Exploring Text Recombination for Automatic Narrative Level Detection

LREC 2022, Marseille

Nils Reiter[♣], Judith Sieker[♢], Svenja Guhr[♠], Evelyn Gius[♠], Sina Zarrieß[♢]

Universities of Cologne[♣], Bielefeld[♢], Darmstadt[♠]

June 2022

INSTITUT FÜR DIGITAL HUMANITIES UNIVERSITÄT ZU KÖLN

CRETA CENTER FOR REFLECTED TEXT ANALYTICS

Narrative Level

- Embedded stories, told by characters of a story
- Widely used phenomenon in narrative texts (and other media)
- Crucial for content-driven narrative analysis
- Important for subsequent NLP tasks (e.g., coreference resolution)

Narrative Level

- Embedded stories, told by characters of a story
- Widely used phenomenon in narrative texts (and other media)
- Crucial for content-driven narrative analysis
- Important for subsequent NLP tasks (e.g., coreference resolution)

Example

[...] "Tell on," quoth the King who chanced to be sleepless and restless and therefore was pleased with the prospect of hearing her story. So Shahrazad rejoiced; and thus, on the first night of the Thousand Nights and a Night, she began with the
Tale of the Trader and the Jinni. 🔽
It is related, O auspicious King, that there was a merchant of the merchants who had much wealth [...]

Arabian Nights, archive.org

Annotating Narrative Levels

- No annotated corpora are available
- Shared task on guideline development Gius et al. (2021)
  - Task: Establish a guideline for annotating levels in English texts
  - Evaluation by looking at theory, applicability (IAA), usefulness
  - Extremely challenging annotation task, due its length

Annotating Narrative Levels

- No annotated corpora are available
- Shared task on guideline development (Gius et al. (2021))
  - Task: Establish a guideline for annotating levels in English texts
  - Evaluation by looking at theory, applicability (IAA), usefulness
  - Extremely challenging annotation task, due its length

Contents of this Talk

- Establish method to induce training data and
- Evaluate that it does help a BERT-NSP-model

Text Recombination

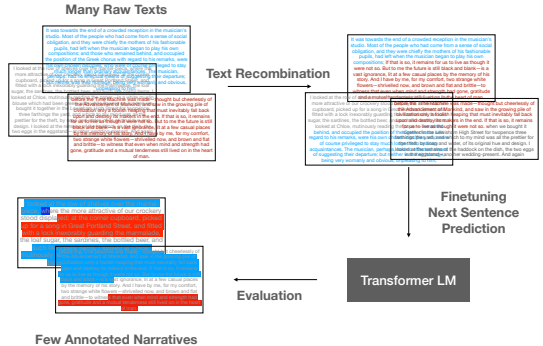
- 38 shortest English texts from ELTeC corpus Burnard/Odebrecht (2021)
  - 14 002 to 68 607 words long
- Split into training (70 %) and test (30 %)
- Concatenate n randomly selected texts, with $n \sim N(\mu = 3, \sigma = 1)$
  - …and tag the point of concatenation
- Remove duplicates

= Synthetic stories dataset

Text Recombination

- 38 shortest English texts from ELTeC corpus — Burnard/Odebrecht (2021)
  - 14 002 to 68 607 words long
- Split into training (70 %) and test (30 %)
- Concatenate n randomly selected texts, with $n \sim N(\mu = 3, \sigma = 1)$
  - …and tag the point of concatenation
- Remove duplicates

= Synthetic stories dataset



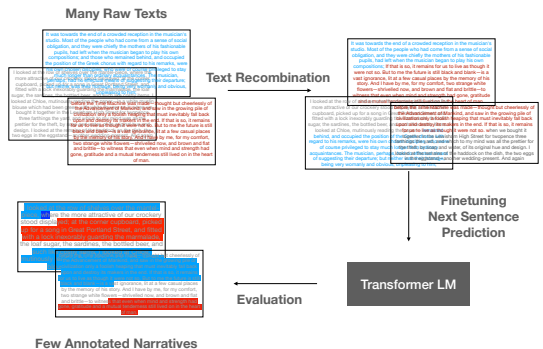
Text Recombination

- 38 shortest English texts from ELTeC corpus — Burnard/Odebrecht (2021)
  - 14 002 to 68 607 words long
- Split into training (70 %) and test (30 %)
- Concatenate n randomly selected texts, with $n \sim N(\mu = 3, \sigma = 1)$
  - …and tag the point of concatenation
- Remove duplicates

