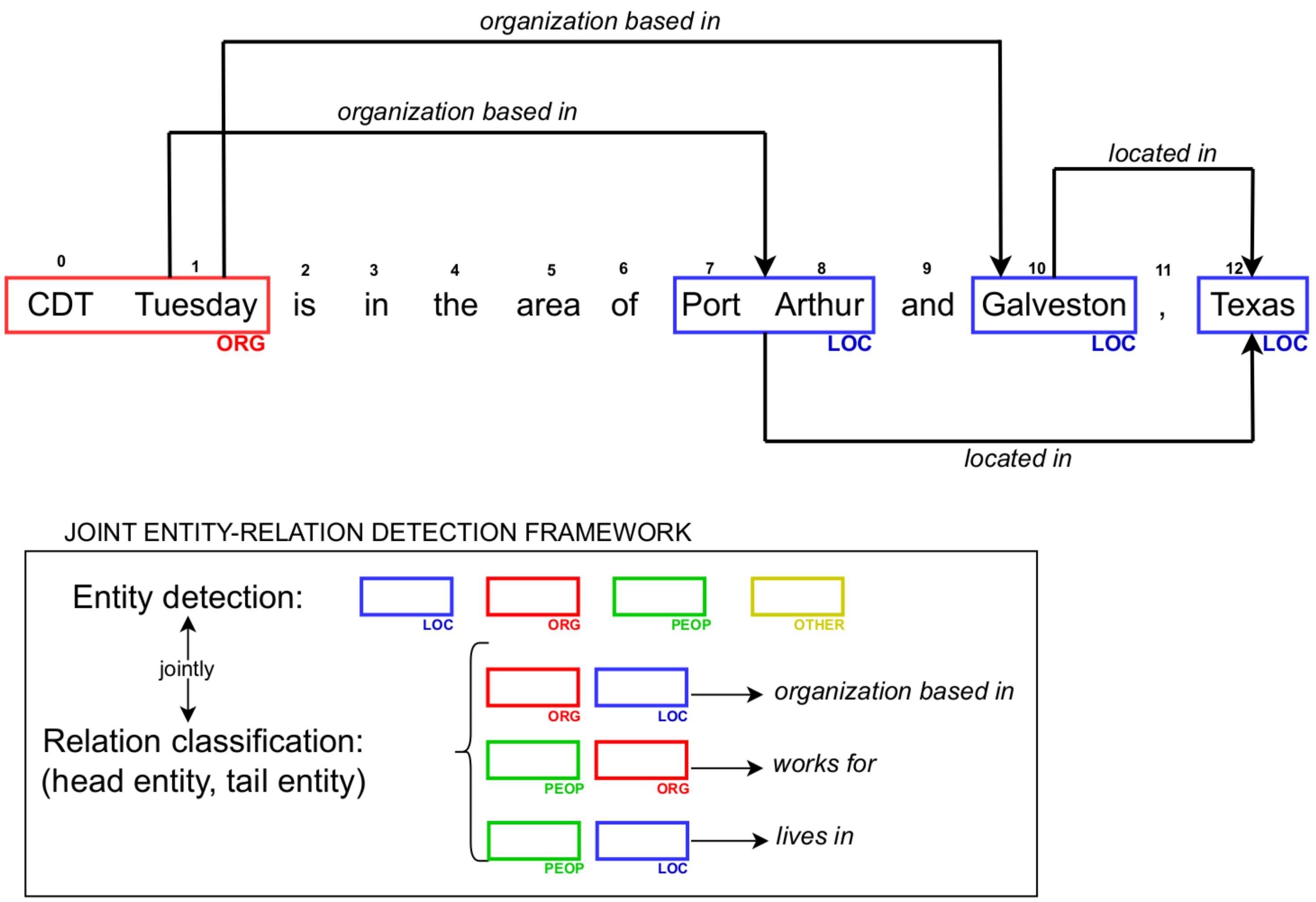
ASPER: Answer Set Programming Enhanced Neural Network Models for Joint Entity-Relation Extraction

Trung Hoang Le, Huiping Cao, Tran Cao Son

New Mexico State University

Research partially funded by NSF awards #1757207, #1914635, and #1812628.

