Distantly Supervised Entity Relation Extraction with Adapted Manual Annotations

Changzhi Sun

East China Normal University

changzhisun@stu.ecnu.edu.cn

November 12, 2018

Overview

1. Introduction
2. Model
3. Experiments
4. Conclusions

Introduction

1. Entity Relation Extraction.
   - **supervised joint** learning algorithm.
   - annotations on entities and relations are limited.
2. Distant Supervision.
   - generating training data automatically.
   - aligned triples in KB and free texts.
   - **noisy** samples.
3. **Motivation**: adapt a small manually labelled dataset to the large automatically generated dataset.
4. **Challenge**: heterogeneous on datasets.
   - different types of entities and relations.
   - different guidelines.

Heterogeneous

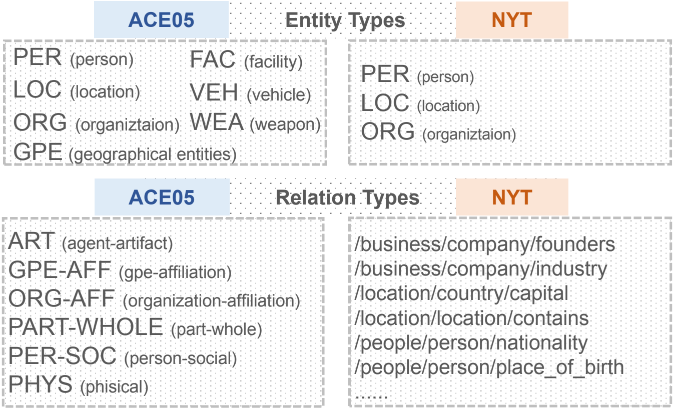


Figure: Datasets with different annotation schemas.

Heterogeneous

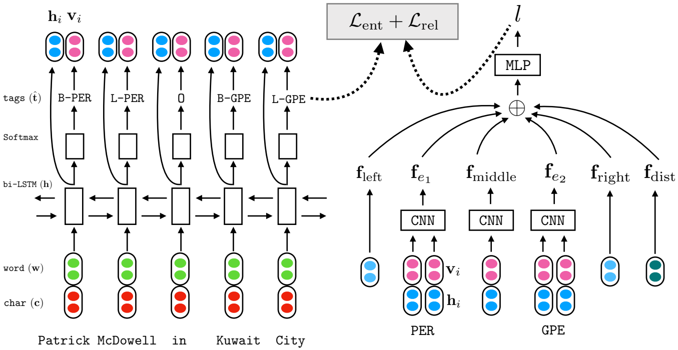
1. Merging datasets.
    - ignores inherent differences between the two schemas.
2. Building (approximate) mapping.
    - inflexible to extend to more schemas.

In this paper, we propose a novel fine-grained framework to incorporate manually labelled dataset to the distantly supervised learning process

1. Introduce two shared tasks.
    - bridge discrepancies between two heterogeneous annotation schemas.
2. consistency constraints between shared tasks and original tasks.

Contributions

1. We first investigate the problem of using manually labelled datasets to help distantly supervised entity relation extraction.
2. We design a novel adaptation framework which transfers heterogeneous knowledge through new shared tasks.
3. Our proposed model significantly outperforms the state-of-the-art methods on benchmark NYT dataset.



- biLSTM-based sequence labelling model for detecting entities ($\mathcal{M}_{\text{seq}}$).
- CNN-based multi-class classifier for detecting relations ($\mathcal{M}_{\text{rel}}$).

Adapting via Shared Representations

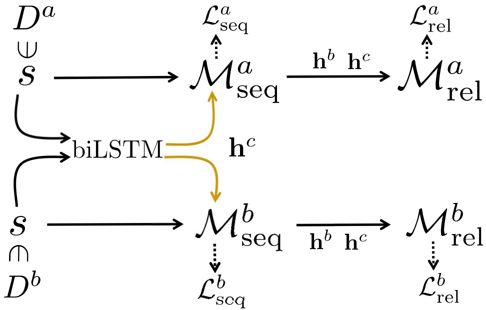


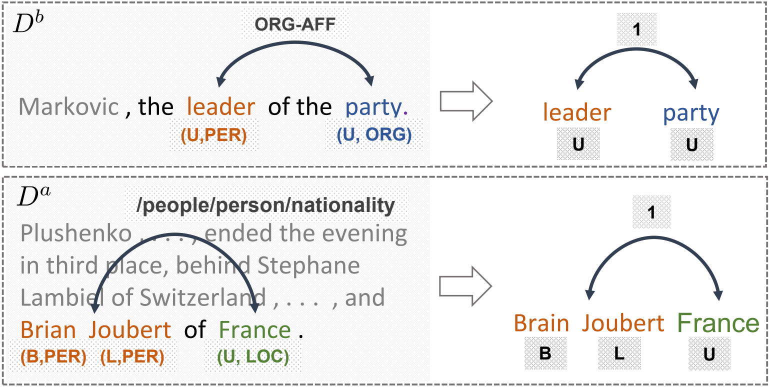
Figure: Adapting via shard representations $\mathbf{h}^c$.

- superscript a: large automatically labeled dataset.
- superscript b: small manually labeled dataset.
- superscript c: shared.