= Synthetic stories dataset

Experiments

1. Evaluation on synthetic stories
2. Evaluation on real level-annotated stories



Experiment 1: Evaluation on Synthetic Stories

- Original BERT model provided by HuggingFace — Devlin et al. (2019)
- Use next sentence prediction head for level boundary detection
- Evaluation with and without fine-tuning on synthetic data set
- Context window of 54 tokens in both directions
- Metrics: precision, recall, boundary similarity — Fournier (2013)
  - Averaged over test set (300 texts)
  - Boundary similarity: Transposition window of $n_t = 100$ characters

Experiment 1: Evaluation on Synthetic Stories

- Original BERT model provided by HuggingFace — Devlin et al. (2019)
- Use next sentence prediction head for level boundary detection
- Evaluation with and without fine-tuning on synthetic data set
- Context window of 54 tokens in both directions
- Metrics: precision, recall, boundary similarity — Fournier (2013)
  - Averaged over test set (300 texts)
  - Boundary similarity: Transposition window of $n_t = 100$ characters

Finetuning	Precision	Recall	Boundary sim.
No	2.61± 1.8	55.41±25.2	2.51 ± 1.8
Yes	32.39±36.09	25.68±27.12	19.20 ±24.25

Experiment 2: Evaluation on Real Level-Annotated Stories

- Re-use of guideline development shared task
- Evaluation on all annotations for all guidelines
  - I.e.: 2 annotators for each of 7 guidelines
- With and without fine-tuning

Experiment 2: Evaluation on Real Level-Annotated Stories

- Re-use of guideline development shared task
- Evaluation on all annotations for all guidelines
  - I.e.: 2 annotators for each of 7 guidelines
- With and without fine-tuning

Guideline	Without finetuning		With finetuning		Gain by finetuning	
	Precision	Recall	Precision	Recall	Precision	Recall
Ketschik et al. (2021)	12.27	11.08	33.33	10.76	21.06	−0.32
	7.69	7.13	10.26	2.98	2.56	−4.15
Barth (2021)	12.18	9.79	17.95	7.40	5.77	−2.39
	15.13	9.70	10.26	2.75	−4.87	−6.95

Table: Prediction results for narrative level boundaries (see paper for full table)

Conclusions and Outlook

- Annotating narrative levels classically does not scale
  - Mostly because it's a non-local phenomenon
- Even crudely generated training data helps
- Shared task on generating the best training data in preparation


Thank you!

Context Window

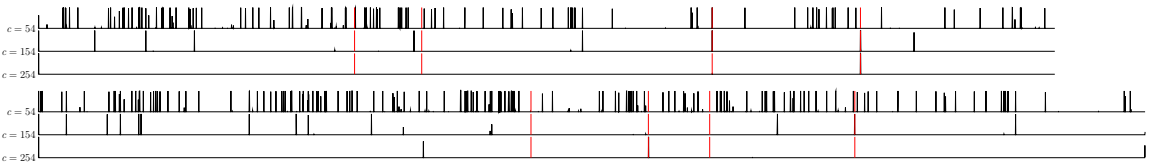


Figure: Predicted break probabilities in two randomly selected texts for different context windows. Red lines indicate true boundaries.

References I

- Barth, Florian (2021). "Annotation Guidelines for Narrative Levels and Narrative Acts v2". In: *Journal of Cultural Analytics* 6.4. DOI: 10.22148/001c.30701.
- Burnard, Lou/Carolin Odebrecht (2021). *English Novel Corpus (ELTeC-eng) April 2021 release (v1.0.1)*. DOI: 10.5281/zenodo.4662490.
- Devlin, Jacob/Ming-Wei Chang/Kenton Lee/Kristina Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *arXiv:1810.04805 [cs]*. arXiv: 1810.04805. URL: http://arxiv.org/abs/1810.04805 (visited on 11/21/2021).
- Fournier, Chris (2013). "Evaluating Text Segmentation using Boundary Edit Distance". In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 1702–1712. URL: http://aclweb.org/anthology/P13-1167.
- Gius, Evelyn/Marcus Willand/Nils Reiter (2021). "On Organizing a Shared Task for the Digital Humanities – Conclusions and Future Paths". In: *Cultural Analytics* 6.4.

References II

Ketschik, Nora/Benjamin Krautter/Sandra Murr/Yvonne Zimmermann (2021). "On the Theory of Narrative Levels and Their Annotation in the Digital Context". In: *Journal of Cultural Analytics* 6.4. DOI: 10.22148/001c.30700.