实体关系抽取研究进展

——Research progress on relation extraction

胡伟龙

huweilong@whu.edu.cn

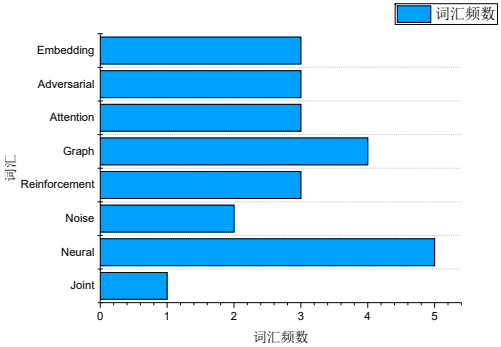
https://github.com/knightBoy

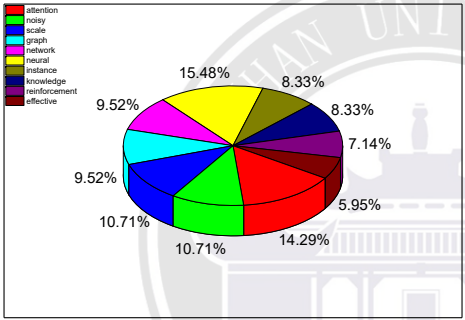
小米-武汉大学人工智能联合实验室
武汉大学语言与信息研究中心
武汉大学计算机学院

2019 年 4 月 18 日

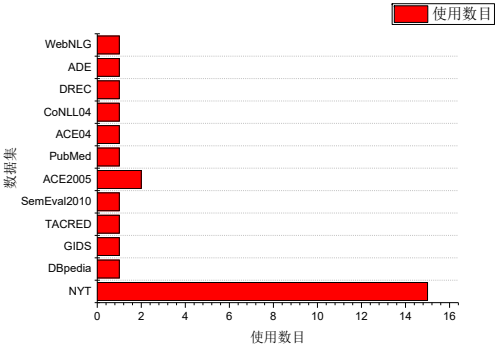
Contents

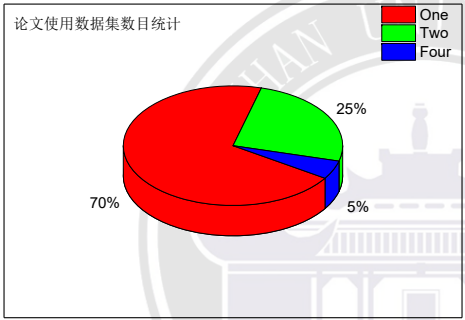
Topics from titles and abstracts





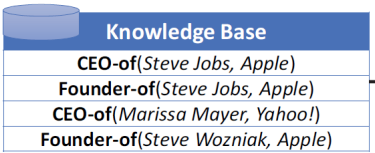
Dataset


使用数目
WebNLG
ADE
DREC
CoNLL04
ACE04
PubMed
ACE2005
SemEval2010
TACRED
GIDS
DBpedia
NYT
数据集
0
2
4
6
8
10
12
14
16
使用数目


