LSE at DEFT 2018: Classification of tweets based on Deep Learning

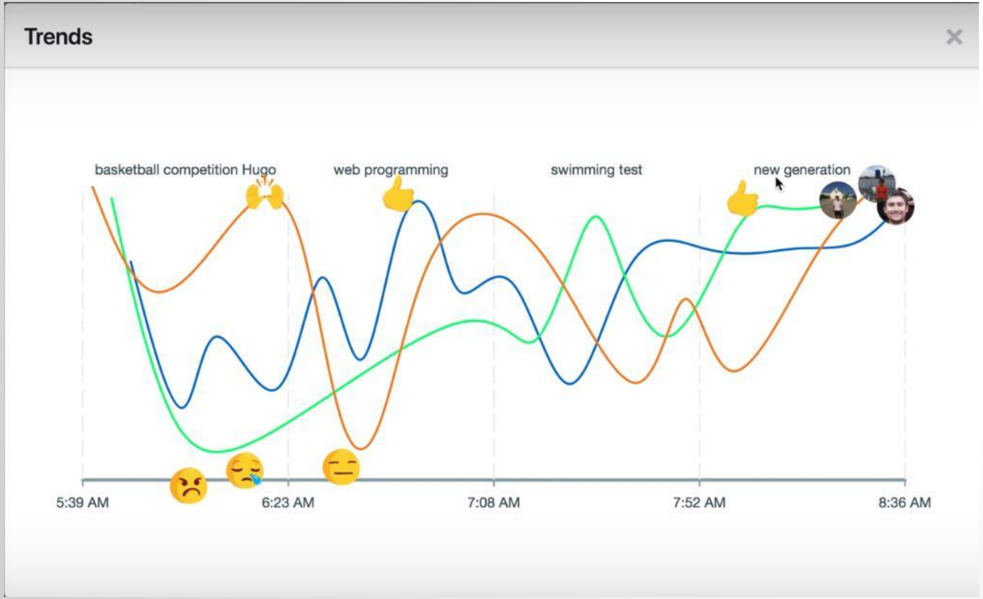
Antoine Sainson
Hugo Linsenmaier
Alexandre Majed
Xavier Cadet
Abdessalam Bouchekif

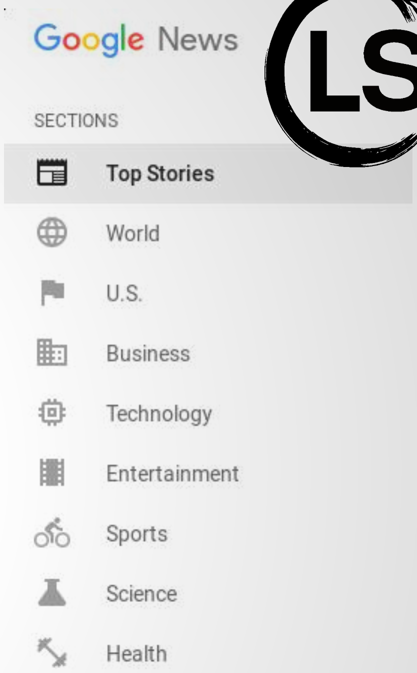
Introduction

- ➢ Tasks 1 (0.894) and 2 (0.793).
- ➢ Text Classification.
- ➢ Deep Learning models: CNN, LSTM, BLSTM, GRU.

Applications

- ➢ Google news
- ➢ Feedback of customers





Pipeline

Preprocessing → Word Embeddings → Learning

Preprocessing

- ➔ Lemmatization
- ➔ Normalization

Word Embeddings

- ➔ One Hot
- ➔ Word2vec

Learning

- ➔ CNN
- ➔ LSTM
- ➔ BLSTM
- ➔ GRU
- ➔ Combination

Preprocessing

Lemmatization

- TreeTagger
- Get rid of gender and number agreement
- Transform word into its canonical form
  - Verbs to infinitive form:
    - “ai” -> “avoir”

Normalization

- Lower Case
- Remove useless information
  - URLs, emails, dates
  - “le”, “de”, “te”, “ce”
- Keep negation words for sentiment analysis
- Group smiley by feelings

