

Extraction par modèles de langue

Extraction 2

Damien Nouvel

Inalco

Utilisation de modèles de langue

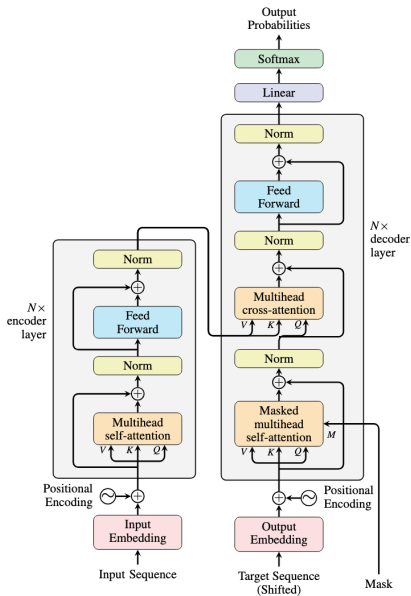
Modèle de langue

- Wikipedia
 - **Modèle de langue** : En traitement automatique des langues, un modèle de langage, modèle de langue ou modèle linguistique est un modèle statistique de la distribution de symboles distincts (lettres, phonèmes, mots) dans une langue naturelle ou un langage formel. Un modèle de langage vise fondamentalement à prédire le mot suivant dans une séquence de mots.
 - **Grand modèle de langage** : modèle de langage possédant un grand nombre de paramètres (généralement plus d'un milliard).

Composantes d'un modèle de langue

- **encodeur** : convertit les entrées textuelles en entrée (symboles, tokens) en paramètres (vecteurs, matrices)
- **décodeur** : convertit les paramètres (vecteurs, matrices) pour générer la sortie (symboles, tokens)

Architecture BERT



Extraction assistée par modèles de langue

Tâches pour l'extraction assistée par modèles de langue

- annotation de textes (apprentissage)
- génération de textes (prompt)

Types de modèles de langues

- monolingue ou multilingue
- généraliste ou spécialisé (affiné) dans le domaine
- apport d'informations complémentaires (RAG)

Évaluation des extractions

Métriques classiques en TAL (jeu de données de référence)

- précision
- rappel
- f-mesure

LLM-as-a-judge

- évaluation par un autre LLM
- définition de critères d'évaluation (prompt)
- évaluation directe ou par comparaisons