# Robust Learning and Inference for IE

## New Frontiers of Information Extraction (Part III)

Muhao Chen
Department of Computer Science / Information Sciences Institute
University of Southern California

**July 2022**

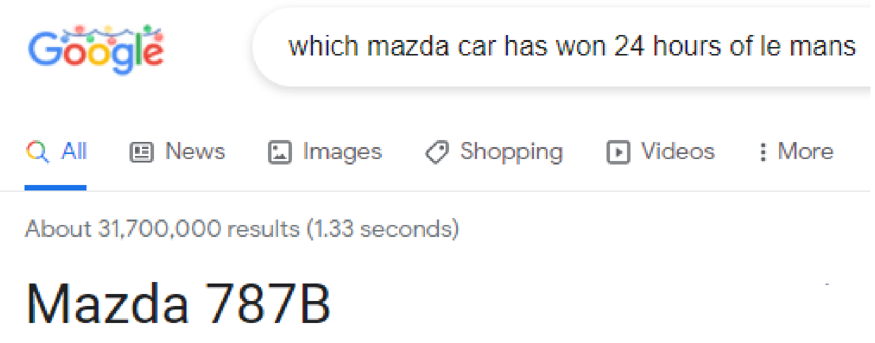
**NAACL Tutorials**

**New Frontiers of Information Extraction**

How do we make IE models ***reliable***?

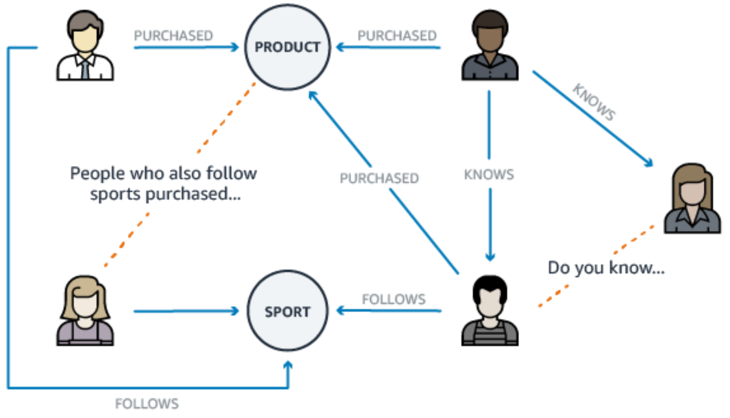
# AI Needs to Understand Relations of Concepts

## QA & Semantic Search



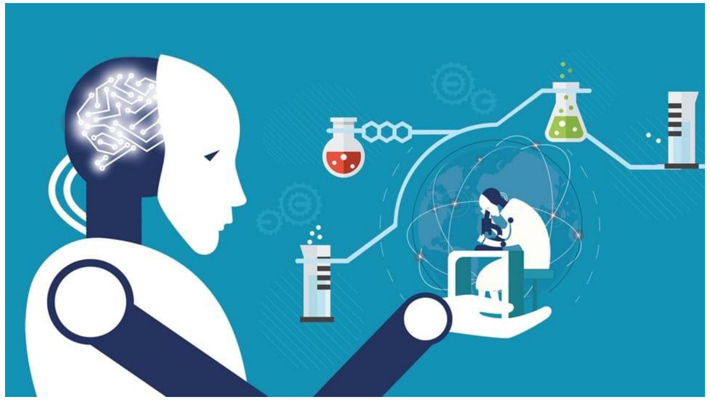

Relations of Entities

## E-Commerce

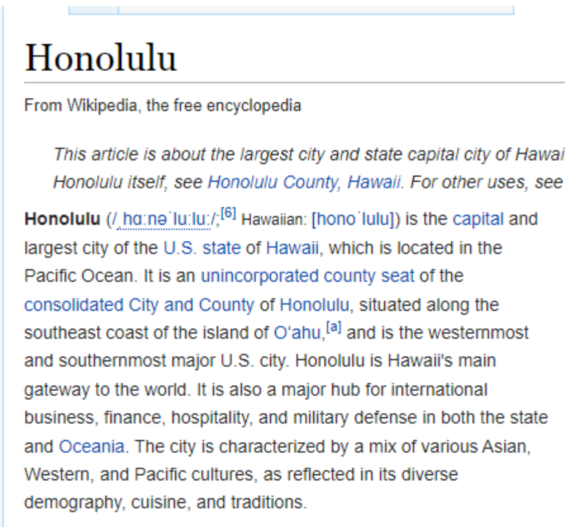



Relations of Products and Users

## Comp. Bio. Med.



Interactions of (bio)molecules
Relations of diseases and drugs

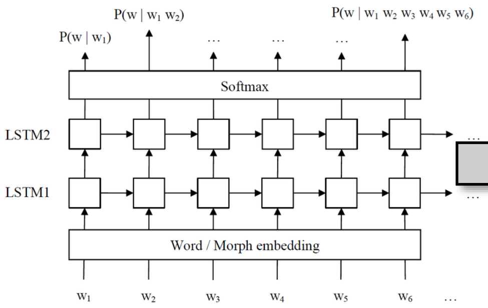


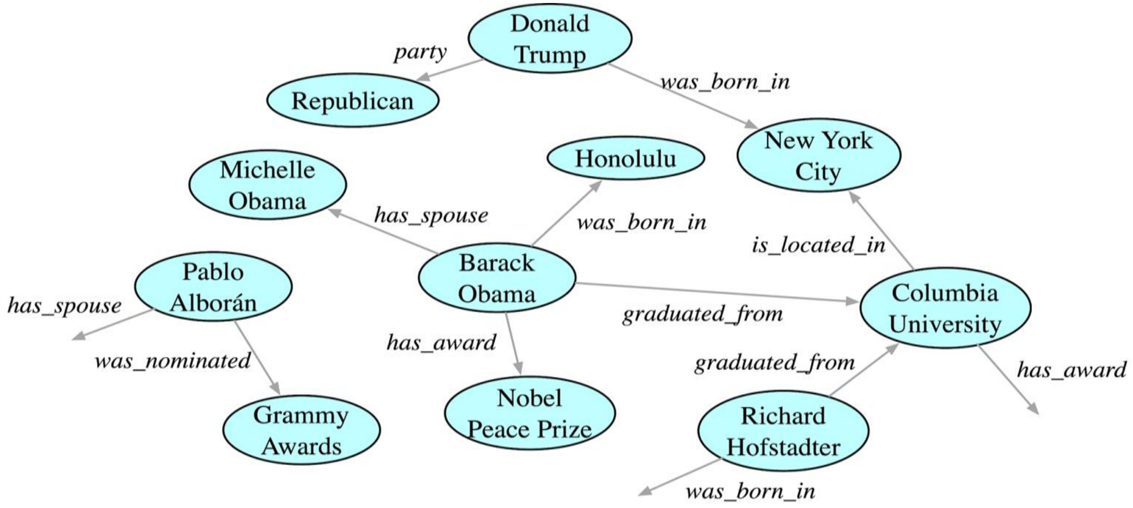
Honolulu

From Wikipedia, the free encyclopedia

*This article is about the largest city and state capital city of Hawai... Honolulu itself, see Honolulu County, Hawaii. For other uses, see*

**Honolulu** (/ˌhɑːnəˈluːluː/;[6] Hawaiian: [honoˈlulu]) is the capital and largest city of the U.S. state of Hawaii, which is located in the Pacific Ocean. It is an unincorporated county seat of the consolidated City and County of Honolulu, situated along the southeast coast of the island of Oʻahu,[a] and is the westernmost and southernmost major U.S. city. Honolulu is Hawaii's main gateway to the world. It is also a major hub for international business, finance, hospitality, and military defense in both the state and Oceania. The city is characterized by a mix of various Asian, Western, and Pacific cultures, as reflected in its diverse demography, cuisine, and traditions.




IE automatically extracts structural knowledge about concepts and relations

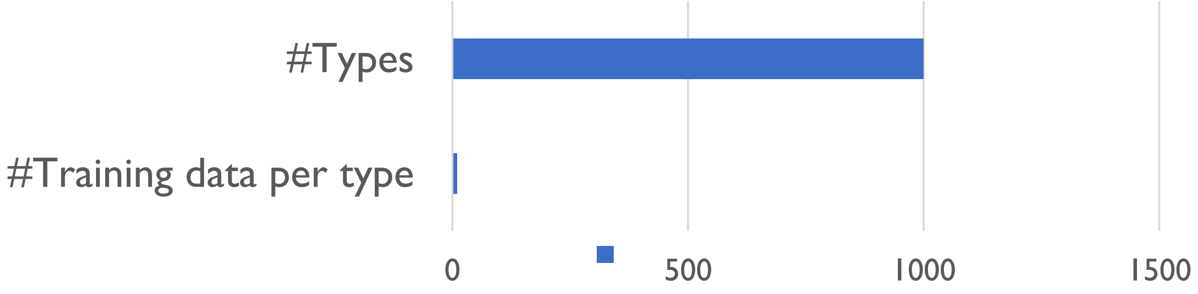
# Fragility of IE Models

## Fragility in Learning

| | | | |
|---|---|---|---|
| Wrong Args | Authorities said they ordered the detention of <u>Bruno 's wife</u> , [Dayana Rodrigues]$_{tail:per}$ , who was found with [Samudio]$_{head:per}$'s baby . | *per:spouse* | 109 |
| Relation Def. | [Zhang Yinjun]$_{tail:per}$ , <u>spokesperson</u> with one of China 's largest charity organization , the [China Charity Federation]$_{head:org}$ | *org:top_mem.* | 96 |
| Entity Type | [Christopher Bollyn]$_{head:per}$ is an [<u>independent</u>]$_{tail:religion}$ journalist | *per:religion* | 31 |

Noisy Training Data

IE (structural) annotation is difficult and often noisy
- 5-8% errors in TACRED & CoNLL03
- <70% IAA in HiEve & IC
- etc.




Data distribution in UFET

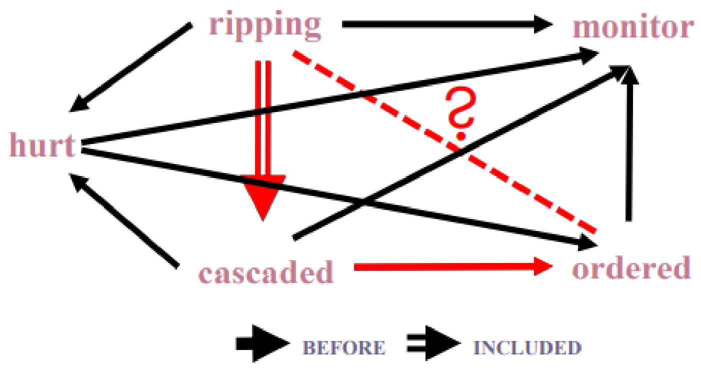
Ultra Diverse Labels and Low Training Resources

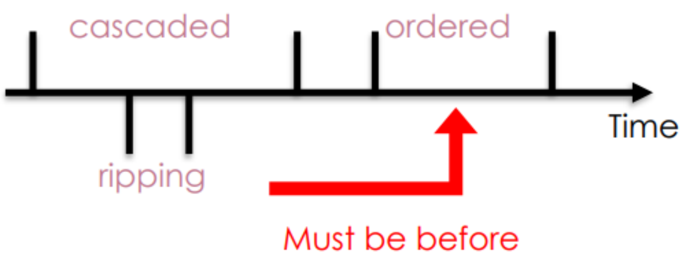
The extracts are often:
- Diverse and unbalanced
+ Expensive and insufficient

# Fragility of IE Models

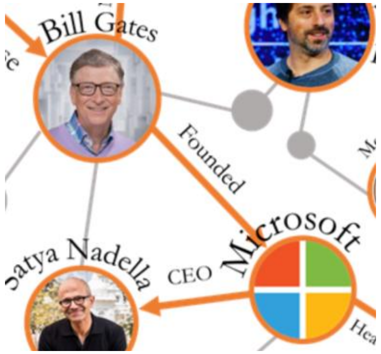
## Fragility in Inference

*In Los Angeles that lesson was brought home Friday when tons of earth **cascaded** down a hillside, **ripping** two houses from their foundations. No one was **hurt**, but firefighters **ordered** the evacuation of nearby homes and said they'll **monitor** the shifting ground until March 23rd.*




How do we ensure the extracts are **globally consistent**?

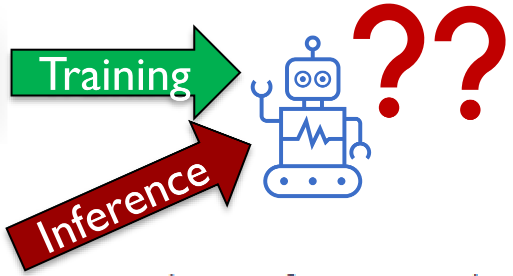



*Visited?* or *founded?*

Bill Gates [PER] paid a visit to Building 99 [MISC] of Microsoft [ORG] yesterday .

And **faithful**?

Michael Jordan [PERSON] is a professor at Berkeley [ORG]




SARS - CoV-2 ORF3a interacts with VSP39 -- a core subunits of HOPS complex

What about **out-of-distribution** Inputs?

Michael Jordan [PER] is an expert in machine learning .

Statistician? Comp. neuroscientist?

**Unknown** Extract?

# Goal: Robust IE

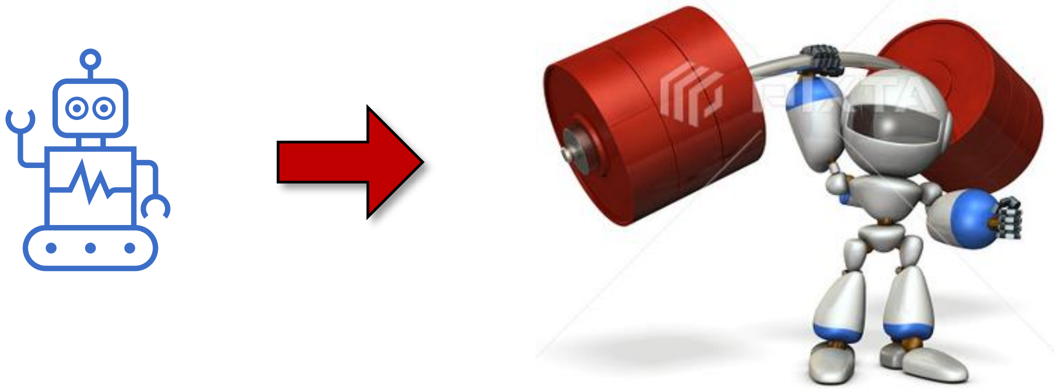
## The goal of developing a robust IE system

**Robustness in Learning**
- **Noise robustness**: proactively identifying and mitigating training noise
- **Constraint learning**: capturing logical constraints of labels
- **Debiased training**: mitigating feature shortcuts and balancing training signals

Overcome minimal, noisy and biased supervision

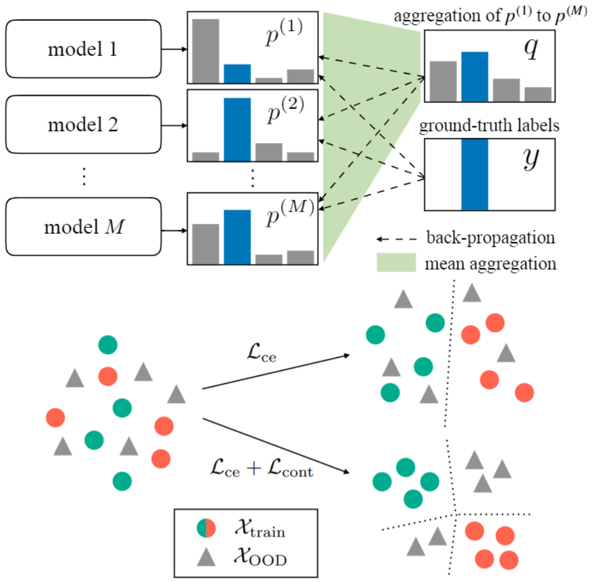
**Robustness in Inference**
- **Selectiveness**: knowing what is extractable, what is not
- **Constrained inference**: ensuring logically consistent extracts
- **Faithfulness**: does not rely on spurious correlation

Self-contained, selective and faithful extraction.

