INF554 : Data challenge

Dimitri DELPECH de SAINT GUILHEM
Matthieu SOUDA

Introduction

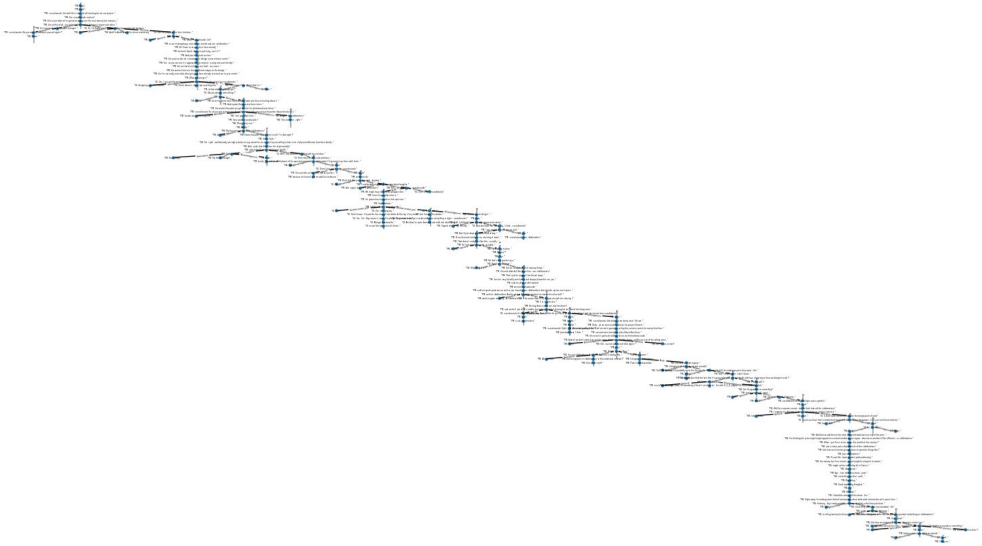


Table of contents

I. Features utilization

Features

Neural Networks

Text features :

- Tokenization
- Embeddings of words or sentences

Type feature :

