Multi-Operation Automatic Text Simplification

Fernando Alva Manchego

@feralvam

BeyondMinds

05 July 2021
Outline

● What is (Automatic) Text Simplification?

● Latest Contributions:
  ○ A new dataset for evaluation of multi-operation simplifications
  ○ A hybrid approach for controllable simplification
What is Text Simplification?

To modify the content and structure of a text so that it is easier to understand while preserving its main idea and as much as possible of its meaning.

- **Lexical Paraphrasing**: Uncommon words are replaced by simpler synonyms

  Original: Owls are the order Strigiformes, comprising 200 bird of prey species. Owls hunt mostly small mammals, insects, and other birds though some species specialize in hunting fish.

  Simplification: An owl is a bird. There are about 200 kinds of owls. Owls’ prey may be birds, large insects (such as crickets), small reptiles (such as lizards) or small mammals (such as mice, rats, and rabbits).

- **Sentence Splitting**: A long sentence is divided into several smaller ones.

- **Compression**: “Unimportant” information is removed.

- **Elaboration**: Unusual concepts are explained.

From: http://videolectures.net/esslli2011_lapata_simplification/
What is Text Simplification useful for?

- **Information Accessibility**
  - Comprehension in low-ability readers (Mason and Kendall, 1978)
  - Adults suffering from aphasia (Shewan, 1985)
  - People with dyslexia (Rello et al., 2013)
  - Non-native speakers and ESL learners (Crossley et al., 2007)

- **NLP Tasks**
  - Parsing (Chandrasekar et al., 1996)
  - Summarisation (Siddharthan et al., 2004; Silveira and Branco, 2012)
  - Machine Translation (Štajner and Popovic, 2016)
  - ...

Simplification Scope

- **Word-Level** (a.k.a Lexical Simplification)
  
  The cat *perched* on the mat. → The cat *sat* on the mat.

- **Sentence-Level**
  
  The second *largest* city of Russia *and one of the world’s major cities*, St. Petersburg has played a *vital* role in Russian history.

  St. Petersburg is the second *biggest* city in Russia.

  St. Petersburg has played an *important* role in Russian history.

- **Document-Level**

  (a) Facebook Chief Executive Mark Zuckerberg announced Tuesday that he plans to eventually donate 99 percent of the Facebook stock owned by him and his wife, Priscilla Chan, shares that are worth about $45 billion today.

  (b) That amount would make it one of the largest philanthropic commitments ever.

  (a) Facebook Chief Executive Mark Zuckerberg announced that he and his wife, Priscilla Chan, will donate 99 percent of their Facebook stock to charity.

  (b) Their promised gift would be one of the largest charitable donations ever made.

  (c) Together, the couple's shares are currently worth about $45 billion.
Slightly more fourth-graders nationwide are reading proficiently compared with a decade ago, but only a third of them are now reading well, according to a new report.

Fourth-graders are better readers than 10 years ago. But few of them read well.

How do you determine the quality of an automatic simplification?

Sequence-to-Sequence Model
- Machine Translation
- Summarization
- Caption Generation
...
Metrics used in Machine Translation

- **BLEU** (Papineni et al., 2002)

\[
p_n = \frac{\sum_{S \in C} \sum_{ngram \in S} \text{Count}_\text{matched}(\text{ngram})}{\sum_{S \in C} \sum_{ngram \in S} \text{Count}(\text{ngram})}
\]

\[
BP = \begin{cases} 
1 & \text{if } c > r \\
1 - \frac{r}{c} e & \text{if } c \leq r 
\end{cases}
\]

\[
\text{BLEU} = BP \times \exp \left( \sum_{n=1}^{N} w_n \log p_n \right)
\]

- **BERTScore** (Zhang et al., 2020)
SARI (Xu et al., 2016)

\[
SARI = d_1 F_{add} + d_2 F_{keep} + d_3 P_{del}
\]
\[
d_1 = d_2 = d_3 = 1/3
\]

**Input:** About 95 species are currently accepted.

**Output-1:** About 95 you now get in .

**Output-2:** About 95 species are now agreed .

**Output-3:** About 95 species are currently agreed.

**REF-1:** About 95 species are currently known .
**REF-2:** About 95 species are now accepted .
**REF-3:** 95 species are now accepted .

