

Sujet :

Réalisation d'une Intelligence Artificielle pour synthétiser les contributions issues du grand débat.

Données :

Les données sont issues du site <https://granddebat.fr/pages/donnees-ouvertes> et l'utilisation des données sont conforme aux dispositions de la licence ouverte d'Etalab.

Problème type

Les contributions, suites de mots du langage naturel, sont ici traitées comme des séries temporelles.

Les résultats des traitements sont des séquences de mots.

Le problème décrit s'inscrit dans le traitement dit "séquence à séquence".

Méthode de Baseline

La méthode Latent Dirichlet Allocation (?) servira de baseline en tant qu'extracteur de thèmes d'un corpus déjà thématisé.

Algorithmes utilisés

L'algorithme proposé est basé sur un modèle encodeur / décodeur de réseau de neurones récurrents avec un mécanisme d'attention.

Le module tensor2tensor est utilisé dans ce cadre.

Il est proposé ici d'utiliser un réseau de Neurones récurrents avec des cellules dites "Gated Recurrent Unit" (GRU)

Contexte technique

Le framework tensorflow est utilisé dans l'environnement Google Computing Platform (GCP).

Documents de référence

- Titre : Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond
Source : <https://arxiv.org/pdf/1602.06023.pdf>
Date de publication : Aout 2016
- Titre : Tensor2Tensor for Neural Machine Translation
Source : <https://arxiv.org/pdf/1803.07416.pdf>
Date de publication : Mars 2018
- Titre : Text Summarization Using Keras Models
Source : <https://hackernoon.com/text-summarization-using-keras-models-366b002408d9>
Date : Décembre 2018

Justification des choix

de par leur efficacité, les réseaux de neurones gagnent en popularité pour le traitement naturel du langage.

La recherche dans le domaine des réseaux de neurones appliqués au traitement du langage progresse à grand rythme en combinant des techniques diverses.

Les données issues du langage naturel sont produites par une intelligence humaine. Elles posent un défi de taille pour le paradigme de l'intelligence artificielle. Cette dernière s'inspire largement des travaux les plus récents en linguistique et en neurobiologie.

Ces travaux s'inscrivent dans la démarche de mettre les techniques issues de la recherche en Intelligence Artificielle au service du bien public.