Concrete Example

Si vs avez pas le bac svp ne pensez pas a vous suicider sur la ligne du RER D merci

vs pas bac svp ne penser pas suicider ligne rer merci

Word Embeddings

Word Embeddings

- ➢ Scalar Vector Representation of a Word
- ➢ Context Sensitive
- ➢ One hot representation
- ➢ FastText

$$
\begin{array}{lc}
 & \begin{array}{ccccc} \text{Le} & \text{Bus} & & & \end{array} \\
\text{Le} & = [\ 1,\ 0,\ 0,\ 0,\ 0\ ] \\
\text{Bus} & = [\ 0,\ 1,\ 0,\ 0,\ 0\ ] \\
\text{Est} & = [\ 0,\ 0,\ 1,\ 0,\ 0\ ] \\
\text{En} & = [\ 0,\ 0,\ 0,\ 1,\ 0\ ] \\
\text{Retard} & = [\ 0,\ 0,\ 0,\ 0,\ 1\ ]
\end{array}
$$

Skip-Gram

➢ Predict the context of words

Source Text	Training Samples
Le bus est souvent en retard le matin.	(Le, bus) (Le, est)
Le bus est souvent en retard le matin.	(bus, Le) (bus, est) (bus, souvent)
Le bus est souvent en retard le matin.	(est, Le) (est, bus) (est, souvent) (est, en)
Le bus est souvent en retard le matin.	(souvent, bus) (souvent, est) (souvent, en) (souvent, retard)

N-Grams

P. Bojanowski et al, 2017

- Way to deal with noisy data
- Each word is treated as a ‘list’ of subwords
- Example for the french word ‘retard’:

[<re, ret, eta, tar, ard, rd>, <retard>]

Word2vec Learning

- #RATP (169K)
- #SNCF (453K)
- #IleDeFrance (90K)
- External Corpus (~50K)

- Context Window: 4
- Min Count: 5
- Skip-Gram

Thematic/Polarity Value

- Probability of given word to belong to a class.
- “Bonheur”:
  - 85% positive
  - 12% negative
  - 0% neutral
  - 3% mixed
- “Bus”:
  - 77% transport
  - 23% unknown

Input Features

- Word Embedding (100)
- “Thematic/Polarity” value (2 or 4)