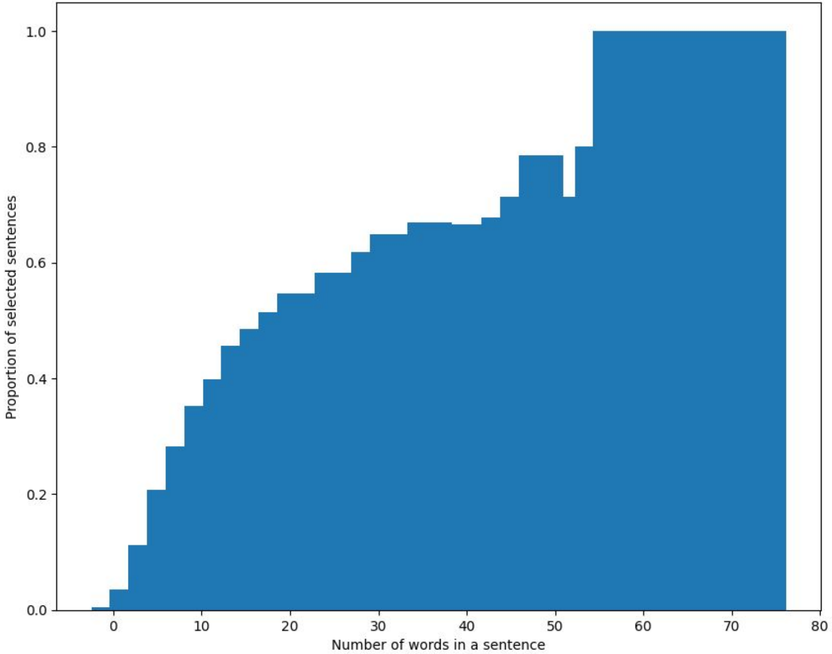
- Encoded as a 16 dimensions vector

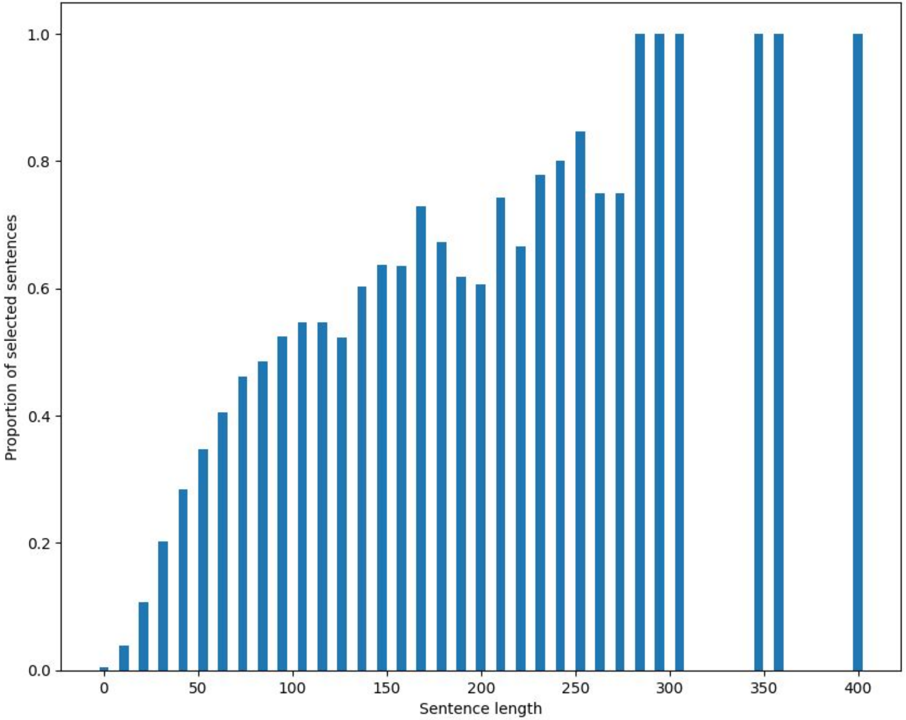
For Naive Bayes and Random Forest

- sentence length (character wise)
- sentence length (word wise)
- discourse type (as a number or as weight based on proportions)
- speaker
- number of child in graph

