`enunlg`:
a Python library for reproducible neural data-to-text experimentation

David M. Howcroft and **Dimitra Gkatzia**
Edinburgh Napier University
{D.Howcroft,D.Gkatzia}@napier.ac.uk

Abstract

Over the past decade, a variety of neural architectures for data-to-text generation (NLG) have been proposed. However, each system typically has its own approach to pre- and post-processing and other implementation details. Diversity in implementations is desirable, but it also confounds attempts to compare model performance: are the differences due to the proposed architectures or are they a byproduct of the libraries used or a result of pre- and post-processing decisions made? To improve reproducibility, we re-implement several pre-Transformer neural models for data-to-text NLG within a single framework to facilitate direct comparisons of the models themselves and better understand the contributions of other design choices. We release our library at `https://github.com/NapierNLP/enunlg` to serve as a baseline for ongoing work in this area including research on NLG for low-resource languages where transformers might not be optimal.

1 Introduction

Dozens of different models for neural data-to-text generation have been proposed in the last decade, before we even consider recent efforts to repurpose large language models for data-to-text natural language generation (NLG). However, these models vary greatly with respect to both low-level and high-level design choices, requiring different kinds of *delexicalisation* and normalisation processes, different ways of encoding and tracking meaning, and using a variety of neural network libraries, among other differences. While we can use the outputs of individual models released by their authors to assess the relative performance of these implementations, there is little work aiming to explore which performance differences are due to the proposed architectures themselves as opposed to other implementation details. In order to explore these differences, encourage reproduction experiments, and provide tools for teaching data-to-text NLG, we developed a Python library implementing several of these models in a common framework.

Datasets	Models
WEN	SCLSTM (described)
E2E Challenge	SCLSTM (released)
Cleaned E2E	TGEN
WEBNLG	CHECKLIST *
NMETHODIUS	CHARSCLSTM *

Table 1: Datasets & models implemented in `enunlg`.
*Indicates a model whose implementation is in-progress.

2 `enunlg`: extensible NLG library

Our `enunlg` library is developed for Python 3.9 with PyTorch 1.9.1. In addition to implementing the models themselves, we provide a variety of file readers & writers to consume different corpora and convert them into appropriate representations for each model. At present, we have tools in place to work with the WEN datasets (dialog system responses for restaurant, hotel, laptop, and TV descriptions: Wen et al., 2016), cleaned data from the E2E Challenge (restaurant descriptions: Novikova et al., 2017; Dušek et al., 2019), the NMETHODIUS corpus (museum exhibits: Stevens-Guille et al., 2020), and WEBNLG (Gardent et al., 2017).

Meaning representation (MR) parsers are included for CUED dialogue acts, E2E slot-value pairs, and RDF triples. Supported neural representations for these MR types include bit-vectors, flattened trees, and unbracketed sequences of triples. Word embeddings can be randomly initialised or loaded from existing vectors.

We reimplement the SCLSTM model proposed by (Wen et al., 2015), originally implemented using Theano and Python 2. During reimplementation, we found that the codebase released with the paper implemented a different architecture from what was described in the paper, so we provide both versions

Proceedings of the 16th International Natural Language Generation Conference: System Demonstrations, pages 4–5
September 11–15, 2023.

in our library. We also provide a reimplementation of TGEN (Dušek and Jurčíček, 2016), originally implemented using Tensorflow 0.6. Kiddon et al. (2016) implemented their CHECKLIST model in Lua with Torch and Deriu and Cieliebak (2018) used Tensorflow 1.10.0 for their CHARSCLSTM.

3 Planned uses

Our goals in developing `enunlg` fall into three broad categories: reproducibility, pedagogy, and easy experimentation. By enabling the use of a single framework with consistent reference implementations of multiple models, the library promotes reproducibility and facilitates fair comparisons, controlling for differences in, e.g., delexicalisation, tokenisation, neural network libraries, etc. A small, consistent codebase that addresses the different elements of implementing neural data-to-text systems also serves a pedagogical function, providing a starting point for student projects. Finally, our design choices aim to make engineering experiments trivial (e.g. hyperparameter search, changing tokenisation, etc) and scientific experiments easy (e.g. developing new end-to-end and pipeline systems for neural NLG) and promote work in low-resource NLG (Howcroft and Gkatzia, 2022).

4 Conclusions

We present `enunlg`, a library for reproducible experimentation in neural data-to-text generation. The code is available from `https://github.com/NapierNLP/enunlg`. We hope that the availability of an extensible library for neural NLG will improve reproducibility in our research community and provide a new set of reference implementations for baseline models.

Acknowledgements

The research was supported under the EPSRC project 'NLG for low-resource domains' (EP/T024917/1).

References

Jan Milan Deriu and Mark Cieliebak. 2018. End-to-end trainable system for enhancing diversity in natural language generation. In *Proc. of the E2E NLG Challenge System Descriptions*.

Ondřej Dušek, David M. Howcroft, and Verena Rieser. 2019. Semantic noise matters for neural natural language generation. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 421–426, Tokyo, Japan. Association for Computational Linguistics.

Ondřej Dušek and Filip Jurčíček. 2016. Sequence-to-sequence generation for spoken dialogue via deep syntax trees and strings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 45–51, Berlin, Germany. Association for Computational Linguistics.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. Creating training corpora for NLG micro-planners. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 179–188, Vancouver, Canada. Association for Computational Linguistics.

David M. Howcroft and Dimitra Gkatzia. 2022. Most NLG is low-resource: here's what we can do about it. In *Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, pages 336–350, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Chloé Kiddon, Luke Zettlemoyer, and Yejin Choi. 2016. Globally coherent text generation with neural checklist models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 329–339, Austin, Texas. Association for Computational Linguistics.

Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. 2017. The E2E dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 201–206, Saarbrücken, Germany. Association for Computational Linguistics.

Symon Stevens-Guille, Aleksandre Maskharashvili, Amy Isard, Xintong Li, and Michael White. 2020. Neural NLG for methodius: From RST meaning representations to texts. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 306–315, Dublin, Ireland. Association for Computational Linguistics.

Tsung-Hsien Wen, Milica Gašić, Nikola Mrkšić, Lina M. Rojas-Barahona, Pei-Hao Su, David Vandyke, and Steve Young. 2016. Multi-domain neural network language generation for spoken dialogue systems. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 120–129, San Diego, California. Association for Computational Linguistics.

Tsung-Hsien Wen, Milica Gašić, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young. 2015. Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1711–1721, Lisbon, Portugal. Association for Computational Linguistics.