Instructions for ACL 2019 Proceedings

Anonymous ACL submission

Abstract

Automatic summarization research has made substantial progress thanks to novel methods and datasets.

Manual evaluation approaches so far either ignore content and focus on fluency, or require expert annotators but nevertheless suffer from low inter-annotator agreement due to the complexity of the task. In the few cases where the contents of the summary are evaluated, the evaluation is biased due to using a single reference summary, which results in different summaries of equal quality being rated according to their similarity to the reference.

In this paper, we propose a Highlight-bAsed Evaluation of Single document Summarization (HArnESS). Our proposal assesses summaries against the original document, facilitated through manually highlighted salient content which can be reused in future studies.

Furthermore it does not require expert annotators, avoids reference bias and provides absolute instead of ranked evaluation of systems.

1 Introduction

Automatic summarization research has made headway over the years with single document summarization as the front-runner thanks to novel methods (See et al., 2017; Kryściński et al., 2018; Narayan et al., 2018) and large datasets (Sandhaus, 2008; Hermann et al., 2015; Narayan et al., 2018) and .

However, these approaches lack a often focus on evaluating contents of summary often require expert annotator e.g. Pyramid (Nenkova and Passonneau, 2004) but nevertheless suffer from low inter-annotator agreement. (Celikyilmaz et al., 2018)

Acknowledgments

The acknowledgments should go immediately before the references. Do not number the acknowledgments section. Do not include this section when submitting your paper for review.

Preparing References:

Include your own bib file like this:
`\bibliographystyle{acl_natbib}`
`\bibliography{acl2019}`

where `acl2019` corresponds to a acl2019.bib file.

References

Asli Celikyilmaz, Antoine Bosselut, Xiaodong He, and Yejin Choi. 2018. Deep Communicating Agents for Abstractive Summarization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*.

Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching Machines to Read and Comprehend. In *Neural Information Processing Systems*, pages 1–14.

Wojciech Kryściński, Romain Paulus, Caiming Xiong, and Richard Socher. 2018. Improving Abstraction in Text Summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1808–1817.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't Give Me the Details, Just the Summary! Topic-aware Convolutional Neural Networks for Extreme Summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

A Nenkova and R Passonneau. 2004. Evaluating content selection in summarization: The pyramid method. *Proceedings of HLT-NAACL*, 2004:145–152.

Evan Sandhaus. 2008. The New York Times Annotated Corpus.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks. In *Proceedings of the 55th Annual Meeting of the ACL.*

A Appendices

Appendices are material that can be read, and include lemmas, formulas, proofs, and tables that are not critical to the reading and understanding of the paper. Appendices should be **uploaded as supplementary material** when submitting the paper for review. Upon acceptance, the appendices come after the references, as shown here. Use `\appendix` before any appendix section to switch the section numbering over to letters.

B Supplemental Material

Submissions may include non-readable supplementary material used in the work and described in the paper. Any accompanying software and/or data should include licenses and documentation of research review as appropriate. Supplementary material may report preprocessing decisions, model parameters, and other details necessary for the replication of the experiments reported in the paper. Seemingly small preprocessing decisions can sometimes make a large difference in performance, so it is crucial to record such decisions to precisely characterize state-of-the-art methods.

Nonetheless, supplementary material should be supplementary (rather than central) to the paper. **Submissions that misuse the supplementary material may be rejected without review.** Supplementary material may include explanations or details of proofs or derivations that do not fit into the paper, lists of features or feature templates, sample inputs and outputs for a system, pseudo-code or source code, and data. (Source code and data should be separate uploads, rather than part of the paper).

The paper should not rely on the supplementary material: while the paper may refer to and cite the supplementary material and the supplementary material will be available to the reviewers, they will not be asked to review the supplementary material.