Feature selection

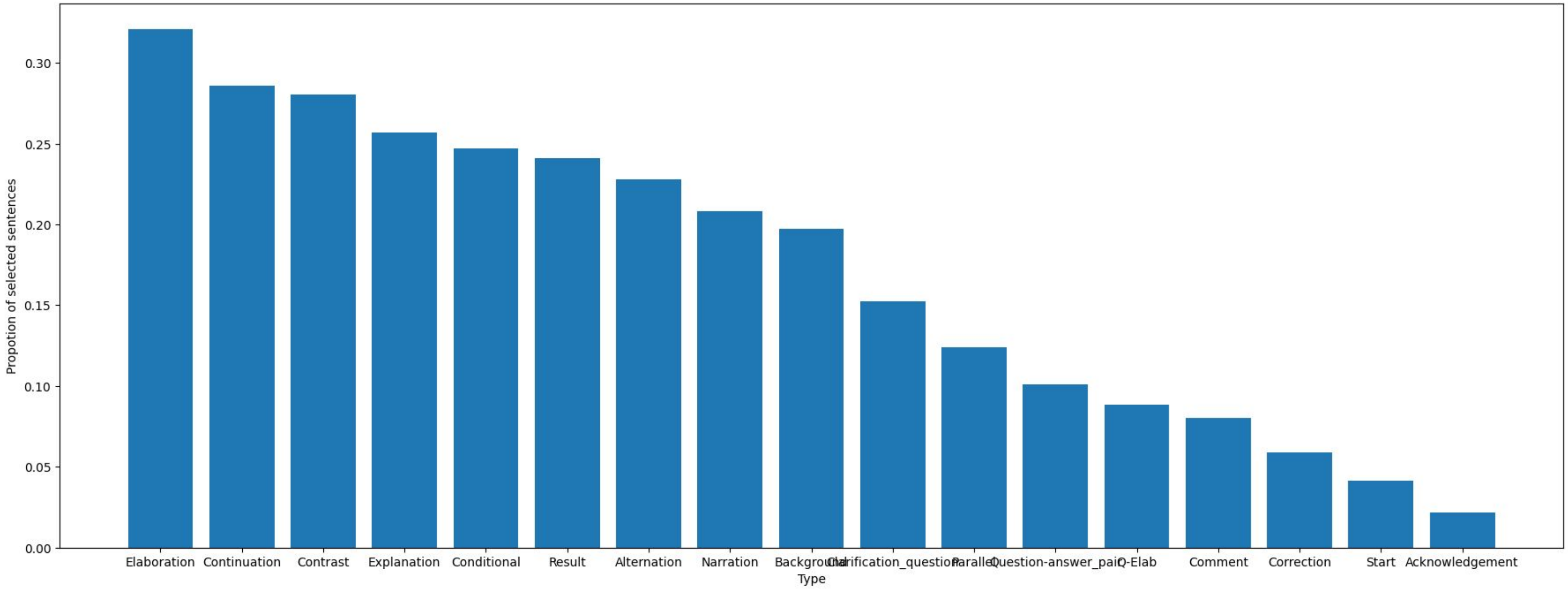
n_words & sentence length





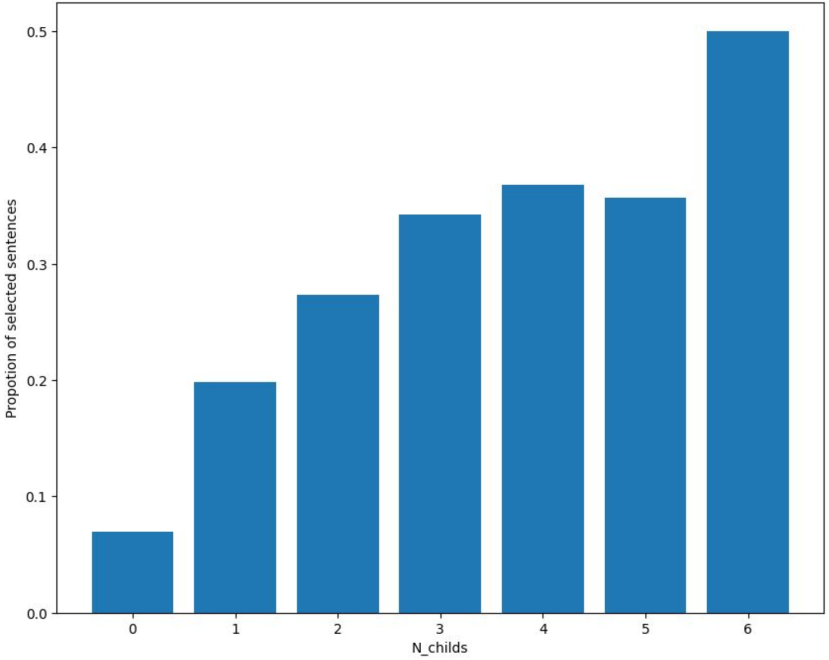
Feature selection

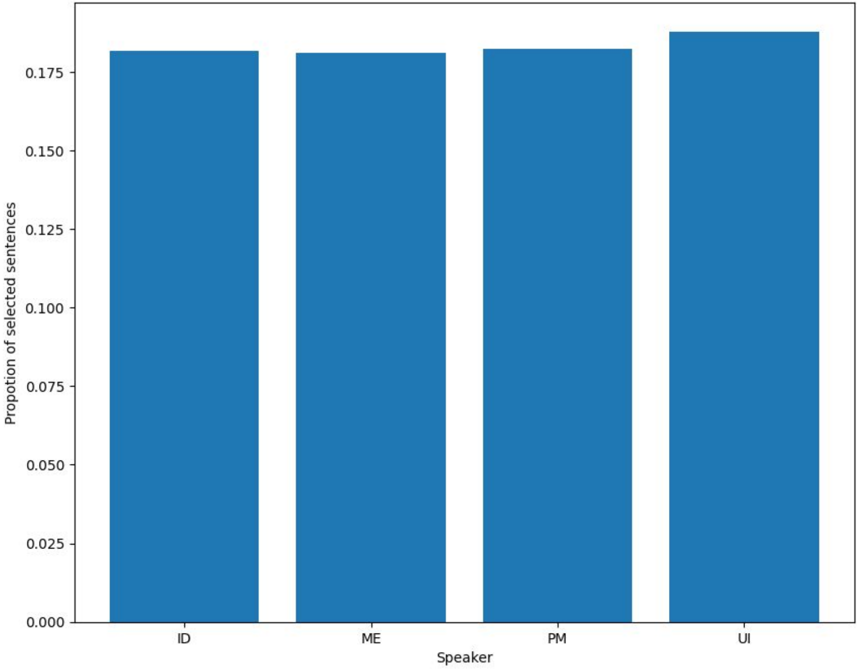
type of sentence (from the graph)



