

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

Safa AlSaidi, Amandine Decker, Stephanie Monteiro

M2 — Software Project



- 1 Motivations
- 2 Our model
- 3 Results
- 4 Issues encountered
- 5 Future work

Motivations

- 1 Train a decoder based on word embeddings

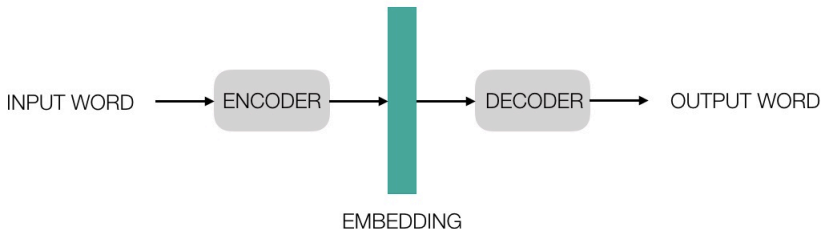
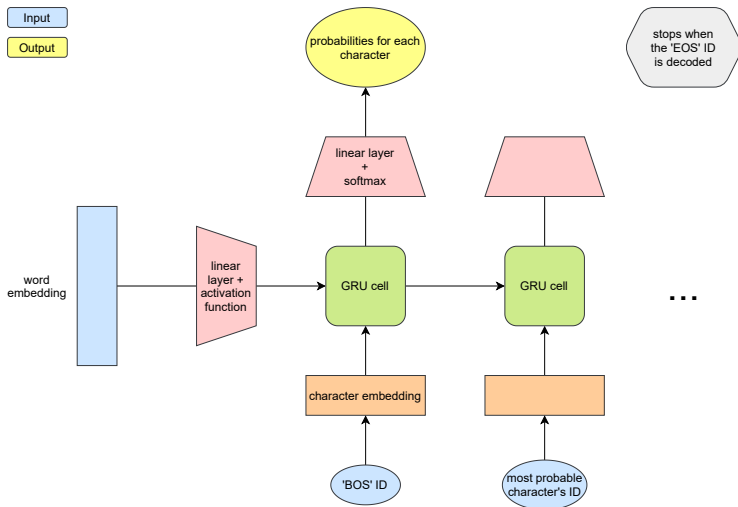


Figure: Encoder-decoder framework

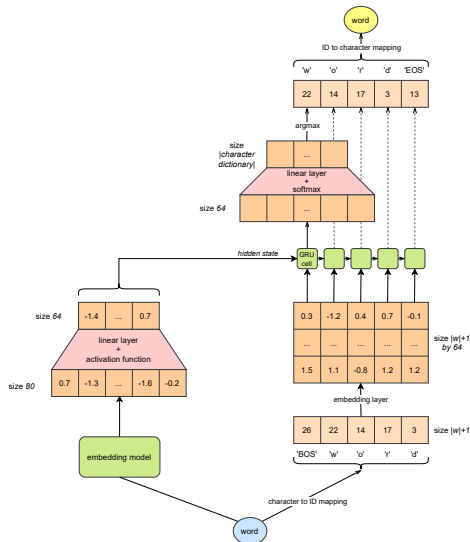
- 2 Aim: input word = output word
 - SIGMORPHON 2016 [Cotterell et al., 2016] and the Japanese Bigger Analogy Test Set [Karpinska et al., 2018].

- 1 Motivations
- 2 Our model**
- 3 Results
- 4 Issues encountered
- 5 Future work

Our model (1)



Our model (2)



Inspired by this blogpost <https://rajatvd.github.io/Generating-Words-From-Embeddings/>

- 1 Motivations
- 2 Our model
- 3 Results**
- 4 Issues encountered
- 5 Future work

First Successful Trial on Finnish

Input word	Output word
-sopimuksella	osottelustella
aamuluennon	aallalaisen
airueina	averin
alaisin	alantan
alein	alainen
ansaitsevia	ansista
automaatio	odunneista
automatisoin	onostavasta
balettitanssijan	balistella
bizetin	bisen

(a) Hidden size: 20

Input word	Output word
aatteistani	aattavansa
äitiänikään	äidän
alkeellinenkin	alkelleenin
antibioottikuureilla	antonpurkkiella
arkussaan	askusraan
armeijaankin	armentajannen
asmine	asmininti
aukealla	aukella
autoimmuunisairaudesta	auntoistusvairoista
bailando	badelo

(b) Hidden size: 64

Figure: Finnish results

Dataset from [Huovilainen, 2018]

- 1 Motivations
- 2 Our model
- 3 Results
- 4 Issues encountered**
- 5 Future work

Issues encountered

- size of the dataset ?
 - ▶ Wikipedia dumps, multilingual model
- romanised Arabic
 - ▶ change to formal Arabic, use an automatic romanising tool
- training \neq usage of the model
- relevant evaluation metric
 - ▶ Levenshtein distance

- 1 Motivations
- 2 Our model
- 3 Results
- 4 Issues encountered
- 5 Future work**

Future work

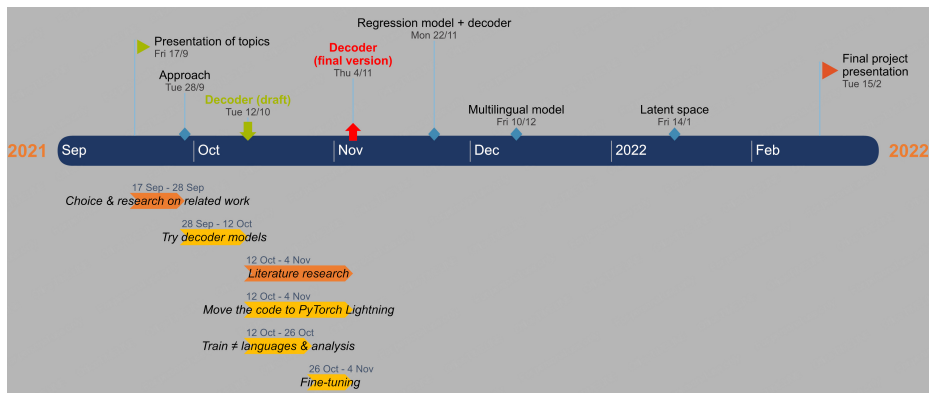


Figure: Project timeline

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.
-  Huovilainen, T. (2018).
Psycholinguistic Descriptives.
-  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.