Keep Calm and Switch On!

Steven Y. Feng, Aaron W. Li, Jesse Hoey
David R. Cheriton School of Computer Science, University of Waterloo

1. Summary

- **Semantic text exchange (STE)**: adjust the semantics of text while preserving its sentiment and fluency
- **Use cases**: text data augmentation and the semantic correction of text generated by chatbots/virtual assistants
- **SMERTI**: a pipeline for STE combining entity replacement, similarity masking, and text infilling
- **Semantic Text Exchange Score (STES)**: a single score to evaluate a model's ability to perform STE
- **Masking (replacement) rate threshold (MRT/RRT)**: a parameter to control the amount of semantic change

2. What is Semantic Text Exchange?

- Original Text: *It is **sunny** outside! That means I must **wear sunscreen**. I hate being **sweaty** and **sticky all over.***
- Replacement Entity: ***rainy***
- Desired Text: *It is **rainy** outside! That means I must **bring an umbrella**. I hate being **wet** and **carrying it around**.*

3. Entity Replacement Module (ERM)

- **Stanford Parser**: determine possible words/phrases to be replaced by the replacement entity (*RE*) using grammatical structure of the input text and *RE*
- **Universal Sentence Encoder (USE)** [1]: identify most similar word/phrase to *RE* (which becomes the replaced entity) by computing semantic similarity between their embeddings

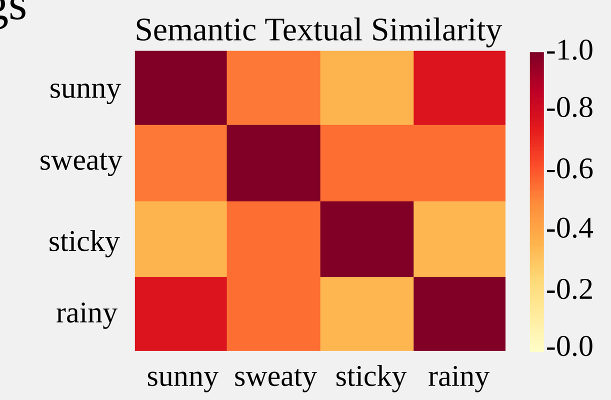

Figure 1: Semantic similarity heat map

4. Similarity Masking Module (SMM)

- Replace semantically similar words to the replaced entity in the input text (above a threshold) with a *[mask]*
- Group adjacent *[mask]* tokens into a single *[mask]*
- Masking (replacement) rate threshold (MRT/RRT): maximum percentage of text that can be masked

5. Text Infilling Module (TIM)

- Two seq2seq models to fill in the *[mask]* tokens:
  - **Bidirectional RNN** with attention
  - **Transformer** with multi-head self-attention