Feature selection

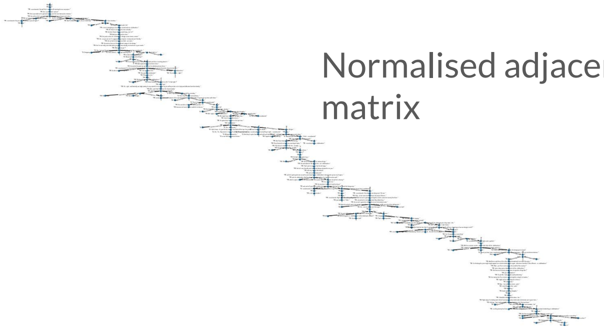
n_childs in the graph & speaker



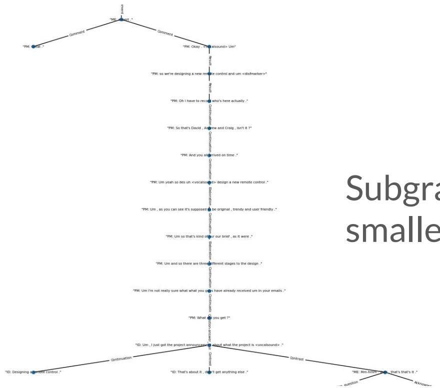


Graph feature

Neural Networks only



Normalised adjacency matrix



Subgraph for smaller batch

Sentence embeddings as node features.

II. Models

Naive Bayes and Random Forest

→ Our own baseline models (fast to train)

→ See the influence of the extracted features

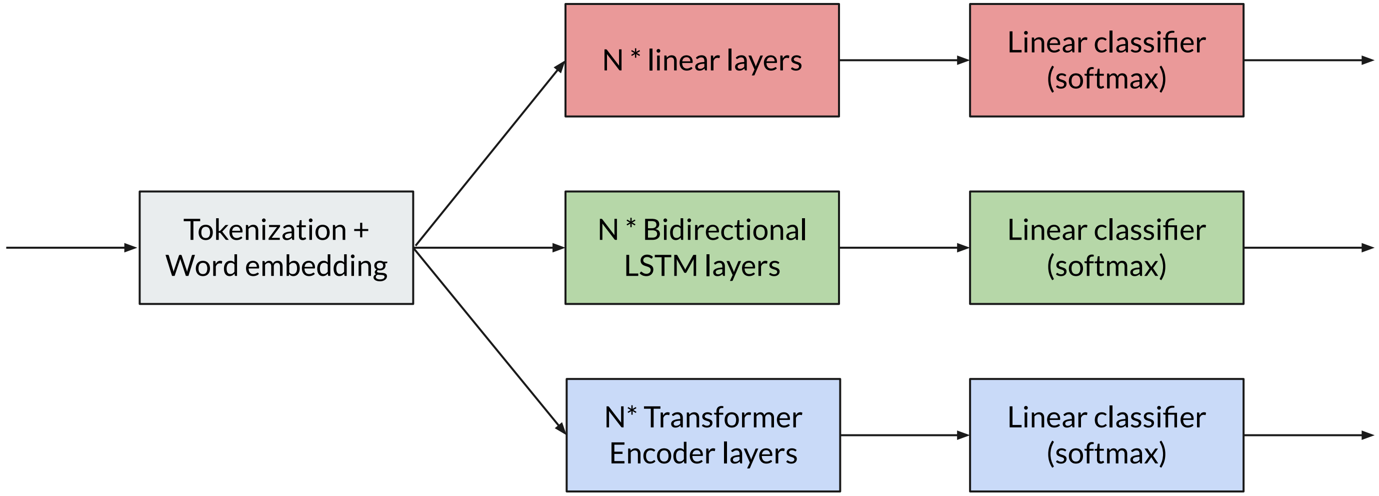
→ Thought about combining it with NN models