Models

Recurrent Models : LSTM, GRU, BLSTM

- LSTM is capable of learning long term dependencies

- GRU is a simplified version of LSTM

- BLSTM consists in running 2 LSTMs in parallel

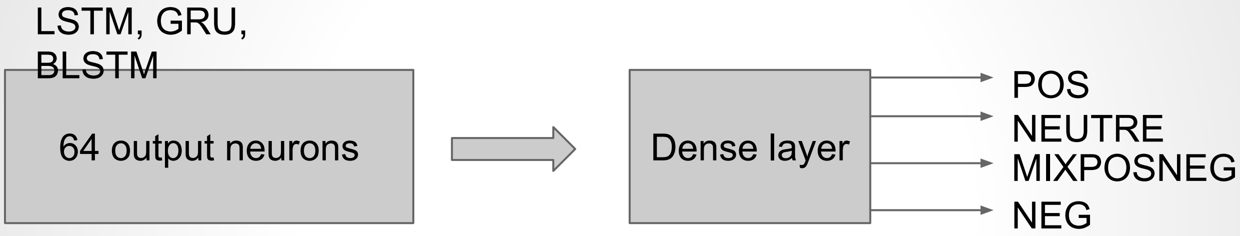
Convolutional Neural Networks

- Obtain good results on text classification tasks

- Composed of convolutional layers, pooling layers and fully connected layers

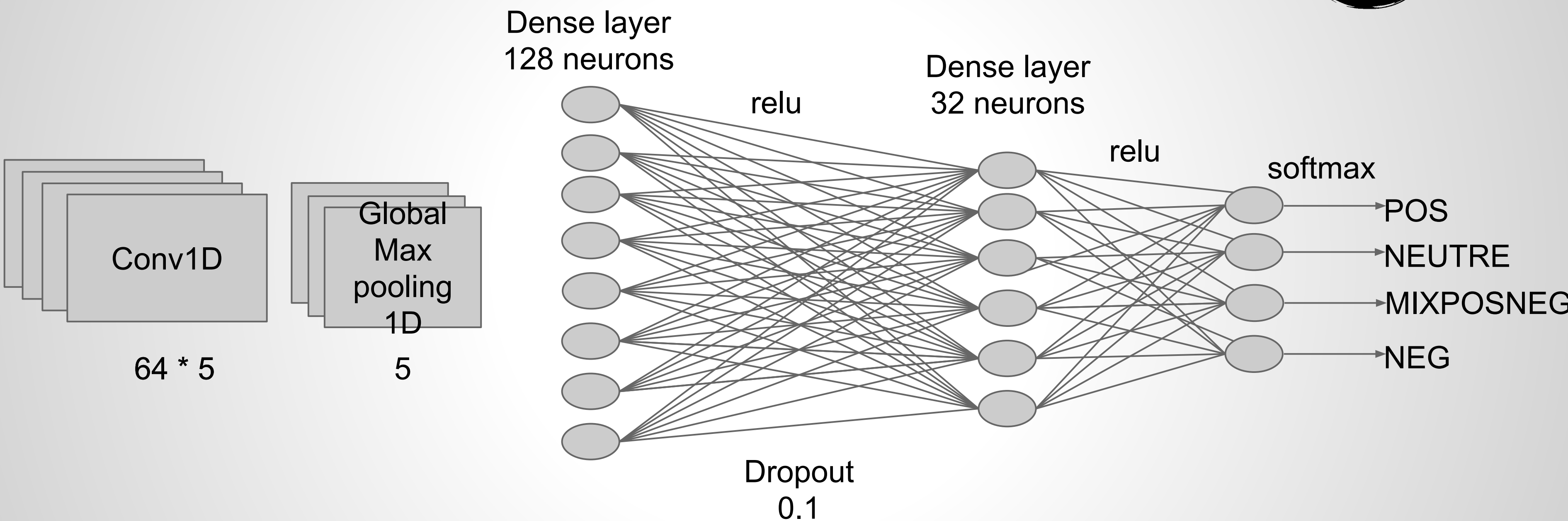
Results

Parameters for RNNs



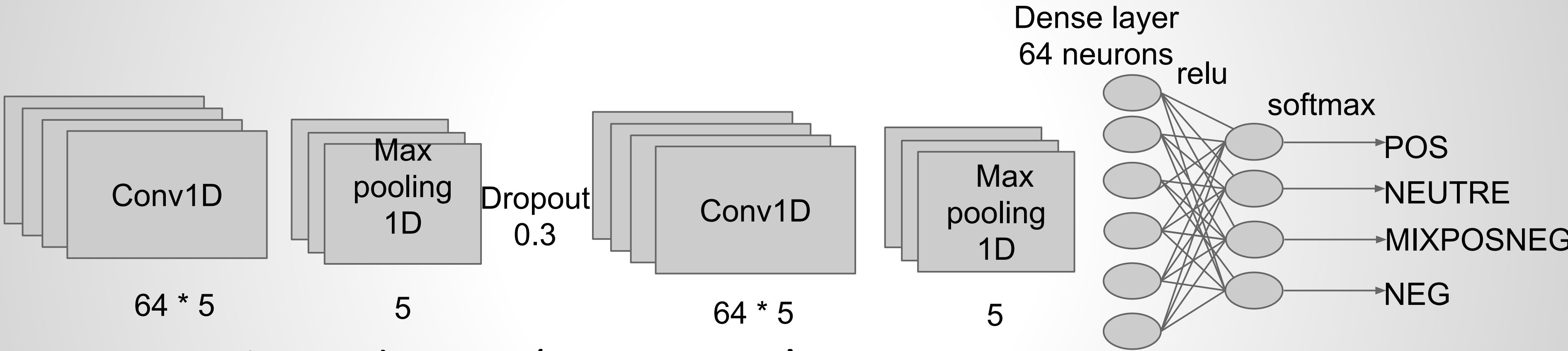
- Loss: binary/categorical crossentropy
- LSTM: 8 epochs
- GRU: 7 epochs
- BLSTM: 10 epochs