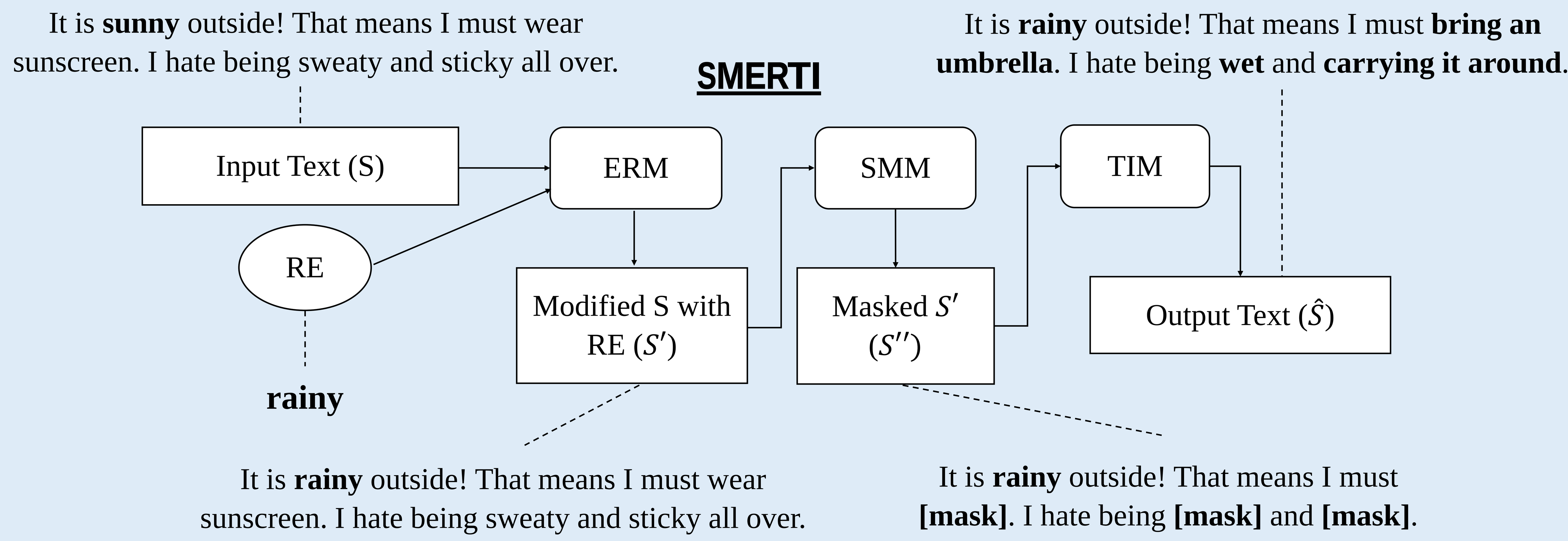

Figure 2: Illustration of the SMERTI pipeline architecture with an example

6. Experiment Details

Datasets:
- Amazon reviews
- Yelp reviews
- Kaggle news headlines

Chosen Evaluation *RE*s:
- 10 nouns per dataset
- 10 verbs per dataset
- 10 adjectives per dataset
- 5 phrases per dataset

Baselines:
- Noun WordNet Model (NWN-STEM) [2]
- General WordNet Model (GWN-STEM)
- Word2Vec Model (W2V-STEM)

7. Human Evaluation

- Eight participants from the University of Waterloo
- 54 total pieces of text rated on following criteria [1-5]:
  - **RE match**: How related is the text to the *RE*?
  - **Fluency**: Does the text make sense and flow well?
  - **Sentiment**: How do you think the author of the text was feeling? (1 – very negative, 5 – very positive)

8. Automatic Evaluation

- For each evaluation *RE*, select one-hundred lines from the test set that does not already contain the *RE*
- Output text evaluated with metrics below:
  - **Fluency (SLOR)** [3]: syntactic log-odds ratio for sentence level fluency, rescaled to [0,1]:

$$SLOR_c(S) = \frac{1}{|S|}\left(\ln(p_M(S)) - \frac{\ln\left(\prod_{w \in S} p_M(w)\right)}{\sum_{w \in S}|w|}\right)$$

  - **Sentiment Preservation Accuracy (SPA)** [0-1]: % of outputs carrying the same sentiment (negative, neutral, or positive) as input
  - **Content Similarity Score (CSS)** [0-1]: semantic similarity between generated text and *RE*; higher values indicate stronger semantic exchange
  - **Semantic Text Exchange Score (STES)** [0,1]: harmonic mean of SLOR, SPA, and CSS; higher scores represent higher overall STE performance:

$$STES = \frac{3 * SLOR * SPA * CSS}{SLOR * SPA + SLOR * CSS + SPA * CSS}$$

9. Example Outputs

Input Text: *great food , large portions ! my family and i really enjoyed our saturday morning breakfast .*
Replacement Entity (RE): *pizza*

Model	MRT/RRT	Generated Output
SMERTI-Transformer	20%	great pizza , large slices ! my family and i really enjoyed our saturday morning lunch .
	80%	great pizza , chewy crust ! nice ambiance and i really enjoyed it.
SMERTI-RNN	20%	great pizza , large delivery ! my family and i really enjoyed our saturday morning place .
	80%	great pizza , amazing pizza ! reasonable and i really enjoyed everyone .
W2V-STEM	20%	great pizza , large portions ! my family and i really enjoyed our saturday morning breakfast .
	80%	awesome pizza, slices slices ! my mom dough we crust liked our sunday morning bagel .
GWN-STEM / NWN-STEM	20%	great food , large stuff ! my family and i really enjoyed our saturday i breakfast