Joint Entity-relation Extraction

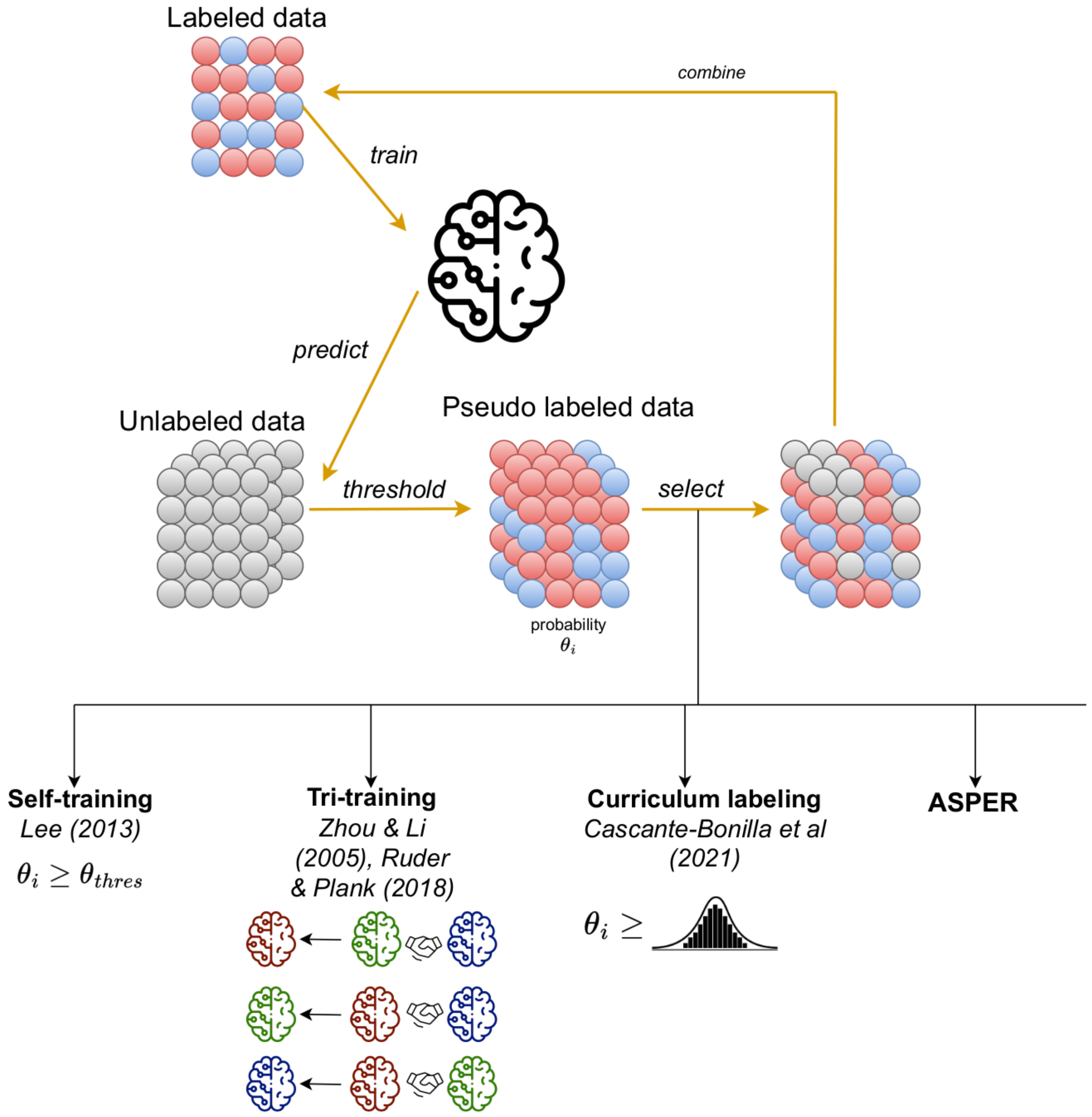


Semi-supervised learning

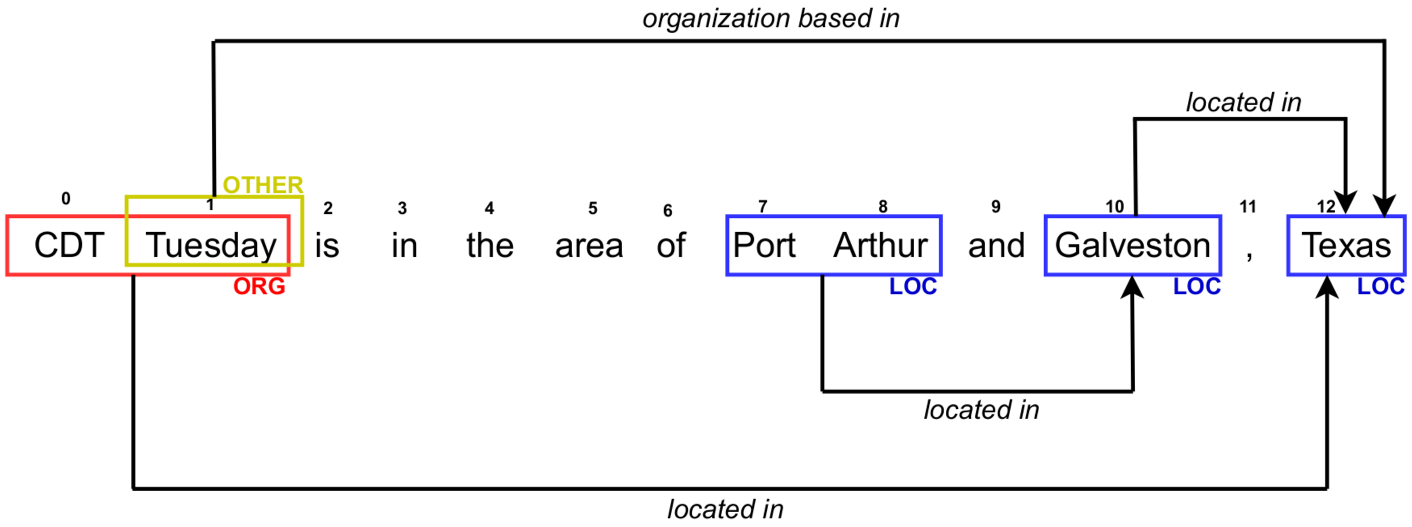
While **Labeled** data is expensive.

Unlabeled data is cheap and plentiful (e.g. Wikipedia dump).

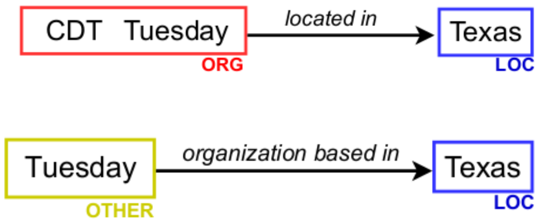
Semi supervised learning (SSL): utilizes both labeled and unlabeled data to improve predictions.



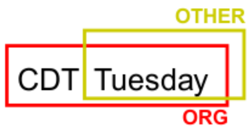
ASPER



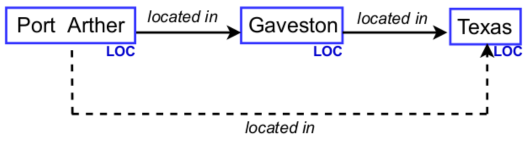
- Inconsistencies:



- Overlapped entities:



- Hidden relations:

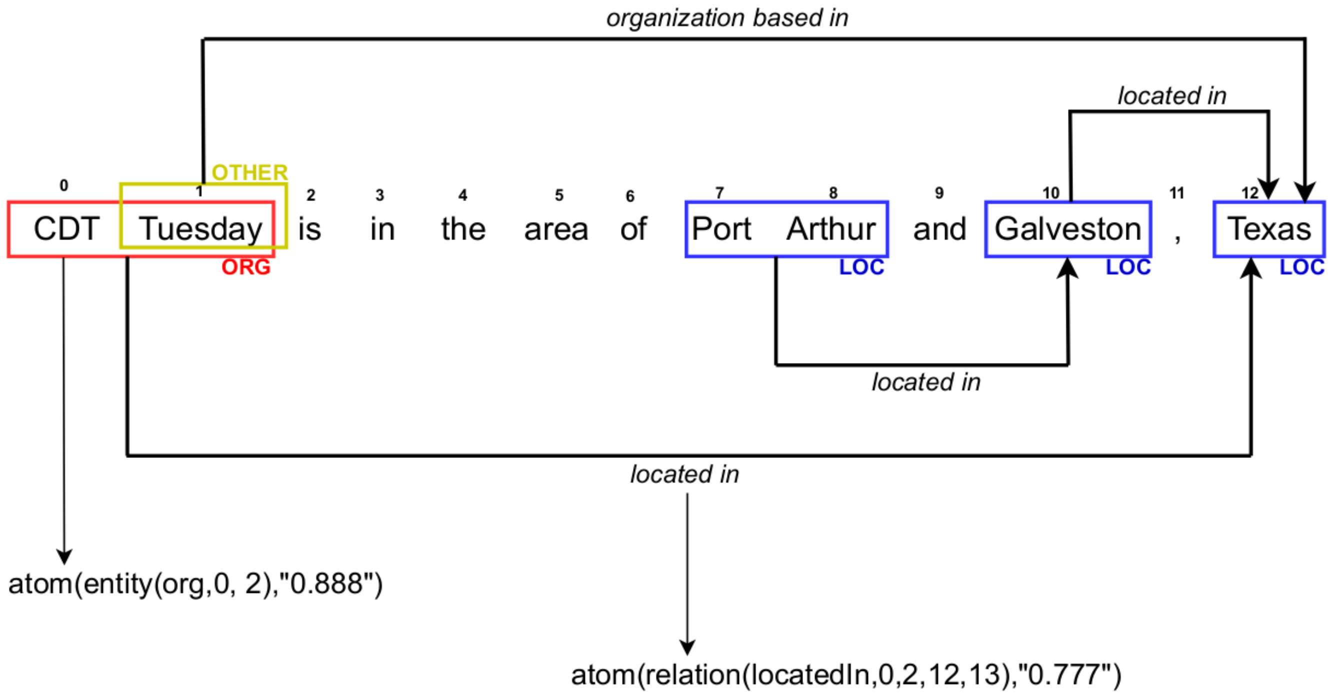


There are many consistent sets of pseudo labels.

1. How to find all of them?
2. Which set is "the best" to include in the training set?

ASPER

- Pseudo labels as atoms
- Three special predicates:
  - ok(X): accept X
  - nok(X): rejects X
  - pi(X): X could be rejected

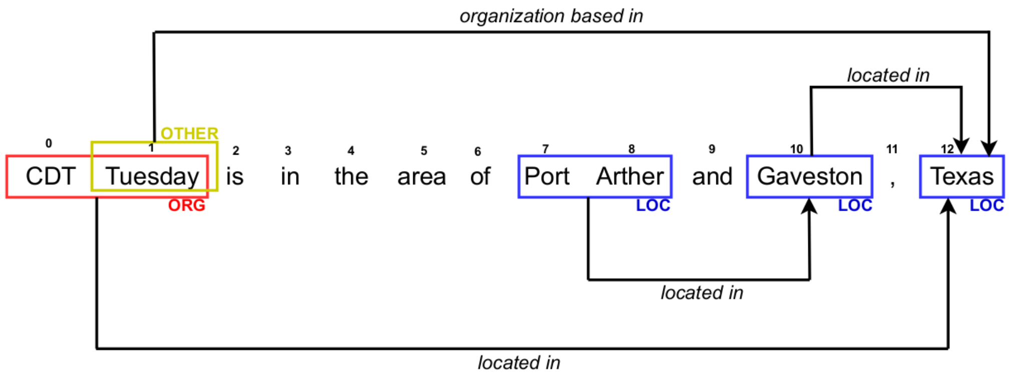


- Knowledge are encoded as rules:

type_def(liveIn, peop, loc).
type_def(locatedIn, peop, loc).
type_def(orgbasedIn, org, loc).
type_def(workFor, peop, org).
type_def(kill, peop, peop).

2{pi(relation(R,B,E,B',E');pi(entity(N,B,E))}<- type_def(R,N1,N2), atom(relation(R,B,E,B',E')), atom(N,B,E)),N1 != N.
<- ok(relation(R,B,E,_,_)), ok(entity(N,B,E)), type_de(R,N',_), N!= N'.