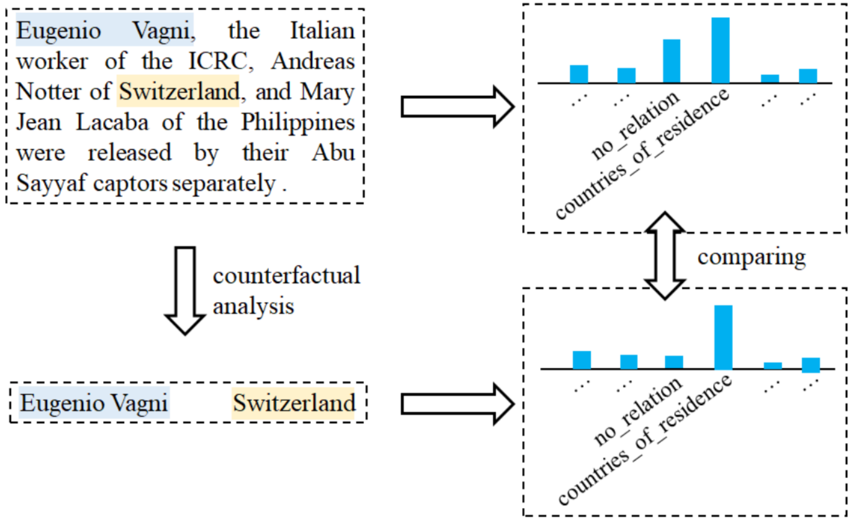


# Agenda

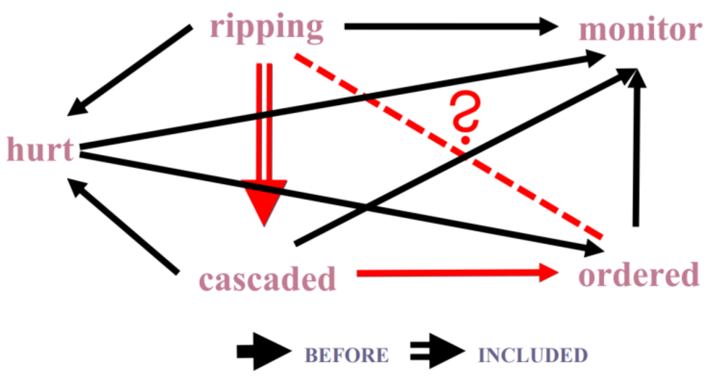
## 1. Noise-robust IE




## 2. Faithful IE



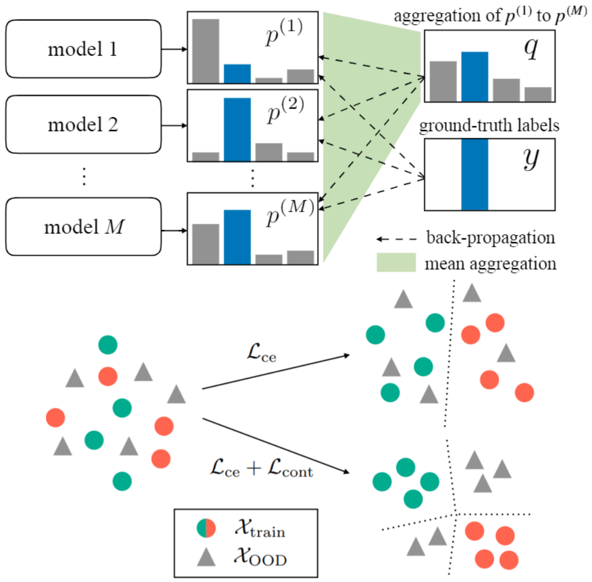

## 3. Logically Consistent IE



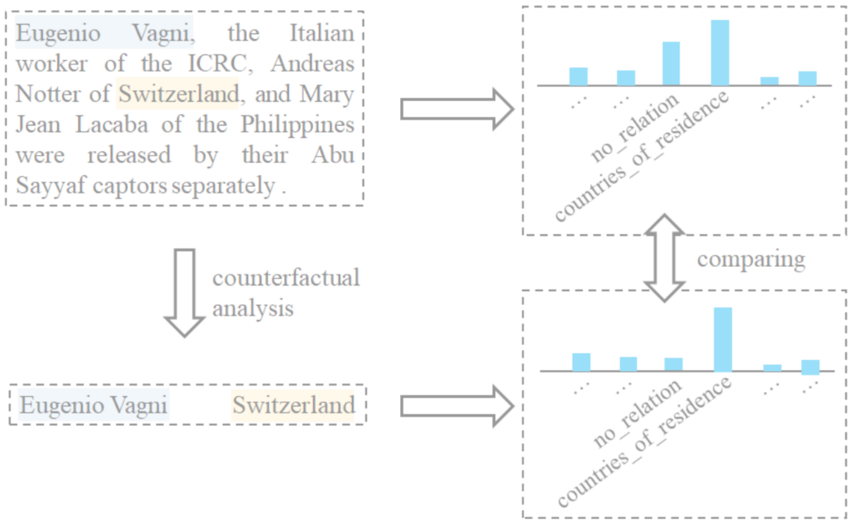

## 4. Open Research Directions

# Agenda

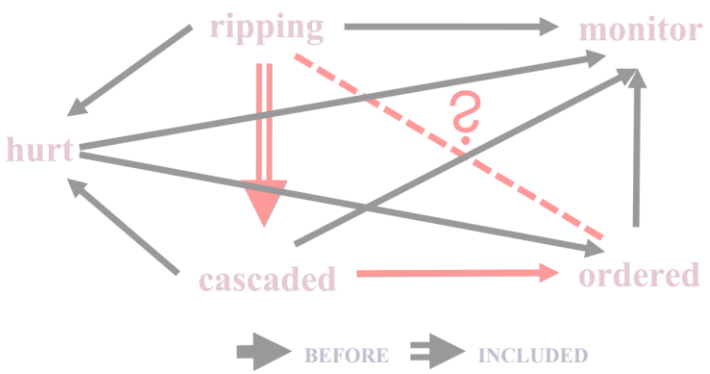
## 1. Noise-robust IE




## 2. Faithful IE




## 3. Logically Consistent IE



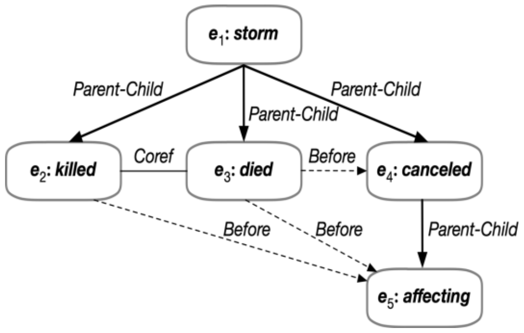

## 4. Open Research Directions

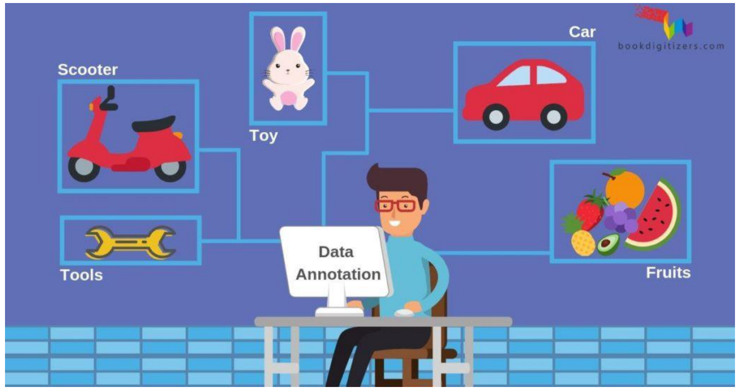
# Noise In Training and Inference

## Training

Annotation for IE is **difficult** and **expensive**

On Tuesday, there was a typhoon-strength ($e_1$:***storm***) in Japan. One man got ($e_2$:***killed***) and thousands of people were left stranded. Police said an 81-year-old man ($e_3$:***died***) in central Toyama when the wind blew over a shed, trapping him underneath. Later this afternoon, with the agency warning of possible tornadoes, Japan Airlines ($e_4$:***canceled***) 230 domestic flights, ($e_5$:***affecting***) 31,600 passengers.




Reading long documents, annotating complex structures

Costs \$2-\$6 and >3 minutes for just 1 relation [Paulheim+ 2018]

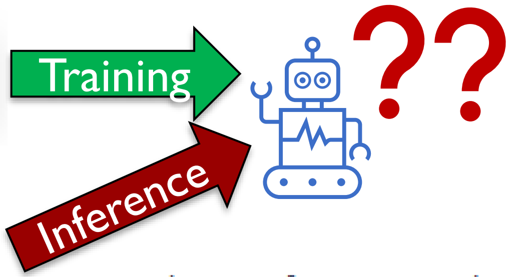
**Hence, IE annotations are inevitably noisy. For example:**

- 5-8% errors in TACRED and CoNLL03
- <70% IAA in HiEve, Intelligence Community, etc.

## Inference

In real application, IE models sees way larger, more diverse and noisy data than in training

Michael Jordan [PERSON] is a professor at Berkeley [ORG]




SARS - CoV-2 ORF3a interacts with VSP39 -- a core subunits of HOPS complex

**Out-of-Distribution** Inputs

Michael Jordan [PER] is an expert in machine learning .

Statistician?

Comp. neuroscientist?

**Unknown** extraction types

Michael Jordan [PER] did not attend UCLA [ORG]

No Rel

**Nothing** to extract

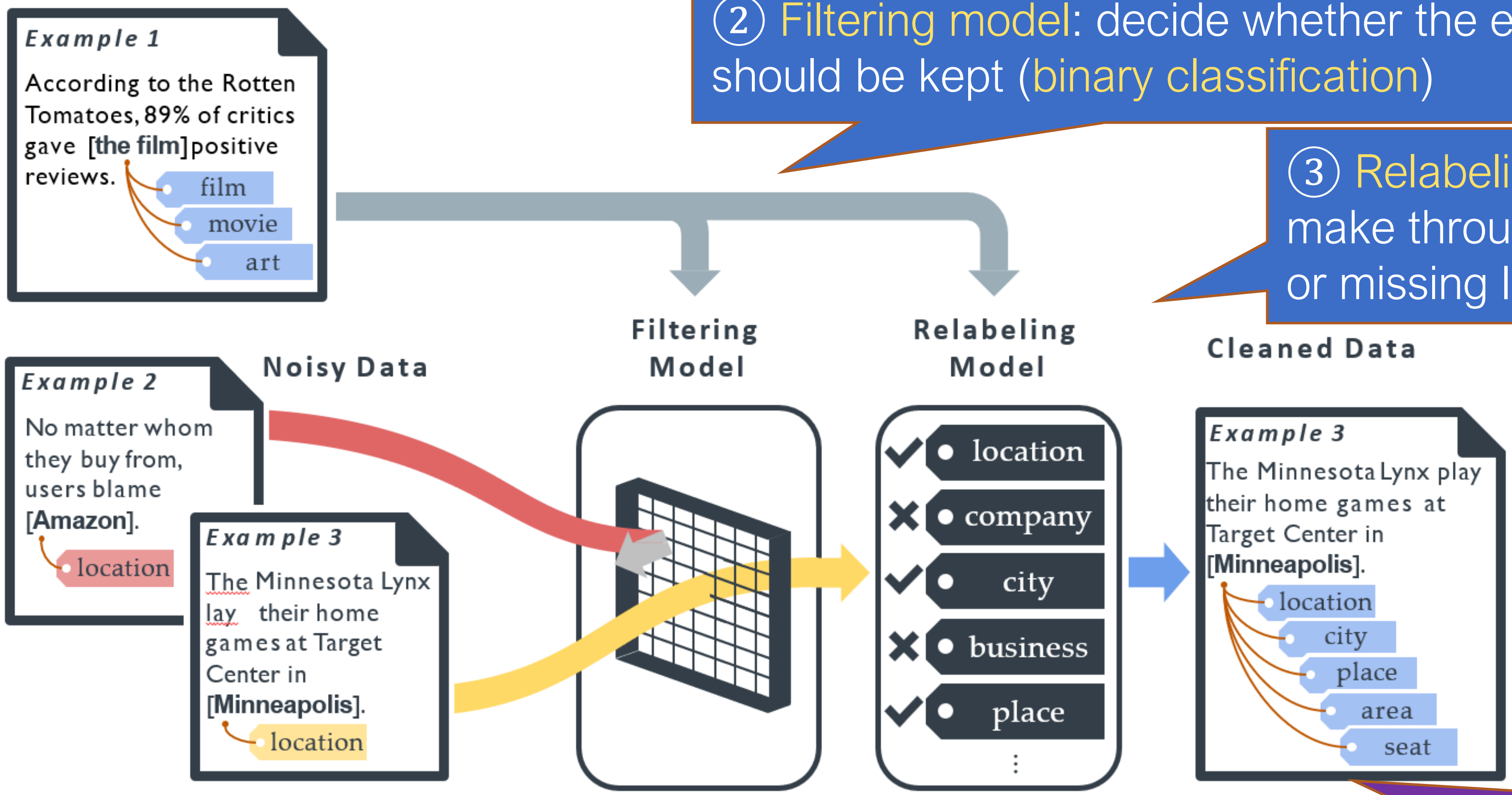
# Supervised Denoising

A noise filtering or relabeling model may be trained, if clean data are available.