论文使用数据集数目统计
One
Two
Four
25%
70%
5%

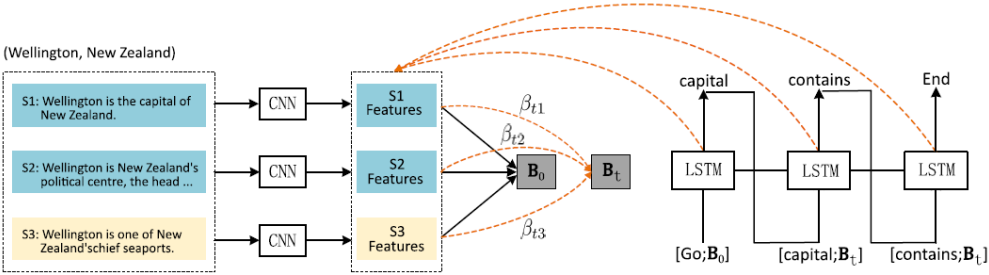
Contents

Background knowledge about distant supervised relation extraction

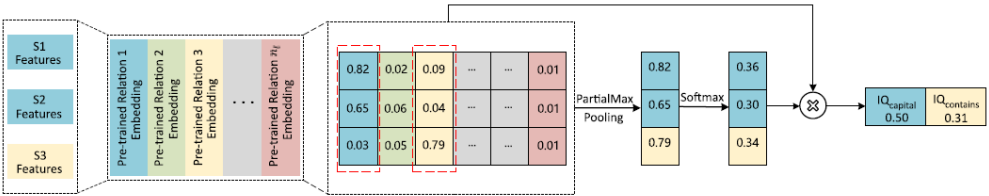


Relation Instances	Entity Pairs	Relation Types
S1: ***Jobs***, *the CEO of* ***Apple*** S2: ***Jobs*** *joins* ***Apple*** *as* S3: ***Jobs*** *co-founded* ***Apple*** *in 1976* S4: ***Jobs*** *launched* ***Apple*** *in 1976*	***(Jobs, Apple)***	CEO-of Founder-of
S5: ***Mayer*** *is the new CEO of* ***Yahoo!*** S6: ***Mayer*** *joins* ***Yahoo!***	***(Mayer, Yahoo!)***	CEO-of
S7: ***Woz*** *co-founded* ***Apple*** *in 1976* S8: ***Woz*** *joins* ***Apple*** *as*	***(Woz, Apple)***	Founder-of

2018-IJCAI-Exploring Encoder-Decoder Model for Distant Supervised Relation Extraction



(a) The architecture of our model.



(b) Information Quantity.

2018-IJCAI-Ensemble Neural Relation Extraction with Adaptive Boosting

The main issue of existing deep models is that their performance may not be stable and could not effectively handle the quite imbalanced, noisy, and wrong labeled data in relation extraction even if a large number of parameters in the model

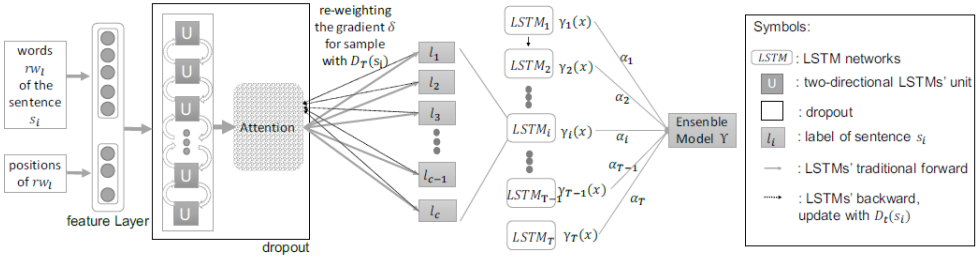
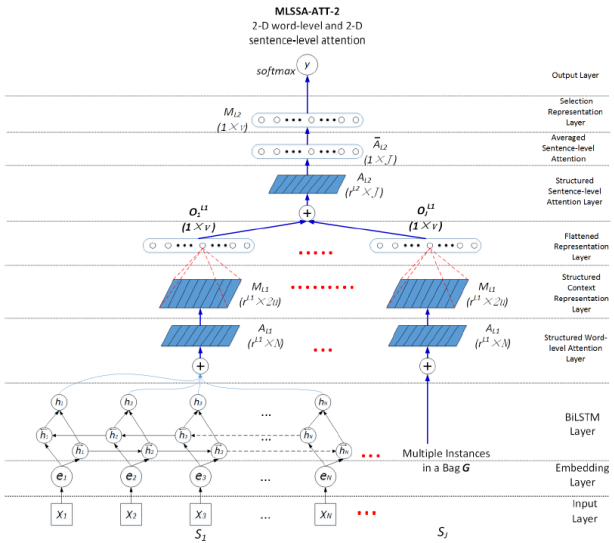


Figure 1: The framework of Ada-LSTMs contains three layers: feature layer, bi-directional Stacked LSTMs' layer with attention and adaptive boosting layer. s_i indicates the original input sentence with a pair of entities and their relation.