CNN1



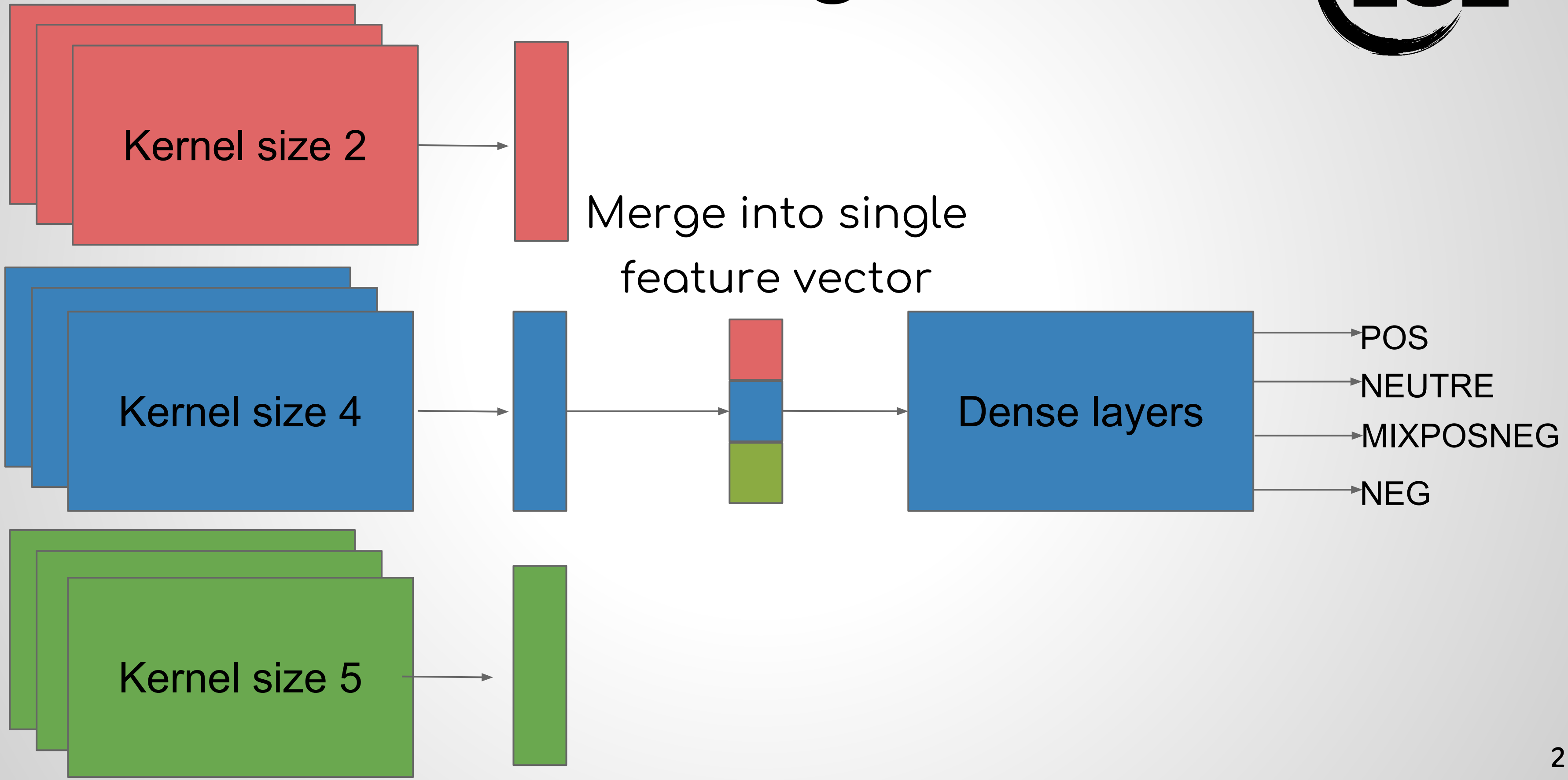
- ➢ Loss: binary/categorical cross-entropy
- ➢ 10 epochs