① Labeled clean data and noisy data

② Filtering model: decide whether the example should be kept (binary classification)

③ Relabeling model: repair examples that make through filtering but which still have errors or missing labels (multi-label classification)




④ Cleaned (task) training data

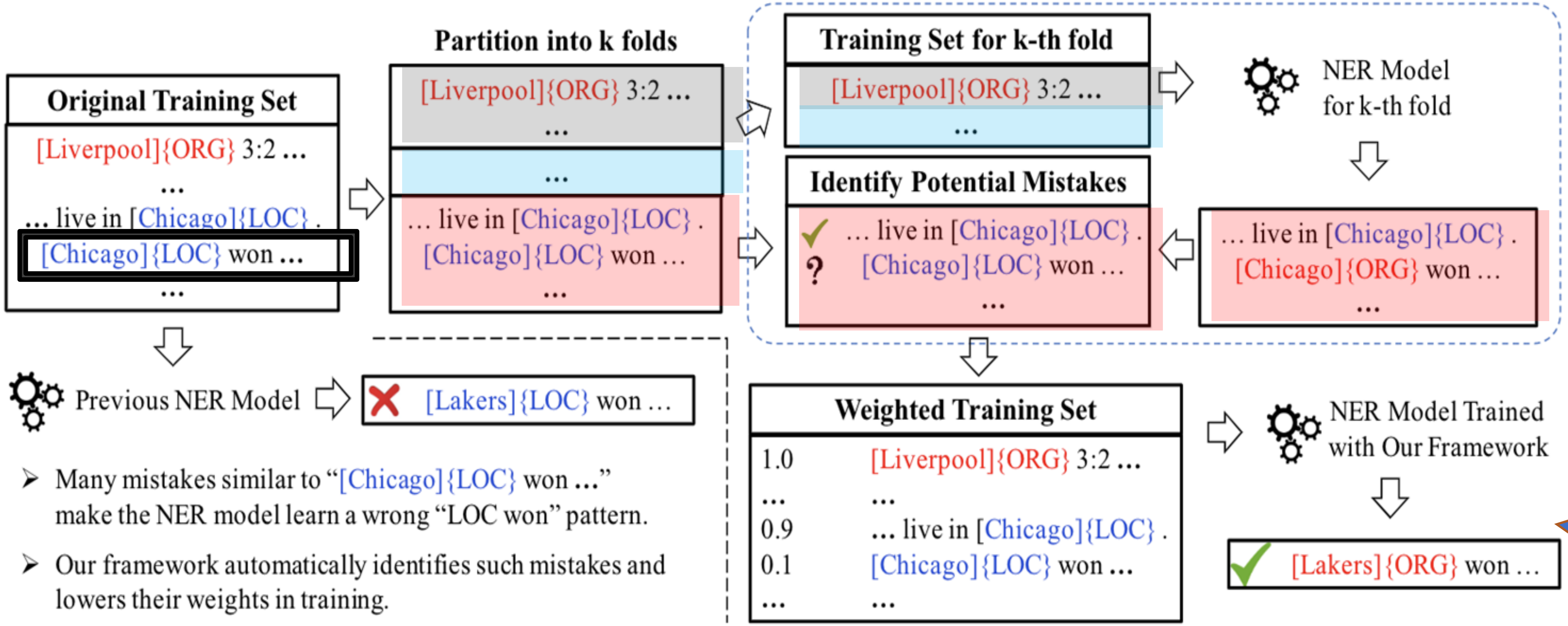
Cost: manually labeling enough clean data can still be expensive.

Onoe and Durrett. Learning to Denoise Distantly-Labeled Data for Entity Typing. **NAACL** 2019

# Unsupervised Denoising: Ensemble

① Partition data into k-folds

② Cross-validate the quality of each fold

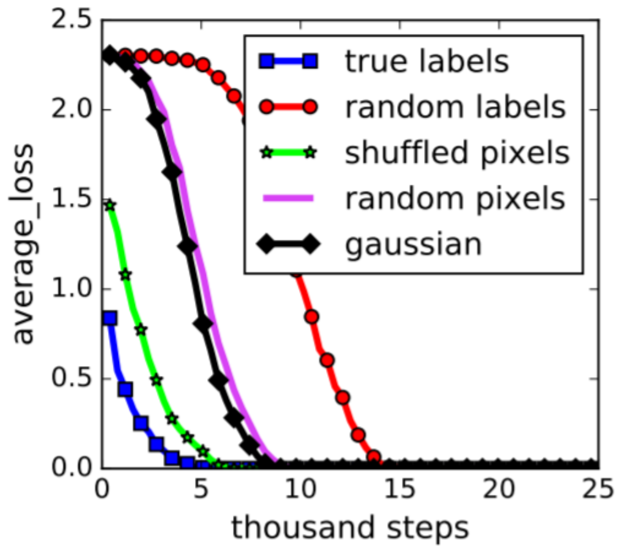


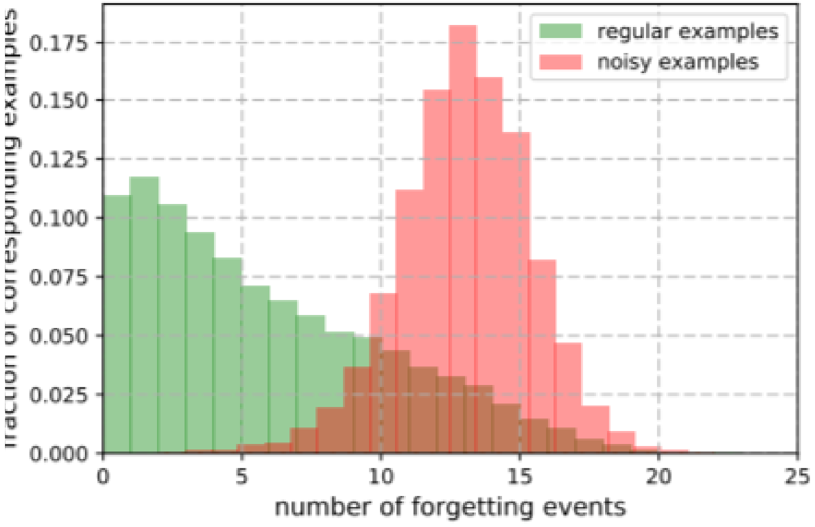

③ Reweight data folds and train the final model

Unsupervised denoising: no longer requires annotated clean data
Cost: needs repeated training and testing of the model for at least $k+1$ times.

Wang et al. CrossWeigh: Training named entity tagger from imperfect annotations. **EMNLP** 2019

# Unsupervised Denoising: Co-regularized Knowledge Distillation



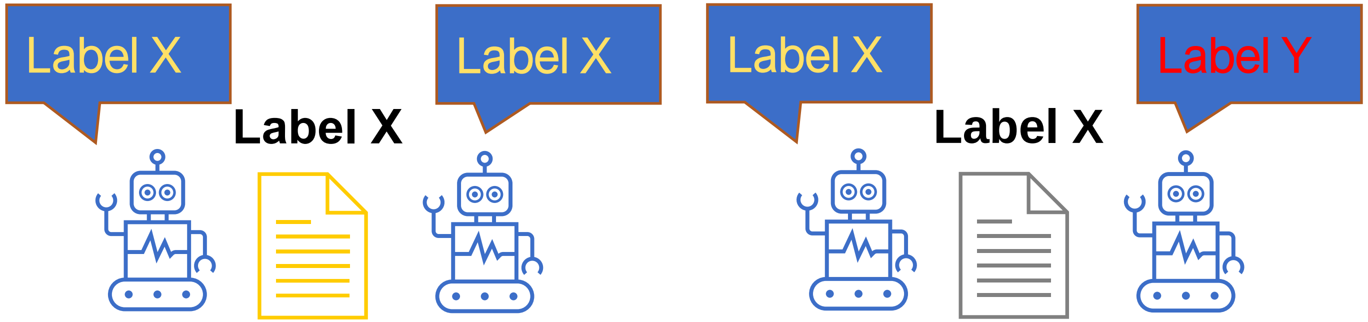

(1) (2)

Noisy labels lead to delayed learning curves [Toneva+ ICLR-19]

Noisy labels are outliers to the task inductive bias.

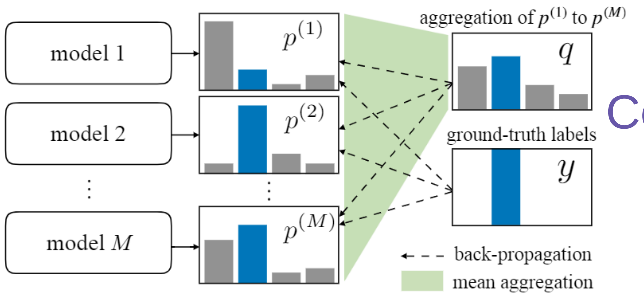
(1) Noisy labels take longer to be learned.
(2) Noisy labels are frequently forgotten.

Model prediction is often inconsistent or oscillates on noisy labels in later epochs.




**Agree: Clean ✓** **Disagree: Noise ✗**

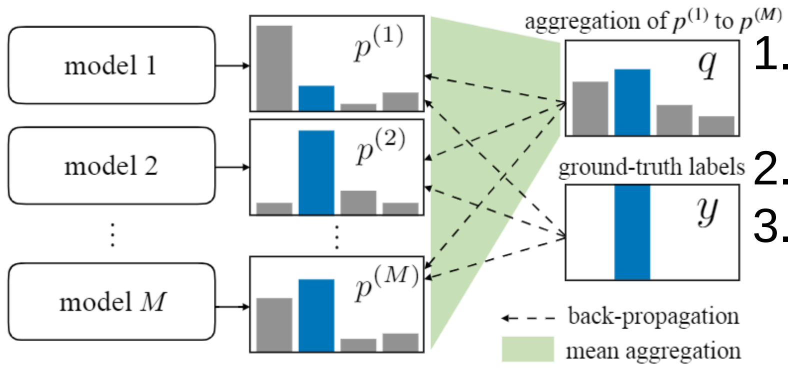
Mutual agreement by models indicates clean/noisy labels




Co-regularization Framework

Zhou and Chen. Learning from Noisy Labels for Entity-Centric Information Extraction. **EMNLP** 2021

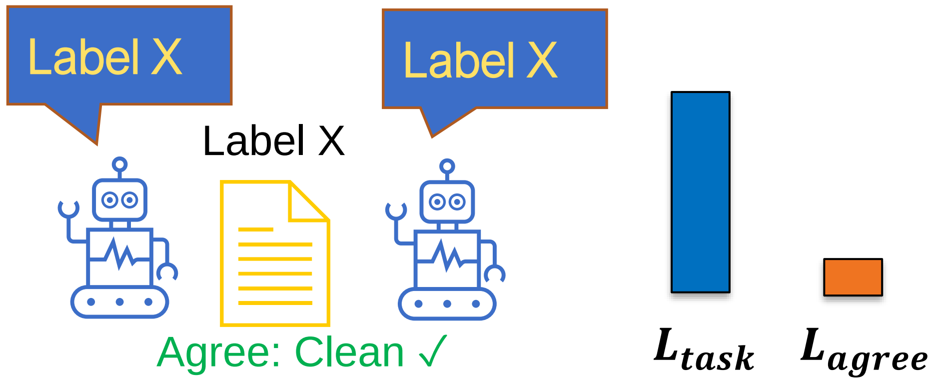
# Unsupervised Denoising: Co-regularized Knowledge Distillation

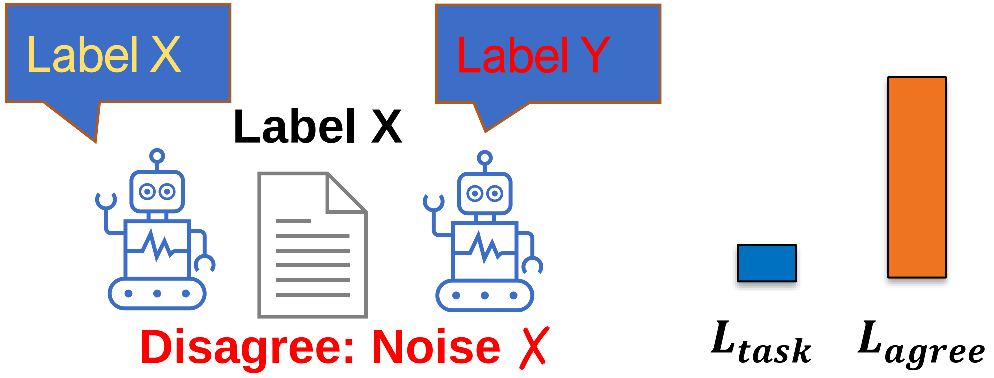



1. Create $\boldsymbol{M}(\geq \mathbf{2}$; 2 is enough) identical neural models with **different initialization**, and **warm up** them using only the **task loss**.
2. Train the models with both **the task loss** and an additional **agreement loss**.
3. Return one of the models.

Cross-entropy $L_{task}$

K-L divergence between model predictions $L_{agree}$




**On clean data**

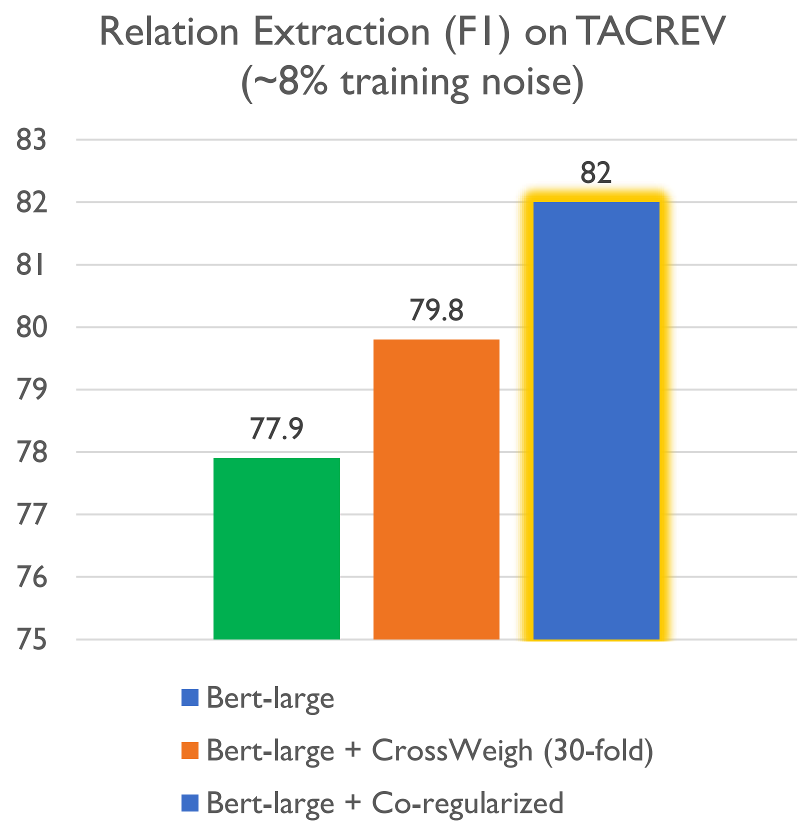
- Lower agreement loss
- Focusing on task optimization

