Deep Learning Approaches to Text Production

Claire Gardent and Shashi Narayan

May 2, 2018

1 Content and Relevance

Text production is a key component of many NLP applications.

In data-driven approaches, it is used for instance, to generate dialogue turns from dialogue moves [Wen et al., 2015, Wen et al., 2016, Novikova et al., 2017], to verbalise the content of Knowledge bases [Gardent et al., 2017a, Gardent et al., 2017b] or to generate natural English sentences from rich linguistic representations, such as dependency trees [Belz et al., 2011, Mille et al., 2018] or Abstract Meaning Representations [May and Priyadarshi, 2017, Konstas et al., 2017, Song et al.,].

In text-driven methods on the other hand, text production is at work in sentence compression [Knight and Marcu, 2000, Cohn and Lapata, 2008, Filippova and Strube, 2008, Pitler, 2010, Filippova et al., 2015, Toutanova et al., 2016]; sentence fusion [McKeown et al., 2010, Filippova, 2010, Thadani and McKeown, 2013]; paraphrasing [Dras, 1999, Barzilay and McKeown, 2001, Bannard and Callison-Burch, 2005, Wubben et al., 2010, Narayan et al., 2016, Dong et al., 2017, Mallinson et al., 2017]; sentence (or text) simplification [Siddharthan et al., 2004, Zhu et al., 2010, Woodsend and Lapata, 2011, Wubben et al., 2012, Narayan and Gardent, 2014, Xu et al., 2015, Narayan and Gardent, 2016, Zhang and Lapata, 2017, Narayan et al., 2017], text summarisation [Wan, 2010, Nenkova and McKeown, 2011, Woodsend and Lapata, 2010, Rush et al., 2015, Cheng and Lapata, 2016, Nallapati et al., 2016, Chen et al., 2016, Tan and Wan, 2017, See et al., 2017, Nallapati et al., 2017, Paulus et al., 2017, Yasunaga et al., 2017, Narayan et al., 2018a, Narayan et al., 2018b, Pasunuru and Bansal, 2018, Celikyilmaz et al., 2018] and end-to-end dialogue systems [Li et al., 2017].

Following the success of encoder-decoder models in modeling sequence-rewriting tasks such as machine translation [Sutskever et al., 2011, Bahdanau et al., 2014], deep learning models have successfully been applied to the various text production tasks. For instance, [Rush et al., 2015] utilize a local attention-based model for abstractive summarisation, [Shang et al., 2015] propose an encoder-decoder model for response generation, [See et al., 2017] uses a hybrid of encoder-decoder model and pointer network [Vinyals et al., 2015] for story highlight generation, [Dong et al., 2017] exploits an encoder-decoder model for question rephrasing and [Konstas et al., 2017] for AMR generation.

In this tutorial, we will cover the fundamentals and the state-of-the-art research on neural models for text production. Each text production task raises a slightly different communication goal (e.g, how to take the dialogue context into account when producing a dialogue turn; how to detect and merge relevant information when summarising a text; or how to produce a well-formed text that correctly capture the information contained in some input data in the case of data-to-text generation). We will outline the constraints specific to each subtasks and examine how the existing neural models account for them.

2 Tutorial Outline

The tutorial will review deep learning approaches to text production. It will consider both text-to-text and data-to-text transformations. It aims to provide the audience with a good knowledge of text production systems, and a roadmap to get them started with the related work.

Proceedings of NAACL-HLT 2018: Tutorial Abstracts, pages 4–9

1. Introduction
   - Relevance of text production
   - Why deep learning for text production
2. Background
   - Deep learning basics
   - Generating Text using RNN LMs
3. Encoding Input Structure
   - Sequential encoders (Sentence compression, simplification, paraphrase generation and dialogue generation)
   - Hierarchical encoders (Document summarization)
   - Graph Encoders (Abstract Meaning Representations to text, structured data such as OWL, RDF and DB, to text)
4. Decoding and Semantic Adequacy
   - Attention and copy mechanism (accuracy)
   - Coverage mechanism (covering all the input)
5. Advanced Topics
   - Deep Reinforcement learning for text production
   - Convolutional Seq2Seq and Tranformer Models
6. Systems, Shared Tasks and Open Challenges

3 Presenters

Claire Gardent
Senior Research scientist
CNRS, LORIA, Nancy
`claire.gardent@loria.fr`
`https://members.loria.fr/CGardent/`

