
INSIGHT

Template-Based Multi-Solution Approach for Data-to-Text Generation

Abelardo Vieira Mota
Ticiana L. Coelho da Silva
José Antônio F. de Macêdo

Departamento de Computação
Universidade Federal do Ceará

AGENDA

1. Introduction
2. Related Work
3. Solution
4. Experiments and Results
5. Conclusion

1. Introduction

Introduction

- Natural Language Generation
- Why text?
- Why automatize?
- Why triples?

Problem Definition

$t = \langle subject, predicate, object \rangle$

Input data: $\{t_1, t_2, \ldots, t_n\}$

Output: $[w_1, w_2, \ldots, w_l]$

- adequate
- grammatically correct
- fluent

c

2. Related Work

Modular + Template

1. **LOD-DEF (DUMA; KLEIN, 2013)**:
    a. Templates of texts;
    b. RDF graphs with max depth 1;
2. **Lemon (CIMIANO et al., 2013)**:
    a. Templates of parse trees;
    b. Domain-specific rules;
3. **Template Translation (GATTI et al., 2018)**:
    a. Translate templates from Dutch to English;

Integrated + Neural

1. **Word vs Character (JAGFELD et al., 2018)**:
    a. Base symbol: word x character
2. **Inverse KL Divergence (ZHU et al., 2019)**:
    a. Different training objective
3. **Deep Graph Encoders (MARCHEGGIANI; PEREZ-BELTRACHINI, 2018) e GTR-LSTM (DISTIAWAN et al., 2018)**:
    a. Graph encoders
4. **Low-Resourced Text Generation (ZANG; WAN, 2019)**:
    a. Reinforcement Learning

Modular + Neural

1. **Neural pipeline vs end-to-end (FERREIRA et al., 2019):**
    a. Neural Networks with modular x end-to-end architectures
2. **Step-by-Step (MORYOSSEF et al., 2019):**
    a. Modular architecture (statistical + neural)

c

3. Solution

Template

Structure

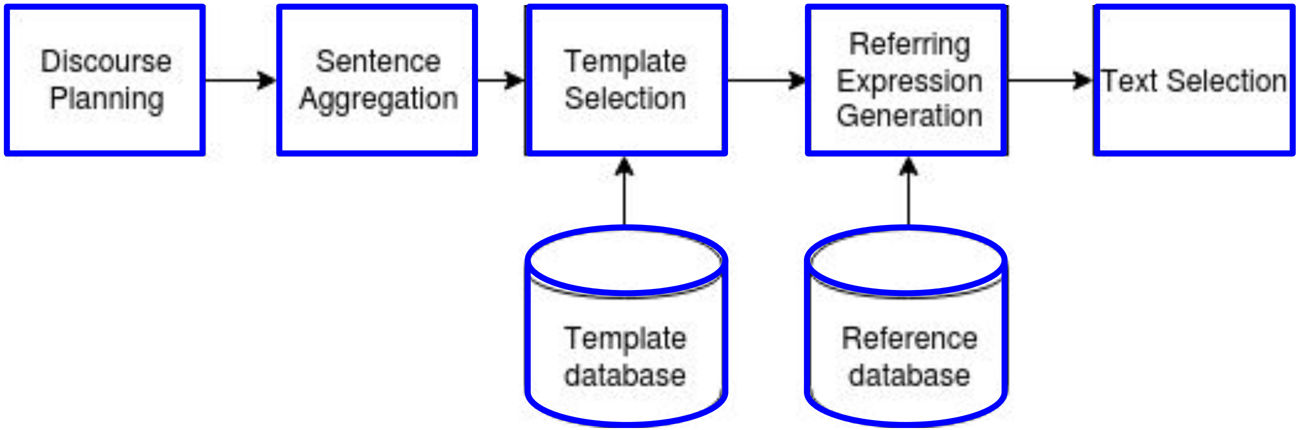
$$\langle s = [\langle \mathbf{slot\_1},\ \mathit{birthDate},\ \mathbf{slot\_2}\rangle,\ \langle \mathbf{slot\_1},\ \mathit{birthPlace},\ \mathbf{slot\_3}\rangle,\ \langle \mathbf{slot\_3},\ \mathit{capital},\ \mathbf{slot\_4}\rangle],$$