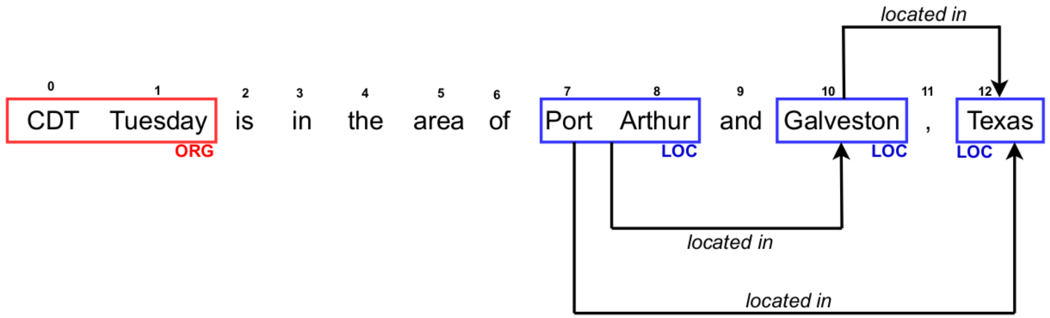
ASPER



- Each answer set contains ok(…) atoms.

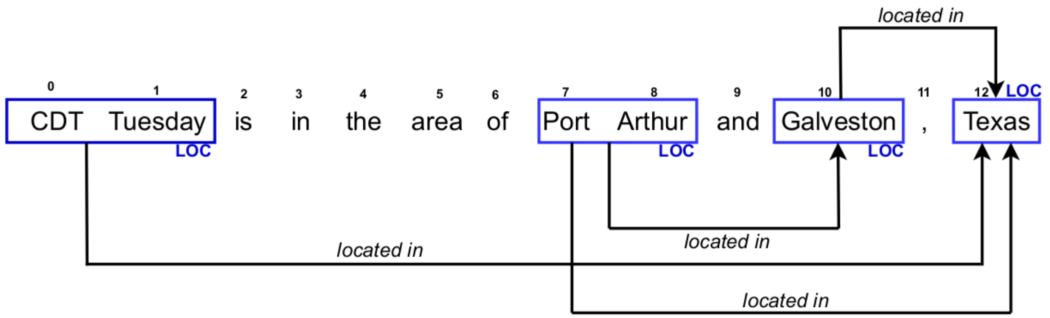
ok(entity(org,0,2))
ok(entity(loc,7,9))
ok(entity(loc,10,11))
ok(entity(loc,12,13))

ok(relation(locatedIn,7,9,10,11))
ok(relation(locatedIn,10,11,12,13))
ok(relation(locatedIn,7,9,12,13))



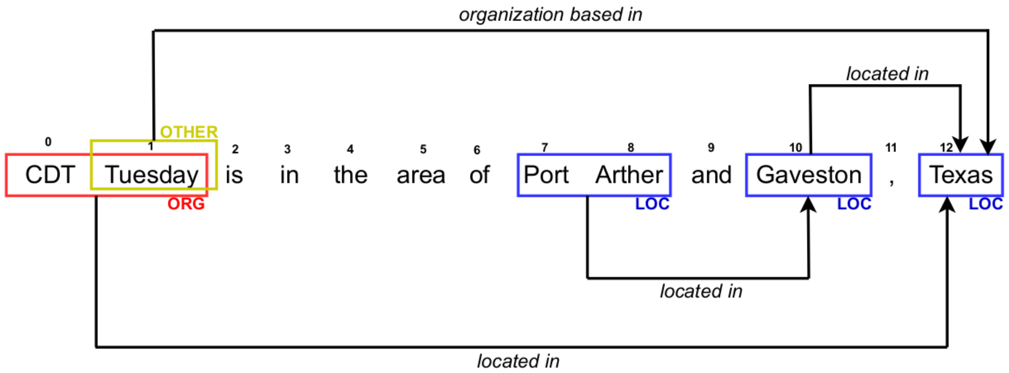
ok(entity(loc,7,9))
ok(entity(loc,10,11))
ok(entity(loc,12,13))
ok(entity(loc,0,2))