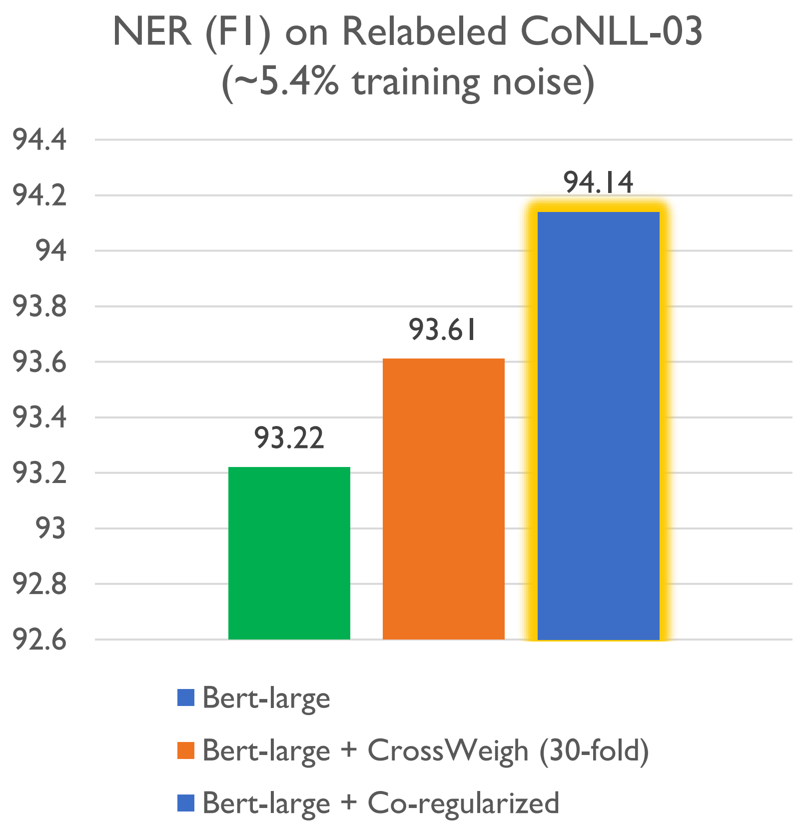
**On noisy data**

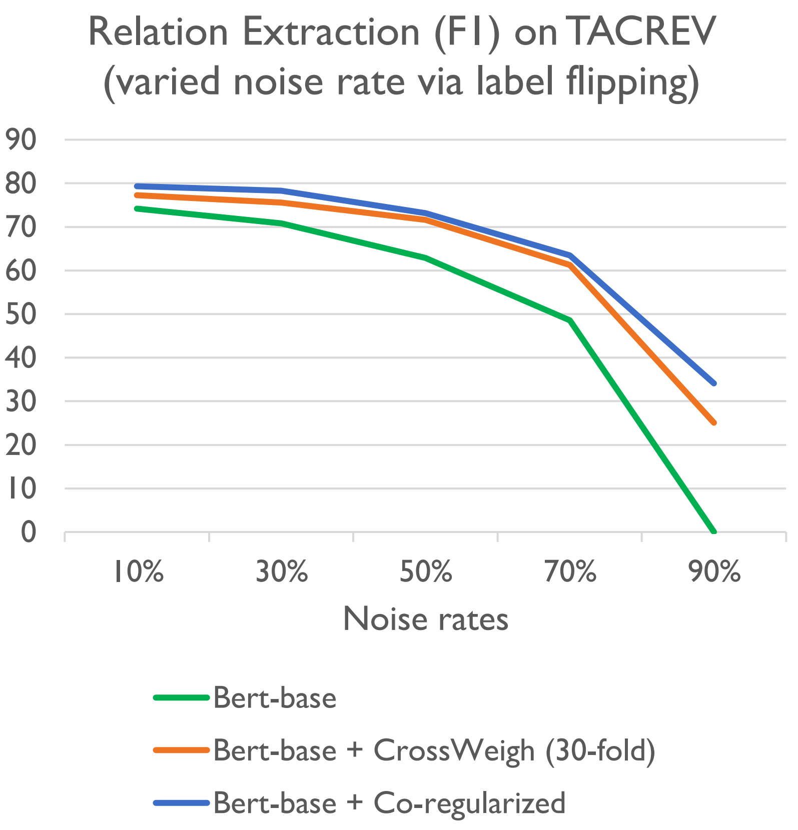
- Higher agreement loss
- Task optimization proactively prevents fitting those data

Zhou and Chen. Learning from Noisy Labels for Entity-Centric Information Extraction. **EMNLP** 2021

# Unsupervised Denoising: Co-regularized Knowledge Distillation



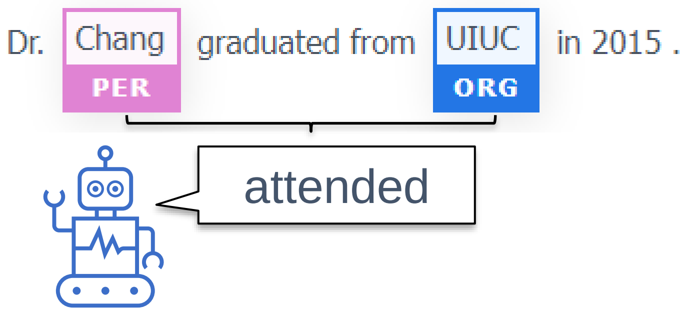

**Merits of co-regularized knowledge distillation**

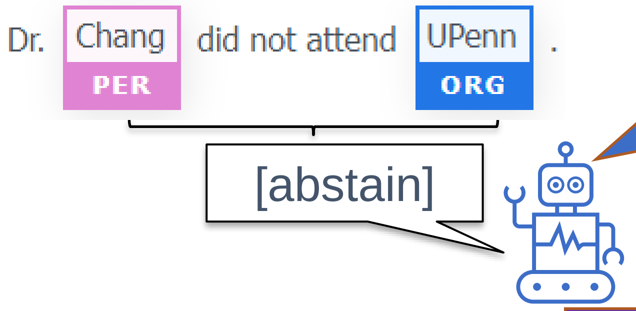
- More robust than ensemble (cross-weight), especially under higher noise rates
- More efficient (only 1-fold of training and no additional inference cost)
- Can be applied to train any backbone IE models (see results w/ LUKE and C-GCN in the paper)

Zhou and Chen. Learning from Noisy Labels for Entity-Centric Information Extraction. **EMNLP** 2021

# Noise in Inference

In inference, IE models need to know when to not extract



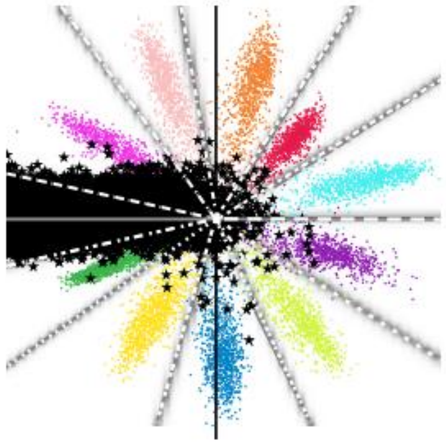

IE models can be exposed to many exception cases in real-world application.

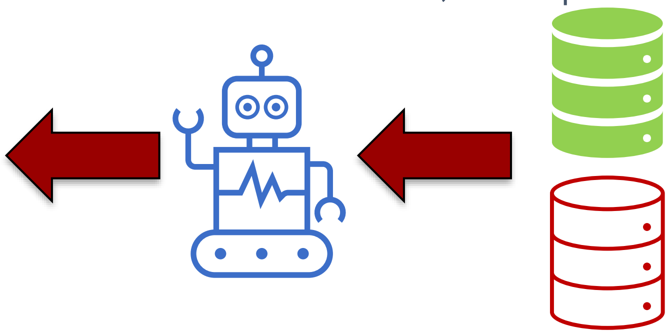
How to make inference more selective?

A supervised approach can be a choice

- Classify exceptions into an open class/background set
- However, exceptions can never be close to exhaustive in training data

Open class





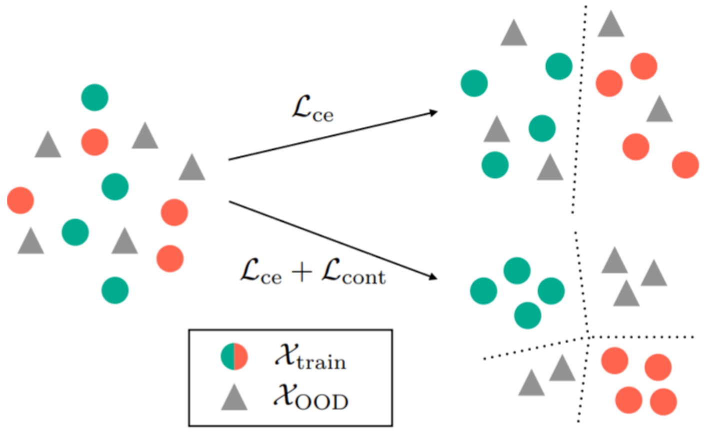
- Task training data
- Annotated exceptions

Dhamija et al. Reducing network agnostophobia. **NeurIPS** 2018

# Learning to Abstain without Annotated "Abstention"?

This is still an underexplored area, but there are at least two lines of strategies

## Unsupervised out-of-distribution (OOD) detection




Increase inter-class discrepancy ⇒ Better OOD detection

Creating compact representations with (margin-based) contrastive learning

- Indirectly making OOD instances as "background" representation

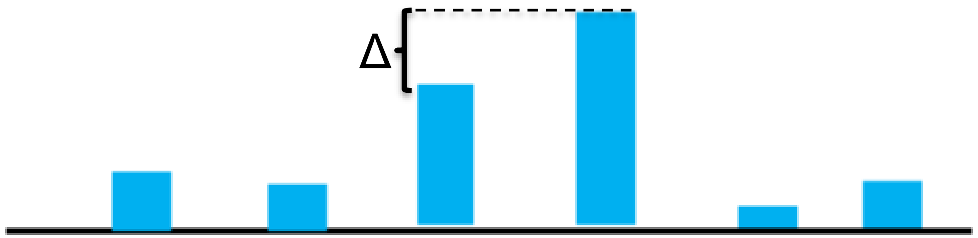
Inference with Mahalanobis distance

- High-order distance measures improve OOD detection

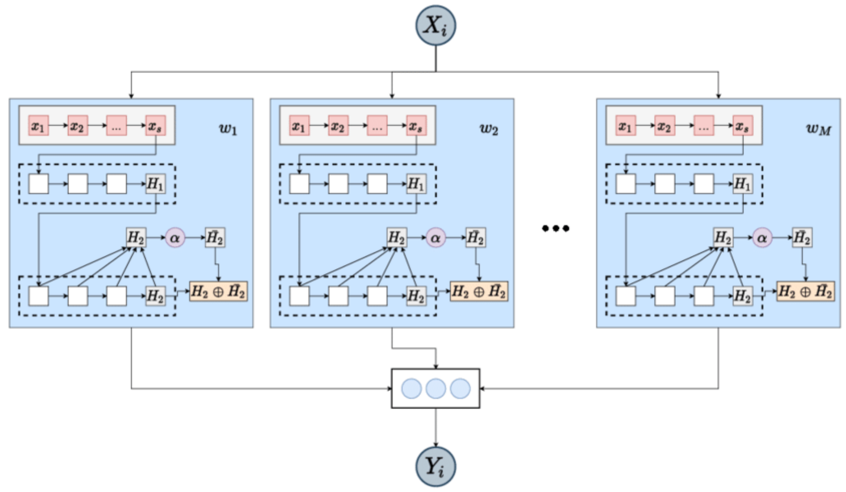
Zhou et al. Contrastive Out-of-Distribution Detection for Pretrained Transformers. **EMNLP** 2021

## Estimating the uncertainty of prediction

Softmax response: difference between top two class predictions



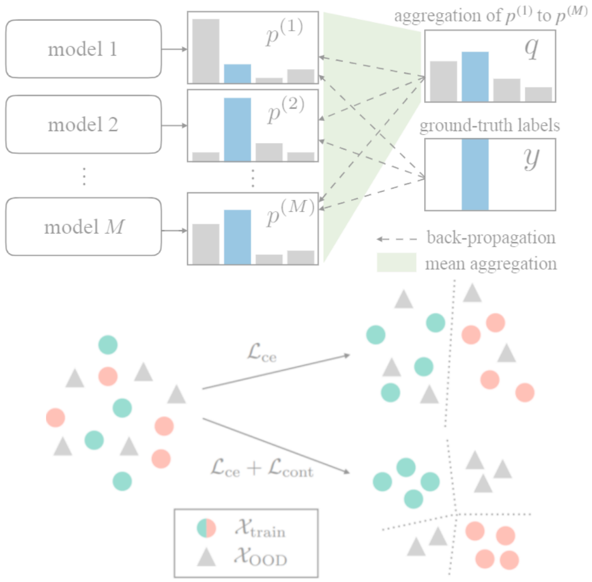

Prediction variance in Monte-Carlo dropout



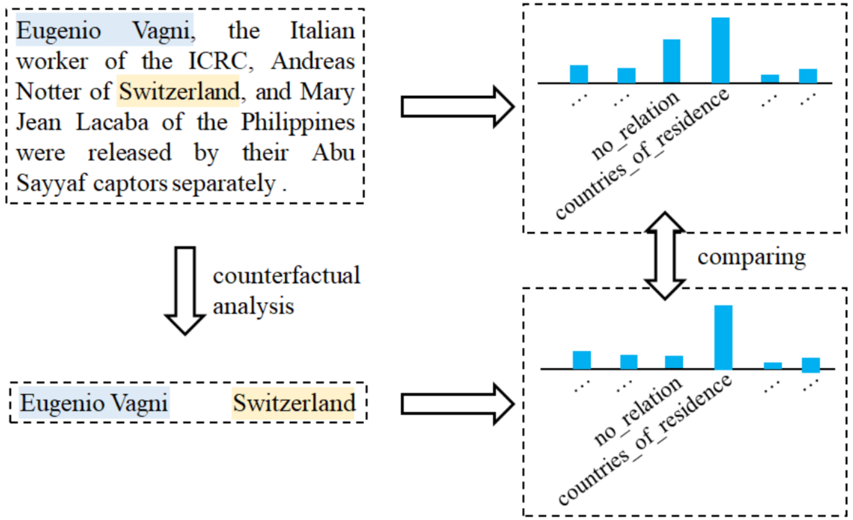

Xin et al. The Art of Abstention: Selective Prediction and Error Regularization for Natural Language Processing. **ACL** 2021

# Agenda

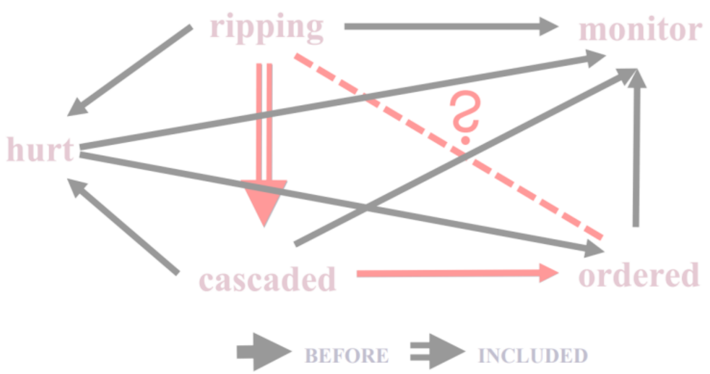
## 1. Noise-robust IE




## 2. Faithful IE




## 3. Logically Consistent IE




## 4. Open Research Directions

# Faithfulness Issues

IE systems may not always **faithfully** extract what is described in the **context**

Entity relation extraction: Bill Gates [PER] paid a visit to Building 99 [MISC] of Microsoft [ORG] yesterday .

Rel?


Visit ✓

FounderOf ✗

According to prior knowledge

Prior knowledge (in PLMs) can lead to biased extraction

Temporal relation extraction: I went to *event1* see the doctor. However, I got more seriously *event2* sick.

Before? After?