$$[\langle \mathit{Barack\_Obama},\ \mathit{birthDate},\ 1961\text{-}8\text{-}4\rangle,\ \langle \mathit{Barack\_Obama},\ \mathit{birthPlace},\ \mathit{Hawaii}\rangle,\ \langle \mathit{Hawaii},\ \mathit{capital},\ \mathit{Honolulu}\rangle]$$

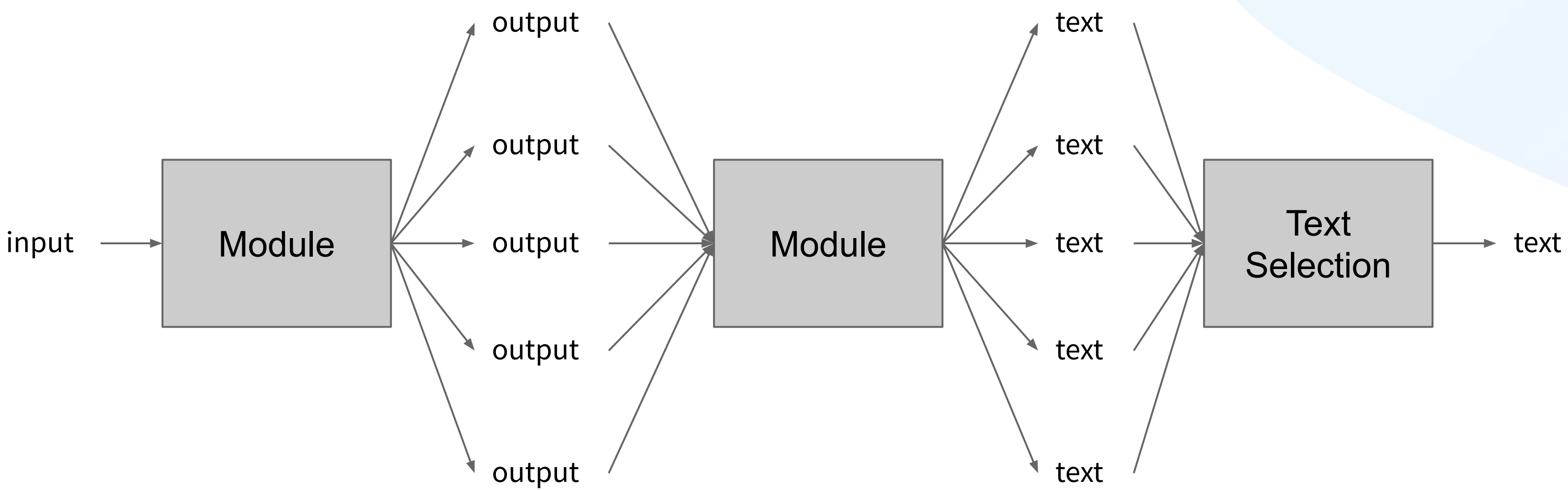
$$Z = \text{“}\mathbf{slot\_1}\ \mathit{was\ born\ in}\ \mathbf{slot\_4},\ \mathbf{slot\_3},\ \mathit{on}\ \mathbf{slot\_2}.\text{”}\rangle$$

Text

Modular Architecture



Multi-Solution



N-gram models

- N-gram models are used as heuristics in each module
- They estimate a probability of a sequence using the **chain rule** and the **Markov assumption**

$$P(w_1 w_2 \ldots w_n) \approx \prod_i P(w_i \mid w_{i-k} \ldots w_{i-1})$$

c

4. Experiments and Results

Setup

- **WebNLG 2017 shared task** {train, dev, test} (COLIN et al., 2016);
- Template and reference database extracted from the enhanced version (Ferreira et al. 2018);
- **Grid-Search** to determine the N-gram models order and the amount of output for each module
  - train -> dev (BLEU)
- **Final evaluation**
  - (train + dev) -> test
  - **methods**: BLEU, METEOR, TER and manual inspection

