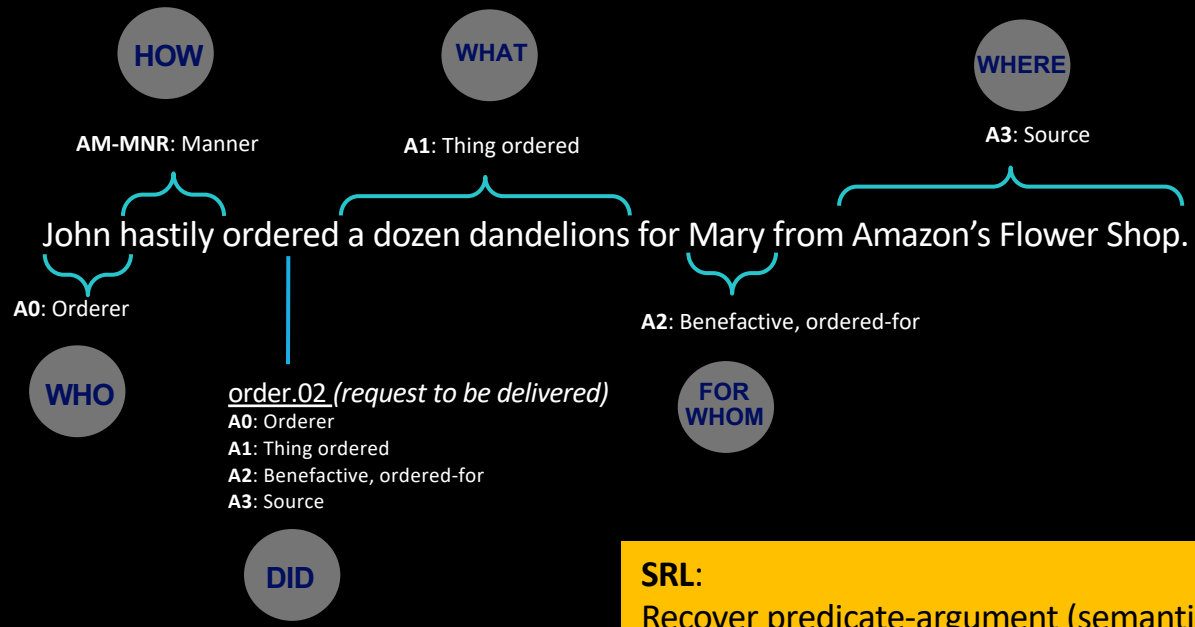
Crowd-in-the-Loop: A Hybrid Approach for Annotating Semantic Roles

Chenguang Wang, Alan Akbik, Laura Chiticariu, Yunyao Li,
Fei Xia, and Anbang Xu

EMNLP'17

What is Semantic Role Labeling (SRL)?



SRL:
Recover predicate-argument (semantic) structure from text

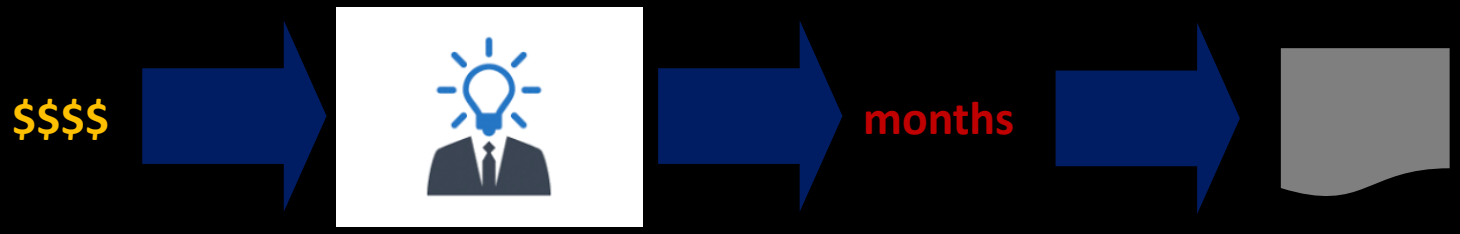
Who did what to whom, when, where and how?

Why Obtaining Labeled Data for SRL is Challenging?

Real world: Language challenge and domain challenge

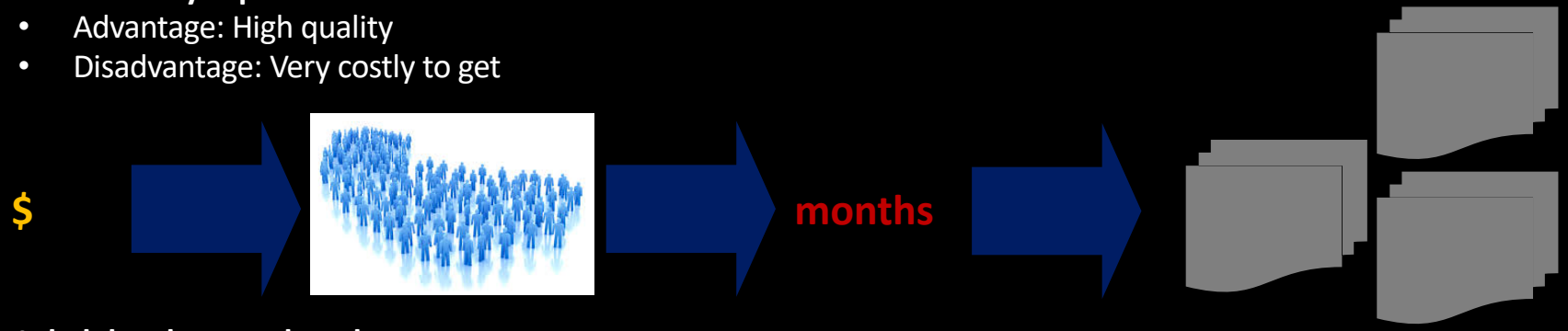
- Lack of labeled data
  - Require linguistic expertise
  - Require language or domain expertise