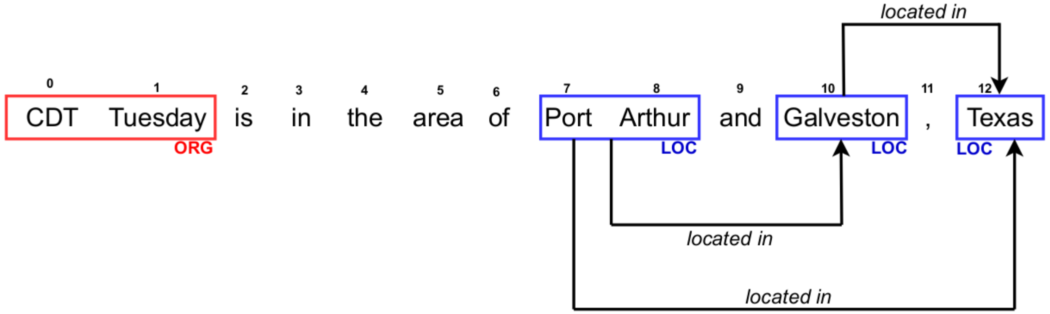
ok(relation(locatedIn,7,9,10,11))
ok(relation(locatedIn,10,11,12,13))
ok(relation(locatedIn,7,9,12,13))
ok(relation(locatedIn,0,2,12,13))

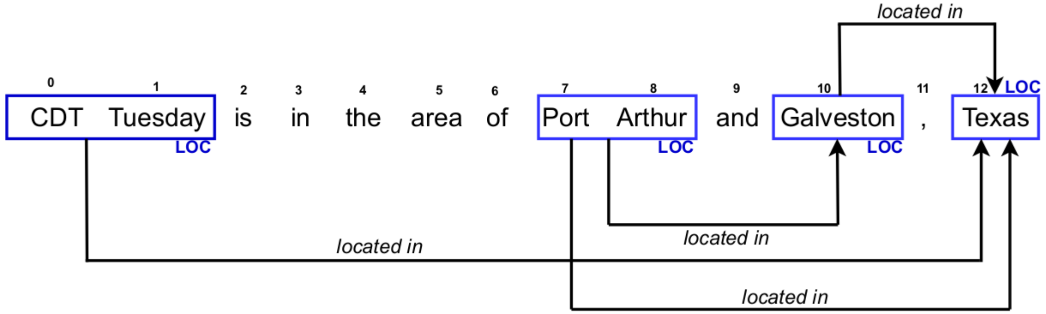


ASPER

- **Which** answer set is better?







While **both** are **consistent** 2nd **seems** to be **better** because it keeps as much as what the model **originally predicted**.

ASPER

- Selection of consistent pseudo labels

  prob(X,P) <- atom(X,P), ok(X)

  invprob(X,P) <- atom(X,P), nok(X).

