CoAnnotating: Uncertainty-Guided Work Allocation between Human and Large Language Models for Data Annotation

Minzhi Li, Taiwei Shi, Caleb Ziems, Min-Yen Kan, Nancy F Chen, Zhengyuan Liu, Diyi Yang

National University of Singapore, Institute for Inforcomm Research, A*STAR
University of Southern California, Stanford University

1. Introduction

- A novel paradigm for human-Large Language Model (LLM) collaboration for data annotation: CoAnnotating

2. Expertise Estimation

Prompt LLM in diverse ways

- Basic Instruction
- Sequence Swapping
- Paraphrase
- Question Answering

Are they paraphrases?

3. Work Allocation

- Ours: Entropy based Allocation: Sort by entropy (asc.)
- Text 100: yyyy
- Text 500: yyyy
- Text 10: yyyy

Baseline 1: Random Allocation

Baseline 2: Self-evaluation based Allocation: Sort by self-reported confidence (desc.)

4. Allocation Analysis

- Pareto Efficiency: an optimal strategy that it is impossible to further improve the annotation quality with a lower annotation cost.

5. Allocative Efficiency

<table>
<thead>
<tr>
<th>% LLM</th>
<th>0</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>57.5</td>
<td>55.9</td>
<td>53.2</td>
<td>46.2</td>
<td>50.3</td>
<td>42.0</td>
</tr>
<tr>
<td>Self-Eval</td>
<td>57.5</td>
<td>57.8</td>
<td>55.9</td>
<td>51.8</td>
<td>52.9</td>
<td>43.0</td>
</tr>
<tr>
<td>Entropy</td>
<td>57.5</td>
<td>58.4</td>
<td>56.9</td>
<td>55.9</td>
<td>53.8</td>
<td>42.0</td>
</tr>
</tbody>
</table>

6. Conclusion + Future Directions

- Consider variations within human and LLM annotators; aim for superhuman-level performance