Crowd-in-the-Loop Learning for Annotating SRL Labels



Label data by experts

- Advantage: High quality
- Disadvantage: Very costly to get



Label data by crowd workers

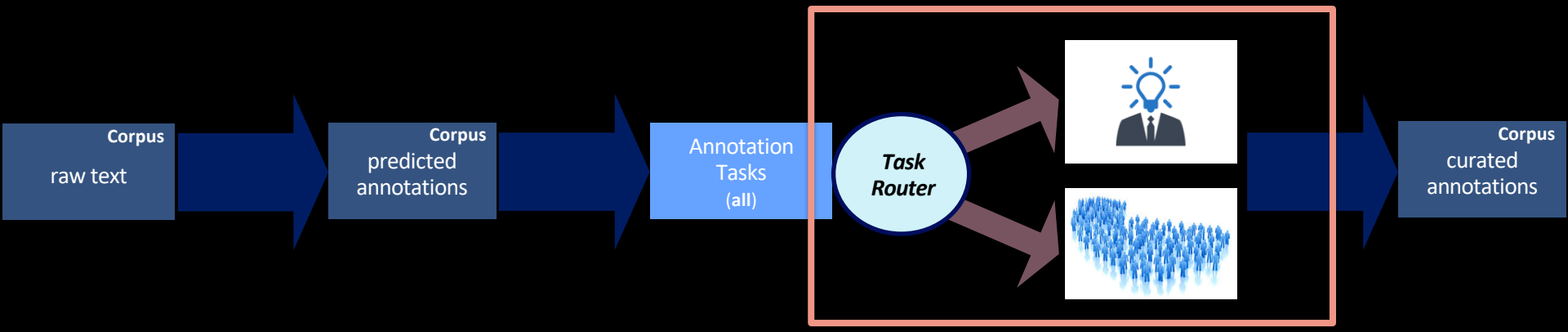
- Advantage: Relatively less costly to get
- Disadvantage: Relatively low quality

SRL shown to be difficult to crowdsource

Crowd-in-the-Loop Learning for Annotating SRL Labels

Crowd-in-the-Loop Learning

- Automatically determine difficulty of curation task
- Difficult tasks are curated by experts
- Easy tasks are curated by crowd



Results

Approach	Annotation Quality			Workload		Correctness	
	P	R	F1	crowd	expert	crowd-only	hybrid
Baseline without curation	0.86	0.83	0.85	0%	0%	-	-
Crowd$_{min3}$	0.92	0.88	0.90	100.0%	0%	0.84	0.84
Crowd$_{min4}$	0.89	0.85	0.87	100.0%	0%	0.84	0.84
Crowd$_{all5}$	0.87	0.84	0.85	100.0%	0%	0.84	0.84
Hybrid$_{min3}$	0.90	0.86	0.88	100.0%	2.2%	0.84	0.84
Hybrid$_{min4}$	0.91	0.87	0.89	100.0%	9.9%	0.84	0.86
Hybrid$_{all5}$	0.93	0.89	0.91	100.0%	27.3%	0.84	0.88
Crowd-in-the-Loop$_{Random}$	0.92	0.88	0.90	**66.4%**	**33.6%**	0.83	0.89
Crowd-in-the-Loop$_{TaskRouter}$	**0.96***	**0.92***	**0.94***	**66.4%**	**33.6%**	**0.92***	**0.95***

+9% F1 improvement compared to SRL output

Results

Approach	Annotation Quality			Workload		Correctness	
	P	R	F1	crowd	expert	crowd-only	hybrid
Baseline without curation	0.86	0.83	0.85	0%	0%	-	-
Crowd$_{min3}$	0.92	0.88	0.90	100.0%	0%	0.84	0.84
Crowd$_{min4}$	0.89	0.85	0.87	100.0%	0%	0.84	0.84
Crowd$_{all5}$	0.87	0.84	0.85	100.0%	0%	0.84	0.84
Hybrid$_{min3}$	0.90	0.86	0.88	100.0%	2.2%	0.84	0.84
Hybrid$_{min4}$	0.91	0.87	0.89	100.0%	9.9%	0.84	0.86
Hybrid$_{all5}$	0.93	0.89	0.91	100.0%	27.3%	0.84	0.88
Crowd-in-the-Loop$_{Random}$	0.92	0.88	0.90	**66.4%**	**33.6%**	0.83	0.89
Crowd-in-the-Loop$_{TaskRouter}$	**0.96***	**0.92***	**0.94***	**66.4%**	**33.6%**	**0.92***	**0.95***

Crowd-in-the-Loop works well! Saving 66.4% expert efforts