2018-EMNLP-Multi-Level Structured Self-Attention for Distantly Supervised Relation Extraction



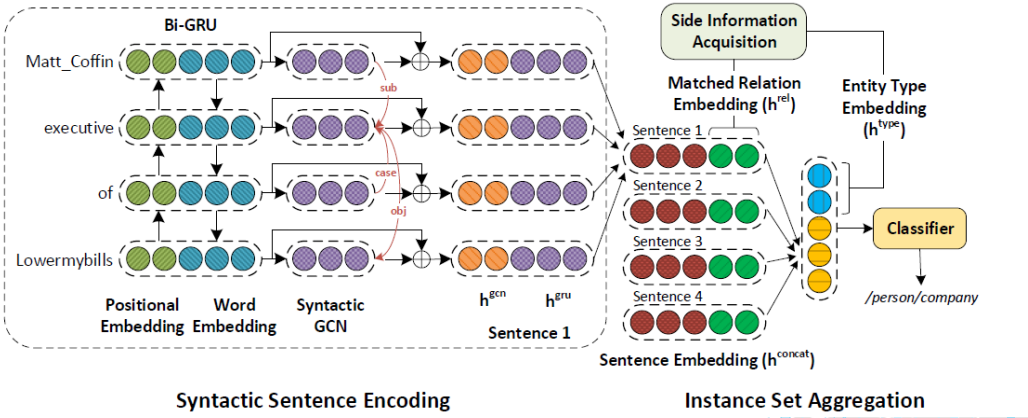
Two important representation learning problems:

- Entity pair-targeted context representation learning from an instance
- Valid instance selection representation learning over multiple instances

The deficiency of the 1-D attention vector is that it only focuses on one or a small number of aspects of the sentence, or one or a small number of instances

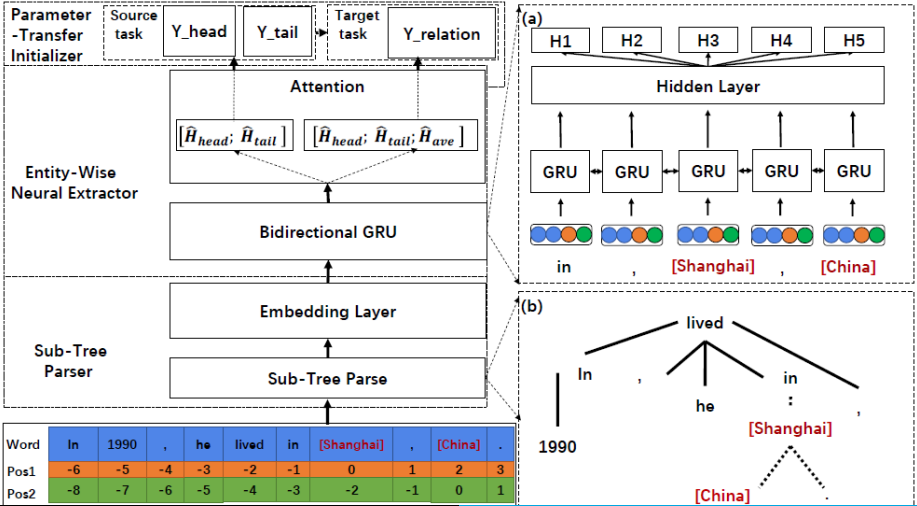
2018-EMNLP-RESIDE: Improving Distantly-Supervised Neural Relation Extraction using Side Information ⋆

Relevant side information can be effective for improving RE. For instance, in the sentence, 'Microsoft was started by Bill Gates.', the type information of entity can be helpful in predicting the correct relation *founderOfCompany*.



2018-EMNLP-Neural Relation Extraction via Inner-Sentence Noise Reduction and Transfer Learning

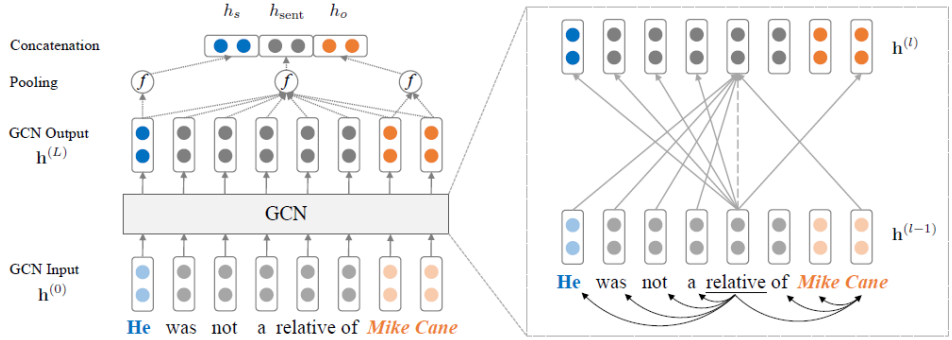
(1) Word-level noise within sentences (2) The robustness of RE against noise.