Before ✓

After ✗

According to statistics

Data

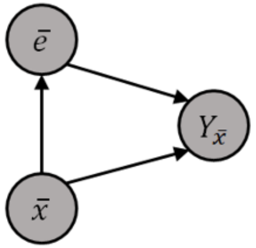
(Statistically) Biased training can lead to biased extraction

# Shortcut Prediction: Take Relation Extraction as An Example

## What we hope the IE model to do

Bill Gates paid a visit to Building 99 of Microsoft yesterday.

Comprehend the *context*, and induce the mentioned *relation* of *entities*.

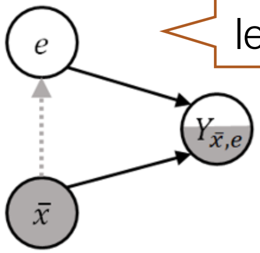



Relations should be inferred based on both mentions and the context

## What it may actually do

Bill Gates [illegible] Microsoft [illegible]

Read the *entities* and guess the *relation* without understanding the *context*.




Context is not captured, leading to entity bias

Overly relying on entity mentions lead to a shortcut for RE

How to do we mitigate this spurious correlation?

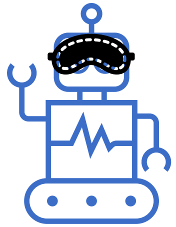
Wang et al. *Should We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis*. **NAACL** 2022

# Strategy 1: Debiased Training

Mention masks: mask out entity names with their types

Person paid a visit to Building 99 of Org yesterday.

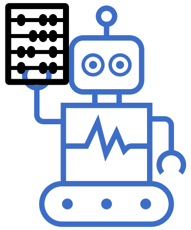
Similarly for *event RE*, we can mask using trigger types and tense



Mask mentions in both training and inference

- Pro: reduces mention biases
- Con: loses semantic information about entities ⇒ performance drop

Reweighting instances: FoCal loss, resampling, two-stage optimization, etc.



$$\mathrm{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

Upweight hard instances

- Pro: reduces training biases by (indirectly) upweighting “underrepresented” instances
- Con: hard instances are not always “underrepresented” instances

Lin et al. *Focal loss for dense object detection*. **CVPR** 2017
Liu et al. Just Train Twice: Improving Group Robustness without Training Group Information. **ICML** 2021

# Strategy 2: Counterfactual Inference

Measure the biases using counterfactual instances, then deduct the biases

① Original Instance ($x$)

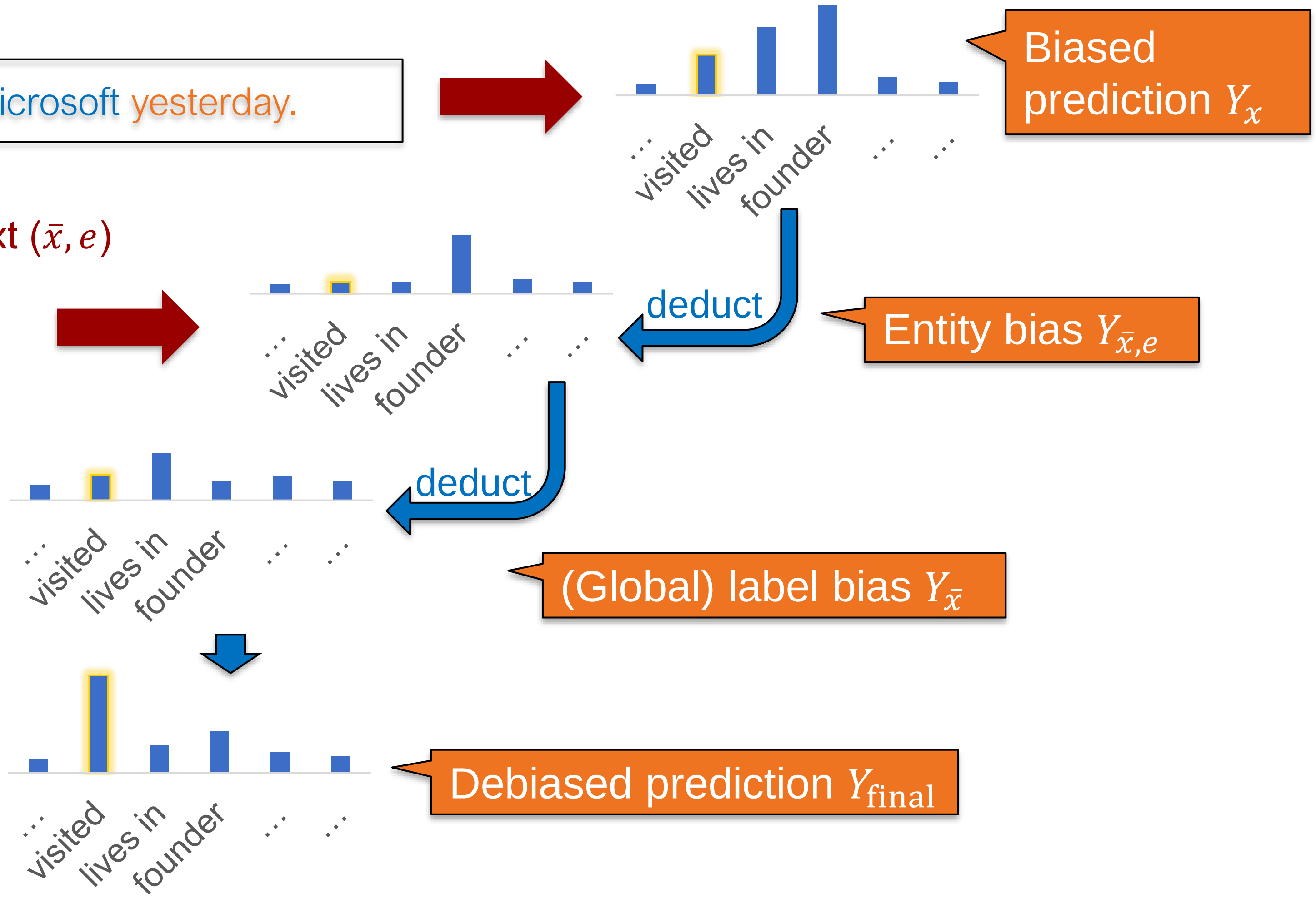
Bill Gates paid a visit to Building 99 of Microsoft yesterday.

② Counterfactual instance w/o context ($\bar{x}, e$)

Bill Gates Microsoft

③ Empty counterfactual instance ($\bar{x}$)

∅




$$Y_{\text{final}} = Y_x - \lambda_1 Y_{\bar{x},e} - \lambda_2 Y_{\bar{x}}$$

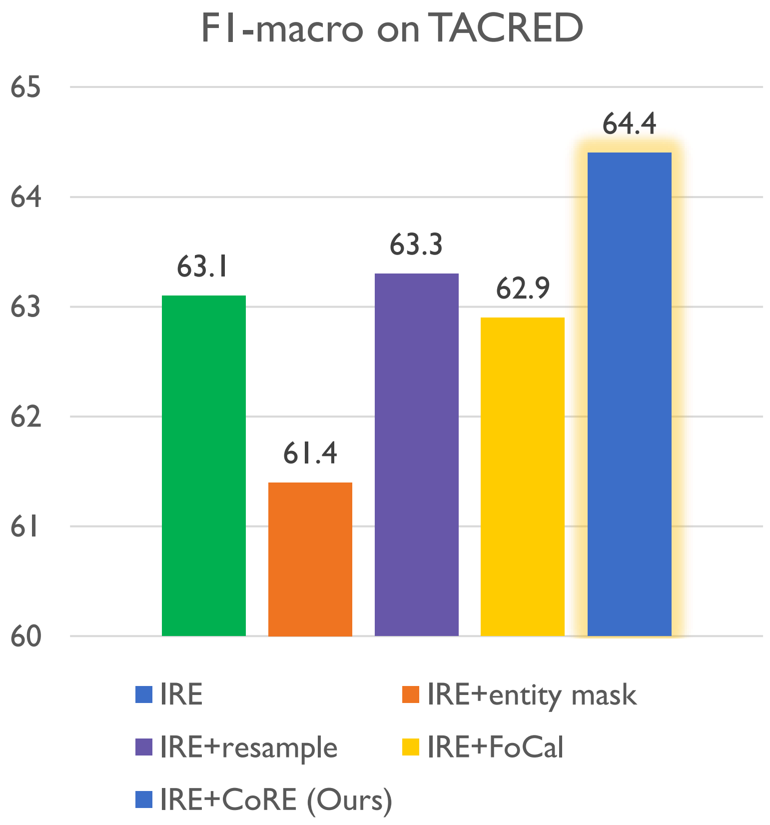
$$\lambda_1^\star, \lambda_2^\star = \arg\max_{\lambda_1,\lambda_2} \psi(\lambda_1, \lambda_2) \quad \lambda_1, \lambda_2 \in [a, b]$$

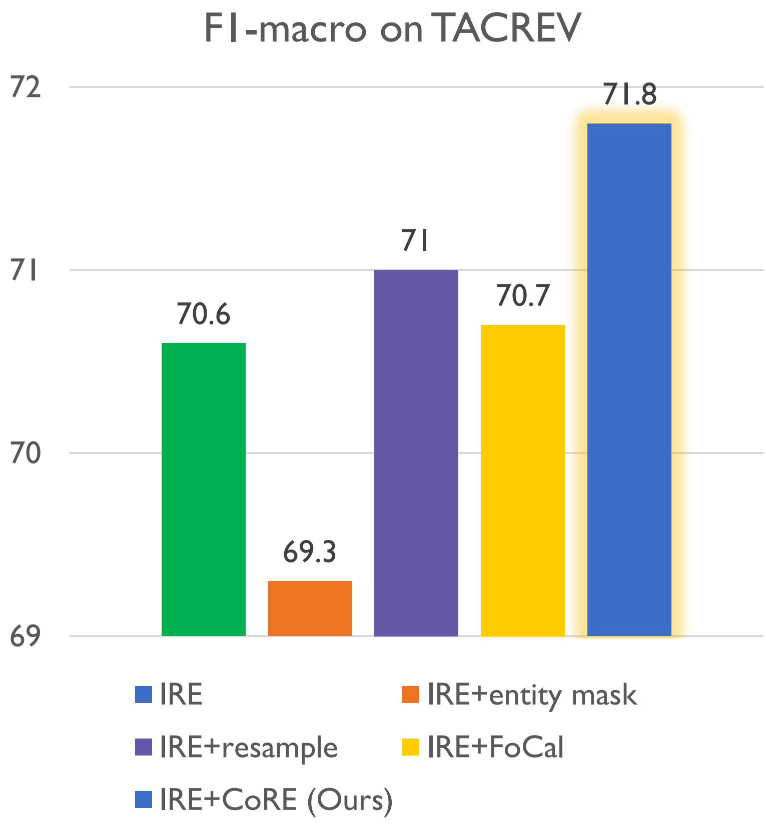
Obtained on dev set

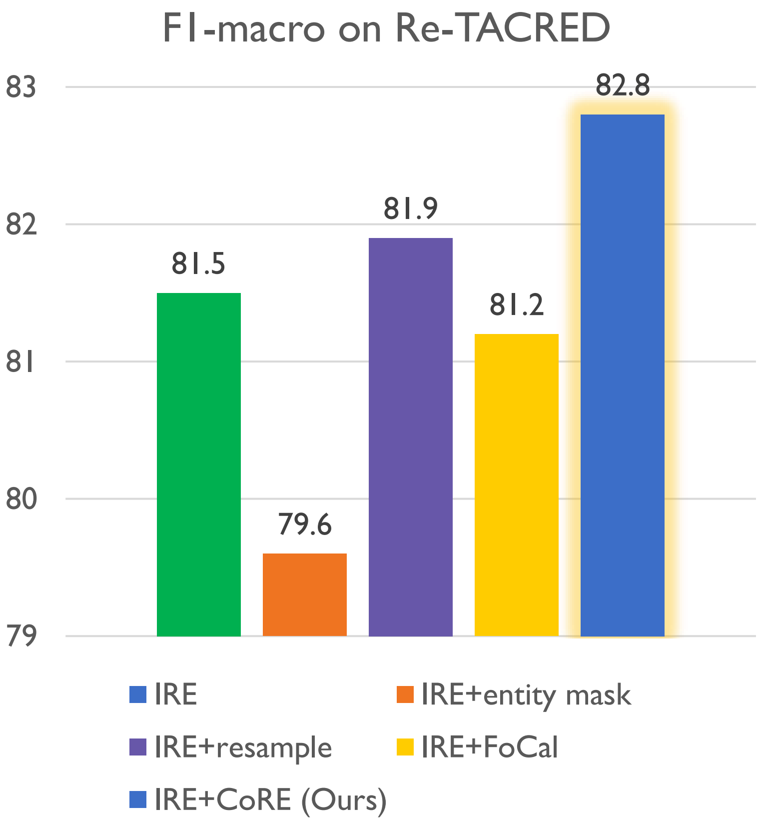

Chen et al. *Counterfactual Inference for Text Classification Debiasing*. **ACL** 2021

Wang et al. *Should We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis*. **NAACL** 2022

# Counterfactual Inference




Counterfactual inference leading to more precise and fairer relation extraction.

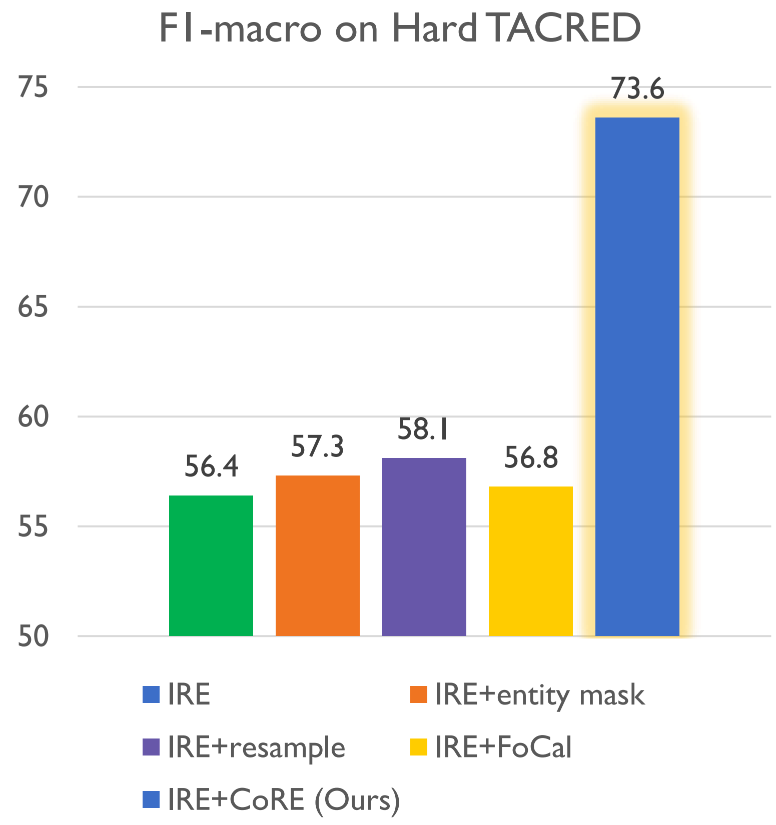
*$IRE_{RoBERTa}$ is one of the best-performing sentence-level RE model (Zhou and Chen 2021). Results also available for LUKE.

