

Évaluation des apports d'un reseau de neurones antagoniste pour une classification semi-supervisé de textes français issue d'une base de données faiblement labellisés

Dylan Baptiste, Stéphane Cormier, and Shufan Jiang

Université de Reims Champagne-Ardenne
dylan.baptiste@etudiant.univ-reims.fr
stephane.cormier@univ-reims.fr
shufan.jiang@univ-reims.fr

Résumé Les tâches de classifications de textes se heurtent souvent au problème de bases de données peu voir pas labellisé. Ici nous nous intéressons au apports de l'ajout d'un GAN afin d'artificiellement créer des données labellisés. Le but est d'améliorer les performances d'un classifieur qui ne se serait entraîné que sur des données labellisé issue d'une petite base de données.

Keywords: réseau de neurones, réseaux antagonistes génératifs, traitement du langage, classification, semi-supervisé

Table des matières

Évaluation des apports d'un reseau de neurones antagoniste pour une classification semi-supervisé de textes français issue d'une base de données faiblement labellisés	1
<i>Dylan Baptiste, Stéphane Cormier, and Shufan Jiang</i>	
1 Contexte	2
2 État de l'art	2
2.1 Bidirectional Encoder Representations from Transformers	2
2.2 Réseaux antagonistes génératifs	3
3 Architecture	3
3.1 Générateur	3
3.2 Discriminateur	3
3.3 GAN	4
4 Entraînement	5
4.1 Supervisé	5
4.2 Semi-supervisé	5
Compétition ou coopération ?	5
Stabilisation du Générateur et Discriminateur	5
5 Evaluation	5
5.1 Mesures utilisées	5
5.2 Comparaison des mesures	5
6 Conclusion et perspectives	5

Table des figures

Liste des tableaux

1 Contexte

Ce travail à été réalisé dans le cadre d'un TER (Travail d'Étude et de Recherche) par Dylan Baptiste, étudiant en 2^e année de Master, supervisé par Shufan Jiang, doctorante, et Stéphane Cormier, enseignant-chercheur, à l'Université de Reims Champagne-Ardenne.
présentation des données...

2 État de l'art

2.1 Bidirectional Encoder Representations from Transformers

en fr "Représentations d'encodeur bidirectionnel à partir de transformateurs"
BERT, CamemBERT...explications des ouputs de tels modeles

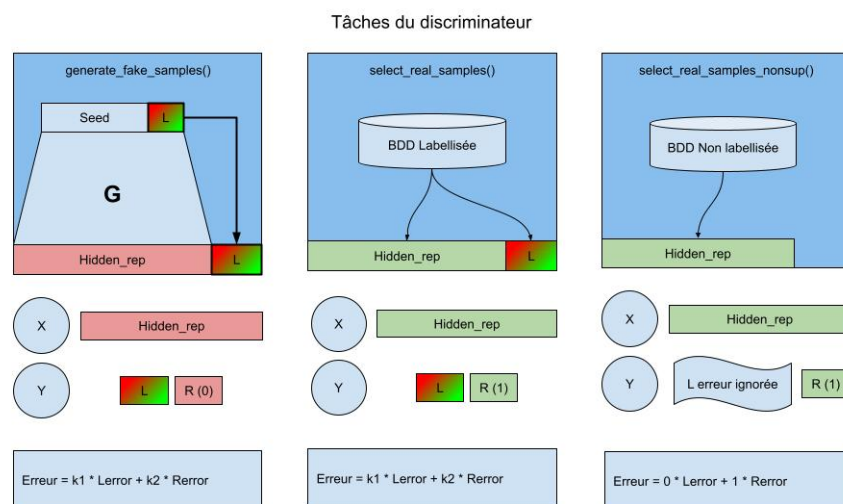
2.2 Réseaux antagonistes génératifs

GAN, Conditional GAN...

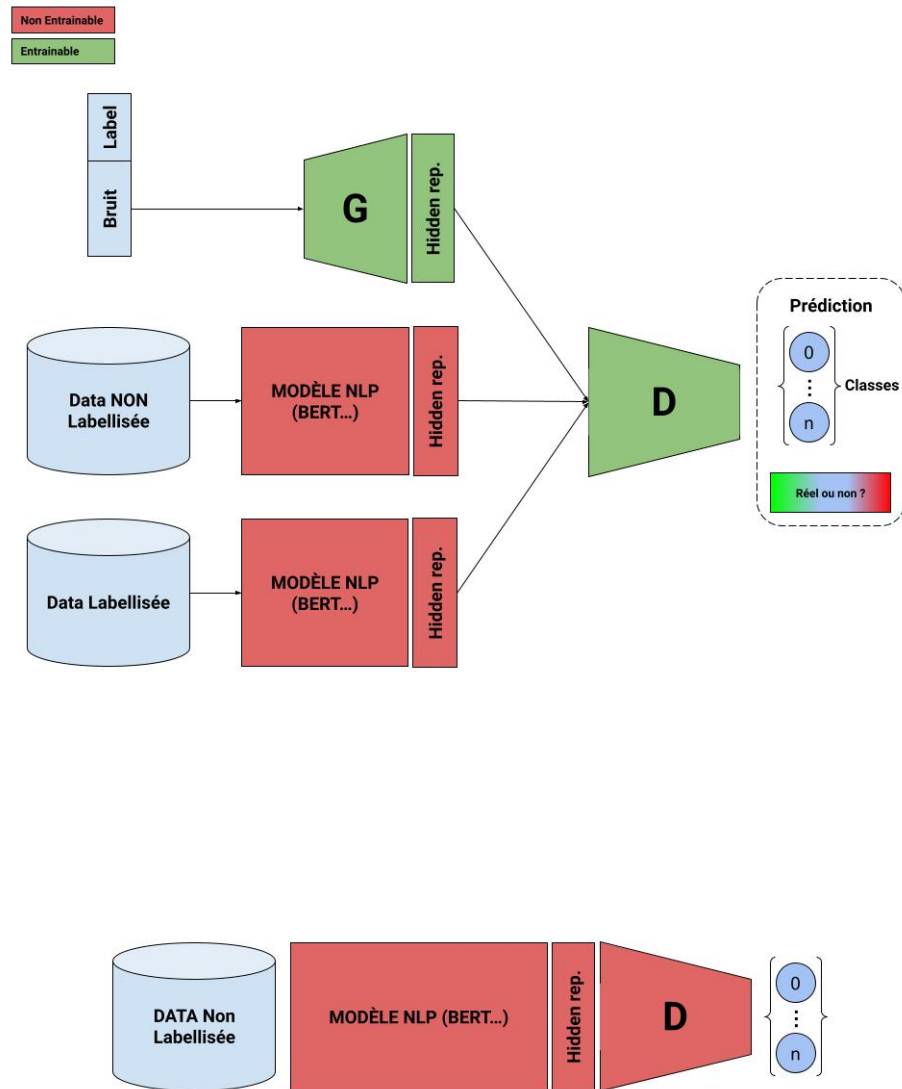
3 Architecture

3.1 Générateur

3.2 Discriminateur



3.3 GAN



4 Entraînement

- somme de 2 bce (label, réalité)

4.1 Supervisé

4.2 Semi-supervisé

Compétition ou coopération ?

Stabilisation du Générateur et Discriminateur

5 Evaluation

5.1 Mesures utilisées

recall, precision, f1 score , APS,AUC au sein d'une cross validation (5folds, équilibre pour l'entraînement)

5.2 Comparaison des mesures

6 Conclusion et perspectives

perspectives : generation au niveau des token, utilisation d'un modèle NLP entraîné sur le corpus que l'on souhaite classer