Word	In	1990	,	he	lived	in	[Shanghai]	,	[China]	.
Pos1	-6	-5	-4	-3	-2	-1	0	1	2	3
Pos2	-8	-7	-6	-5	-4	-3	-2	-1	0	1

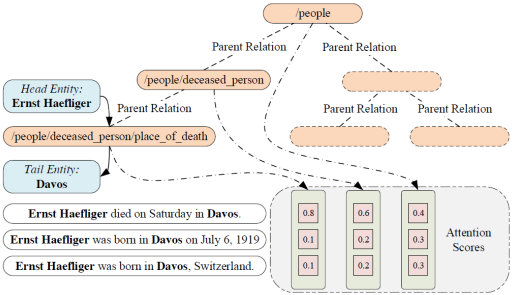
2018-EMNLP-Graph Convolution over Pruned Dependency Trees Improves Relation Extraction ⋆

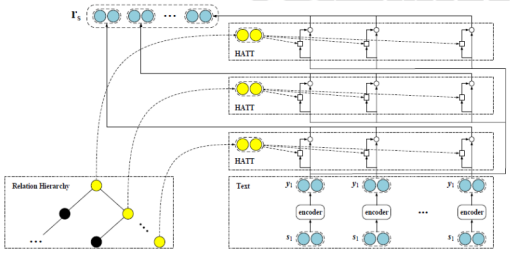
Existing dependency-based models either neglect crucial information by pruning the dependency trees too aggressively, or are computationally inefficient because it is difficult to parallelize over different tree structures.



2018-EMNLP-Hierarchical Relation Extraction with Coarse-to-Fine Grained Attention ⋆

Most existing methods handle each relation in isolation, regardless of rich semantic correlations located in relation hierarchical.

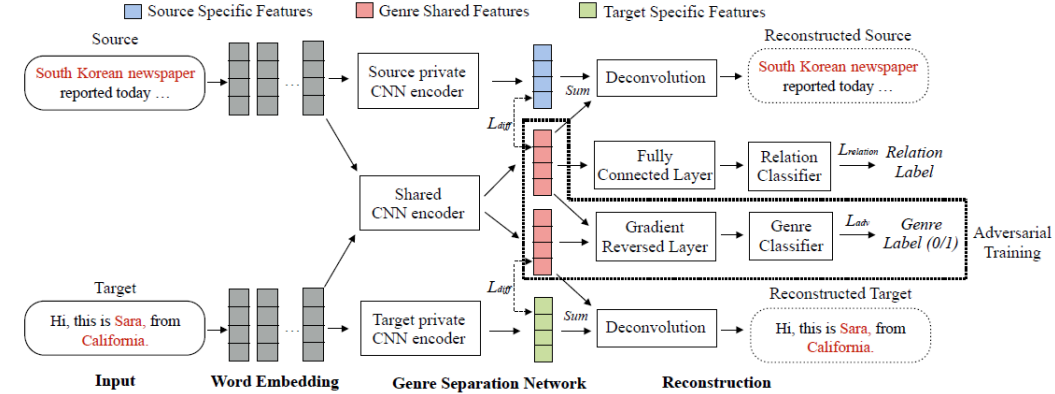




The attention on the bottom layer can capture more specific features of the relation, and the attention on the top-layer can capture the common features shared by several related sub-relations.

