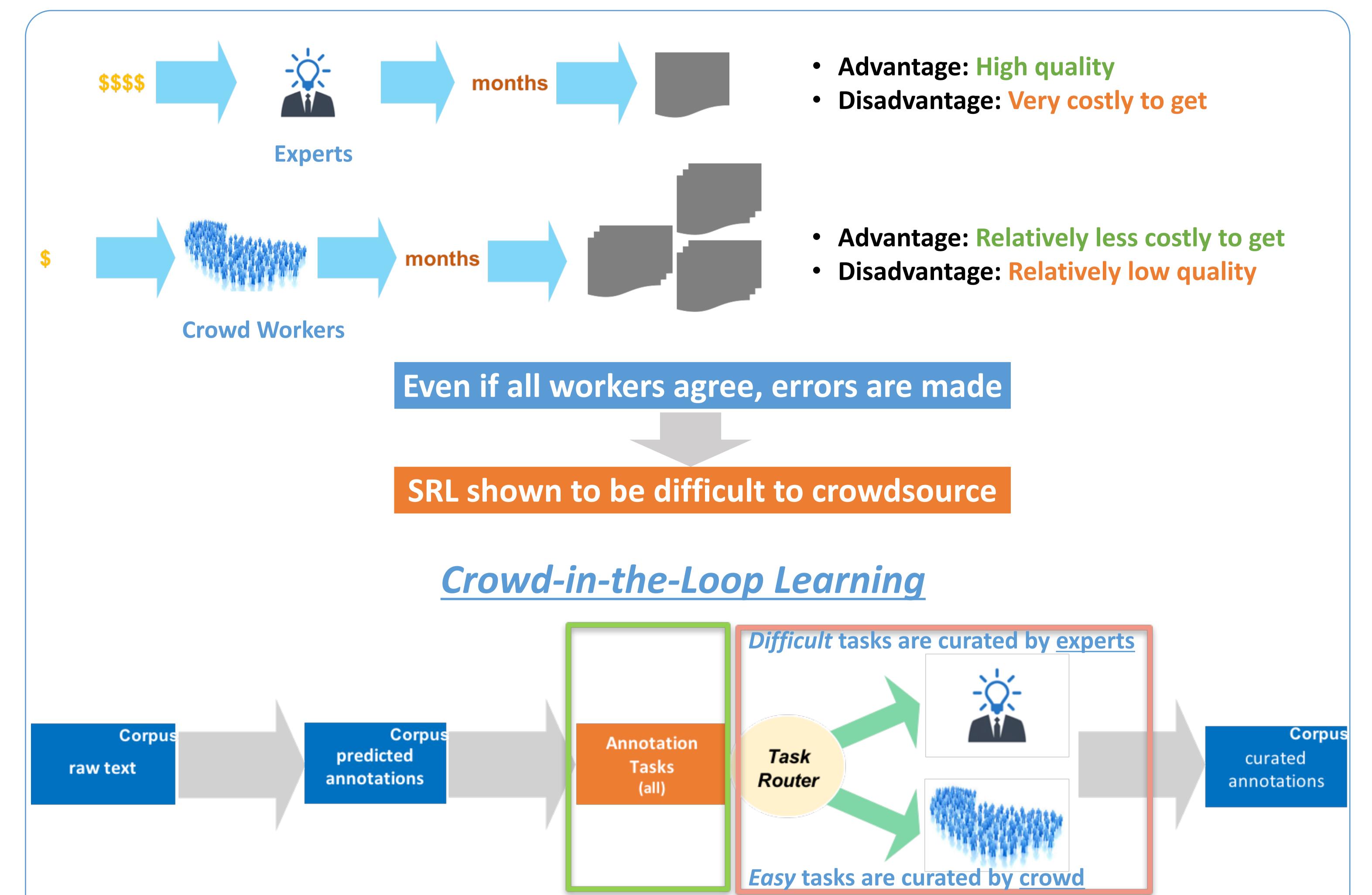
## **Crowd-in-the-Loop: A Hybrid Approach for Annotating Semantic Roles**

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





# **Task Router Classifier**

	Frame	e: BUY.01 (purchase)
Role	Description	Question
<b>A0</b>	buyer	Who is buying something?
<b>A1</b>	thing bought	What is being bought?
<b>A2</b>	seller	From whom is something bought?
<b>A3</b>	price paid	What is the price paid?
<b>A4</b>	benefactive	For whom is something bought?

#### Sentence

And many fund managers have built up cash levels and say they will be buying stock this week.

Question What is being **bought** in this sentence? Is it: "stock"?

**Answer Options** O No, right answer *not* mentioned OYes ○ No, right answer mentioned ○ Unsure

#### **Annotation tasks**

## **Sentence-level features**

*# of questions for sentence,* 

# of question types for sentence...
Local-layer SVM classifier

**Task-level features** 

Frame label, role label, question type

### **Global-layer SVM classifier**

 $\max_{\alpha} \mathbf{1}^{T} \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^{T} \mathbf{Y} \mathbf{K} (\mathbf{X}^{f^{T}} \mathbf{X}^{f}) \mathbf{Y} \boldsymbol{\alpha}$ s.t.  $\boldsymbol{y}^{T} \boldsymbol{\alpha} = 0, 0 \leq \boldsymbol{\alpha} \leq C \mathbf{1}.$ 

#### **Fuzzy two-layer SVM classifier**

Approach	ANNOTATION QUALITY		WORKLOAD		Correctness		
	Р	R	F1	crowd	expert	crowd-only	hybrid
Baseline without curation	0.86	0.83	0.85	0%	0%	-	-
CROWD <sub>min3</sub>	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 <sub>all5</sub>	0.87	0.84	0.85	100.0%	0%	0.84	0.84
HYBRID <sub>min3</sub>	0.90	0.86	0.88	100.0%	2.2%	0.84	0.84
HYBRID <sub><math>min4</math></sub>	0.91	0.87	0.89	100.0%	9.9%	0.84	0.86
HYBRID <sub>all5</sub>	0.93	0.89	0.91	100.0%	27.3%	0.84	0.88
CROWD-IN-THE-LOOP <sub>Random</sub>	0.92	0.88	0.90	66.4%	33.6%	0.83	0.89
$CROWD-IN-THE-LOOP_{TaskRouter}$	<b>0.96</b> *	0.92*	0.94*	66.4%	33.6%	0.92*	0.95*

Results

rowd-in-the-Loop radeoff Quality and Cost!

- +9% F1 improvement
- -66.4% expert efforts