Filling the Gap: Decoding of Word Embeddings for Generation of Coherent New Words

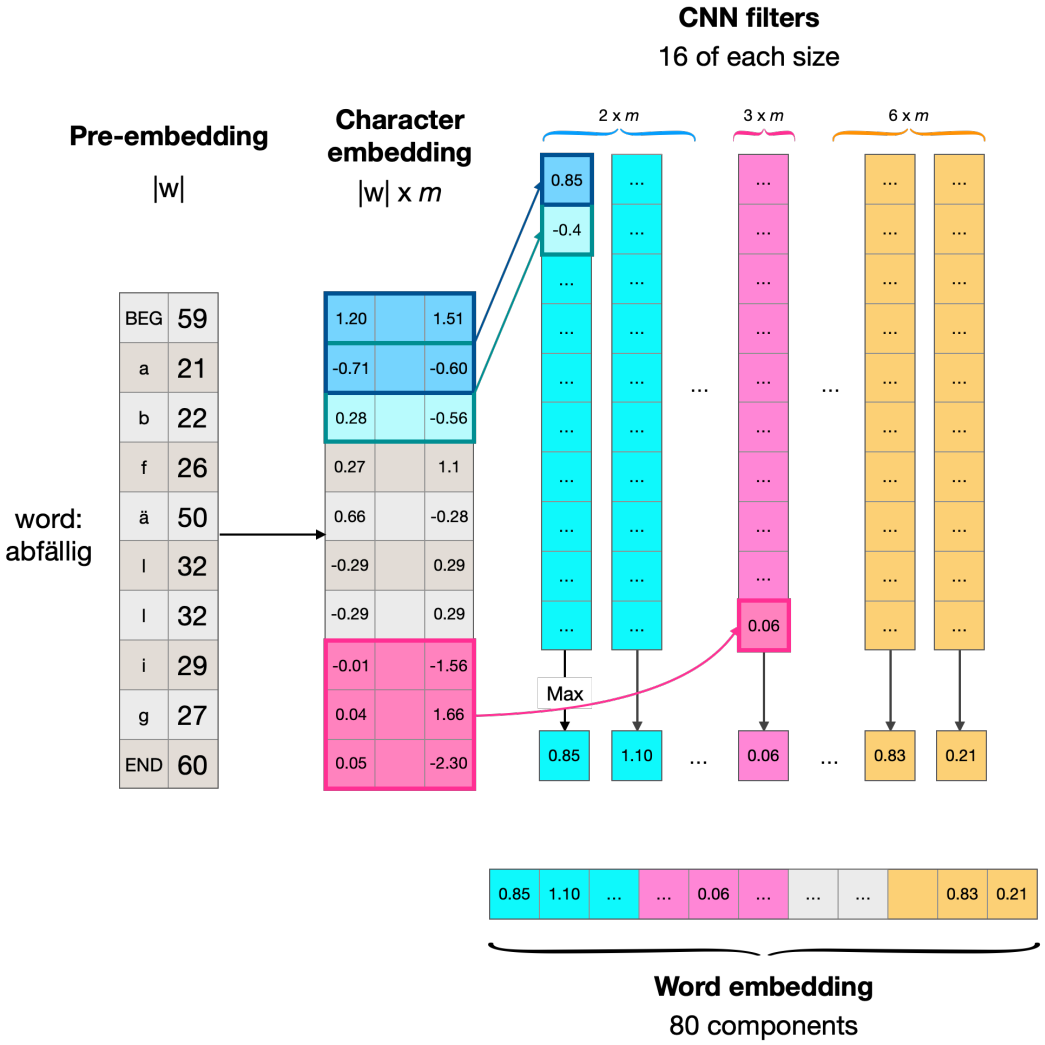
Safa AlSaidi, Amandine Decker, Stephanie Monteiro

M2 – Software Project

What we want to work on

1. Build a decoder that transforms *embedded vectors* to *words*
2. Work on morphological inflection/derivation
   - root, affixes: prefixes, suffixes..
3. Apply it to 11 different languages
   - based on two datasets [Cotterell et al., 2016, Karpinska et al., 2018]
   - improve regression task
     - $A : B :: C : X \xrightarrow{X=?} A : B :: C : D$

Pre-existing embedding model



This model was inspired by [Kim et al., 2016]

Method

Datasets:

- SIGMORPHON 2016 [Cotterell et al., 2016] and the Japanese Bigger Analogy Test Set [Karpinska et al., 2018].

Tools:

- Python libraries (mainly PyTorch)
- Grid'5000

Our repository on Github:
`https://github.com/Safa-98/morphological_embeddings_decoder`

Our Approach

1. Initial intuition: find a method to reverse the embedding model
2. Usual decoders: RNN, GRU, LSTM
3. One possible approach: *Teacher Forcing*
   - `https://rajatvd.github.io/Generating-Words-From-Embeddings/`

Future Leads

1. Auto-encoder
2. Multilingual decoder/auto-encoder
3. Decoder based on subwords/morphemes
   - `https://github.com/colingoldberg/morphemes`

Where can it be applied?

- Analogy solving (educational application)
  - "play":"build"::"replay":$X \rightarrow X =$ "rebuild"
- General structure: apply to other types of data
- Multilingual model: underrepresented languages

شكرا جزيلا

Thank you

Merci

អរគុណ

Obrigado

References I

Cotterell, R., Kirov, C., Sylak-Glassman, J., Yarowsky, D., Eisner, J., and Hulden, M. (2016).
The sigmorphon 2016 shared task—morphological reinflection.
In *Proceedings of the 2016 Meeting of SIGMORPHON*, Berlin, Germany. Association for Computational Linguistics.

Karpinska, M., Li, B., Rogers, A., and Drozd, A. (2018).
Subcharacter Information in Japanese Embeddings: When Is It Worth It?
In *Proceedings of the Workshop on the Relevance of Linguistic Structure in Neural Architectures for NLP*, pages 28–37, Melbourne, Australia. Association for Computational Linguistics.

Kim, Y., Jernite, Y., Sontag, D., and Rush, A. (2016).
Character-aware neural language models.
In Proceedings of the AAAI Conference on Artificial Intelligence, 30(1).