→ 0.2683
→ 0.7594
→ 0.5890
ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations

Fernando Alva-Manchego\textsuperscript{1,*} and Louis Martin\textsuperscript{2,3,*} and Antoine Bordes\textsuperscript{3}
Carolina Scarton\textsuperscript{1} and Benoît Sagot\textsuperscript{2} and Lucia Specia\textsuperscript{1,4}
\textsuperscript{1}University of Sheffield, \textsuperscript{2}Inria, \textsuperscript{3}Facebook AI Research, \textsuperscript{4}Imperial College London

ACL 2020

https://github.com/facebookresearch/asset
Automatic Evaluation in Sentence Simplification

- Human editors perform **multiple operations** simultaneously
  - *Shouldn’t automatic systems be expected to do the same?*
  - *Are we evaluating them for that?*

- **Multi-reference evaluation datasets** focus on only one operation:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Dev</th>
<th>Test</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TurkCorpus (Xu et al., 2016)</td>
<td>✔</td>
<td>✔</td>
<td>Lexical Paraphrasing</td>
</tr>
<tr>
<td>HSplit (Sulem et al., 2018)</td>
<td>✔</td>
<td></td>
<td>Sentence Splitting</td>
</tr>
</tbody>
</table>
Introducing ASSET

- A multi-reference dataset for sentence simplification in English
- Human editors were instructed to perform 3 types of operations:
  - Lexical Paraphrasing
  - Compression
  - Sentence Splitting

<table>
<thead>
<tr>
<th>Original</th>
<th>He settled in London, devoting himself chiefly to practical teaching.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSET</td>
<td>He lived in London. He was a teacher.</td>
</tr>
<tr>
<td>TurkCorpus</td>
<td>He rooted in London, devoting himself mainly to practical teaching.</td>
</tr>
<tr>
<td>HSplit</td>
<td>He settled in London. He devoted himself chiefly to practical teaching.</td>
</tr>
</tbody>
</table>
How was ASSET created?

Simplification Instructions

Qualification Test

Manual Verification

Selected Workers

Original Sentences

Simplification Task

Random Sample

Manual Verification

TurkCorpus

ASSET

<table>
<thead>
<tr>
<th></th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Sentences</td>
<td>2,000</td>
<td>359</td>
</tr>
<tr>
<td>Simplification References</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Quality of Simplifications in ASSET

Do humans prefer multi-operation over single-operation simplifications?

Instructions → Qualification Test → Manual Verification → Selected Workers → Preference Task

Which sentence is more fluent?
Which sentence expresses the original meaning the best?
Which sentence is easier to read and understand?

359 preference judgements per aspect per dataset pair
Quality of Simplifications in ASSET

- **ASSET**’s simplifications are preferred (or similar) in terms of **fluency** and **simplicity** over TurkCorpus or HSplit.

- Simplifications from **TurkCorpus** or **HSplit** are more meaning preserving:
  - Compression was not allowed when creating simplifications.

<table>
<thead>
<tr>
<th></th>
<th>Fluency</th>
<th>Meaning</th>
<th>Simplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASSET</strong></td>
<td>38.4%*</td>
<td>23.7%</td>
<td>41.2%*</td>
</tr>
<tr>
<td>TurkCorpus</td>
<td>22.8%</td>
<td>37.9%*</td>
<td>20.1%</td>
</tr>
<tr>
<td><strong>Similar</strong></td>
<td>38.7%</td>
<td>38.4%</td>
<td>38.7%</td>
</tr>
<tr>
<td><strong>ASSET</strong></td>
<td>53.5%*</td>
<td>17.0%</td>
<td>59.0%*</td>
</tr>
<tr>
<td>HSplit</td>
<td>19.5%</td>
<td><strong>51.5%</strong>*</td>
<td>14.8%</td>
</tr>
<tr>
<td><strong>Similar</strong></td>
<td>27.0%</td>
<td>31.5%</td>
<td>26.2%</td>
</tr>
</tbody>
</table>

Percentages of judges who preferred simplifications in ASSET or TurkCorpus/HSplit.

- **ASSET**’s simplifications are preferred (or similar) in terms of **fluency** and **simplicity** over TurkCorpus or HSplit.