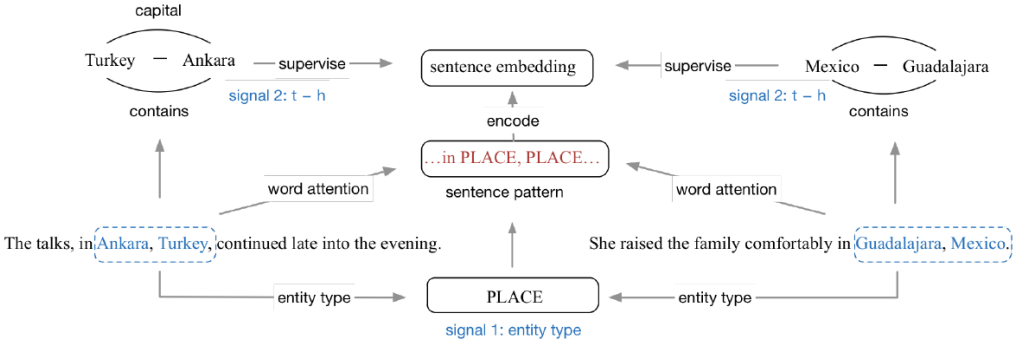
2018-EMNLP-Genre Separation Network with Adversarial Training for Cross-genre Relation Extraction

Relation Extraction suffers from dramatical performance decrease when training a model on one genre and directly applying it to a new genre, due to the distinct feature distributions.



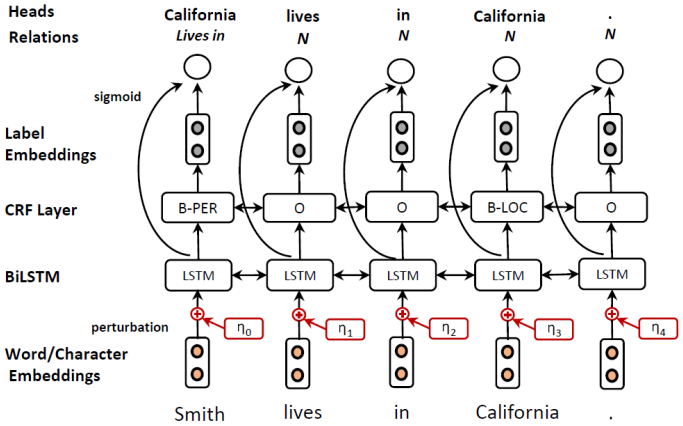
2018-EMNLP-Label-Free Distant Supervision for Relation Extraction via Knowledge Graph Embedding

Previous works have tried different ways to address the noise problem. However, stacking extra model does not fundamentally solve the problem of inadequate supervision signals of distant supervision, and will introduce expensive training costs. Besides, they still used the hard relation label derived from distant supervision, which also brought in much noise.



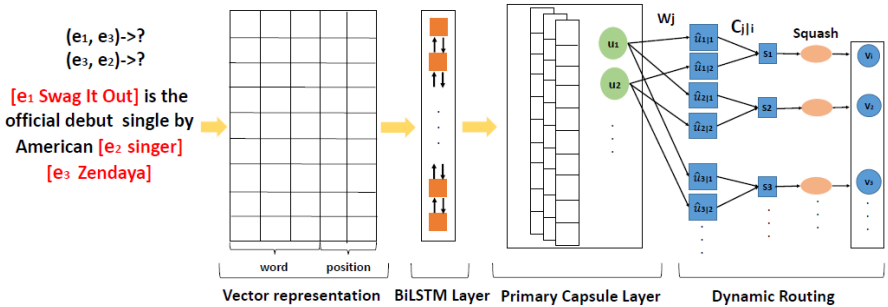
2018-EMNLP-Adversarial training for multi-context joint entity and relation extraction ⋆

Intentional small scale perturbations (i.e., adversarial examples) to the input of such models may lead to incorrect decisions.



2018-EMNLP-Attention-Based Capsule Networks with Dynamic Routing for Relation Extraction

- First, the existing models focus on, and heavily rely on, the quality of instance representation.
- Second, CNN subsampling fails to retain the precise spatial relationships between higher-level parts



2018-AAAI-SEE: Syntax-aware Entity Embedding for Neural Relation Extraction

Some work using syntax concerns mainly on the connections between entity pairs, paying much attention on the words that link the two entities semantically, while neglects the representation of entities themselves.