Results :

- NB : around 0.48 f1 score
- RandomForest : around 0.40 f1 score

A range of models

Simple MLP
LSTM (Bidirectional)
Transformer Encoder



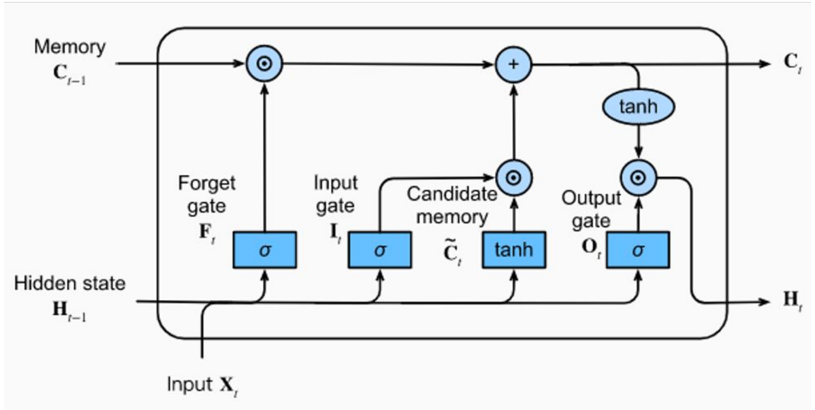
LSTM

→ Recurrent Neural Network

→ Bidirectional layers : understanding the context before and after each word

→ Recurrent to understand sequences of sentences

→ LSTM specific : cell state to understand long term trends in the discourse



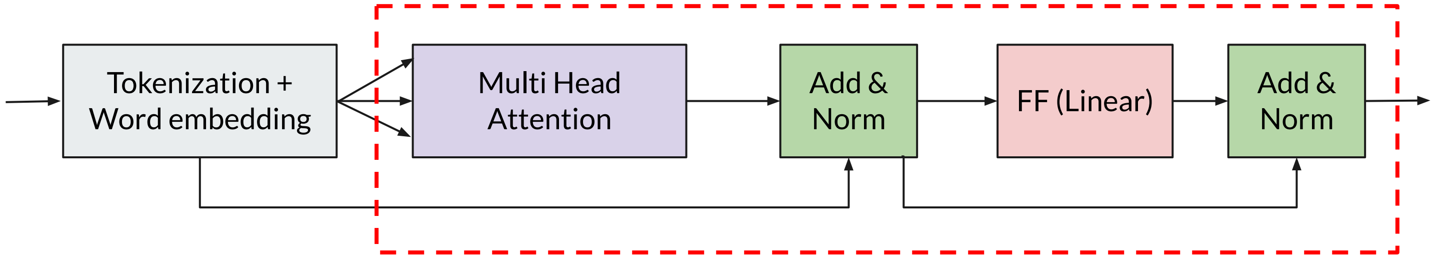
Transformer Encoder

→ From *Attention is all you need (2017)* Transformer implementation

→ Multiple Encoder layers

→ Frequently used in NLP tasks

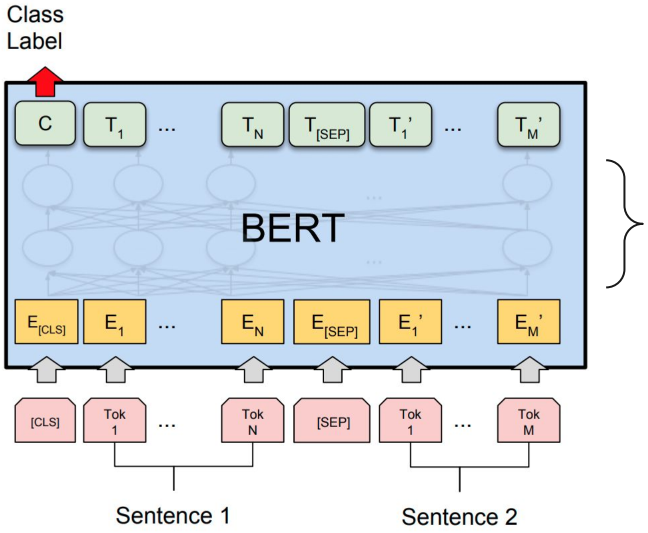
→ Self-Attention



Encoder Layer architecture

BERT

Bidirectional Encoder Representations from Transformers.