- Simplifications from **TurkCorpus** or **HSplit** are more meaning preserving:
  - Compression was not allowed when creating simplifications.
Are standard multi-reference automatic evaluation metrics reliable when using multi-operation simplifications?
ASSET for Automatic Evaluation

**BLEU** (Papineni et al., 2002):

- Strong **correlation with Meaning Preservation** using simplifications from ASSET or TurkCorpus.
- Some **correlation with Fluency** judgements, but that is **not always the case for Simplicity**.
  - In line with previous work that has shown that BLEU is not a good estimate for simplicity.

<table>
<thead>
<tr>
<th>Metric</th>
<th>References</th>
<th>Fluency</th>
<th>Meaning</th>
<th>Simplicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU</td>
<td>ASSET</td>
<td>0.42*</td>
<td>0.61*</td>
<td>0.31*</td>
</tr>
<tr>
<td></td>
<td>TurkCorpus</td>
<td>0.35*</td>
<td>0.59*</td>
<td>0.18</td>
</tr>
<tr>
<td>SARI</td>
<td>ASSET</td>
<td>0.16</td>
<td>0.13</td>
<td>0.28*</td>
</tr>
<tr>
<td></td>
<td>TurkCorpus</td>
<td>0.14</td>
<td>0.10</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Pearson correlation of human ratings with automatic metrics on automatic simplifications. (*) p < 0.05

**SARI** (Xu et al., 2016):

- **Low correlation** with all criteria and significant only for **simplicity** with ASSET’s references.

SARI may not be suitable to evaluate simplicity in multi-operation simplifications. Better metrics are needed!
Takeaways

● **ASSET**, a new **multi-reference dataset** for evaluation of Sentence Simplification in English
  - Simplifications contain **multiple rewriting transformations**
  - Simplifications are judged **simpler than those in other evaluation corpora**

● **Multi-reference automatic evaluation** metrics show **low correlation** for human judgements of **Simplicity** when using multi-operation simplifications
  - **New metrics are required** for automatic evaluation of simplifications with multiple rewriting operations
Controllable Text Simplification with Explicit Paraphrasing

Mounica Maddela¹, Fernando Alva-Manchego², Wei Xu¹
¹School of Interactive Computing, Georgia Institute of Technology
²Department of Computer Science, University of Sheffield

NAACL 2021

https://github.com/mounicam/controllable_simplification
Previously: Automatic Text Simplification

- **Rewrite** complex text **into simpler language** while retaining the original meaning
- Often involves **three operations**: delete, split and paraphrase

According to Ledford, Northrop executives said they would build substantial parts of the bomber in Palmdale, creating about 1,500 jobs.

According to Ledford, Northrop said they would build most of the bomber parts in Palmdale. It would create 1,500 jobs.
Previously: Automatic Text Simplification

- **Rewrite** complex text into simpler language while retaining the original meaning
- Often involves **three operations**: delete, split and paraphrase

According to Ledford, Northrop executives said they would build *substantial parts of the bomber* in Palmdale, creating about 1,500 jobs.

According to Ledford, Northrop said they would build *most of the bomber parts* in Palmdale. It would create 1,500 jobs.
Problems with SotA Simplification Systems

- Perform **mostly deletion**;

  According to Ledford, Northrop executives said they would build substantial parts of the bomber in Palmdale, creating about 1,500 jobs.

- **Lack controllability** in terms of edit operations

  ```
  ledford is a big group of bomber in palmdale.
  ledford is northrop.
  , said they would build palmdale parts of the substantial in creating.
  ```
Our Work - Controllable Text Generation

- **Control over 3 edit operations** - deletion, splitting and paraphrasing
- Incorporate **linguistic rules with neural** generation models
- **New setup to evaluate** models’ capability over these edit operations
Step 1: Candidate Generation

- **Split + Delete**: Rule-based method \cite{niklaus-etal-2019-transforming} + a seq2seq model

  - 35 hand-crafted grammar rules for English based on Stanford’s parser
  - successfully split 92% of sentences with $\geq 20$ words and make only 6.8% errors.

Step 1: Candidate Generation

- **Split + Delete**: Rule-based method (Niklaus et al., 2019) + a seq2seq model

**INPUT SENTENCE:**

*The exhibition, which opened Oct. 8 and runs through Jan. 3, features 27 self-portraits.*

**SPLITS:**

- The exhibition features 27 portraits.
- The exhibition opened Oct. 8.
- The exhibition runs through Jan. 3.