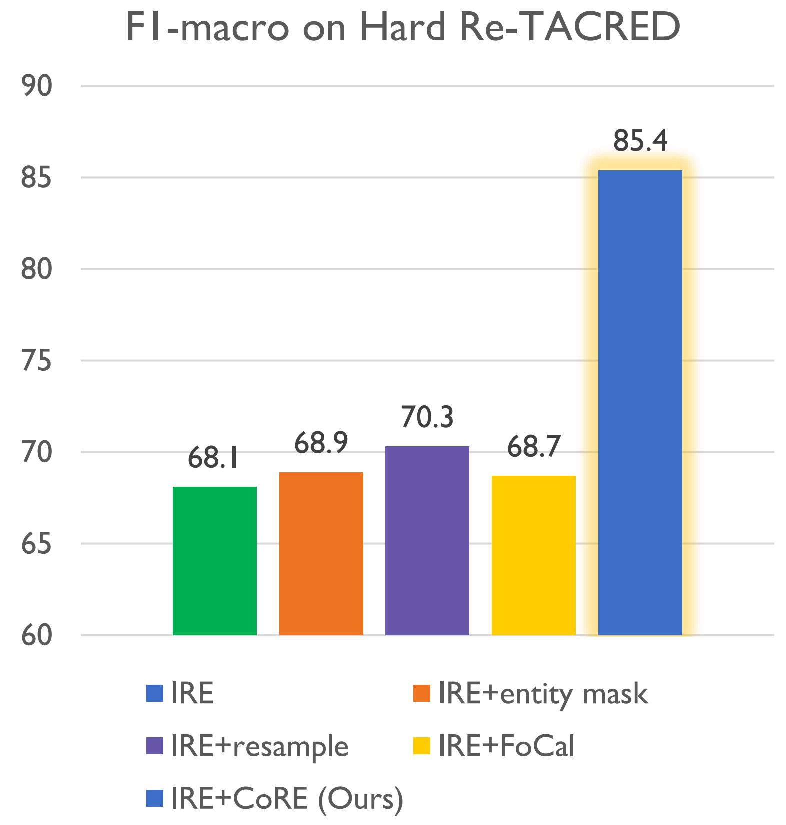
Wang et al. *Should We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis*. **NAACL** 2022

# Counterfactual Inference

Evaluation on out-of-distribution versions of TACRED and Re-TACRED.

- Filtered test sets where combinations of entities and relations have not appeared in training sets.
- Models cannot guess the relations trivially based on entity mentions.




Counterfactual inference leads to significantly more faithful relation extraction.

Wang et al. *Should We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis*. **NAACL** 2022

# Faithfulness Issues in Other IE Tasks

Faithfulness in IE is still an underexplored research direction.

## Entity Typing and Linking

**Mention-Context bias**

**Input:** Last week I stayed in ***Treasure Island*** for two nights when visiting Las Vegas.
**Gold labels:** hotel, resort, location, place
**Pred labels:** island, land, location, place


**Dependency bias**

**Input:** *Most car spoilers* are made from polyurethane, while some are made from lightweight steel or fiberglass.
**Gold labels:** part, object
**Pred labels:** object, car, vehicle

Xu et al. Does Your Model Classify Entities Reasonably? Diagnosing and Mitigating Spurious Correlations in Entity Typing. 2022

## NER

Original NER Examples

I thank my **Beijing** [GPE] friends and wish everyone a Happy **New Year** [EVENT] .

WIKIDATA — Entity-level Attack

Natural Adversarial Examples (*Entity-only*)

I thank my **Bari** [GPE] friends and wish everyone a Happy **Casimir Pulaski Day** [EVENT] .

Context-level Attack

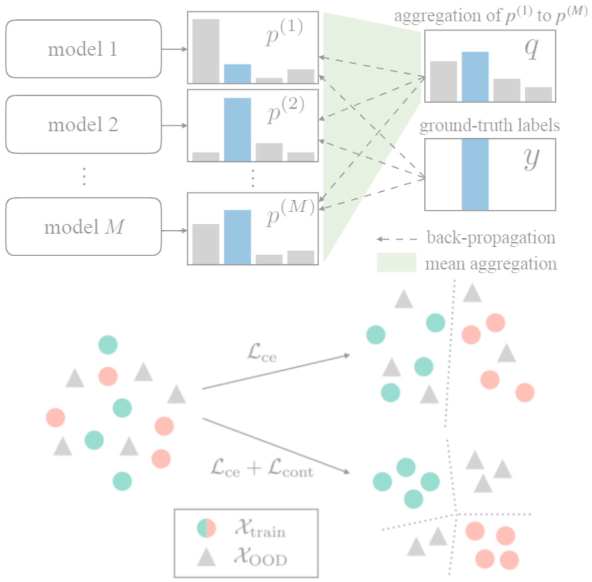
Natural Adversarial Examples (*Entity* + *Context*)

I ***admire*** my **Bari** [GPE] ***roommates*** and wish everyone a Happy **Casimir Pulaski Day** [EVENT] .

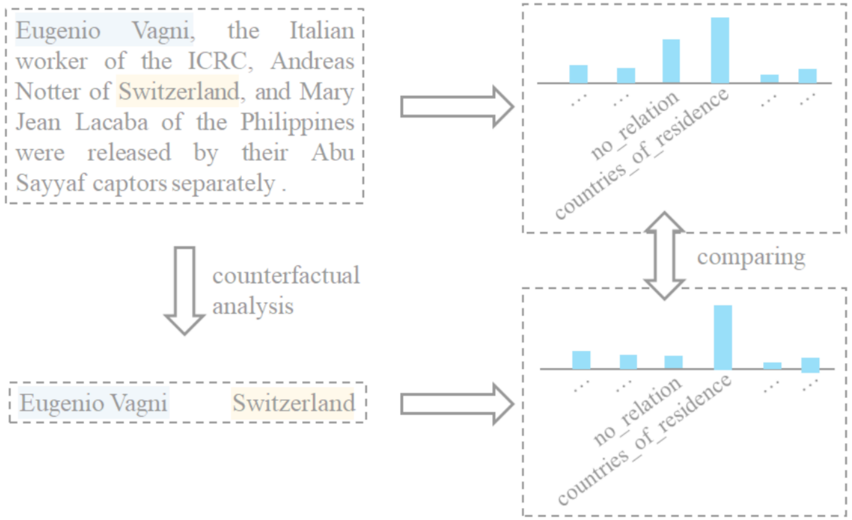
Lin et al. RockNER: A Simple Method to Create Adversarial Examples for Evaluating the Robustness of Named Entity Recognition Models. **EMNLP-21**

# Agenda

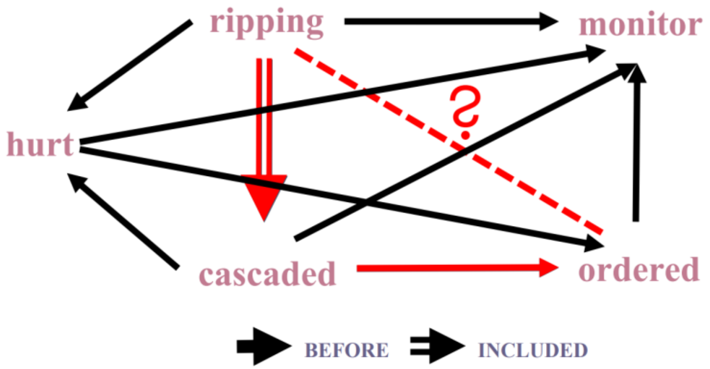
## 1. Noise-robust IE




## 2. Faithful IE




## 3. Logically Consistent IE

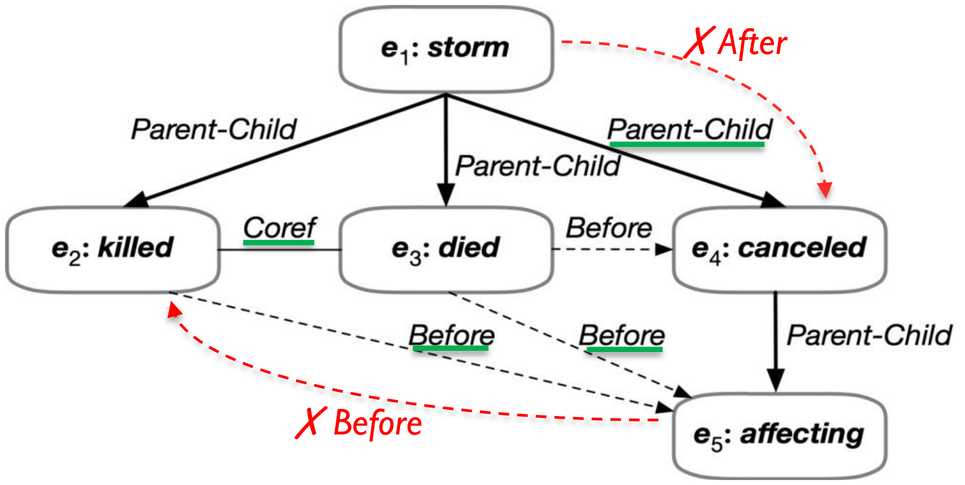



## 4. Open Research Directions

# Consistency of IE

How do we ensure the extracts are **globally consistent**?

On Tuesday, there was a typhoon-strength (*$e_1$:storm*) in Japan. One man got (*$e_2$:killed*) and thousands of people were left stranded. Police said an 81-year-old man (*$e_3$:died*) in central Toyama when the wind blew over a shed, trapping him underneath. Later this afternoon, with the agency warning of possible tornadoes, Japan Airlines (*$e_4$:canceled*) 230 domestic flights, (*$e_5$:affecting*) 31,600 passengers.




Take event-event relation extraction as an example

- Temporal Relations
- Subevent Relations (Memberships)
- Event Coreference

Extracts are not independent, but a structure with dependencies

- E.g., Temporal relations cannot be a loop
- A main event cannot happen after a subevent

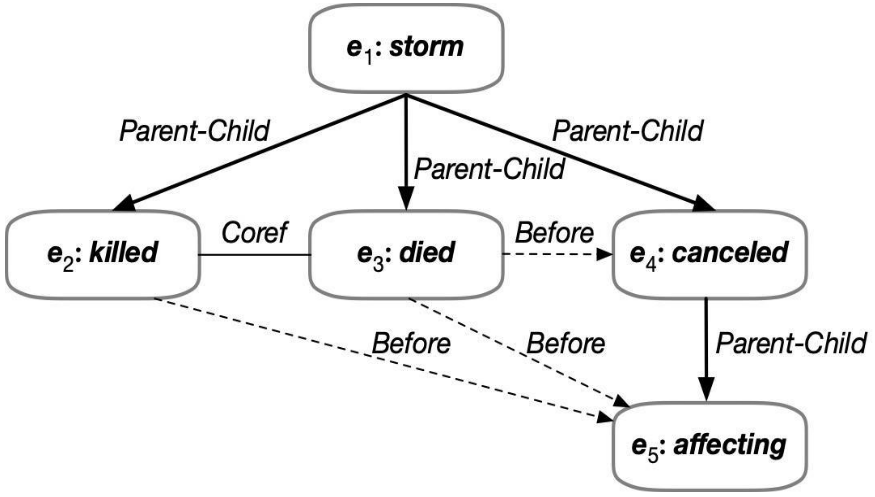
Wang et al. Joint Constrained Learning for Event-Event Relation Extraction. **EMNLP 2020**

# Logical Constraints Of Relations

*Symmetry*
*e3:died* is BEFORE *e4:canceled*
=> *e4:canceled* is AFTER *e3:died*

*Conjunction*
*e3:died* is BEFORE *e4:canceled*
∧*e4:canceled* is a PARENT of *e5:affecting*
=> *e3:died* BEFORE *e5:affecting*

*Transitivity*
*e1:storm* is PARENT of *e4:canceled*
∧*e4:canceled* is a PARENT of *e5:affecting*
=> *e1:storm* is a PARENT of *e5:affecting*




(we also consider *Implication* and *Negation*)

Why adding logical constraints in learning?

- Learning to provide **globally consistent** predictions
- Providing **indirect supervision** across tasks/learning resources


Wang et al. Joint Constrained Learning for Event-Event Relation Extraction. **EMNLP 2020**
Li et al. A Logic-Driven Framework for Consistency of Neural Models. **EMNLP 2019**

# Incorporating Logical Constraints in A Neural Architecture

Symmetry and negation are captured by implication loss; Transitivity is captured by conjunction loss.

Using **product *t*-norm** model constraints as differentiable functions

- $L_A$ Task Loss: $\top \rightarrow r(e_1, e_2)$ $\quad -w_r \log r_{(e_1,e_2)}$
- $L_S$ Implication Loss: $\alpha(e_1, e_2) \leftrightarrow \bar{\alpha}(e_2, e_1)$ $\boxed{\rightarrow}$ $|\log \alpha_{(e_1,e_2)} - \log \bar{\alpha}_{(e_2,e_1)}|$
- $L_C$ Conjunction Loss: $\alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \gamma(e_1, e_3)$ $\boxed{\rightarrow}$ $\log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log \gamma_{(e_1,e_3)}$

  $\alpha(e_1, e_2) \wedge \beta(e_2, e_3) \rightarrow \neg\delta(e_1, e_3)$ $\boxed{\rightarrow}$ $\log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log(1 - \delta_{(e_1,e_3)})$
- Training Objective: $L = L_A + \lambda_S L_S + \lambda_C L_C$

Constraints become regularizers

| $\alpha \backslash \beta$ | PC | CP | CR | NR | BF | AF | EQ | VG |
|---|---|---|---|---|---|---|---|---|
| PC | PC, ¬AF | – | PC, ¬AF | ¬CP, ¬CR | BF, ¬CP, ¬CR | – | BF, ¬CP, ¬CR | – |
| CP | – | CP, ¬BF | CP, ¬BF | ¬PC, ¬CR | – | AF, ¬PC, ¬CR | AF, ¬PC, ¬CR | – |
| CR | PC, ¬AF | CP, ¬BF | CR, EQ | NR | BF, ¬CP, ¬CR | AF, ¬PC, ¬CR | EQ | VG |
| NR | ¬CP, ¬CR | ¬PC, ¬CR | NR | – | – | – | – | – |
| BF | BF, ¬CP, ¬CR | – | BF, ¬CP, ¬CR | – | BF, ¬CP, ¬CR | – | BF, ¬CP, ¬CR | ¬AF, ¬EQ |
| AF | – | AF, ¬PC, ¬CR | AF, ¬PC, ¬CR | – | – | AF, ¬PC, ¬CR | AF, ¬PC, ¬CR | ¬BF, ¬EQ |
| EQ | ¬AF | ¬BF | EQ | – | BF, ¬CP, ¬CR | AF, ¬PC, ¬CR | EQ | VG, ¬CR |
| VG | – | – | VG, ¬CR | – | ¬AF, ¬EQ | ¬BF, ¬EQ | VG | – |

