Sentence Fusion

Andrés Herrera Cepero, Laurine Jeannot, Quentin Pouvreau and Imane Zouhri

Université de Lorraine

Objectives

Starting from a project which split complex sentences into several simple ones, this project aims to do the inverse, *i.e* to merge a set of related sentences into a unique complex sentence.

Challenge: to produce a deep learning model that makes predictions which are:

- are **well-formed**: the sentences produced need to be a sentence grammatically correct.
- maximise **the informations saved**: we need to keep as much informations from the simple sentences as possible.
- are **coherent** with the originals: all the informations that are in the complex sentence should be in the set of simple sentences.

Experimental Setup

Open-NMT

- Neural network framework
- PyTorch
- GPU needed

Grid5000

- Testbed for research
- Usable by SSH process
- GPU config which is needed for our training

Acknowledgements

Special thanks to Claire Gardent for the support and the guidance provided.
Thanks to Loria for the opportunity to work in a stimulating research environment on an current natural language processing topic.

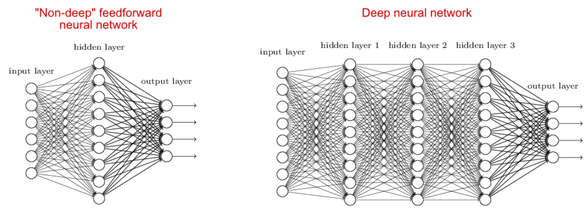
Introduction

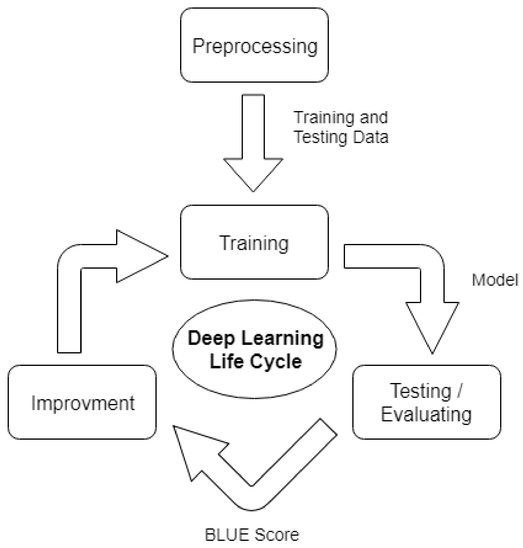
Typical applications of **sentence fusion** are its use in the context of multi-document summarization, for example it can allow to produce a **summary** from multiple news articles that have information in common.

The aim of this supervised project is to investigate **deep learning approaches** to sentence fusion using the Split-and-Rephrase dataset [Narayan et al., 2017]. and to **evaluate** them.

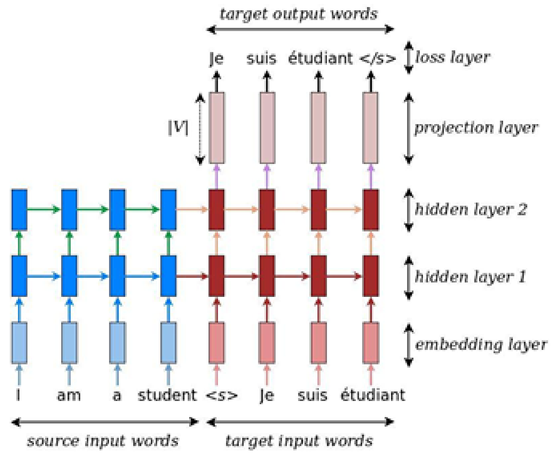
Simple sentences	**"The 11th Mississippi Infantry Monument is a contributing property."; "The 11th Mississippi Infantry Monument was established in 2000."; "The 11th Mississippi Infantry Monument is located in Gettysburg and Adams County in Pennsylvania ."**
Resulting sentence	**"The 11th Mississippi Infantry Monument , established in 2000 , is a contributing property and located in Gettysburg and Adams County in Pennsylvania ."**

Approach





Because of the sequence based nature of the sentence fusion task, a sequence to sequence model was the best suited to utilize in the investigation.
Since its beginnings, different approaches have been used to sequence to sequence tasks, some focusing on statistics.
Recently, however, Neural Machine Translation(NMT) apply a sequence to sequence translation which produces higher quality results.



Conclusions

The first point to conclude is that further investigation is recommended with an improved dataset, specific to Sentence Fusion task.
On the other hand, we could conclude that even deep learning had good results as presented in the "Split-and-Rephrase Better Evaluation and a Stronger Baseline" paper, it is not as good a fit for sentence fusion since the attention model didn't provide better scores whereas it did it on the Split-and-Rephrase task.
By all means, deep learning for Sentence Fusion is a hard and time consuming task.

Results: BLEU score

Metric to evaluate how good is a translation
Basically for translation
Compare a reference corpus to the predictions made by the machine
Here compare the predictions to the complex sentence in the reference corpus

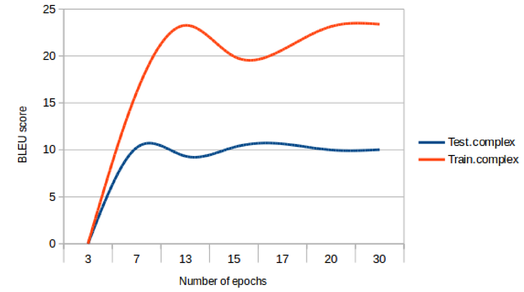


Figure 1: BLEU score for fullmodel S&R v.0.1

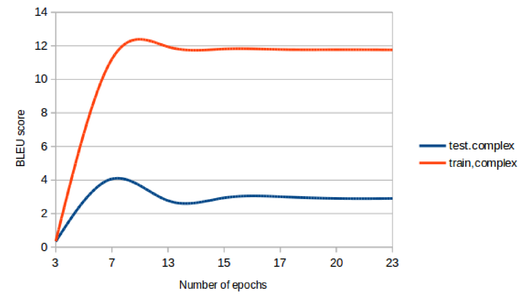


Figure 2: BLEU scores for fullmodel_copy_attn v.0.1

Results: Human Evaluation

Ideally made by two different persons on a significant part of the results.
Criteria:

- Fluency of the produced sentence
- Grammatical correction
- Is it a complex sentence ?
- Can we guess the theme of the sentence ?

Example: *Ahmet_Davutoglu is Turkey 's leader where the currency is located in Izmir .*
Evaluation: Grammar: 2; Fluency: 5; Complex: 1; Theme guessed: 1

Contact Information

Web https://members.loria.fr/CGardent/
Email claire.gardent@loria.fr