Step 1: Candidate Generation

- **Split + Delete**: Rule-based method (Niklaus et al., 2019) + a seq2seq model

![Diagram of Candidate Generation]

**COMBINE SPLITS AS CANDIDATES:**

The exhibition features 27 portraits. The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition features 27 portraits.

The exhibition opened Oct. 8. The exhibition runs through Jan. 3. .. (and more)

Step 2: Candidate Ranking

- Rank all candidate outputs after (splitting and deletion)

```
“Gold” Scoring Function

\[ g^*(v_i, y) = e^{-\lambda \|\phi v_i - \phi y\|} \times BERTScore(v_i, y) \]
```
Step 2: Candidate Ranking

- Rank all candidate outputs after (splitting and deletion)

Features:
- number of words in $v_i$ and $x$, compression ratio of $v_i$ with respect to $x$,
- Jaccard similarity between $v_i$ and $x$, the rules applied on $x$ to obtain $v_i$, and the number of rule applications.

Loss function:

$$L_{MR} = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{n_k^2} \sum_{i=1}^{n_k} \sum_{j=1, i \neq j}^{n_k} \max(0, 1 - l_{ij}^k d_{ij}^k)$$

$$d_{ij}^k = g(v_i^k) - g(v_j^k)$$

$$l_{ij}^k = \text{sign} \left( g^*(v_i^k, y^k) - g^*(v_j^k, y^k) \right)$$

Ranker score  Length-penalized BERTScore
Step 3: Paraphrase Generation

- Paraphrase **top-ranked** candidate
  - Data Augmentation with additional training data that focuses on lexical paraphrasing
  - Copy-control token as a soft constraint to control paraphrasing
  - Auxiliary task (if a word should be copied)
**Example Output**

**Input:** Since 2010, project researchers have uncovered documents in Portugal that have revealed who owned the ship.

**Reference:** Scientists have found documents in Portugal. They have also found out who owned the ship.

<table>
<thead>
<tr>
<th>Hybrid-NG</th>
<th>since 2010, project researchers have uncovered documents in Portugal that have about who owns the ship.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>since 2010, scientists have uncovered documents in Portugal that have revealed who owned the ship.</td>
</tr>
<tr>
<td>Transformer</td>
<td><strong>they discovered</strong> that the ship <strong>had been important.</strong></td>
</tr>
<tr>
<td>EditNTS</td>
<td>since 2010, project researchers have uncovered documents in Portugal. have revealed who owned the ship</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Our Model (split, cp = 0.6)</th>
<th>scientists have found a secret deal. they have discovered who owned the ship.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Model (split, cp = 0.7)</td>
<td>scientists have found documents in portugal. they have also found out who owned the ship.</td>
</tr>
<tr>
<td>Our Model (split, cp = 0.8)</td>
<td>scientists have found a documents in portugal. they have discovered who owned the ship.</td>
</tr>
</tbody>
</table>
Experiments - Human Evaluation

- Trained on Newsela-Auto (Jiang et al., 2020)
- 259,778 train + 32,689 valid + 33,391 test pairs
- Human evaluation on 100 random simplifications

Edit-Focused Automatic Evaluation

- Evaluation on sections of the Newsela-Auto test set (Jiang et al., 2020)
- We report SARI, the main automatic metric for Sentence Simplification

Takeaways

- Novel **hybrid approach** for Sentence Simplification that offers **control over 3 types of edit operations**
- Inject **linguistic knowledge** into neural models
- **New dataset** to evaluate lexical paraphrasing
- **Evaluation setup** for edir operations
Conclusions

● Text Simplification is a text generation tasks whose goal is to rewrite a complex sentence into an easier to understand version

● Types of Rewriting: delete, paraphrase, split, etc.

● In this talk:
  ○ A new dataset for evaluation with multi-operation simplification references
  ○ A new hybrid approach for controllable simplification

● Challenges:
  ○ Develop new metrics that evaluate multi-operation capabilities
  ○ Design models that perform more variety of operations (e.g. elaboration)
Thanks!

Fernando Alva Manchego

✉️ f.alva@sheffield.ac.uk
🐦 @feralvam