

DGoT: Dynamic Graph of Thoughts for Scientific Abstract Generation

Xinyu Ning, Yutong Zhao, Yitong Liu*, Hongwen Yang
Beijing University of Posts and Telecommunications, China
{nxybupt, zhaoyutong, liuyitong, yanghong}@bupt.edu.cn

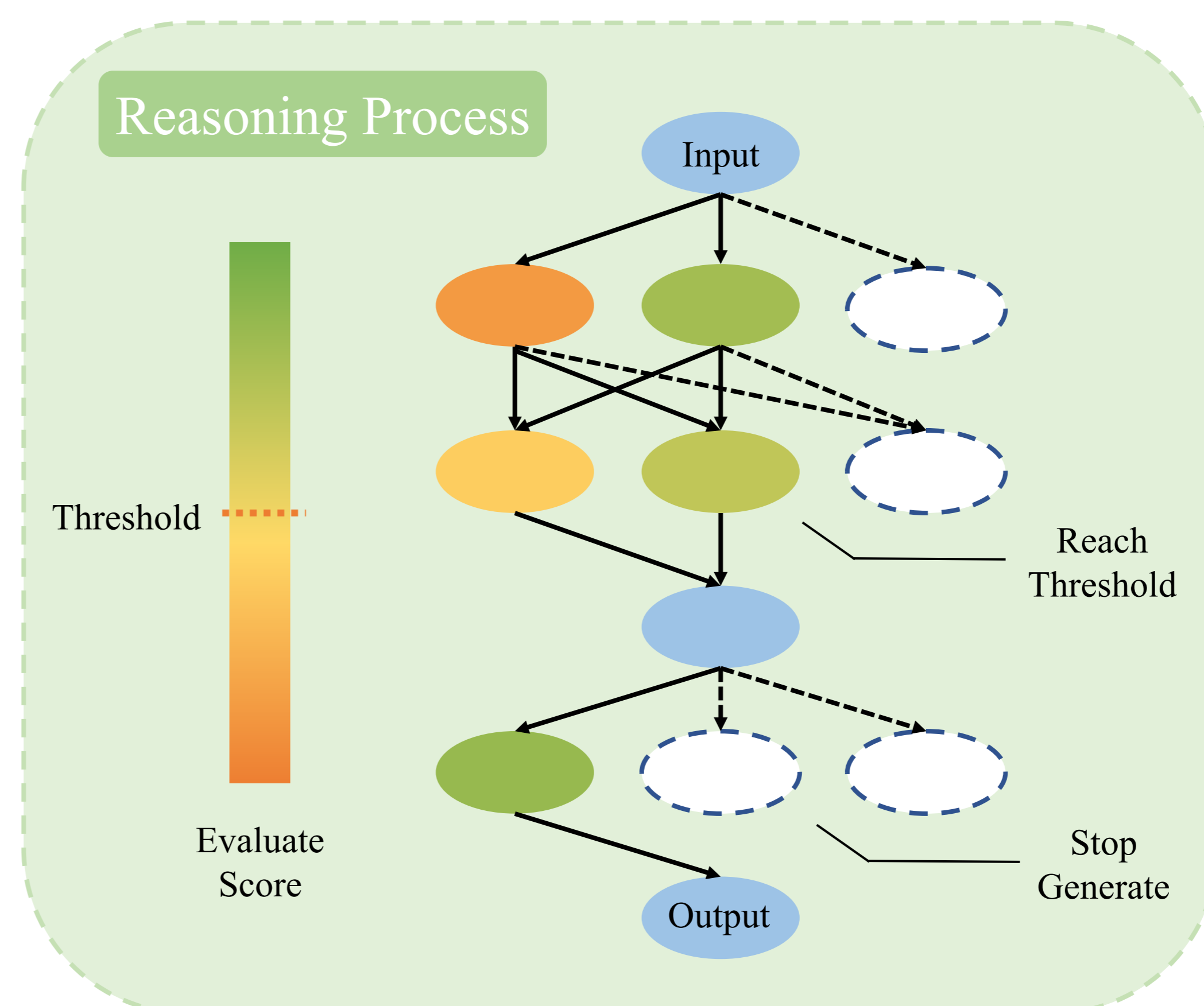
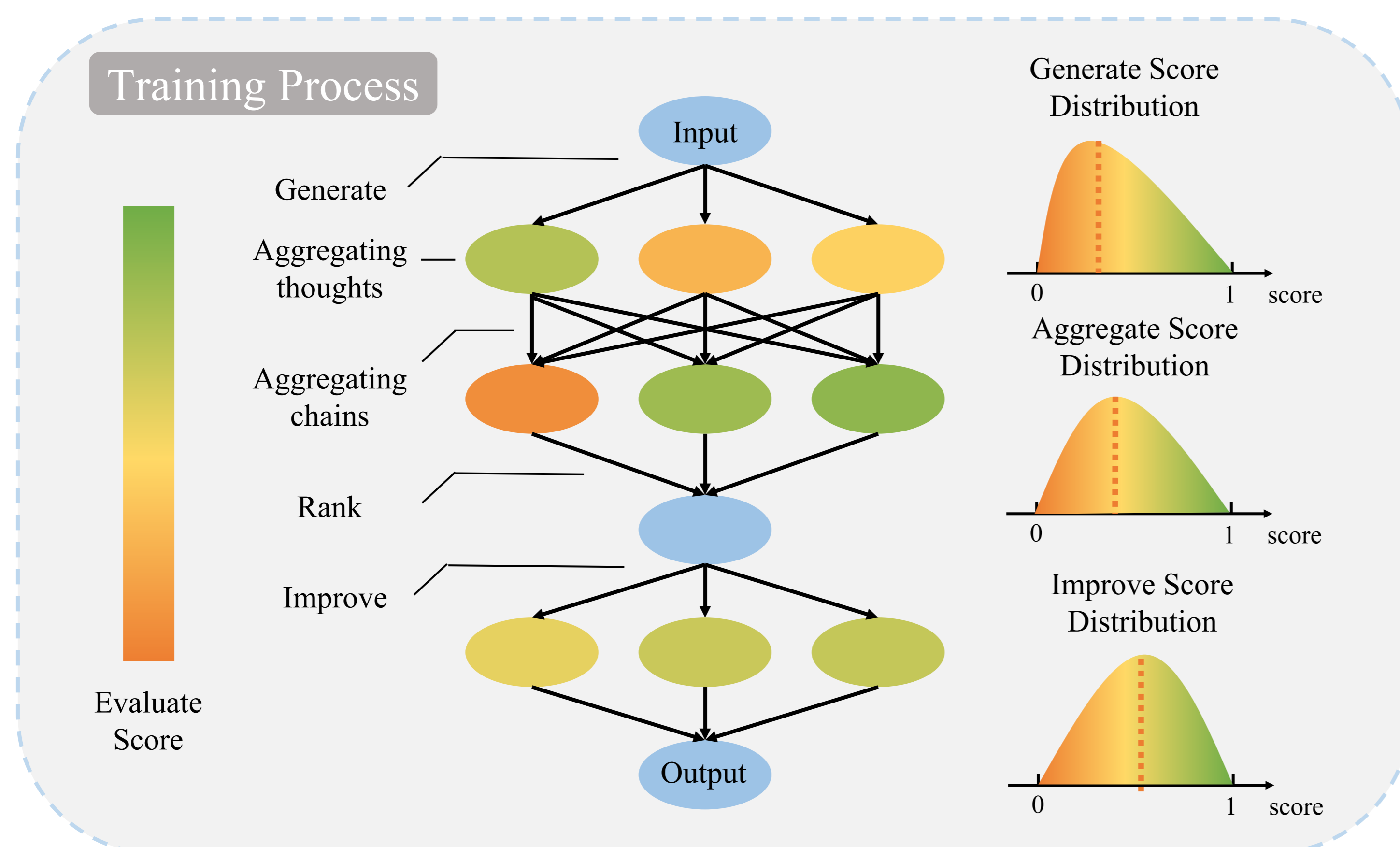
Challenges: Generating good abstracts