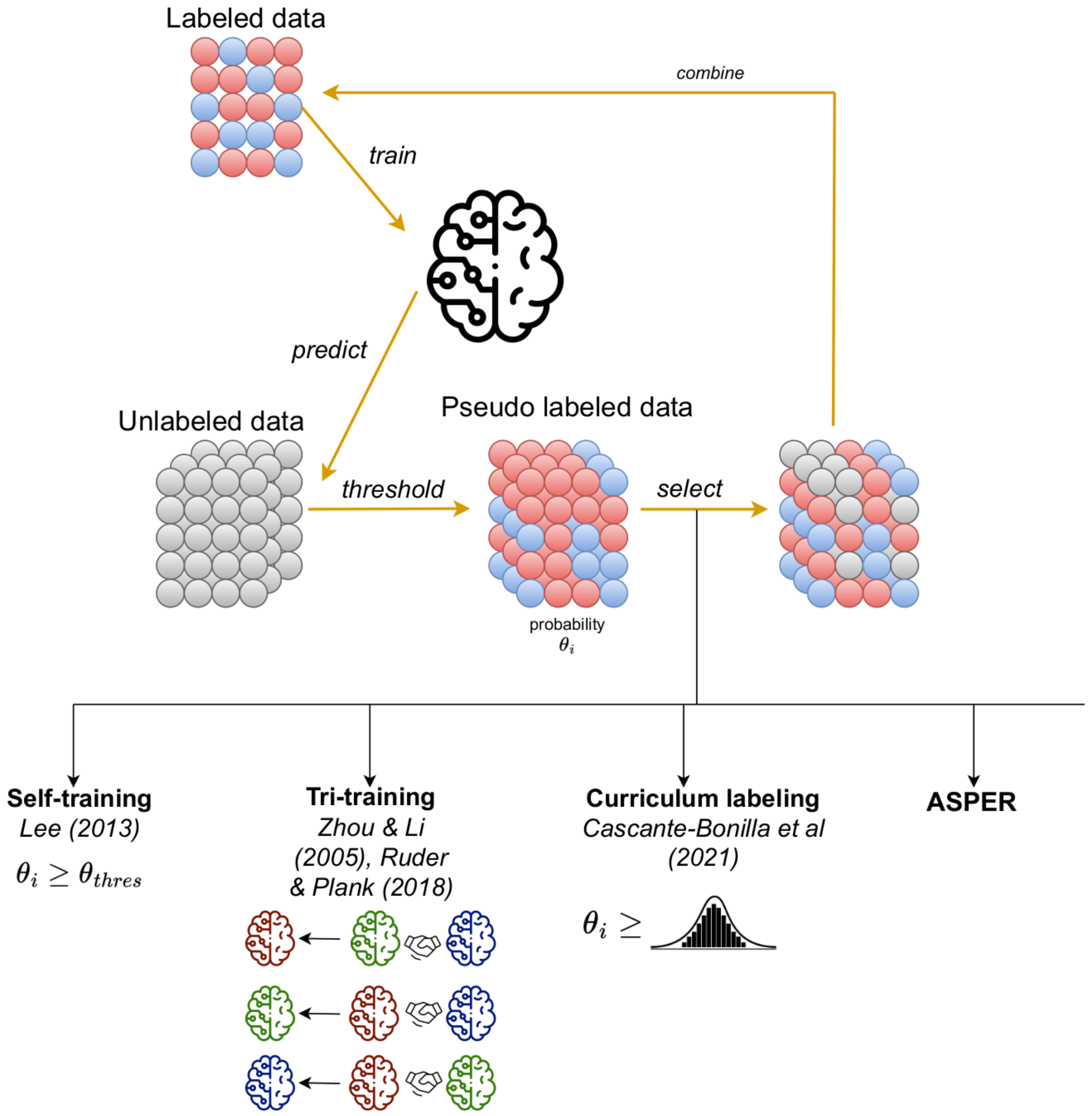
- Each answer set has two scores
  - Preference:

    $\prod_{prob(a,p) \in W} p \times \prod_{invprob(a,p) \in W} (1-p)$

  - Confidence:

    min{p | prob(l,p) ∈ W}

- Answer sets are selected based on preference.
- Sentences are selected based on confidence.



Experiments

- We use two datasets, **CoNLL04** and **SciERC** which have been utilized in other entity/relation extraction work.
- The **CoNLL04** dataset extracted from **newspapers** with train/dev/test split as **922/231/288** sentences.
- The **SciERC** dataset extracted from **artificial intelligence research papers** with train/dev/test split as **1861/275/551** sentences.
- We use a portion of training set (10%) as labeled data and the rest as unlabeled data.
- To get stable results, we randomly choose **five subsets** from the training data and train five models and report the average result.
- Comparison methods: **Self-training** (Lee 2013), **Curriculum labeling** (Cascante-Bonilla et al 2021) and **Tri-training** (Zhou&Li 2005, Ruder and Plank 2018).

