

Présentation du projet

Objectifs

- Manipuler un modèle de génération «deep»
- Examiner précisément les productions
- Evaluer les métriques d'évaluation

Matériel

- Modèle de génération Yseop sur HuggingFace basé sur T5 -
 - https://huggingface.co/yseop/FNP_T5_D2T_complete
- Des fichiers de test sur le drive (bit.ly/M2-Diderot)
 - 41 fichiers de 100 lignes
- Un article qui décrit le modèle (data2text_EMNLP_2021.pdf)

Contenu des fichiers de test

- Colonne 1: nom de l'intention (DescrValue, DescrVariation, etc.)
 - (pour information)
- Colonne 2: la phrase, telle que trouvée dans un rapport financier
 - *The net costs for tails provisions in 2019 were €8.2 million more than those for 2018.*
- Colonne 3: les triplets correspondants (qui devraient générer la phrase en colonne 2)
 - The net costs for tails provisions | valls | €8.2 million && €8.2 million | dTime | in 2019

Travail attendu

- Choisir un fichier de triplets
- Pour chacun des triplets, faire générer une phrase au modèle
 - Voir mode opératoire sur HuggingFace
- Valider manuellement la fidélité du texte généré:
 - toutes les informations sont-elles présentes (ie pas d'omission d'un fait) ?
 - y a-t-il des informations «en trop» dans la phrase générée (hallucination) ?
 - calcul de score BLEU (*Bilingual evaluation understudy*) entre la phrase générée et l'attendu
- ... pour chacune des 100 phrases
 - Module `nltk.align.bleu_score` en python

Restitution

Triplets	Phrase générée	Phrase attendue	Omission	Hallucination	BLEU
corporation tax rate decFrom 20% corporation tax rate valls 19% 19% dTime 2017	Corporation tax rate was 19% for the year.	Corporation tax rate was 19% in 2017, decreasing from 20% last year	1	1	0.33

Avec un résumé:

- nombre d'erreurs
- nombre de phrases «conformes»
- votre estimation sur BLEU: est-ce que ce score est utile pour évaluer la génération ? (spoiler)

Modalités

- A faire seul ou en groupe (≤ 4 p.)
- Prenez des fichiers différents !
 - Ou si vous prenez le même fichier qu'un autre groupe, vous devez ajouter à votre rapport un calcul d'accord inter-annotateurs (voir [ici](#)) avec l'autre groupe

Des questions ?
