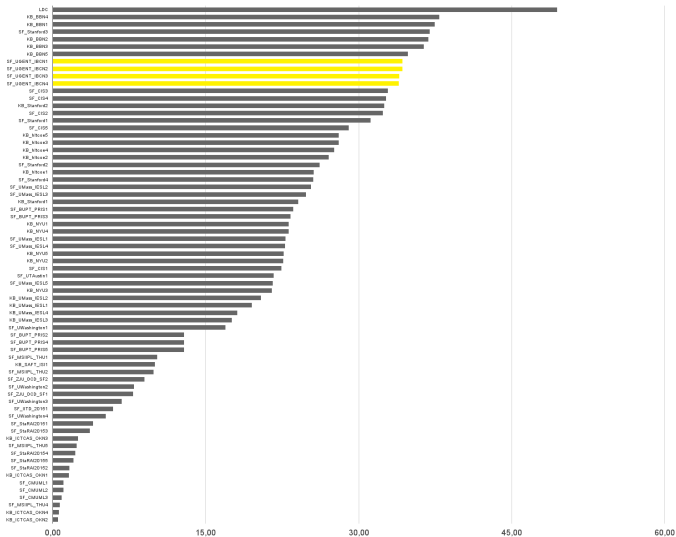
Research and Future Directions

- Neural Network sentence classifiers for relation extraction ((Kim, EMNLP 2014), (Liu, ACL 2015))

- Distant supervision for neural sentence classifiers ((Zeng, EMNLP 2015), (Lin, ACL 2016))

- Column-less, Row-less Universal Schema ((Verga, AKBC 2016),(Verga, NAACL 2016))

- Standardization of candidate sets ((Adel, NAACL 2016), (Verga , NAACL 2016))

- Distant Supervision + Partial Supervision ((Angeli, EMNLP 2014), (Pershina , ACL 2014))

Research and Future Directions

CNN

RNN
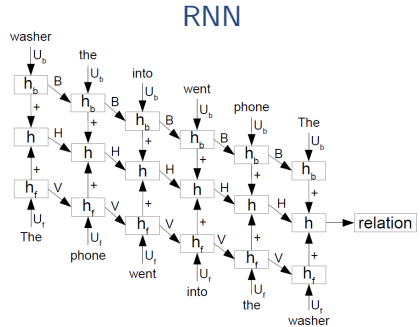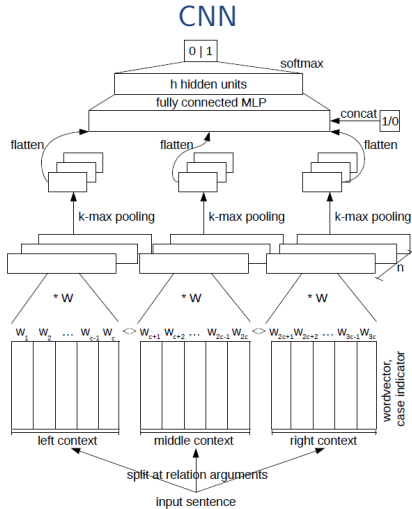
Research and Future Directions

- Attention Mechanism
- Using pattern-filtered trainingdata
- Use labeled patterns for biased training objective
- Neural bootstrapping
- Vector representations of relations
- ...

RNN

Thank you!
Questions?