GECko Software project, work update 5

Eduardo Calò, Léo Jacqmin, Thibo Rosemplatt

NLP M2 IDMC, Université de Lorraine

December 18th, 2020





Progress

Where We Are At





Progress

Where We Are At



What did not Work

Implementation of a Czech model:

- Found a dataset annotated in M2 format [1]
- Trained GECToR model using this dataset

• Train accuracy: 0.8452

Validation accuracy: 0.8121

• **Test** F_{0.5}: 0.4871

Issues: not pre-trained with synthetic data, not enough training





What Worked - Delta Function

Highlight changes at token-level

Making use of GECToR's sequence tagging formulation e.g. ThereSEPL \parallel SEPR \parallel REPLACE_It

- 1. **predict** correction
- 2. **annotate** original text with tags given prediction
- looping through tokens, retrieve changes for input and output





What Worked - Environments

Separation of DEV and PROD environments:

- waitress serves our application online
- Basic **Flask** serving utility **offline**

Bringing in advantages in security, stability, speed...





What Worked - UX

Various **UX** improvements (thanks to cognitive sciences mates!)

- Type Enter to predict, Shift + Enter to line-break
- Paste without formatting
- Added a Tagline



What Worked - New Model

Integration of a **Sentence Reordering model** [2]

Input	Output
This chemical is widly used in the swimming	Chlorine is well known for its sanitizing
pools market. Chlorine is well known for its	properties. This chemical is widely used in
sanatizing properties.	the swimming pools market.

Figure: Example of expected sentence reordering





What Worked - New Model cont.

Addressed **discourse error correction**, but simplified to a less fine-grained task

How it works:

- 1. Input: list of sentences
- 2. Creates **pairs** of sentences
- 3. **Probabilities** assigned to each sentence of each pair
- 4. Graph computation of the best path
- 5. Output: reordered paragraph





Progress

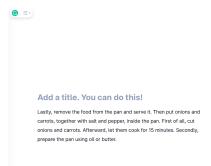
Where We Are At





Live Demo

Back to the origins...









Our Pipeline

- 1. User input text
- 2. spaCy sentencizer(input) $\rightarrow \#S$ from 1 to n
- 3. if S = 1:
 - GECToR(S)

else:

- ∀ S: do GECToR(S)
- store all grammatically corrected but possibly un-ordered S in L
- sentence reordering model on L
- 4. done



Progress

Where We Are At





Improvements

- Deploy updated web app
- Improve performances of sentence reordering
- Minor UX improvements





Optional Improvements

If we have time...

- Visualize swapped sentences
- Mobile layout





References I

- [1] J. Naplava and M. Straka. Grammatical error correction in low-resource scenarios, 2019.
- [2] S. Prabhumoye, R. Salakhutdinov, and A. W. Black. Topological sort for sentence ordering. arXiv preprint arXiv:2005.00432, 2020.





Thanks for listening! Questions?



