MCTS: A Multi-Reference Chinese Text Simplification Dataset

Ruining Chong[1], Luming Lu[1], Liner Yang[1] , Jinran Nie[1], Zhenghao Liu[2],
Shuo Wang, Shuhan Zhou[1], Yaoxin Li[1] and Erhong Yang[1]

[1]School of Information Science, Beijing Language and Culture University
[2]Department of Computer Science and Technology, Northeastern University
[3]Department of Computer Science and Technology, Institute for AI, Tsinghua University

Text Simplification

- Make the text easier to understand!
  - For reading assistance
  - For nonnative speakers
  - For language disorders
  - For data augmentation of other NLP tasks

Complex Text

Another major coal transportation corridor in China - the Jigang Railway connecting Tianjin Jixian County and Tianjin Port - has recently broken ground.

Text Simplification System

Simplified Text

Jigang Railway connects Tianjin Jixian County and Tianjin Port for coal transportation, and construction began a few days ago.

ICALL

Introduction

- Very little research on Chinese text simplification.
- How to evaluate the Chinese text simplification systems?

- **MCTS**, a multi-reference dataset for the evaluation of Chinese text simplification models.
  - Largest
  - First Published
  - Human-annotated
  - A by-product training dataset
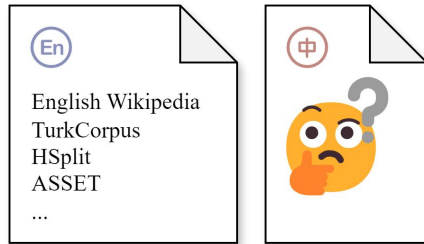
Creating MCTS

- Filtered Penn Chinese Treebank (CTB)

- Annotate
  - Undergraduate Students with **Linguistic** or **CS** background
  - Training course
  - Qualification test

Creating MCTS

- MCTS Simplification Instructions
  - **Paraphrasing**
  - Compression
  - **Structure Changing**
- ASSET Simplification Instructions
  - lexical paraphrasing
  - Compression
  - Sentence Splitting

- 723 sentences with 5 reference. (3615 Simplified Sentences)
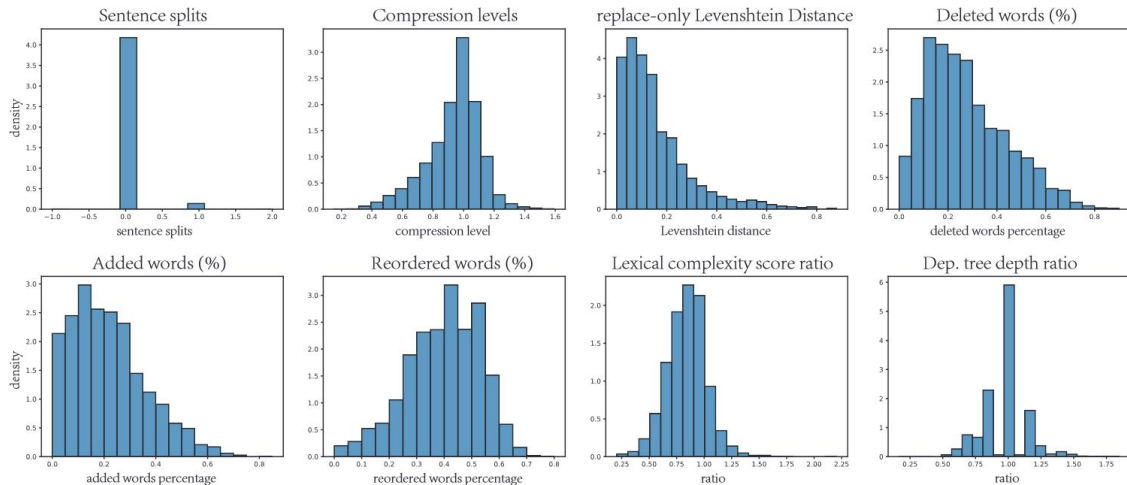 - 366 valid, 357 test.