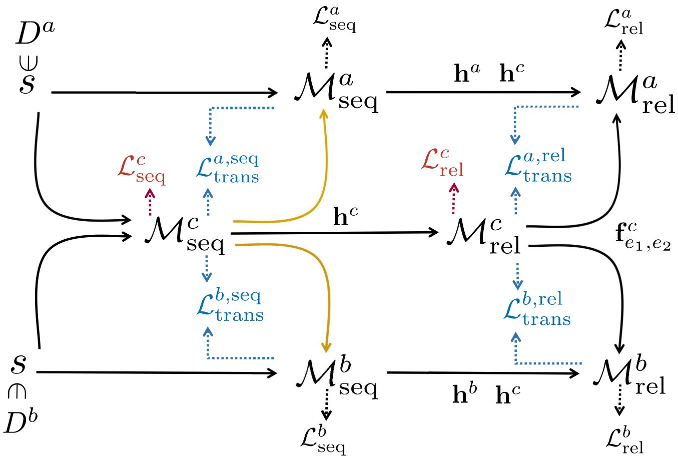
Adapting via Shared Tasks

1. **Entity span detection**.
    - focus on locating boundaries of entities.
2. **Binary relation detection**.
    - predict whether a certain relation exists between an entity pair.

Adapting via Shared Tasks



Adapting via Shared Tasks



1. **Margin constraint**:
   - a natural consistency requirement between P^a_{seq} and P^c_{seq}
   - $\mathcal{T}^a_e$: the set of possible entity types

$$P^c_{\mathsf{seq}}(\star|s) = \sum_{*\in\mathcal{T}^a_e} P^a_{\mathsf{seq}}((\star,*)|s), \quad \star \in \{\mathtt{B},\mathtt{I},\mathtt{L},\mathtt{O},\mathtt{U}\}.$$

2. **Transition loss**:
   - $M^a_{\mathrm{seq}}, M^b_{\mathrm{seq}}$ are transition matrices which convert discrete distributions $P^a_{\mathrm{seq}}, P^b_{\mathrm{seq}}$ to P^c_{seq}.

$$\mathcal{L}^{\nabla,\mathsf{seq}}_{\mathsf{trans}} = \|P^c_{\mathsf{seq}} - M^{\nabla}_{\mathsf{seq}} P^{\nabla}_{\mathsf{seq}}\|_2, \quad \nabla \in \{a,b\}.$$

   - similarly

$$\mathcal{L}^{\nabla,\mathsf{rel}}_{\mathsf{trans}} = \|P^c_{\mathsf{rel}} - M^{\nabla}_{\mathsf{rel}} P^{\nabla}_{\mathsf{rel}}\|_2, \quad \nabla \in \{a,b\}.$$

Experiments

Model	Relation P	R	F
Gormley (2015)	55.3	15.4	24.0
Mintz (2009)	25.8	39.3	31.1
Tang (2015)	33.5	32.9	33.2
Hoffman (2011)	33.8	32.7	33.3
L&J (2014)	57.4	25.6	35.4
Ren (2017)	42.3	**51.1**	46.3
Zheng (2017)	61.5	41.4	49.5
Wang (2018)	64.3	42.1	50.9
Sun (2018)	67.4	42.0	51.7
Our Model	**70.4**	45.6	**55.4**
Sun (2018)(exactly match)	65.2	40.6	50.0
Our Model(exactly match)	68.3	44.2	53.7

Table: Results on the NYT dataset.

Experiments

Model	Relation P	R	F
Gormley (2015)	55.3	15.4	24.0
Mintz (2009)	25.8	39.3	31.1
Tang (2015)	33.5	32.9	33.2
Hoffman (2011)	33.8	32.7	33.3
L&J (2014)	57.4	25.6	35.4
Ren (2017)	42.3	**51.1**	46.3
Zheng (2017)	61.5	41.4	49.5
Wang (2018)	64.3	42.1	50.9
Sun (2018)	67.4	42.0	51.7
Our Model	**70.4**	45.6	**55.4**
Sun (2018)(exactly match)	65.2	40.6	50.0
Our Model(exactly match)	68.3	44.2	53.7

Table: Results on the NYT dataset.

Model	Entity P	Entity R	Entity F	Relation P	Relation R	Relation F
only D^a	82.6	91.2	86.7	61.8	43.3	50.9
$D^a \cup D^b$	84.5	91.9	88.1	60.7	42.3	49.8
only $\mathbf{h}^c$	83.5	93.1	88.0	65.6	42.0	51.2
$\mathbf{h}^c + \mathcal{L}^c$	86.2	92.5	89.2	66.5	45.4	53.9
$+ \mathcal{L}_{\mathrm{trans}}$	82.8	89.6	86.1	64.8	46.2	54.0
$+ \mathcal{L}_{\mathrm{trans}}^{\mathrm{seq}}$	86.6	92.9	89.6	70.4	**45.6**	**55.4**
$+ \mathcal{L}_{\mathrm{trans}}^{\mathrm{rel}}$	**87.2**	**93.9**	**90.4**	**72.9**	40.9	52.4

Table: Results on the NYT dataset in different settings. "only D^a" uses basic models trained on D^a; "$D^a \cup D^b$" denotes the adaptation model via merging datasets; "only $\mathbf{h}^c$" denotes adaptation model via shared representations; "$\mathbf{h}^c + \mathcal{L}^c$" is the adaptation model via shared task. "$+ \mathcal{L}_{\mathrm{trans}}$" is the "$\mathbf{h}^c + \mathcal{L}^c$" with entity and relation tracontent...nsition losses; Similarly, "$+ \mathcal{L}_{\mathrm{trans}}^{\mathrm{seq}}$" and "$+ \mathcal{L}_{\mathrm{trans}}^{\mathrm{rel}}$" keep only transition losses on entity and relation respectively.

Conclusions

1. We propose a novel adaptation framework for distantly supervised joint entity relation extraction using the high quality heterogeneous dataset.
2. Introducing two shared tasks and imposing consistency constraints between shared tasks and original task.
3. Our framework could control the adaptation process in a more careful and interpretable way.
4. Experiments on benchmark NYT dataset show the effectiveness of the proposed methods.

Thanks and Q & A