Predicting and interpreting embeddings for out of vocabulary words in downstream tasks

Nicolas Garneau†, Jean-Samuel Leboeuf†, Luc Lamontagne

Department of Computer Science and Software Engineering, Université Laval

†Authors contributed equally to this work.

Introduction

We propose a novel way to handle out of vocabulary (OOV) words in downstream natural language processing (NLP) tasks. We implement a network that predicts useful embeddings for OOV words based on their morphology and the context in which they appear.

Motivations :

- OOV words handling in NLP task is an **underestimated problem**.
- Few learned, end-to-end, solutions proposed.

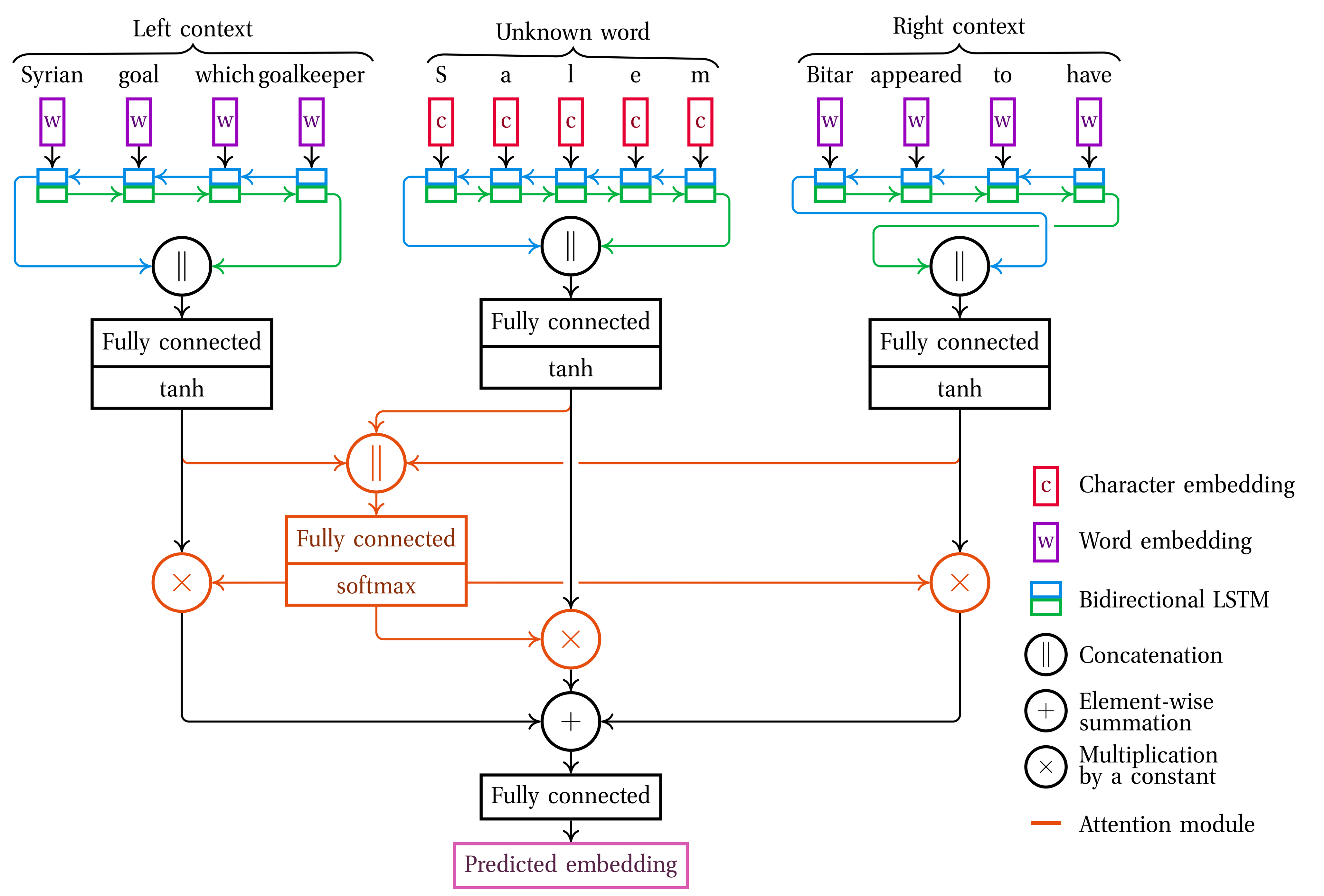
Related work :

- Pinter et al. (2017) : Predict OOV embedding using the characters.
- Bahdanau et al. (2017) : Learn OOV representation from their definition in a dictionary.

Goals :

- Evaluate the impact of OOV words in labeling tasks.
- Provide a more meaningful way to handle OOV words using **context** and **morphology**.
- Understand when it is important and what is relevant to model OOV embeddings.
- **Interpret the predicted embeddings** according to the surrounding linguistic elements.
- Provide a “drop-in”, “end-to-end” module.

OOV handling net



The net consists in 3 bi-LSTM taking as input the left context, the right context and the word characters. An attention module ponderates their outputs which are then combined in a last fully connected layer.

Experiments

Set up :

- Labeling tasks :
  - **Named Entity Recognition** (NER).
  - **POS tagging** (POS).
- Dataset : **CoNLL 2003**

Training details :

- Tensors sizes :
  - Char. emb. : 20.
  - Word emb. : 100 (**GloVe**).
  - LSTMs hidden state : 128.
- Context size from 2 words to the whole sentence.
- Standard learning rate on the labeling task parameters, reduced learning rate on Comick using SGD (0.01, 0.001).

Examples

Entity	Ponderation Word	Left	Right	Examples
PER	0.19	**0.49**	0.32	**in sentencing darrel** *voeks* , 38 , to a 10-year prison term on thursday
PER	0.15	**0.59**	0.26	<BOS> **australian parliamentarian john** *langmore* has formally resigned from his lower house
PER	0.15	**0.61**	0.24	had received today **from mr john vance** *langmore* , a letter resigning his place as
PER	0.15	**0.69**	0.16	<BOS> **rtrs - australian mp john** *langmore* formally resigns . <EOS>
ORG	0.22	**0.46**	0.32	the number of plastic surgeries in [...] the **brazilian plastic surgery society** (*sbcp*) , said ,
ORG	0.28	0.23	**0.49**	to increase them in the united states , " *sbcp* **vice-president oswaldo saldanha said**
LOC	0.16	0.22	**0.62**	some residents of the *kazanluk* **area are moslems who** converted to islam during
LOC	0.20	**0.47**	0.33	at a mosque in the **central bulgarian town of** *kazanluk* , causing damage but no injuries
MISC	**0.68**	0.11	0.21	freestyle ***skiing-world*** cup aerials results .
MISC	**0.42**	0.18	**0.40**	the ***franco-african*** **summit decided to send** a mission bangui [...] civil war .

Qualitative example on several OOV words (underlined). We can see that depending on the context and the target, the weights may shift drastically.

Experiments

Set up :

- Labeling tasks :
  - **Named Entity Recognition** (NER).
  - **POS tagging** (POS).
- Dataset : **CoNLL 2003**

Experiments

Set up :

- Labeling tasks :
  - **Named Entity Recognition** (NER).
  - **POS tagging** (POS).
- Dataset : **CoNLL 2003**