CNN2



- Loss: binary/categorical cross-entropy
- optimizer :
- 6 epochs

CNN Kernel Merge



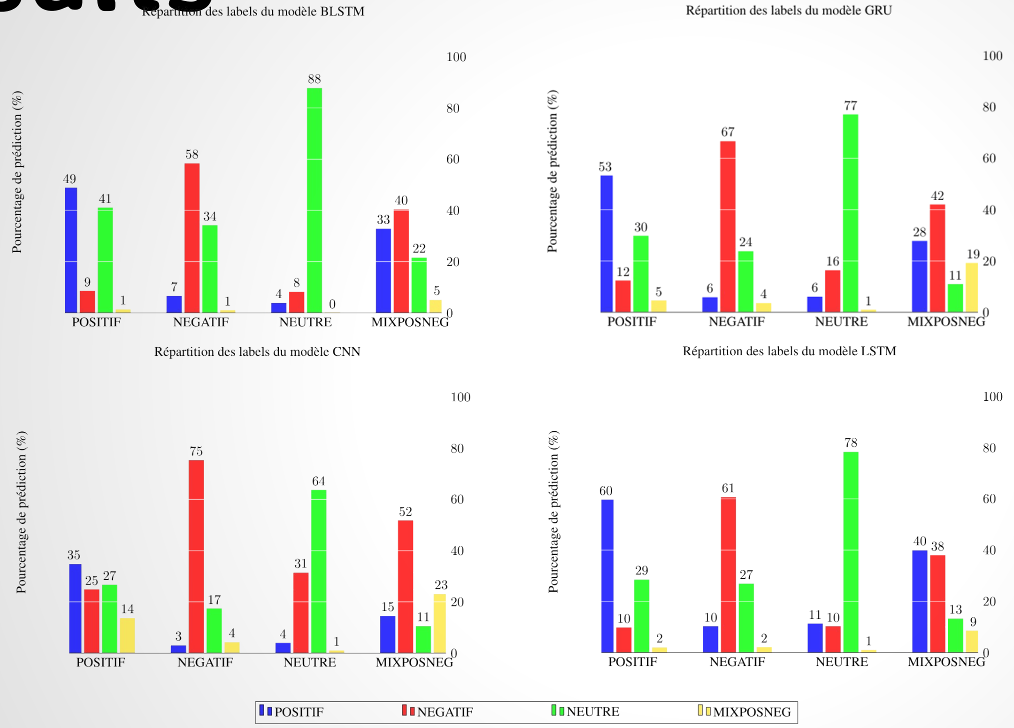
Results - Task 1

Type de réseau	Precision	F1-measure
CNN 1	0.799	0.888
CNN 2	0.798	0.888
CNN Kernel Merge	0.803	0.890
BLSTM	0.810	0.892
GRU	0.800	0.889
LSTM	0.795	0.886
Combination with NN	0.767	0.868
Average	**0.808**	**0.894**

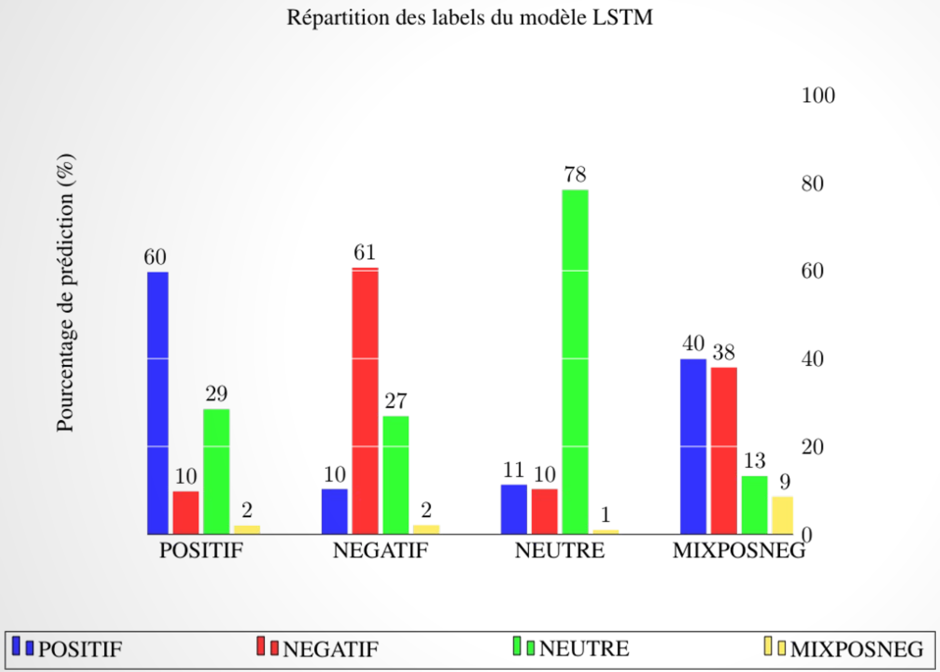
Results - Task 2

Type de réseau	Precision	F1-measure
CNN 1	0.605	0.754
CNN 2	0.610	0.757
CNN Kernel Merge	0.639	0.780
BLSTM	0.623	0.768
GRU	0.641	0.781
LSTM	0.632	0.774
Combination with NN	0.610	0.757
Average	**0.657**	**0.793**