- **Generalization:** Transferring knowledge to unseen domains.
- **Training Cost:** Data, time, and compute-intensive training.
- **Domain-Specific:** Scientific jargon complicates comprehension.
- **Capturing Complexity:** Extracting the problem, methods, and conclusions from the entire text.

Solution: Pre-trained LLMs

- **Generalization Power:** Pre-trained with vast data.
- **Cost-Effectiveness:** Leveraging extensive pre-training.
- **Few-Shot Learning:** Adapting output by examples.
- **Prompt Engineering:** Mitigating hallucinations via CoT and combining contents through DoT.

DGoT for Scientific Abstract Generation



Improved graph-based prompt approach

Objectives

- Enhancing abstract effectiveness
- Reducing prompt costs

① Transformations for improving summarization result

Generation $\mathcal{T}_{\text{Gen}}(G, p_{\theta})$ | Aggregation $\mathcal{T}_{\text{Agg}}(G, p_{\theta})$ | Improving $\mathcal{T}_{\text{Impr}}(G, p_{\theta})$

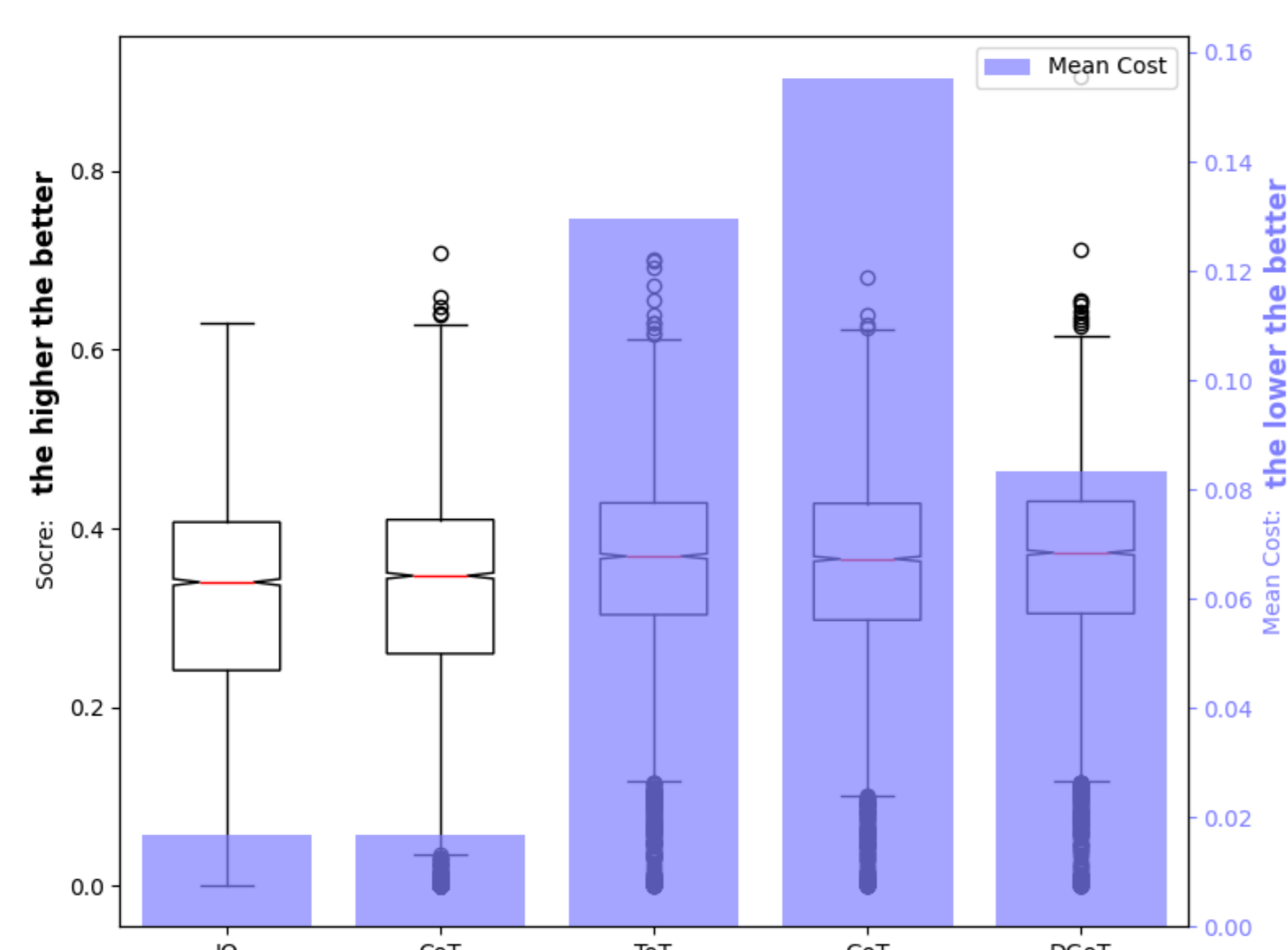
② Evaluating transformations with $\varepsilon(p_{\theta}, S)$ and obtaining their distribution.

③ Threshold-based dynamic transformation.

Empirical Analysis

➤ Main Experimental Result

Method	R-1	R-2	R-L	Prompt Tokens	Response Tokens	Cost	Cost-effectiveness
IO	0.303	0.081	0.166	10660.79	402.79	0.0167	
CoT	0.314	0.083	0.171	10644.81	358.77	0.0166	
ToT	0.356(0.042)	0.098	0.190	82850.63	2606.48	0.1294 (0.1128)	2.686
GoT	0.354(0.040)	0.099	0.190	99184.15	3219.40	0.1552 (0.1386)	3.465
DGoT	0.358(0.044)	0.099	0.192	53414.97	1565.12	0.0833 (0.0667)	1.516



43.7% to 56.4% cost-effectiveness compared to other multi-round query prompt approaches.

DGoT achieves the highest ROUGE score.

➤ Discussions and Future Work

We explore other potential factors influencing the results.

Longer input is not necessarily better; truncating input may lose citation information.

The ROUGE scores of introduction and abstract are in trade-off.

Improving scores through additional inquiries demonstrates marginal utility.