Experiments

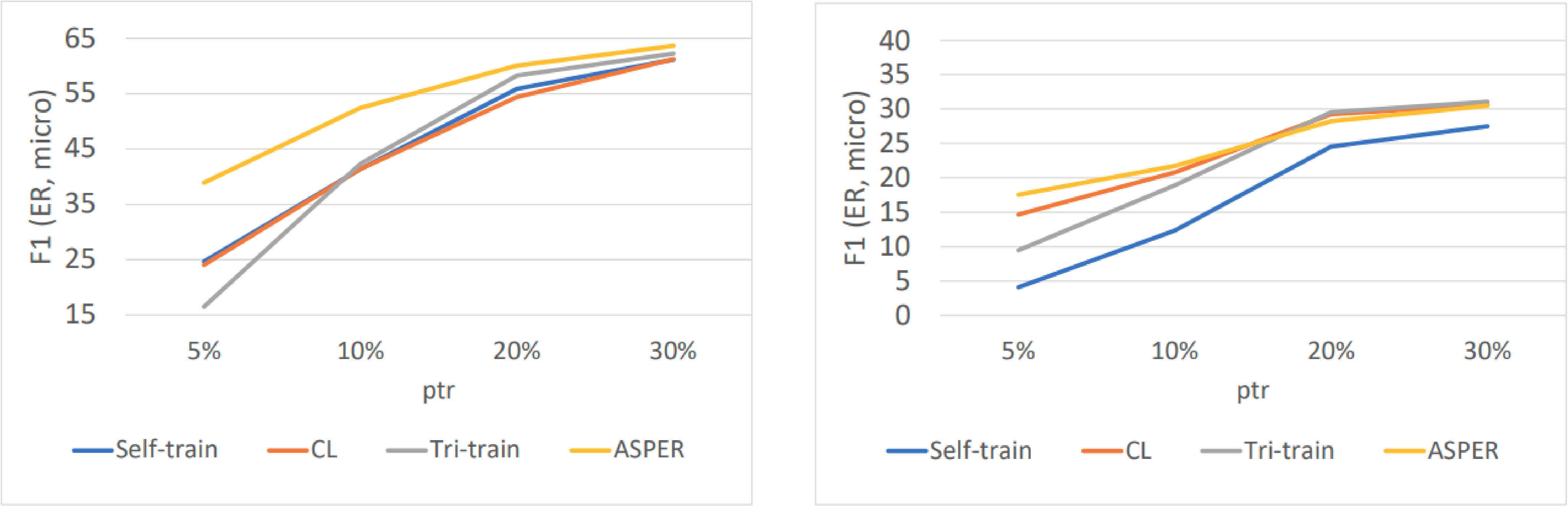
Performance of ASPER with 10% labeled data

	CoNLL04 dataset					
	F_1 (micro)			F_1 (macro)		
method	E	R	ER	E	R	ER
Self-train	77.74±1.7	41.76±5.7	41.39±5.7	72.50±1.9	43.19±6.0	42.82±6.0
CL	77.49±1.1	41.61±3.0	41.35±3.2	72.03±1.6	43.07±3.8	42.77±4.0
Tri-train	78.63±2.4	42.60±6.7	42.29±6.7	72.49±2.5	42.99±7.1	42.64±7.2
ASPER	**81.25**±1.2	**52.47**±3.6	**52.41**±3.6	**75.90**±1.7	**53.32**±4.0	**53.27**±4.0

	SciERC dataset					
	F_1 (micro)			F_1 (macro)		
method	E	R	ER	E	R	ER
Self-train	56.72±1.2	18.60±2.6	12.36±1.7	54.43±1.4	11.07±3.7	6.98±2.3
CL	60.75±0.8	31.00±2.1	20.81±1.0	59.19±0.4	22.00±3.8	15.55±1.8
Tri-train	**60.99**±0.7	27.43±1.9	18.94±1.4	**59.52**±0.4	17.09±3.6	11.59±2.7
ASPER	60.34±0.6	**32.30**±1.2	**21.73**±1.2	59.10±0.4	**22.72**±3.1	**16.06**±2.3

Experiments

ASPER's performance when we vary the portion of labeled data



(a) CoNLL04

(b) SciERC

Conclusions

- ASPER leverages ASP to improve NN models in the joint recognition of entities and relation task when limited amount of training data is available.
- The ASP program encodes different types of commonsense rules by taking advantage of the commonsense domain knowledge.
- The experiments on two real datasets show that ASPER can report significantly better results than the other baselines in most cases.
- ASPER is a framework that can be extended for Semi-supervised learning when pseudo labels have clear semantic.

References

1. D.-H. Lee, “Pseudo-label: The simple and efficient semi supervised learning method for deep neural networks,” in Workshop on challenges in representation learning, ICML, vol. 3, 2013.

2. ZHOU, Z. AND LI, M. 2005. Tri-training: Exploiting unlabeled data using three classifiers. IEEE Trans. Knowl. Data Eng. 17, 11, 1529–1541.

3. RUDER, S. AND PLANK, B. 2018. Strong baselines for neural semi-supervised learning under domain shift. In ACL 2018. Association for Computational Linguistics, 1044–1054.

4. CASCANTE-BONILLA, P., TAN, F., QI, Y., AND ORDONEZ, V. 2021. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In AAAI. AAAI Press, 6912–6920.