Findings of the Second Workshop on Automatic Simultaneous Translation

Ruiqing Zhang, Chuanqiang Zhang, Zhongjun He, Hua Wu, Haifeng Wang

Baidu Inc.
Shared Task of Automatic Simultaneous Translation (AST)

1. Text-to-text Track

<table>
<thead>
<tr>
<th>Source</th>
<th>今天</th>
<th>上午</th>
<th>我</th>
<th>要</th>
<th>去趟</th>
<th>公司。</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>This</td>
<td>morning,</td>
<td>I</td>
<td>will</td>
<td>go to</td>
<td>the company.</td>
</tr>
</tbody>
</table>

2. Speech-to-text Track

Source

Target

This morning, I will go to the company.
Evaluating AST Systems

Two examples of the results submitted by two teams

Translation Quality: BLEU
Latency: Consecutive Wait (CW)
Average Lagging (AL)

(a)                                                                                             (b)
Evaluating AST Systems
——Monotonic Optimal Sequence

**Optimal Point:**
One result is considered optimal if there is no other point or line above it at an identical latency. In this case, the result is of the highest translation quality at that latency and we define it as an Optimal Point.

![Graph showing BLEU scores against AL (Average Lagging) with Monotonic Optimal Points indicated.](image-url)
Evaluating AST Systems

—— Optimal Points

Optimal Point:
One result is considered optimal if there is no other point or line above it at an identical latency. In this case, the result is of the highest translation quality at that latency and we define it as an Optimal Point.
Evaluating AST Systems

---

Monotonic Optimal Sequence

\[ S_{T_i} = \frac{N(p^*_t)}{N(p_t)} \]
Evaluating AST Systems
——Iterative Monotonic Optimal Sequence

(a). Plotting all teams
(b). Step 1 of I-MOS
(c). Step 2 of I-MOS
(d). Step 3 of I-MOS

Team1 S: 2.5
Team2 S: 2.4
Team3 S: 1.75
Team4 S: 1.67
Team5 S: 1.0

MOS-1

Team1 S: 1.5
Team2 S: 1.4
Team3 S: 0.75
Team4 S: 0.67
Team5

MOS-2

Team1 S: 0.5
Team2 S: 0.4
Team3 S: 0.5
Team4
Team5

MOS-3

Team1 S: 0.5
Team2 S: 0.4
Team3
Team4
Team5

BLEU
Results of the shared task
——— 6 systems evaluated on the two tracks

(a). Track1. Text-to-Text AST
(b). Track2. Speech-to-Text AST
Discussion
——important challenges for AST

• Data Scarcity
  – BSTC is still insufficient to the data-hungry E2E speech translation models.

• Evaluation Dilemma
  – It remains open to question whether it is reasonable to compare two systems with no intersection in latency.

• Applications
  – Robustness and controllability