Shashi Narayan
Research Associate
School of Informatics, University of Edinburgh
`shashi.narayan@ed.ac.uk`
`http://homepages.inf.ed.ac.uk/snaraya2/`

Claire Gardent is a research scientist at CNRS (the French National Center for Scientific Research). Prior to joining the CNRS, she worked at the Université de Clermont-Ferrand, Saarbrücken Universität and Amsterdam Universiteit. She received her Ph.D. degree from the University of Edinburgh. Her research interests include (executable) semantic parsing, natural language generation and simplification and, more recently, the use of computational linguistics for linguistic analysis. She was nominated Chair of the EACL and acted as program chair for various international conferences, workshops and summer schools (EACL, ENLG, SemDIAL, SIGDIAL, ESSLLI, *SEM). She currently heads the WebNLG project (Nancy, Bolzano, Stanford SRI) and is the chair of SIGGEN, the ACL Special Interest Group in Natural Language Generation. Recently she co-organised the WebNLG Shared Task, a challenge on generating text from RDF data.

Shashi Narayan is a research associate at the School of Informatics at the University of Edinburgh. His research focuses on natural language generation, understanding and structured predictions. A major aim of his research is to build on the hypothesis that tailoring a model with knowledge of the task structure and linguistic requirements, such as syntax and semantics, leads to a better performance. The questions raised in his research are relevant to various natural language applications such as question answering, paraphrase generation, semantic and syntactic parsing, document understanding and summarization, and text simplification. He mostly rely on machine learning techniques such as deep learning and spectral methods to develop NLP frameworks. His research has appeared in computational linguistics journals (e.g., TACL, Computational Linguistics and Pattern Recognition Letters) and in conferences (e.g., ACL, EMNLP, NAACL, COLING, EACL and INLG). He was nominated on the SIGGEN board (2012-14) as a student member. He co-organised the WebNLG Shared Task, a challenge on generating text from RDF data. Recently, he was nominated as an area co-chair for Generation at NAACL HLT 2018.

4 Audience, Previous Tutorials and Venue

Based on the recent upsurge of interest in NL generation as witnessed by the increase in submissions in that domain at the major NLP conferences, we target an audience of 60 to 100 students and researchers from both academia and industry. We are not aware of any recent tutorial on the topic of text production. Our preference for the venue is NAACL and EMNLP.

References

[Bahdanau et al., 2014] Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473.

[Bannard and Callison-Burch, 2005] Bannard, C. and Callison-Burch, C. (2005). Paraphrasing with bilingual parallel corpora. In *Proceedings of ACL*.

[Barzilay and McKeown, 2001] Barzilay, R. and McKeown, K. R. (2001). Extracting paraphrases from a parallel corpus. In *Proceedings of ACL*.

[Belz et al., 2011] Belz, A., White, M., Espinosa, D., Kow, E., Hogan, D., and Stent, A. (2011). The first surface realisation shared task: Overview and evaluation results. In *Proceedings of ENLG*.

[Celikyilmaz et al., 2018] Celikyilmaz, A., Bosselut, A., He, X., and Choi, Y. (2018). Deep communicating agents for abstractive summarization. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, US.

[Chen et al., 2016] Chen, Q., Zhu, X., Ling, Z., Wei, S., and Jiang, H. (2016). Distraction-based neural networks for modeling documents. In *Proceedings of IJCAI*, pages 2754–2760, New York, USA.

[Cheng and Lapata, 2016] Cheng, J. and Lapata, M. (2016). Neural summarization by extracting sentences and words. In *Proceedings of ACL*, pages 484–494, Berlin, Germany.

[Cohn and Lapata, 2008] Cohn, T. and Lapata, M. (2008). Sentence compression beyond word deletion. In *Proceedings of COLING*.

[Dong et al., 2017] Dong, L., Mallinson, J., Reddy, S., and Lapata, M. (2017). Learning to paraphrase for question answering. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 875–886, Copenhagen, Denmark.

[Dras, 1999] Dras, M. (1999). *Tree adjoining grammar and the reluctant paraphrasing of text*. PhD thesis, Macquarie University, Australia.

[Filippova, 2010] Filippova, K. (2010). Multi-sentence compression: Finding shortest paths in word graphs. In *Proceedings of COLING*.

[Filippova et al., 2015] Filippova, K., Alfonseca, E., Colmenares, C., Kaiser, L., and Vinyals, O. (2015). Sentence compression by deletion with LSTMs. In *Proceedings of EMNLP*.