ICALL

Analysis

Figure 1: Density of text features in simplifications from MCTS

Experiments

- Direct Back Translation
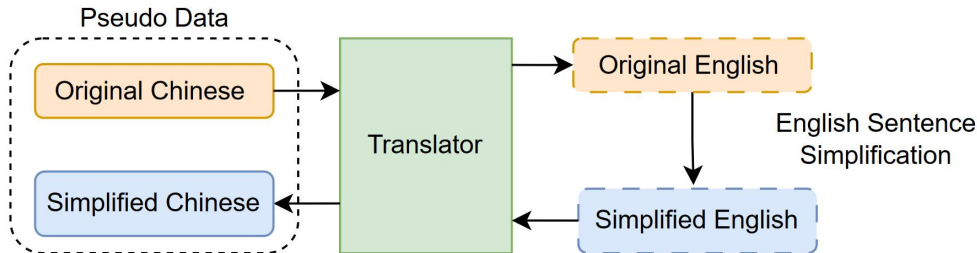- Translated Wiki-Large
- Cross-Lingual Pseudo Data

Fig: Cross Lingual Pseudo Data Method Diagram

Experiments

- Large Language Models
  - Text-davinci-003
  - GPT-3.5-turbo
  - GPT-3.5-turbo-Fewshot

> **我想让你把我的复杂句子替换成简单的句子。你要保持句意不变，但使句子更简单。**
> **复杂句：** 在信息技术领域，被一些专家认为是第三次信息革命的突破口的人工智能技术，目前已被美国、日本和欧洲等列为科研计划的重中之重。
> **简单句：**

> 人工智能技术被认为是信息技术领域的第三次革命，它目前已被美国、日本和欧洲等列为科研计划的重点。

Results

Method	SARI ↑	BLEU ↑	L1-3 (%) ↑	L7+ (%) ↓
Source	22.37	84.75	40.24	44.90
Gold Reference	48.11	61.62	46.25	39.50
Direct Back Translation	40.37	48.72	39.19	45.44
Translated Wiki-Large	28.30	**82.20**	40.32	44.92
Cross-Lingual Pseudo Data	38.49	<u>63.06</u>	41.57	44.24
gpt-3.5-turbo-few-shot	**43.95**	56.46	**44.44**	**40.67**
gpt-3.5-turbo	<u>42.39</u>	49.22	<u>43.68</u>	<u>41.29</u>
text-davinci-003	37.97	36.18	38.80	45.32

Table 5: The automatic evaluation results on the test set of MCTS. Source refers to the score calculated by taking the original sentence directly as the output of the system. ↑ The higher, the better. ↓ The lower, the better. **Bold** means the best result, and <u>underline</u> means the second-best result.

ICALL

Results

Method	Simplicity ↑	Fluency ↑	Adequacy ↑	Avg. ↑	Rank ↓
Direct Back Translation	3.42 ±0.87	4.36 ±0.78	**4.72** ±0.56	4.17	2.88
Cross-Lingual Pseudo Data	4.11 ±0.81	<u>4.46</u> ±0.65	3.88 ±0.96	4.15	2.86
gpt-3.5-turbo	<u>4.17</u> ±0.89	<u>4.46</u> ±0.70	<u>4.43</u> ±0.78	<u>4.35</u>	<u>2.29</u>
Gold Reference	**4.20** ±1.08	**4.68** ±0.55	4.31 ±0.93	**4.40**	**1.97**

Table 6: The human evaluation results. Avg. means the average score of fluency, adequacy, and simplicity. Rank means the subjective ranking for the simplifications. Each annotator annotates all 30 sentences. The values in the table after the ± symbol are the standard deviation.

ICALL

Conclusion

- We introduce the MCTS, a human-annotated dataset for the validation and evaluation of Chinese text simplification systems.
- Simplification ability of some unsupervised methods and advanced large language models were tested on MCTS .
- Even advanced large language models are still inferior to human simplification under the zero-shot and few-shot settings.
- Hope our work can motivate the development of Chinese text simplification systems and provide references for future research.

ICALL

Thanks!