Automatic evaluation

Approach	**BLEU**	**METEOR**	**TER**
adaptcentre	60.59	0.45	0.38
deepnlg-e2ernn	58.36	0.42	0.40
template-pipe-multi	57.87	0.43	0.37
seq2seq-wc-word	55.82	0.41	0.40
gcn	55.35	0.39	0.40
template-pipe-single	55.27	0.42	0.40
BIU-Chimera-v1	53.20	0.44	0.47
upf-forge	40.88	0.41	0.56

Manual Inspection

- “the main ingredient in batagor are fried fish dumpling with tofu and vegetables ***inpeanut*** sauce .”
- “the language spoken in texas is english and one of the languages is english”
- “california gemstone benitoite .”

5. Conclusion

Conclusion

- We proposed and evaluated a template-based multi-solution approach
- The results obtained are competitive when compared to state-of-the-art methods

References

- **DUMA, D.; KLEIN, E.** Generating natural language from linked data: Unsupervised template extraction. In: Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers. [S.l.: s.n.], **2013**. p. 83–94.
- **CIMIANO, P.; LÜKER, J.; NAGEL, D.; UNGER, C.** Exploiting ontology lexica for generating natural language texts from rdf data. In: Proceedings of the 14th European Workshop on Natural Language Generation. [S.l.: s.n.], **2013**. p. 10–19.
- **GATTI, L.; LEE, C. van der; THEUNE, M.** Template-based multilingual football reports generation using wikidata as a knowledge base. In: Proceedings of the 11th International Conference on Natural Language Generation. [S.l.: s.n.], **2018**. p. 183–188.
- **JAGFELD, G.; JENNE, S.; VU, N. T.** Sequence-to-sequence models for data-to-text natural language generation: Word-vs. character-based processing and output diversity. arXiv preprint arXiv:1810.04864, **2018**.
- **ZHU, Y.; WAN, J.; ZHOU, Z.; CHEN, L.; QIU, L.; ZHANG, W.; JIANG, X.; YU, Y.** Triple-to-text: Converting rdf triples into high-quality natural languages via optimizing an inverse kl divergence. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. [S.l.: s.n.], **2019**. p. 455–464.

References

- **DISTIAWAN, B.; QI, J.; ZHANG, R.; WANG, W.** Gtr-lstm: A triple encoder for sentence generation from rdf data. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). [S.l.: s.n.], **2018**. p. 1627–1637.
- **MARCHEGGIANI, D.; PEREZ-BELTRACHINI, L.** Deep graph convolutional encoders for structured data to text generation. arXiv preprint arXiv:1810.09995, **2018**.
- **ZANG, H.; WAN, X.** A semi-supervised approach for low-resourced text generation. arXiv preprint arXiv:1906.00584, **2019**.
- **FERREIRA, T. C.; LEE, C. van der; MILTENBURG, E. van; KRAHMER, E.** Neural data-to-text generation: A comparison between pipeline and end-to-end architectures. arXiv preprint arXiv:1908.09022, **2019**.
- **MORYOSSEF, A.; GOLDBERG, Y.; DAGAN, I.** Step-by-step: Separating planning from realization in neural data-to-text generation. arXiv preprint arXiv:1904.03396, **2019**.
- **COLIN, E.; GARDENT, C.; MRABET, Y.; NARAYAN, S.; PEREZ-BELTRACHINI, L.** The webnlg challenge: Generating text from dbpedia data. In: Proceedings of the 9th International Natural Language Generation conference. [S.l.: s.n.], **2016**. p. 163–167.
- **FERREIRA, T. C.; MOUSSALLEM, D.; KRAHMER, E.; WUBBEN, S.** Enriching the webnlg corpus. In: Proceedings of the 11th International Conference on Natural Language Generation. [S.l.: s.n.], **2018**. p. 171–176.

Thank you!

abevieiramota@gmail.com