[Filippova and Strube, 2008] Filippova, K. and Strube, M. (2008). Dependency tree based sentence compression. In *Proceedings of INLG*.

[Gardent et al., 2017a] Gardent, C., Shimorina, A., Narayan, S., and Perez-Beltrachini, L. (2017a). Creating training corpora for nlg micro-planning. In *Proceedings of ACL*.

[Gardent et al., 2017b] Gardent, C., Shimorina, A., Narayan, S., and Perez-Beltrachini, L. (2017b). The webnlg challenge: Generating text from rdf data. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 124–133.

[Knight and Marcu, 2000] Knight, K. and Marcu, D. (2000). Statistics-based summarization-step one: Sentence compression. In *Proceedings of AAAI-IAAI*.

[Konstas et al., 2017] Konstas, I., Iyer, S., Yatskar, M., Choi, Y., and Zettlemoyer, L. (2017). Neural amr: Sequence-to-sequence models for parsing and generation. In *Proceedings of ACL*, pages 146–157, Vancouver, Canada.

[Li et al., 2017] Li, J., Monroe, W., Shi, T., Jean, S., Ritter, A., and Jurafsky, D. (2017). Adversarial learning for neural dialogue generation. In *Proceedings of EMNLP*, pages 2157–2169.

[Mallinson et al., 2017] Mallinson, J., Sennrich, R., and Lapata, M. (2017). Paraphrasing revisited with neural machine translation. In *Proceedings of EACL*.

[May and Priyadarshi, 2017] May, J. and Priyadarshi, J. (2017). Semeval-2017 task 9: Abstract meaning representation parsing and generation. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 536–545, Vancouver, Canada.

[McKeown et al., 2010] McKeown, K., Rosenthal, S., Thadani, K., and Moore, C. (2010). Time-efficient creation of an accurate sentence fusion corpus. In *Proceedings of NAACL-HLT*.

[Mille et al., 2018] Mille, S., Bohnet, B., Wanner, L., Belz, A., and Pitler, E. (2018). Surface realization shared task 2018: The generation challenges. In *To appear in Multilingual Surface Realization Workshop at ACL 2018*.

[Nallapati et al., 2017] Nallapati, R., Zhai, F., and Zhou, B. (2017). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Proceedings of AAAI*, pages 3075–3081, San Francisco, California USA.

[Nallapati et al., 2016] Nallapati, R., Zhou, B., dos Santos, C. N., Gülçehre, Ç., and Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence RNNs and beyond. In *Proceedings of CoNLL*, pages 280–290, Berlin, Germany.

[Narayan et al., 2018a] Narayan, S., Cardenas, R., Papasarantopoulos, N., Cohen, S. B., Lapata, M., Yu, J., and Chang, Y. (2018a). Document modeling with external attention for sentence extraction. In *Proceedings of the 56st Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia.

[Narayan et al., 2018b] Narayan, S., Cohen, S. B., and Lapata, M. (2018b). Ranking sentences for extractive summarization with reinforcement learning. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, US.

[Narayan and Gardent, 2014] Narayan, S. and Gardent, C. (2014). Hybrid simplification using deep semantics and machine translation. In *Proceedings of ACL*.

[Narayan and Gardent, 2016] Narayan, S. and Gardent, C. (2016). Unsupervised sentence simplification using deep semantics. In *Proceedings of INLG*.

[Narayan et al., 2017] Narayan, S., Gardent, C., Cohen, S. B., and Shimorina, A. (2017). Split and rephrase. In *Proceedings of EMNLP*, pages 606–616, Copenhagen, Denmark.

[Narayan et al., 2016] Narayan, S., Reddy, S., and Cohen, S. B. (2016). Paraphrase generation from latent-variable pcfgs for semantic parsing. In *Proceedings of the 9th International Natural Language Generation conference*, pages 153–162, Edinburgh, UK.

[Nenkova and McKeown, 2011] Nenkova, A. and McKeown, K. (2011). Automatic summarization. *Foundations and Trends in Information Retrieval*, 5(2–3):103–233.

[Novikova et al., 2017] Novikova, J., Dušek, O., and Rieser, V. (2017). The E2E dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Saarbrücken, Germany. arXiv:1706.09254.