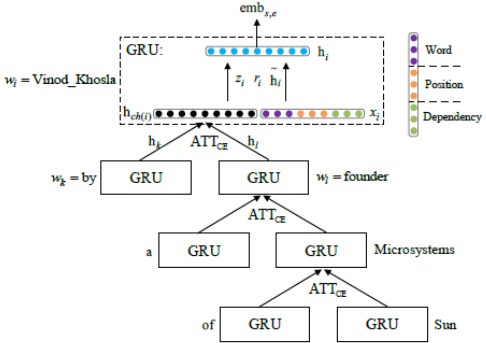


Figure 2: Workflow of entity embedding via tree-GRU.

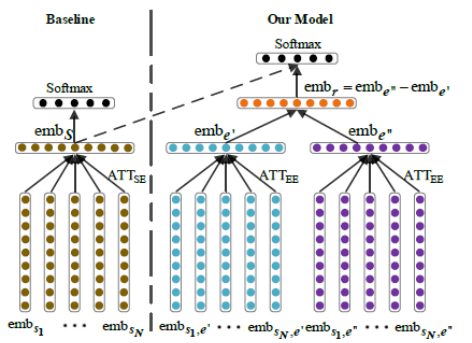
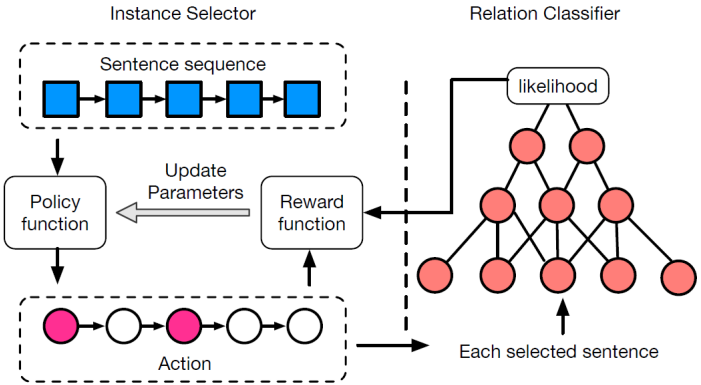


Figure 3: Workflow of the baseline and our approach.

2018-AAAI-Reinforcement Learning for Relation Classification from Noisy Data ⋆

Multi-instance learning suffers from two limitations:

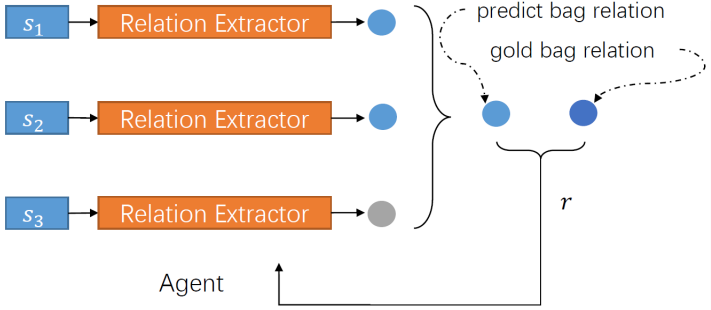
- Unable to handle the sentence-level prediction
- Sensitive to the bags with all noisy sentences which don't describe a relation at all



The instance selector chooses sentences according to a policy function, and then the selected sentences are used to train a better relation classifier. The instance selector updates its parameters, with a reward computed from the relation classifier.

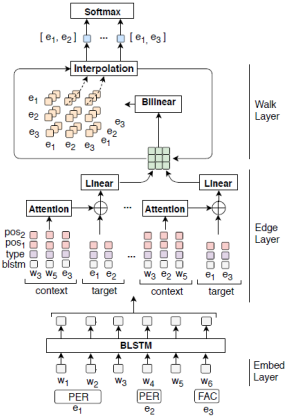
2018-AAAI-Large Scaled Relation Extraction With Reinforcement Learning

The relation extractor is regarded as RL Agent and the goal is to achieve higher long-term reward. The agent reads the bag' s sentences and outputs their extracted relations one by one. We integrate the predicted relations of sentences to predict the relation of the bag, which will be compared with the gold bag relation to determine the long-term reward. We then utilize it to train the relation extractor



2018-ACL-A Walk-based Model on Entity Graphs for Relation Extraction

- A sentence typically contains multiple relations between entity mentions.
- Related pairs are not predefined and consequently all entity pairs need to be considerd to extract relations



