

MorphoDeep: Morfológiai elemzés deep learninggel

Ács Judit
judit@aut.bme.hu

BME AUT

2016. október 27.

Motiváció

- ▶ Morfológia: szavak belső szerkezetének vizsgálata.
 - ▶ $házak = ház + (a)k$
 - ▶ *zúzalékával?*
- ▶ A magyar morfológia kiemelkedően gazdag.
- ▶ Szinte minden nyelvtechnológiai feladathoz kell morfológiai elemzés.
- ▶ Jelenleg kézzel írt szabályok alapján történik.
- ▶ Deep learninget magyarra még nem alkalmaztak.

Irodalomkutatás

1. szegmentálás

- ▶ HMM-alapú, MDL kritérium
- ▶ Morfessor (Creutz and Lagus, 2002)

2. karakteralapú deep learning

- ▶ gépi fordítás (Lee et al., 2016)

3. morfológiai reinflexió

- ▶ SIGMORPHON Shared Task (Cotterell et al., 2016)
- ▶ 12 csapat beadása

Baseline: szófajok osztályozása szóalak alapján

- ▶ szóalak vs. ngram alapú
- ▶ túltanulás! - gyakorlatilag egy memóriát építünk
- ▶ hogyan lehet általánosítani?

Megvalósítás

- ▶ Architektúra
 - ▶ jelenleg FFNN
 - ▶ RNN, LSTM, CNN?
- ▶ Technológia
 - ▶ Python, jupyter
 - ▶ Tensorflow, Keras

Nehézségek

- ▶ nagyon ritka az adat,
- ▶ morfémahatárok nem egyértelműek,
- ▶ kevés gold adat, sok silver,

Nehézségek

- ▶ nagyon ritka az adat,
- ▶ morfémahatárok nem egyértelműek,
- ▶ kevés gold adat, sok silver,
- ▶ magyar morfológia

Hivatkozások



Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden.

The sigmorphon 2016 shared task—morphological reinflection.
In *Proceedings of the 2016 Meeting of SIGMORPHON*, Berlin, Germany, August 2016. Association for Computational Linguistics.



Mathias Creutz and Krista Lagus.

Unsupervised discovery of morphemes.
In *Proc. 6th SIGPHON*, pages 21—30, 2002.



Jason Lee, Kyunghyun Cho, and Thomas Hofmann.

Fully character-level neural machine translation without explicit segmentation.
arXiv preprint arXiv:1610.03017, 2016.