Table 1: Generated output text by model for various masking rates on a Yelp evaluation example

10. Evaluation Results

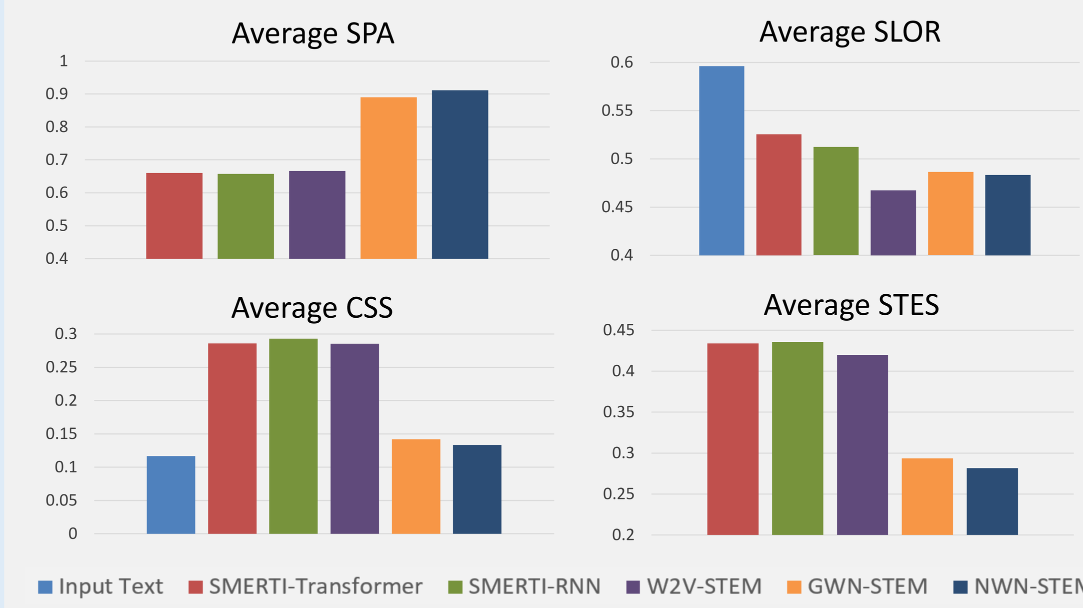

Figure 3: Graphs of automatic evaluation results

Model	RE Match [1-5]	Fluency [1-5]	Sentiment Preservation [0-1]	Harmonic Mean [0-1]
Input Text	1.82	4.13	---	---
SMERTI-Transformer	**3.58**	2.88	0.75	**0.60**
SMERTI-RNN	3.50	2.82	0.58	0.54
W2V-STEM	3.48	2.08	0.67	0.44
GWN-STEM	2.25	2.50	0.83	0.42
NWN-STEM	2.13	**2.96**	**1.00**	0.45

Table 2: Human evaluation results

11. Analysis and Discussion

- **SMERTI** performs best overall (highest STES)
- **SMERTI** performs best on SLOR and CSS
- **WordNet** models perform the worst overall
- **W2V-STEM** achieves the lowest text fluency
- Human and automatic results correlate well
- As **MRT/RRT** increases, SMERTI's SPA and SLOR decrease while CSS increases

12. Conclusion

- **SMERTI** performs strongly on semantic text exchange, outperforming baseline models
- **Trade-off** between semantic exchange against fluency and sentiment preservation, controlled by the masking (replacement) rate threshold
- **Future work**: preservation of personality

13. Acknowledgments

We thank our anonymous reviewers, study participants, and Huawei Technologies Co., Ltd. for financial support.

14. References

[1] Cer et al. 2018. Universal sentence encoder for English. In proceedings of EMNLP 2018: System Demonstrations, pages 169-174.
[2] Yao et al. 2017. Automated crowdturfing attacks and defenses in online review systems. In Proceedings of 2017 ACM SIGSAC CCS, pages 1143–1158.
[3] Kann et al. 2018. Sentence-level fluency evaluation: References help, but can be spared! In Proceedings of the 22nd CoNLL, pages 313–323.

Email: sy2feng@uwaterloo.ca