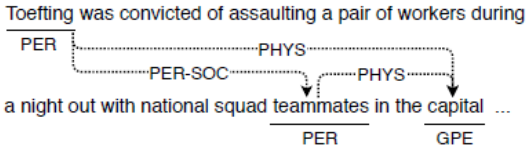
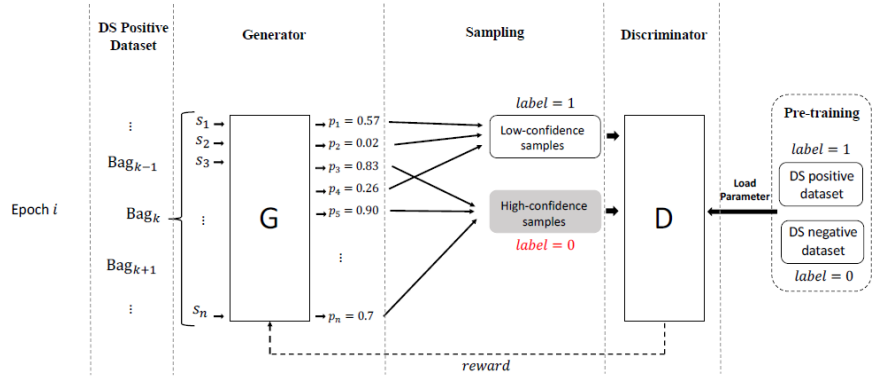


Figure 1: Relation examples from ACE (Automatic Content Extraction) 2005 dataset (Doddington et al., 2004).

2018-ACL-DSGAN: Generative Adversarial Traning for Distant Supervision Relation Extraction ⋆

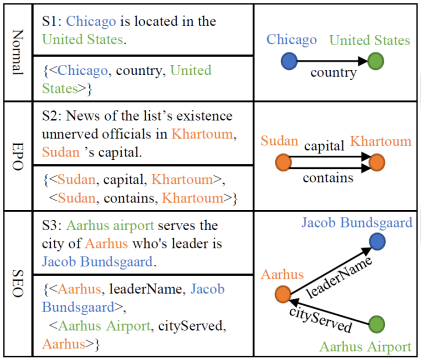
Attention-based methods make the denoising operation in the sentence bag level and overlook the case that all sentences of an entity pair are false positive, which is also the common phenomenon in distant supervsion datasets. Under this consideration, an independent and accurate sentence-level noise reduction strategy is the better choice.



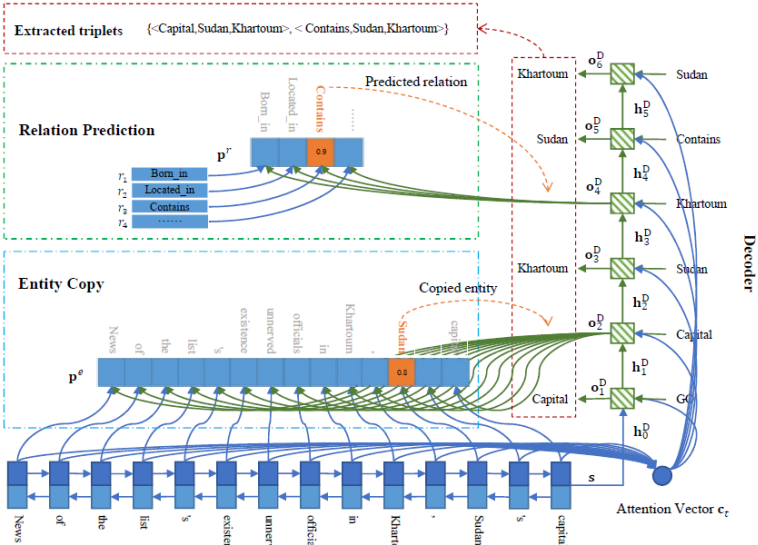
2018-ACL-Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism ⋆

Different relational triplets may have overlaps in a sentence. We divided the sentences into three types according the triplet overlap degree, including Normal, EntityPairOverlap and SingleEntityOverlap.

Existing models mainly focus on Normal class and fail to extract relational triplets precisely.