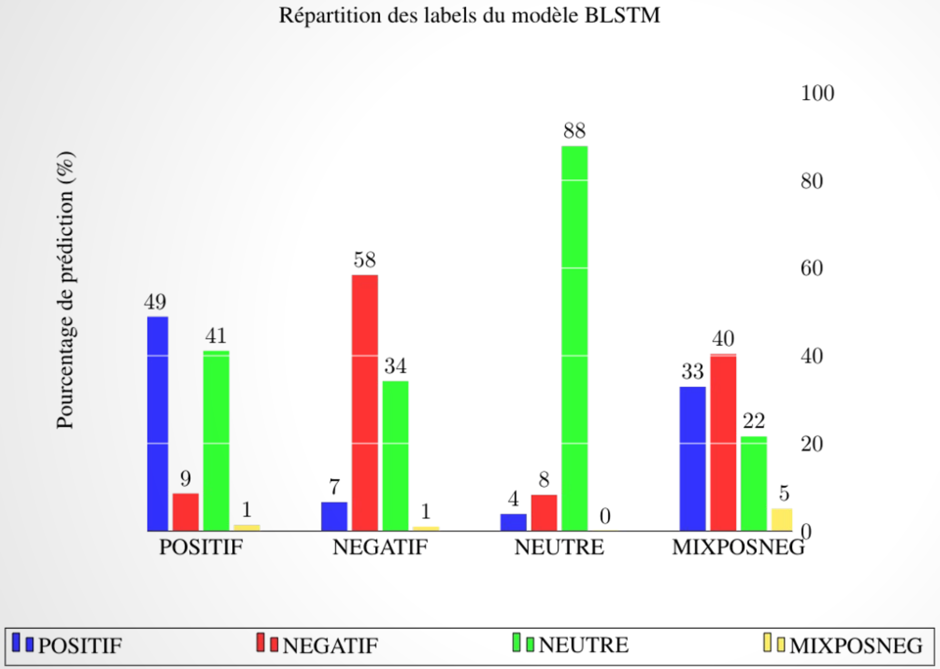
Results


des labels du modèle BLSTM
Répartition des labels du modèle GRU
Répartition des labels du modèle CNN
Répartition des labels du modèle LSTM
Pourcentage de prédiction (%)
0
20
40
60
80
100
POSITIF
NEGATIF
NEUTRE
MIXPOSNEG
BLSTM: 49 9 41 1 | 7 58 34 1 | 4 8 88 0 | 33 40 22 5
GRU: 53 12 30 5 | 6 67 24 4 | 6 16 77 1 | 28 42 11 19
CNN: 35 25 27 14 | 3 75 17 4 | 4 31 64 1 | 15 52 11 23
LSTM: 60 10 29 2 | 10 61 27 2 | 11 10 78 1 | 40 38 13 9
POSITIF NEGATIF NEUTRE MIXPOSNEG

