

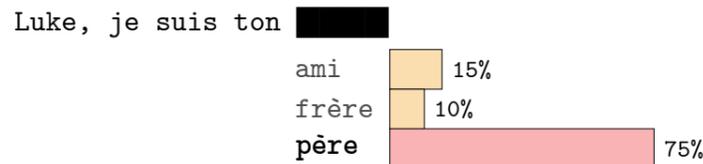
Les Modèles de langage

VOIR LA DÉMONSTRATION

Génération symbolique | phrase : sujet + verbe | mange : verbe | marche : verbe

Des modèles statistiques

Pour donner l'impression de parler, un ordinateur se contente de prolonger des suites de mots. À partir d'un début de texte, il choisit, parmi tous les mots qu'il connaît, celui qui correspond le mieux pour continuer. Précisément, cette correspondance est une probabilité.

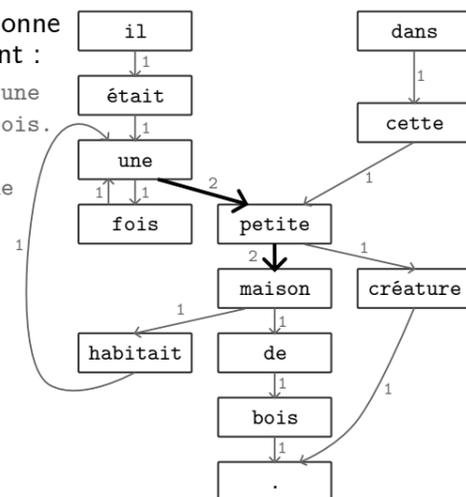


Ces probabilités sont apprises automatiquement par l'ordinateur lors de l'entraînement, une étape où on lui donne une quantité gigantesque de texte à lire. Durant cette lecture, il analyse le contexte dans lesquels les mots apparaissent et en déduit les chances d'apparition des mots.

Par contexte, on entend globalement les mots qui entourent le mot à prédire. Par exemple, pour faire simple, on peut ne considérer que le mot qui précède, ce qui revient à compter le nombre de fois que deux mots se suivent. Étant donné un début de phrase, on peut alors la compléter en prenant parmi les mots qui apparaissent souvent après le dernier mot de la phrase.

Voici ce que cela donne pour le texte suivant :

Il était une fois une petite maison de bois. Dans cette petite maison habitait une petite créature.



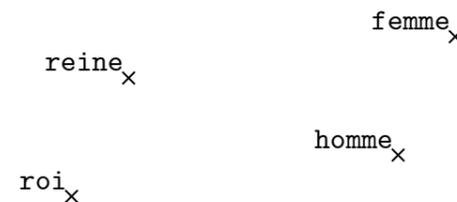
VOIR LA DÉMONSTRATION

Modèle de langage | harry

L'importance du contexte

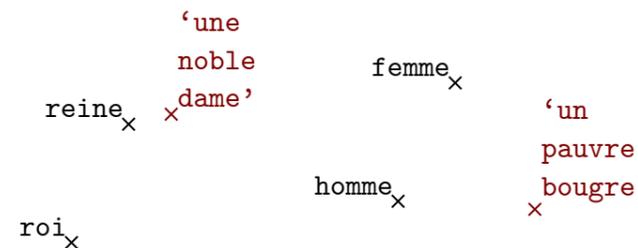
Malheureusement, cette version simpliste du contexte ne permet pas de former des phrases à la syntaxe correcte, encore moins avec un sens. Pour y parvenir, il faut que le contexte contienne beaucoup plus d'information.

L'astuce est de représenter les mots ou leur contexte comme des points sur une carte, bien organisés de façon à refléter des relations sémantiques ou grammaticales.



Remarquez comment l'on transforme « roi » en « reine » de la même façon que l'on transforme « homme » en « femme ».

Si un texte commence par « Une noble dame », on peut trouver sur la carte où placer un point qui représente l'ensemble de ces mots : il se trouvera plutôt vers « reine ». De la même façon, on peut placer l'expression « Un pauvre bougre » plutôt derrière le point « homme ».



Dans ces exemples, la carte n'est qu'en deux dimensions : le genre de haut en bas et la noblesse de gauche à droite. Dans la réalité, les points sont placés sur une carte avec beaucoup plus de dimensions, pour capter toutes les subtilités du langage. Quelques centaines pour les petits modèles jusqu'à plusieurs milliers pour les plus gros.

VOIR LA DÉMONSTRATION

Plongements de mots

Des outils mathématiques

L'ordinateur apprend à placer ces points à l'aide d'outils mathématiques faisant en sorte que des mots qui apparaissent souvent dans les mêmes situations se retrouvent au même endroit. On appelle cela des plongements de mots.

Pour faire le lien entre un point sur la carte des concepts et le prochain mot à générer, on utilise un réseau de neurones. Il s'agit d'un outil mathématique capable d'apprendre à faire des associations entre une entrée (un point sur la carte des concepts) et une sortie (le mot généré).

On entraîne le réseau en lui donnant des entrées et en évaluant sa sortie. Si elle ne nous convient pas, le réseau peut se modifier légèrement de façon à produire la bonne sortie. En répétant ce processus de nombreuses fois, le réseau finit par donner presque toujours la sortie attendue.

La grande force des réseaux de neurones est leur capacité à généraliser ce qu'ils apprennent. Ainsi, même si on leur présente une entrée qu'ils n'ont jamais vu, ils vont pouvoir produire une sortie cohérente.

Les récentes avancées en IA sont dues à une nouvelle architecture de réseau de neurones, publiée en 2017 par Google, baptisés les transformeurs. Ces réseaux portent une attention particulière à la façon dont le contexte influe sur le sens des mots. Par exemple, le point correspondant à « la prise du château » ne sera pas placé au même endroit sur la carte des concepts si l'on parlait d'une bataille médiévale ou d'une alimentation électrique.

VOIR LA DÉMONSTRATION

Génération de texte | système politique plus équilibré et renforçant les droits de

En réalité, les modèles ne génèrent pas le texte mot à mot mais jeton par jeton, un jeton étant un groupe de lettres de taille variable. Cela permet notamment aux modèles de travailler avec des mots qu'ils n'ont jamais rencontrés, du moment qu'ils sont formés de lettres connues.

Il était une fois une petite maison de bois.

Dans cette petite maison habitait une petite créature.