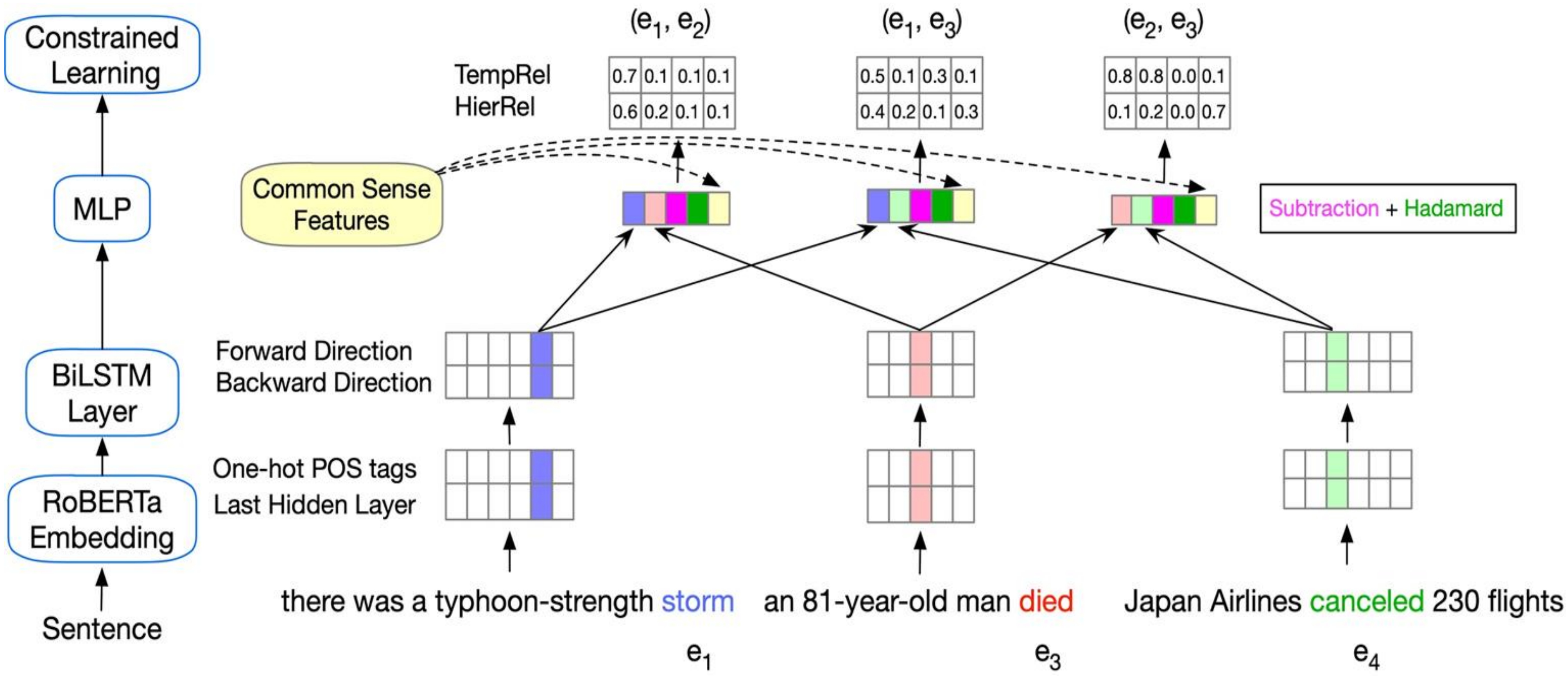
# Joint Constrained Learning

- Temporal Relations
- Subevent Relations (Memberships)
- Event Coreference

Task loss

Implication and conjunction constraint losses

Loss Function: $L = L_A + \lambda_S L_S + \lambda_C L_C$



# The Joint Constrained Learning Architecture

Constrained learning surpasses SOTA TempRel extraction on MATRES [Ning+, ACL-18] by relatively 3.27% in $F_1$.

| Model | $P$ | $R$ | $F_1$ |
|---|---|---|---|
| CogCompTime (Ning et al., 2018c) | 0.616 | 0.725 | 0.666 |
| Perceptron (Ning et al., 2018b) | 0.660 | 0.723 | 0.690 |
| BiLSTM+MAP (Han et al., 2019b) | - | - | 0.755 |
| LSTM+CSE+ILP (Ning et al., 2019) | 0.713 | 0.821 | 0.763 |
| Joint Constrained Learning (ours) | **0.734** | **0.850** | **0.788** |

On HiEve [Glavaš+, LREC-14] for subevent extraction, it relatively surpasses previous methods by at least 3.12% in $F_1$.

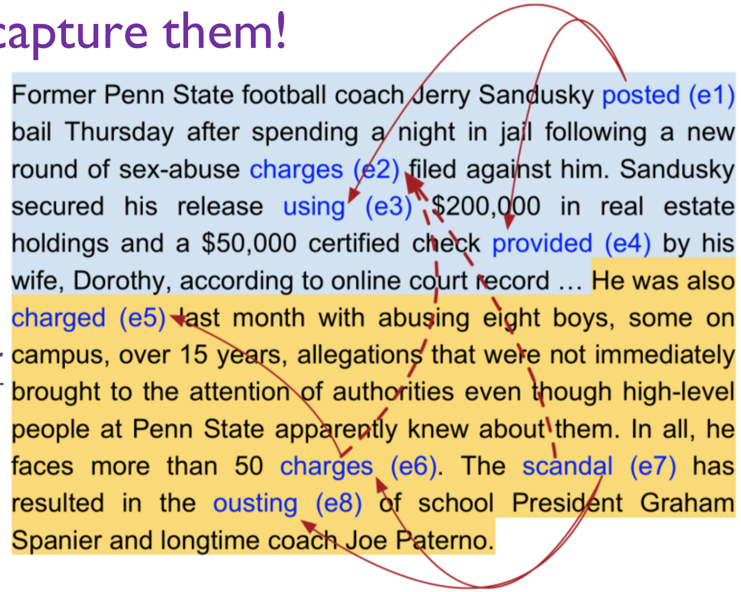
| Model | $F_1$ score | | |
|---|---|---|---|
| | PC | CP | Avg. |
| StructLR (Glavaš et al., 2014) | 0.522 | **0.634** | 0.577 |
| TacoLM (Zhou et al., 2020a) | 0.485 | 0.494 | 0.489 |
| Joint Constrained Learning (ours) | **0.625** | 0.564 | **0.595** |

Key Observations

- Constraints are a natural bridge for learning resources with different sets of relations
- Adding constraints in learning is sufficient to enforce logical consistency of outputs, surpassing ILP in inference (w/ constrained learning) by 2.6-12.3% in ACC

# Automatically Learning Constraints

Some logical constraints can be hard to articulate. We should automatically capture them!




Event-event relations are related to narrative segments

- Text segmentation [Lukasik+ EMNLP-20]: identifying standalone subdocument pieces
- *Subevent relations happen much more often* ***within the same narrative segment***

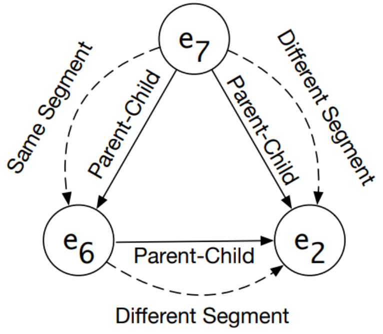
A hard-to-articulate soft probabilistic constraint. How do we capture it?

## Constraint Learning

Training a single-layer rectifier network on all ``triangles'' of the training data

$$\mathbf{w}_k \cdot \mathbf{X} + b_k \geq 0 \quad \Longrightarrow \quad p = \sigma\left(1 - \sum_{k=1}^{K} \mathrm{ReLU}\left(\mathbf{w}_k \cdot \mathbf{X} + b_k\right)\right)$$

Estimates probabilities of conjunctive constraints

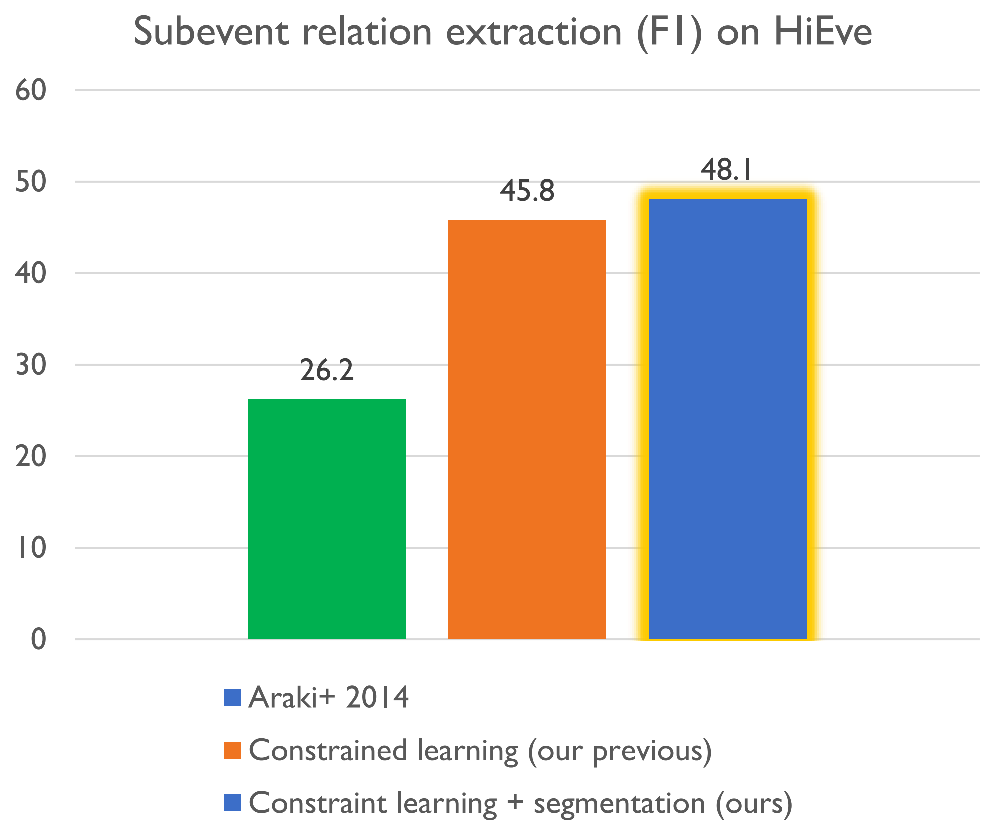


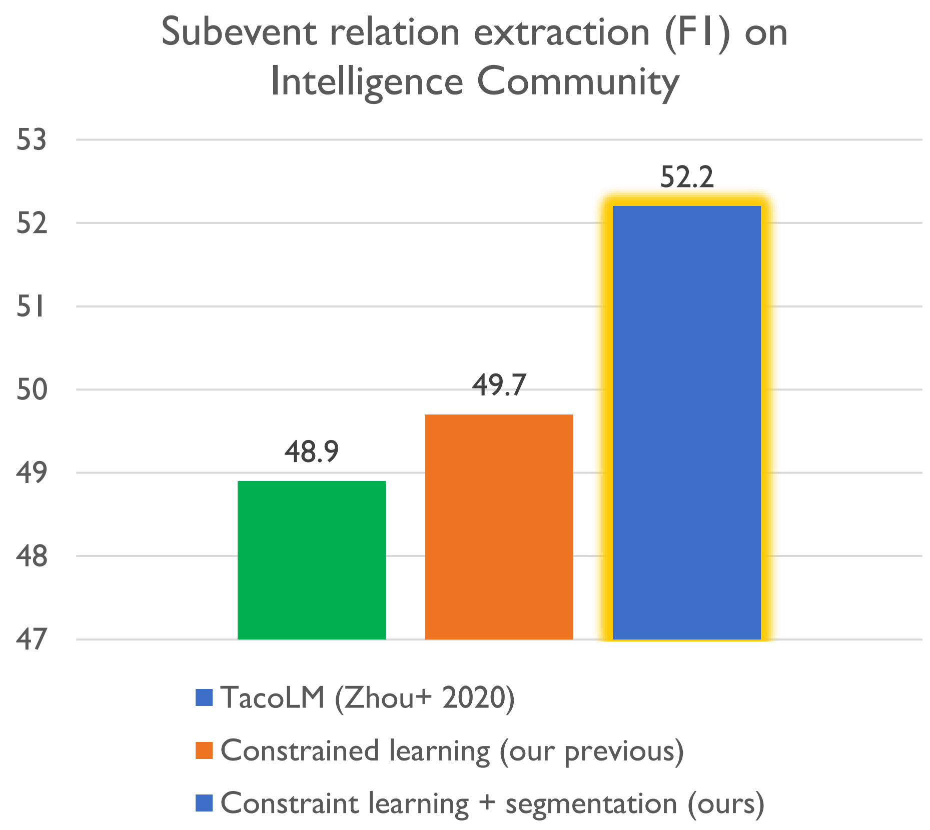

Adding the rectifier estimated constraint probability as a regularization loss in task training

$$L_{cons} = -log\left(Sigmoid(1 - \sum_{k=1}^{N} ReLU(\mathbf{w}_k \cdot \psi + b_k))\right)$$

Pan et al. Learning Constraints for Structured Prediction Using Rectifier Networks. **ACL 2020**
Wang et al. Learning Constraints and Descriptive Segmentation for Subevent Detection. **EMNLP 2021**

# Automatically Learning Constraints

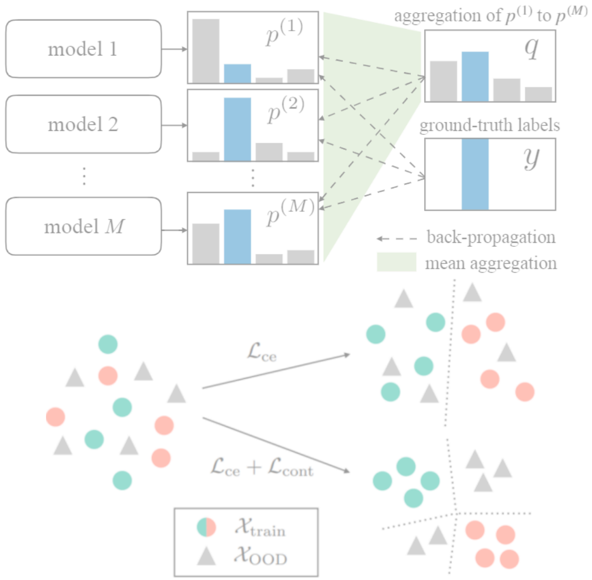


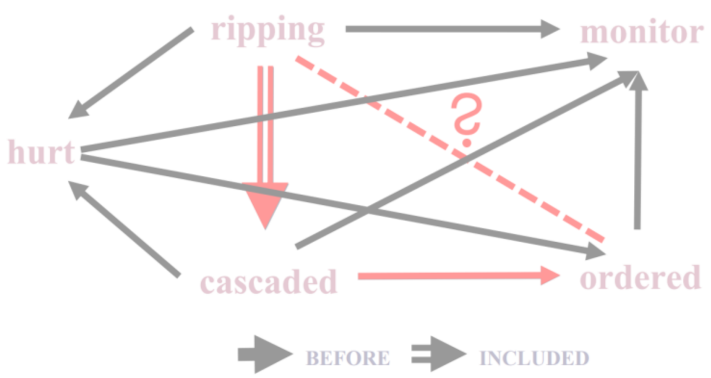

Constraint learning automatically captures soft constraints, and allow narrative segmentation to be introduced as a form of indirect supervision.