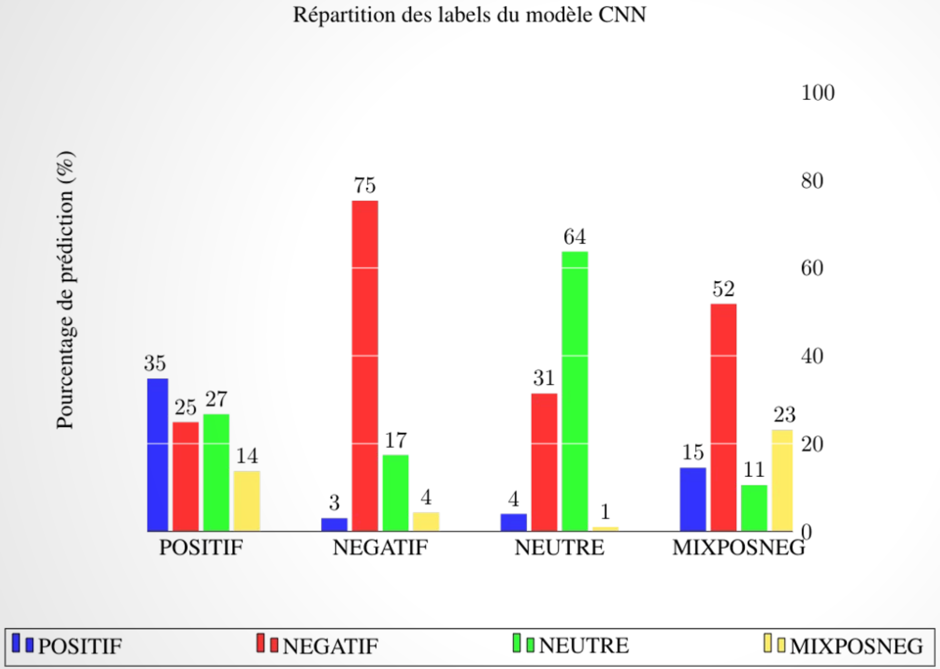
Results - LSTM



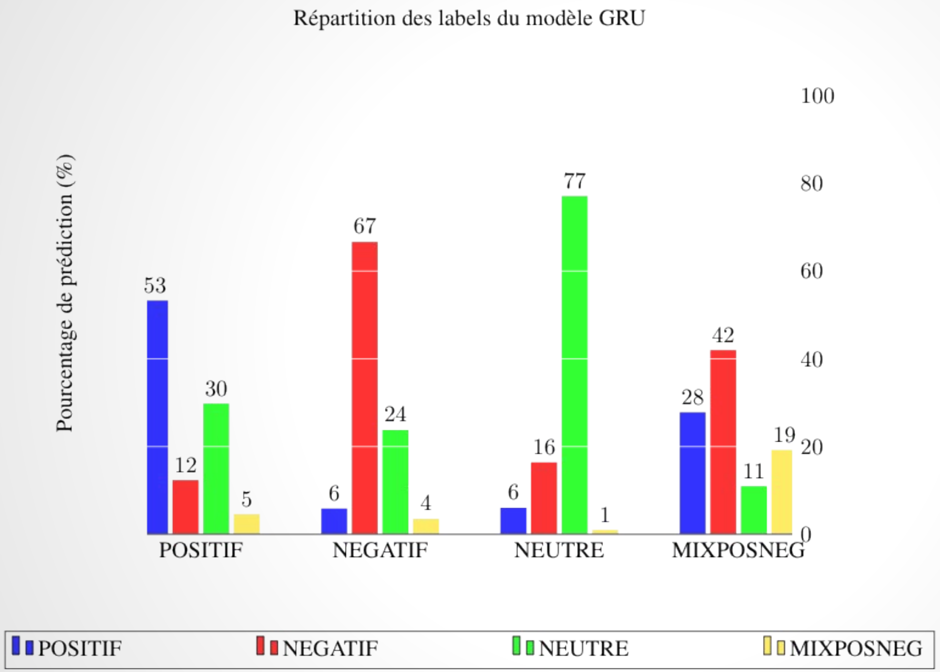
Results - BLSTM



Results - CNN


Répartition des labels du modèle CNN
Pourcentage de prédiction (%)
100
80
60
40
20
0
35
25
27
14
75
3
17
4
64
31
4
1
52
15
11
23
POSITIF
NEGATIF
NEUTRE
MIXPOSNEG
POSITIF
NEGATIF
NEUTRE
MIXPOSNEG

Results - GRU



Répartition des labels du modèle GRU
Pourcentage de prédiction (%)
100
80
60
40
20
0
53
12
30
5
6
67
24
4
6
16
77
1
28
42
11
19
POSITIF
NEGATIF
NEUTRE
MIXPOSNEG
POSITIF
NEGATIF
NEUTRE
MIXPOSNEG

Annotation

Ya intérêt que le bus soit à l'heure parce que en retard le premier jour c'est moyen

Si vs avez pas le bac svp ne pensez pas à vous suicider sur la ligne du RER D merci

Conclusion et perspectives

- ➢ Transfert Learning
- ➢ Combination using logistic regression

Any Questions?

Let the fun begin

$$\sum_{t=1}^{T} \sum_{c \in \mathcal{C}_t} \log p(w_c \mid w_t)$$

$$\log\left(1 + e^{-s(w_t,\, w_c)}\right) + \sum_{n \in \mathcal{N}_{t,c}} \log\left(1 + e^{s(w_t,\, n)}\right)$$

$$\sum_{t=1}^{T} \left[\sum_{c \in \mathcal{C}_t} \ell(s(w_t,\ w_c)) + \sum_{n \in \mathcal{N}_{t,c}} \ell(-s(w_t,\ n)) \right]$$