[Pasunuru and Bansal, 2018] Pasunuru, R. and Bansal, M. (2018). Multi-reward reinforced summarization with saliency and entailment. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, US.

[Paulus et al., 2017] Paulus, R., Xiong, C., and Socher, R. (2017). A deep reinforced model for abstractive summarization. *CoRR*, abs/1705.04304.

[Pitler, 2010] Pitler, E. (2010). Methods for sentence compression. Technical report, University of Pennsylvania.

[Rush et al., 2015] Rush, A. M., Chopra, S., and Weston, J. (2015). A neural attention model for abstractive sentence summarization. In *Proceedings of EMNLP*, pages 379–389, Lisbon, Portugal.

[See et al., 2017] See, A., Liu, P. J., and Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. In *Proceedings of ACL*, pages 1073–1083, Vancouver, Canada.

[Shang et al., 2015] Shang, L., Lu, Z., and Li, H. (2015). Neural responding machine for short-text conversation. *CoRR*, abs/1503.02364.

[Siddharthan et al., 2004] Siddharthan, A., Nenkova, A., and McKeown, K. (2004). Syntactic simplification for improving content selection in multi-document summarization. In *Proceedings of COLING*.

[Song et al.,] Song, L., Zhang, Y., Wang, Z., and Gildea, D.

[Sutskever et al., 2011] Sutskever, I., Martens, J., and Hinton, G. E. (2011). Generating text with recurrent neural networks. In *Proceedings of ICML*.

[Tan and Wan, 2017] Tan, J. and Wan, X. (2017). Abstractive document summarization with a graph-based attentional neural model. In *Proceedings of ACL*, pages 1171–1181, Vancouver, Canada.

[Thadani and McKeown, 2013] Thadani, K. and McKeown, K. (2013). Supervised sentence fusion with single-stage inference. In *Proceedings of IJCNLP*.

[Toutanova et al., 2016] Toutanova, K., Brockett, C., Tran, K. M., and Amershi, S. (2016). A dataset and evaluation metrics for abstractive compression of sentences and short paragraphs. In *Proceedings of EMNLP*.

[Vinyals et al., 2015] Vinyals, O., Fortunato, M., and Jaitly, N. (2015). Pointer networks. In *Advances in Neural Information Processing Systems 28*, pages 2692–2700.

[Wan, 2010] Wan, X. (2010). Towards a unified approach to simultaneous single-document and multi-document summarizations. In *Proceedings of COLING*, pages 1137–1145, Beijing, China.

[Wen et al., 2016] Wen, T.-H., Gašić, M., Mrkšić, N., Rojas-Barahona, L. M., Su, P.-H., Vandyke, D., and Young, S. (2016). Multi-domain neural network language generation for spoken dialogue systems. In *Proceedings of NAACL-HLT*.

[Wen et al., 2015] Wen, T.-H., Gašić, M., Mrkšić, N., Su, P.-H., Vandyke, D., and Young, S. (2015). Semantically conditioned LSTM-based natural language generation for spoken dialogue systems. In *Proceedings of EMNLP*.

[Woodsend and Lapata, 2010] Woodsend, K. and Lapata, M. (2010). Automatic generation of story highlights. In *Proceedings of ACL*, pages 565–574, Uppsala, Sweden.

[Woodsend and Lapata, 2011] Woodsend, K. and Lapata, M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of EMNLP*.

[Wubben et al., 2010] Wubben, S., Van Den Bosch, A., and Krahmer, E. (2010). Paraphrase generation as monolingual translation: Data and evaluation. In *Proceedings of INLG*.

[Wubben et al., 2012] Wubben, S., van den Bosch, A., and Krahmer, E. (2012). Sentence simplification by monolingual machine translation. In *Proceedings of ACL*.

[Xu et al., 2015] Xu, W., Callison-Burch, C., and Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

[Yasunaga et al., 2017] Yasunaga, M., Zhang, R., Meelu, K., Pareek, A., Srinivasan, K., and Radev, D. (2017). Graph-based neural multi-document summarization. In *Proceedings of CoNLL*, pages 452–462, Vancouver, Canada.

[Zhang and Lapata, 2017] Zhang, X. and Lapata, M. (2017). Sentence simplification with deep reinforcement learning. In *Proceedings of EMNLP*.

[Zhu et al., 2010] Zhu, Z., Bernhard, D., and Gurevych, I. (2010). A monolingual tree-based translation model for sentence simplification. In *Proceedings of COLING*.