

GECko

Software project, work update 5

Eduardo Calò, Léo Jacqmin, Thibo Rosemplatt

NLP M2
IDMC, Université de Lorraine

December 18th, 2020

Table of Contents

Progress

Where We Are At

Action Plan

Table of Contents

Progress

Where We Are At

Action Plan

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||SEPR\$REPLACE_It

1. **predict** correction
2. **annotate** original text with tags given prediction
3. 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 widely used in the swimming pools market. Chlorine is well known for its sanitizing properties.	Chlorine is well known for its sanitizing properties. This chemical is widely used in 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



Table of Contents

Progress

Where We Are At


Action Plan

Back to the origins...



Add a title. You can do this!

Lastly, remove the food from the pan and serve it. Then put onions and carrots, together with salt and pepper, inside the pan. First of all, cut onions and carrots. Afterward, let them cook for 15 minutes. Secondly, prepare the pan using oil or butter.



Nothing can stop you now!

Hide Assistant 

Great job! 
See performance

Goals
Adjust goals

All alerts

Correctness 
Looking good

Clarity 
Very clear

Engagement 
Very engaging

Delivery 
Just right

Our Pipeline

1. User input text
2. spaCy sentencizer(input) \rightarrow #S from 1 to n
3. if $S = 1$:
 - GECToR(S)else:
 - $\forall S$: do GECToR(S)
 - store all grammatically corrected but possibly un-ordered S in L
 - sentence reordering model on L
4. done

Table of Contents

Progress

Where We Are At

Action Plan

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. Prabhume, R. Salakhutdinov, and A. W. Black.
Topological sort for sentence ordering.
arXiv preprint arXiv:2005.00432, 2020.

Thanks for listening!
Questions?