

ruBTS: Russian Sentence Simplification Using Back-translation

Farit Galeev, Marina Leushina, Vladimir Ivanov

Innopolis University

Experiments: What have we done?

- Fine-tuning mBART model as monolingual NMT
 - Bad results - 26.38 (SARI)
- Vanilla Transformer learned from scratch as monolingual NMT
 - Better than fine-tuning mBART - 32.50
- Using pretrained mBART on news summarization
 - Better than fine-tuning mBART but worse than Vanilla Transformer - 32.32
- ruBTS
 - The best result among all methods that we have tested - 37.08

Idea behind ruBTS

- Take Transformer model pretrained on NMT task
- Dictionary of the tokenizer is limited
- Initial sentence changing when applying the back-translation method
- Due to translation from one language to another and back we “apply” the limited dictionary twice which tends to simplification

Advantages of the method

- Can be used out of the box
- Doesn't require any training or fine-tuning and show competitive results with other methods
- Reduce carbon footprint
- Gives an opportunity to participate to the researchers to participate in deep learning research
- Either due to the high computations cost or due to the absence of the datasets.
- Powerful data augmentation tool

Disadvantages of the method

- Scalability of the model depends on:
 - the quality of neural machine translator
 - complexity of the sentence
 - memory requirements
 - other factors that sometimes hard to control
- The method can produce meaningless simplifications in terms of human evaluation

Examples

- Worst simplifications

| | Sentence | SARI |
|------------|---|------|
| Original | Дания является одним из мировых лидеров в использовании возобновляемых источников энергии, в частности энергии ветра. | 0.0 |
| Simplified | Дания является одним из мировых лидеров в использовании возобновляемых источников энергии, в частности энергии ветра. | |
| Original | Разделение равнинных и горных районов между двумя государствами лишило бы многочисленных азербайджанских кочевников летних пастбищ. | 9.8 |
| Simplified | Разделение равнинных и горных районов между двумя государствами лишило бы многих азербайджанских кочевников летних пастбищ. | |

Examples

- Best simplifications

| | Sentence | SARI |
|------------|--|-------|
| Original | В вечерне-ночное время могут возникать ощущения нехватки воздуха, сердцебиение, потливость, озноб или приливы жара. | 32.39 |
| Simplified | В вечернее время может возникнуть чувство отсутствия воздуха, сердцебиения, потности, холода или жары. | |
| Original | 1960 году была выпущена модель 172А. Изменения: хвостовое оперение и руль направления с обратной стреловидностью и крепления для поплавкового шасси. | 35.74 |
| Simplified | Модель 172А была выпущена в 1960 году. | |

Results

| Method | SARI |
|---|--------------|
| MBart fine-tuned for translation | 26.38 |
| MBart pretrained on news summarization | 32.32 |
| Transformer | 32.50 |
| Back-translation (MarianMT-based model) | 37.08 |

Q&A