Crowd-in-the-Loop: A Hybrid Approach for Annotating Semantic Roles
Chenguang Wang, Alan Akbik, Laura Chiticariu, Yunyao Li, Fei Xia, and Anbang Xu

---

**Annotation tasks**

- **Frame: *s*r*f (person)**
  - **Role Description**
    - A1: buyer
    - A2: seller
  - **Question**
    - What is being bought?
  - **Answer Options**
    - Yes, right answer mentioned
    - Yes, right answer not mentioned

**Sentence-level features**
- # of questions for sentence
- # of question types for sentence

**Task-level features**
- Frame label, role label, question type

---

**SRL shown to be difficult to crowdsource**

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

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

---

**Results**

- **Crowd-in-the-Loop tradeoff: Quality and Cost**
  - +9% F1 improvement
  - -66.4% expert efforts

---

**Task Router Classifier**

- **Local-layer SVM classifier**
- **Global-layer SVM classifier**
- **Fuzzy two-layer SVM classifier**

---

<table>
<thead>
<tr>
<th>Approach</th>
<th>Annotation Quality</th>
<th>Workload</th>
<th>Correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F1</td>
</tr>
<tr>
<td>Baseline without curation</td>
<td>0.86</td>
<td>0.83</td>
<td>0.85</td>
</tr>
<tr>
<td>CROWD_{min3}</td>
<td>0.92</td>
<td>0.88</td>
<td>0.90</td>
</tr>
<tr>
<td>CROWD_{min4}</td>
<td>0.89</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>CROWD_{min5}</td>
<td>0.87</td>
<td>0.84</td>
<td>0.85</td>
</tr>
<tr>
<td>HYBRID_{min3}</td>
<td>0.90</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>HYBRID_{min4}</td>
<td>0.91</td>
<td>0.87</td>
<td>0.89</td>
</tr>
<tr>
<td>HYBRID_{min5}</td>
<td>0.93</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>CROWD-IN-THE-LOOP Random</td>
<td>0.92</td>
<td>0.88</td>
<td>0.90</td>
</tr>
<tr>
<td>CROWD-IN-THE-LOOP Task/Router</td>
<td><strong>0.96</strong></td>
<td><strong>0.92</strong></td>
<td><strong>0.94</strong></td>
</tr>
</tbody>
</table>