**Background**

**Pipeline Crowdsourcing**

Documents

ENGLISH DOCUMENT ➔ Find, point out the location to be revised

REVISED DOCUMENT ➔ Fix, revise the phrases at these locations

DOCUMENT

REVISED DOCUMENT

Using workflow to crowdsource a task of revising a document

The problems handled by crowdsourcing are becoming more complex, and we need to use workflows that involve more than one type of subtasks with dataflow among them.


**Proposed Framework**

- **Dynamically changing the price setting according to the progress of the workflow**

  **Updating Prices**

  Let \( r_{i,k} \) be the price. Let \( T_{i,k} \) be the number of task instances of \( t_i \) that workers have already completed at time \( k \).

  \[
  r'_{i,k} = \left( \frac{D - T_{i,k}}{\sum_{j=1}^{N} D - T_{i,j}} \right)^p
  \]

  We assign larger weights to tasks with more task instances that have not yet been performed. Parameter \( p \) controls the degree of the influence over \( r'_{i,k} \).

  We normalize \( r'_{i,k} \) so that \( \sum_{i} r'_{i,k} = 1 \), and compute \( r_{i,k} = r_{i,0} \cdot r'_{i,k} \).

  **Processing Flow**

  Given the inputs, the framework generates tasks for \( t_i \) and updates the prices for the tasks according to the progress of execution. The framework monitors the progress status of the tasks at each interval \( k \), and computes the price \( r_{i,k} \) for each \( t_i \) for the \( k \)-th interval according to their status at that time.

**Experiment**

Dynamic pricing was up to 1.8 times faster on average than stepwise batch execution with fixed prices.

Comparing four methods (one stepwise batch execution and three pipeline processing) in a sequential workflow

- **Result**

  **Box plot of total completion time of each method**

  **Cost**

  About Total cost

  Average and maximum

  \( p1 : \$5.51, \$5.70 \)

  \( p3 : \$5.62, \$5.85 \)

  (complete under budget)

  **Quality**

  Accuracy of the four methods was not statistically significant

  81%, 82%, 89%, and 82% for Stepwise, Pipeline_p0, _p1, and _p3, respectively

  Multiple comparison tests with Bonferroni correction revealed significant differences between Stepwise and Pipeline_p1 (p<.01), and between Stepwise and Pipeline_p3 (p<.05) in the total processing time.

  **Proposed Method**

  **Average Time**

  Stepwise: 300.1

  Pipeline_p0: 221.9

  Pipeline_p1: 158.7

  Pipeline_p3: 187.4

  **Total completion time (min)**

  0 100 200 300 400 500

  **Stepwise**

  **Pipeline_p0**

  **Pipeline_p1**

  **Pipeline_p3**

  Tasks in one process: 150 (=3×50) Budget $6.00

<table>
<thead>
<tr>
<th>Method</th>
<th>Stepwise</th>
<th>Pipeline_p0</th>
<th>Pipeline_p1</th>
<th>Pipeline_p3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>p = 0</td>
<td>p = 1</td>
<td>p = 3</td>
<td></td>
</tr>
</tbody>
</table>