# Agenda

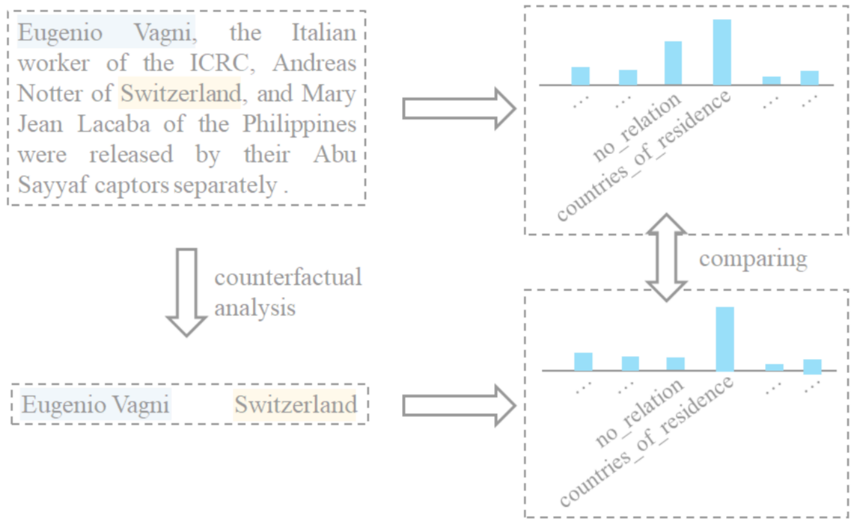
## 1. Noise-robust IE




## 2. Logically Consistent IE




## 3. Faithful IE




## 4. Open Research Directions

# Consolidating Extracts to Knowledge

Extracts are local (differ in contexts), but knowledge is global (unique and consistent)

## Several relevant tasks on text

- Fact verification
- Answer consolidation

Q: Is coffee good for your health?

**Same answers**

Coffee can make you slim down.

Coffee can help with weight loss.

Coffee can relieve headache.

How do those technologies consolidate structural extracts?

## Knowledge alignment across languages


DBpedia

Novel

ヂビペディア

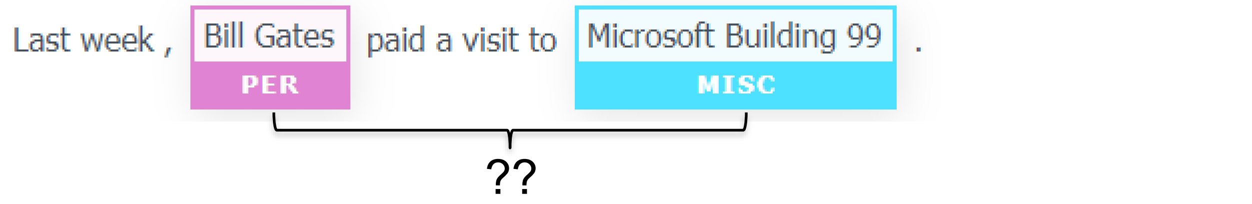
Monogatari (story)
Love story
Royal family story
Realistic novel
Ancient literature

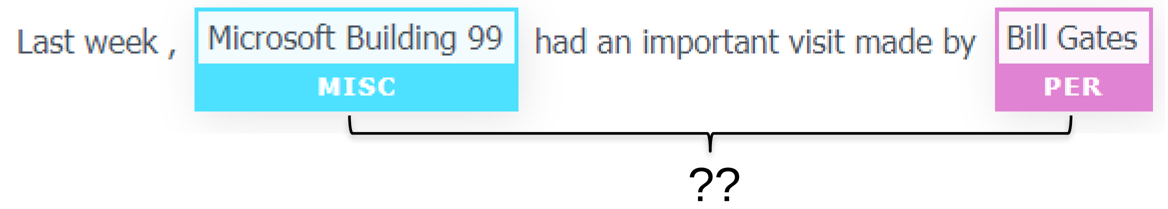
Zhou et al. Answer Consolidation: Formulation and Benchmarking. NAACL 2022
Thorne et al. FEVER: a large-scale dataset for Fact Extraction and VERification. NAACL 2018

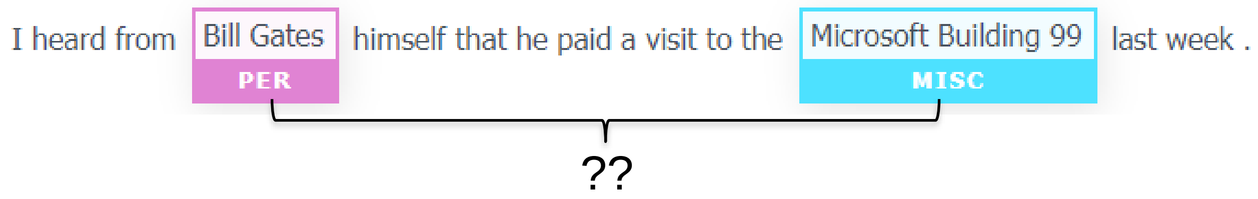
Chen et al. Multilingual Knowledge Graph Completion via Ensemble Knowledge Transfer. EMNLP: Findings 2020
Zhou et al. Prix-LM: Pretraining for Multilingual Knowledge Base Construction. ACL 2022

# Perturbation Robustness

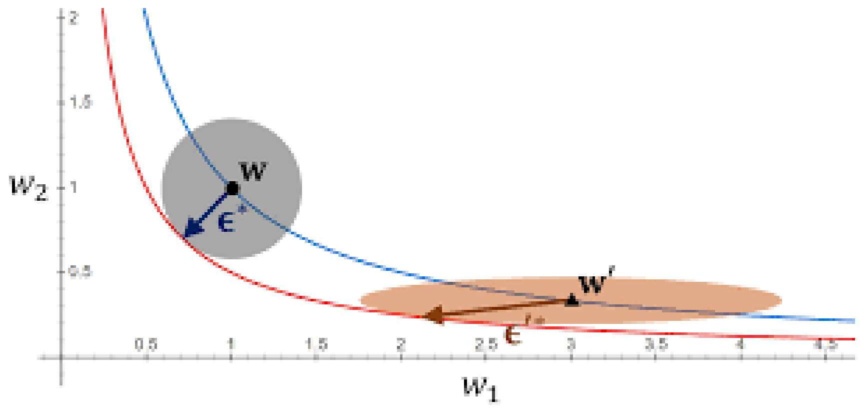
## Semantic Perturbation



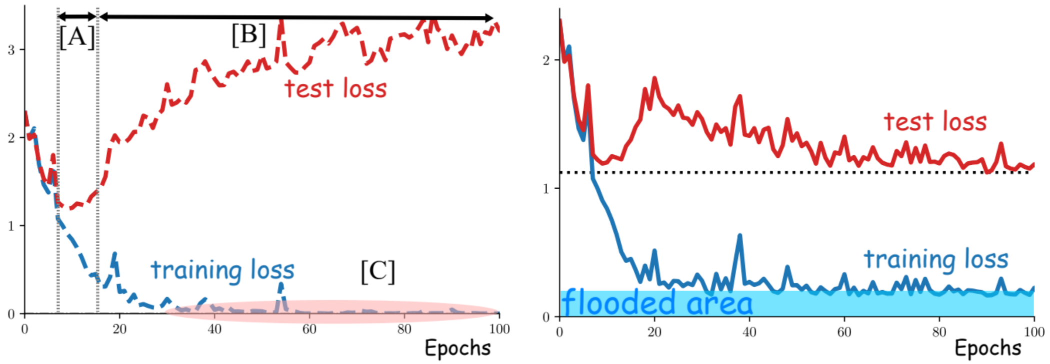




## Parameter Perturbation



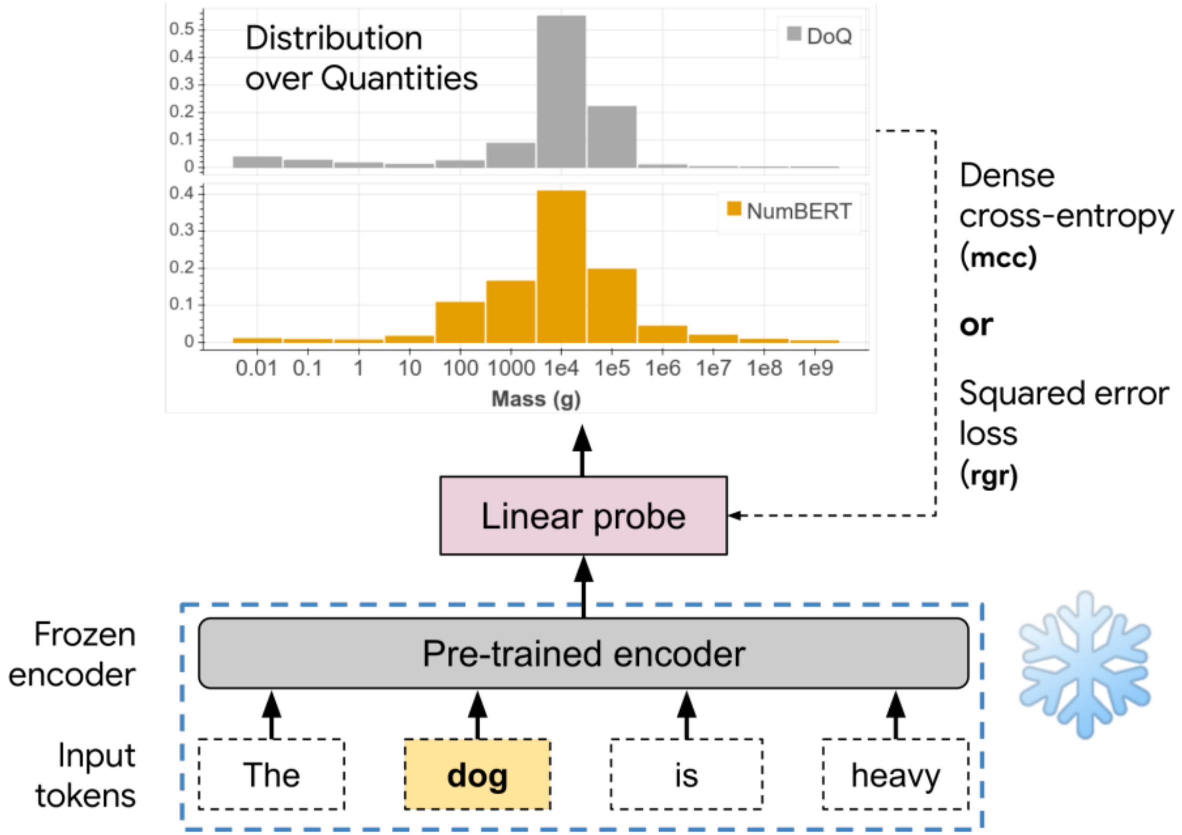

(a) w/o Flooding (b) w/ Flooding




- Qin et al. Improving Entity and Relation Understanding for Pre-trained Language Models via Contrastive Learning. ACL 2021
- Huang et al. Disentangling semantics and syntax in sentence embeddings with pre-trained language models. NAACL 2021
- Foret et al. Sharpness-aware minimization for efficiently improving generalization. ICLR 2020
- Ishida et al. Do We Need Zero Training Loss After Achieving Zero Training Error? ICML 2020

# Quantitative Extraction

## Extracting quantities

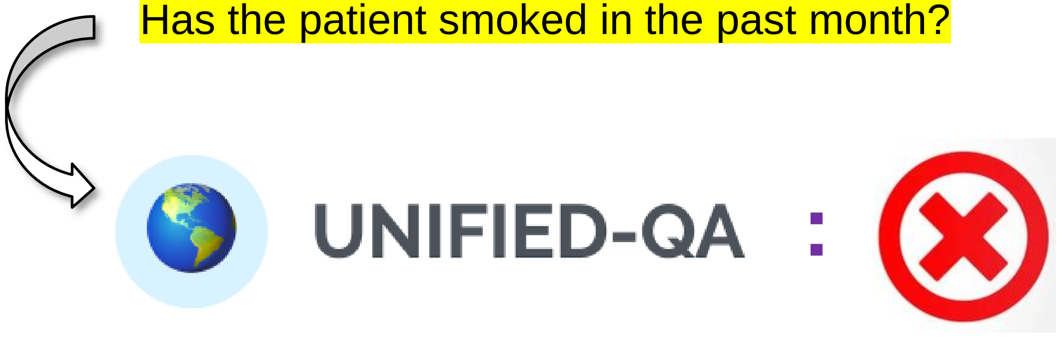



## Temporal verification

Medical Reports

… The patient has been constantly smoking in the past year …

Has the patient smoked in the past month?




Large models still do not support quantitative reasoning well

Zhang et al. Do Language Embeddings Capture Scales? EMNLP: Findings 2020

# References


- Onoe and Durret. Learning to Denoise Distantly-Labeled Data for Entity Typing. NAACL 2019
- Wang et al. CrossWeigh: Training named entity tagger from imperfect annotations. EMNLP 2019
- Zhou and Chen. Learning from Noisy Labels for Entity-Centric Information Extraction. EMNLP 2021
- Toneva et al. An empirical study of example forgetting during deep neural network learning. ICLR 2019
- Dhamija et al. Reducing network agnostophobia. NeurIPS 2018
- Zhou et al. Contrastive Out-of-Distribution Detection for Pretrained Transformers. EMNLP 2021
- Xin et al. The Art of Abstention: Selective Prediction and Error Regularization for Natural Language Processing. ACL 2021
- Qian et al. Counterfactual Inference for Text Classification Debiasing. ACL 2021
- Wang et al. hould We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis. NAACL 2022
- Li et al. A Logic-Driven Framework for Consistency of Neural Models. EMNLP 2019
- Wang et al. Joint Constrained Learning for Event-Event Relation Extraction. EMNLP 2020
- Pan et al. Learning Constraints for Structured Prediction Using Rectifier Networks. ACL 2020
- Wang et al. Learning Constraints and Descriptive Segmentation for Subevent Detection. EMNLP 2021
- Thorne et al. FEVER: a large-scale dataset for Fact Extraction and VERification. NAACL 2018
- Zhou et al. Answer Consolidation: Formulation and Benchmarking. NAACL 2022
- Zhou et al. Prix-LM: Pretraining for Multilingual Knowledge Base Construction. ACL 2022
- Qin et al. Improving Entity and Relation Understanding for Pre-trained Language Models via Contrastive Learning. ACL 2021
- Huang et al. Disentangling semantics and syntax in sentence embeddings with pre-trained language models. NAACL 2021
- Foret et al. Sharpness-aware minimization for efficiently improving generalization. ICLR 2020
- Ishida et al. Do We Need Zero Training Loss After Achieving Zero Training Error? ICML 2020
- Chen et al. Counterfactual Inference for Text Classification Debiasing. ACL 2021
- Wang et al. Should We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis. NAACL 2022
- Lin et al. Focal loss for dense object detection. CVPR 2017
- Zhou and Chen. An Improved Baseline for Sentence-level Relation Extraction. 2021
- Lin et al. A Simple Method to Create Adversarial Examples for Evaluating the Robustness of Named Entity Recognition Models. EMNLP 2021
- Zhang et al. Do Language Embeddings Capture Scales? EMNLP: Findings 2020

# Thank You