2018-ACL-Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism ⋆

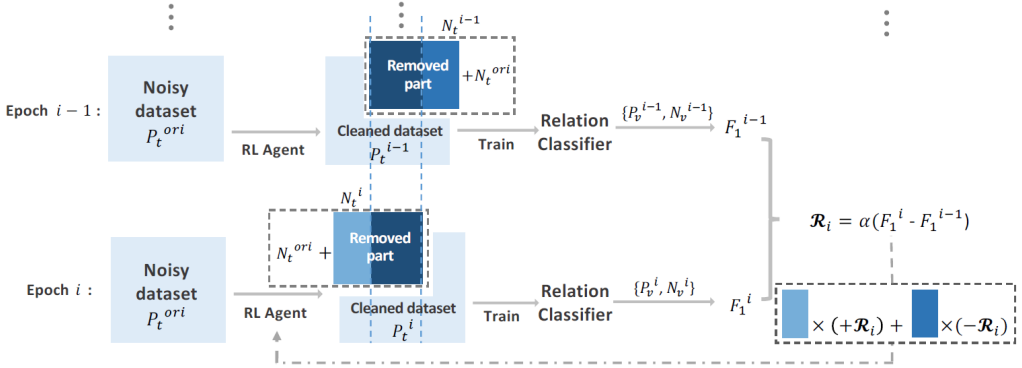


2018-ACL-Ranking-Based Automatic Seed Selection and Noise Reduction for Weakly Supervised Relation Extraction

- Formulation as Ranking Tasks:
  In the seed selection task, we use the k highest ranked instances as the seeds for bootstrapping RE. Likewise, in noise reduction for DS, we only use the k highest ranked triples from the DS-generated data to train a classifier. Note that the value of k in noise reduction may be much larger than in seed selection.
- Approaches to Auto matic Seed Selection and Noise Reduction:
  - K-means-based Approach
  - HITS-based Approach
  - HITS-and K-means-based Approach
  - LSA-based Approach
  - NMF-based Approach

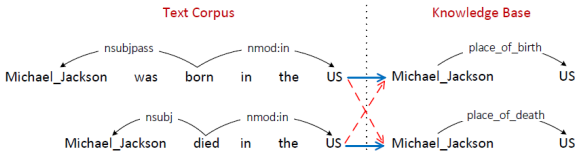
2018-ACL-Robust Distant Supervision Relation Extraction via Deep Reinforcement Learning ⋆

- Most of the state-of-art approaches focus on selecting one-best sentence or calculating soft attention weights over the set of the sentences of on specific entity pair.
- We argue that those incorrectly-labeled candidate sentences must be treated with a hard decision, rather than being dealt with soft attention.

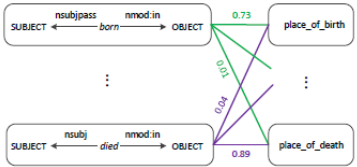


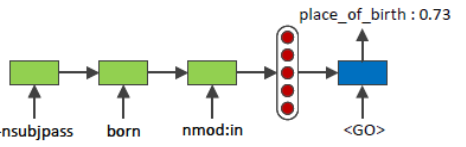
2018-NAACL-Global Relation Embedding for Relation Extraction ⋆

- Textual relations are one of the most discriminative textual signals that lay the foundation of many relation extraction models.
- Traditional embedding methods are based on local statistics, i.e., individual textual-KB relation pairs. Our key hypothesis is that global statistics is more robust to noise than local statistics.



	⟵nsubjpass *born* nmod:in⟶	⟵nsubj *died* nmod:in⟶
place_of_birth	1868	14
nationality	389	20
place_of_death	37	352
...	...	...





Contents

Summarization

- Approachs for noise reduction:
  - Soft decision
    - New Attention mechanism: hierarchical, self-attention, multi-level...
    - Add more information: relation alias, prior hypothesis...
    - New network architectures: copy networks, gcn...
  - Hard decision
    - Ranking-based
    - GAN-based
    - Reinforcement-based
- Experimental tricks:
  - General tricks: pre-training, multi-task, adversarial...
  - Dataset: NYT, DBpedia, SemEval, GIDS, TACRED...
  - Source code: 10/20
- More information:
  - `https://github.com/WHUNLPLab/Papers-to-read`

Thank You for Your Listening

Weilong Hu
School of Computer Science, Wuhan University