Bidirectional Transformer layers

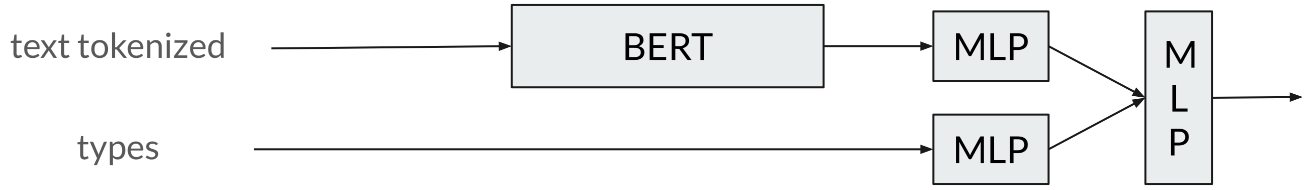
The class Label is used for classification

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018

BERT based models

BERT + MLP: based on the classifier token.

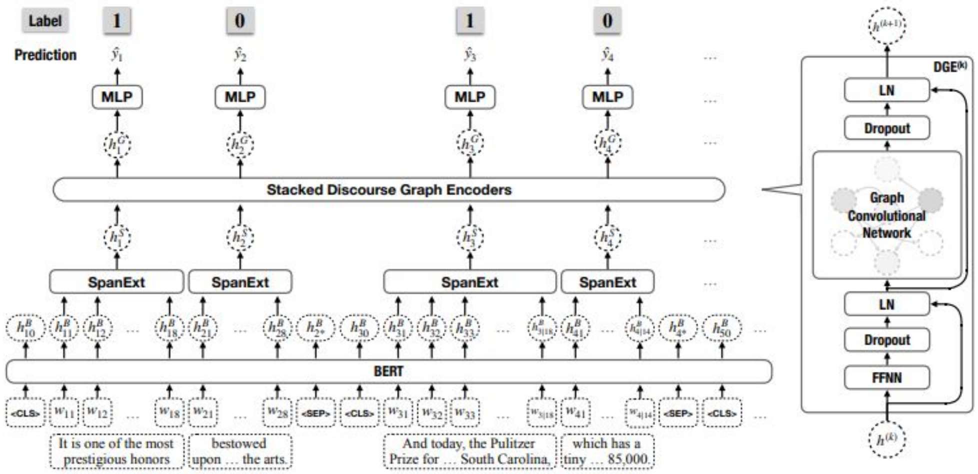
- Best results with one layer of size (Bert hidden size, 2)
- Combined with a simple MLP for types



BERT + LSTM (two versions):

- LSTM layer based on the classification token
- LSTM that uses the sentence embedding created by averaging the word embeddings given.

DiscoBert



SpanExt : create Elementary Discourse Units

We used the sentence embedding as a mean of the word embeddings.

Coreference Graph and Discourse Graph.

We used only the discourse graph.

Discourse-*Aware Neural Extractive Text Summarization, 2020*

III. Results and critics

Results

Model	F1 score
Bert Classifier	0.576
Bert LSTM (on classifier tokens)	0.523
Bert Classifier + types MLP	0.497
Naive Bayes	0.48
LSTM	0.55
Random Forest	0.40
Transformer	0.568

BERT LSTM on sentence embeddings
and
Adaptation of DiscoBERT : GPU out of memory

Critics

Technical difficulties on training.

Not enough variety on GCN implementation. Not adapted to trees ?

If we had to continue :

- Fine tune BERT and then use it in preprocessing (gain in memory and time)
- Combine models and methods
- As in DiscoBert try to create a coreference graph.

Conclusion