

MODÈLES PRÉ-ENTRAÎNÉ DE PROMPT ENGINEERING COMPARATIVEMENT AU MACHINE LEARNING TRADITIONNEL

A

u cours des dernières années, les termes intelligence artificielle[1] et apprentissage machine aussi appelé Machine Learning (ML) sont devenus prédominants.

Cette technologie a révolutionné notre manière d'aborder la résolution de problèmes dans divers secteurs, allant de la santé à la finance, et même dans nos interactions quotidiennes avec la technologie, que ce soit au travail[2] ou dans notre vie personnelle.

Récemment, l'avènement de GPT a fait émerger un nouveau paradigme : nous passons de modèles traditionnels de ML, méticuleusement élaborés et entraînés pour des tâches spécifiques, à des modèles pré-entraînés capables de s'adapter à une multitude d'applications grâce au prompt engineering. Comprendre ces deux approches est essentiel pour débloquer leurs potentiels dans le monde d'aujourd'hui axé sur les données.

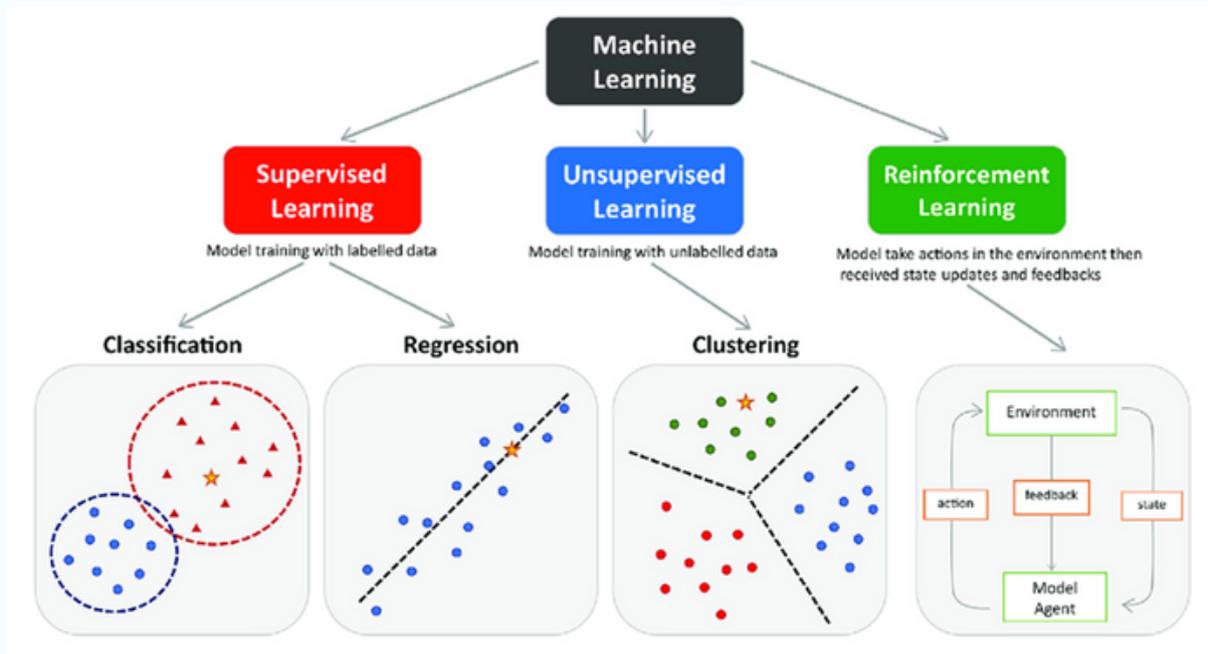


Le Machine Learning traditionnel

Le ML traditionnel est une méthode où les algorithmes apprennent à partir de données pour faire des prédictions ou prendre des décisions. Ce processus implique six étapes principales :

- **Collecte de données**, où les données sont récoltées à partir de diverses sources.
- **Prétraitement des données**, où les données brutes sont nettoyées et structurées.
- **Annotation des données**, où les données reçoivent des étiquettes à prédire.
- **Ingénierie des caractéristiques**, où les aspects les plus pertinents des données sont identifiés et mis en évidence.
- **Entraînement du modèle**, où l'algorithme apprend à partir de ces données préparées pour faire des prédictions futures.
- **Évaluation du modèle**, où le modèle est évalué sur de nouvelles données pour valider sa performance.

Figure 1: Différentes techniques de Machine Learning[1]



Par exemple, si la tâche consiste à traduire du texte de l'anglais vers le français, voici le processus. Tout d'abord, nous devons sélectionner un corpus bilingue dans les deux langues. Ensuite nous pouvons réaliser un prétraitement, tel que la suppression des liens. Ensuite, nous annotons, ce qui dans ce cas signifie appairer chaque phrase anglaise avec sa traduction française connue. Après cela, nous pouvons nous lancer dans l'ingénierie des caractéristiques, comme coder les mots en utilisant des « word embeddings »[3]. Enfin, nous entraînons le modèle et évaluons ses performances.

La force du ML traditionnel réside dans sa spécificité. Les modèles sont conçus sur mesure pour la tâche à accomplir. Par exemple, prédire les tendances du marché boursier ou diagnostiquer des maladies à partir d'images médicales. Cette personnalisation nécessite une compréhension approfondie du problème. Cela implique un investissement important dans la collecte et le traitement des données. Mais cela complexifie aussi le développement et le paramétrage du modèle.

Le Prompt Engineering avec des Modèles Pré-entraînés

Le pré-entraînement est une approche en ML qui a émergé avec l'essor de grands modèles tels que GPT-3, BERT et autres[4]. Ces modèles ont été entraînés sur d'énormes ensembles de données, englobant un large éventail de connaissances et de structures linguistiques. Dans l'approche traditionnelle, un modèle est construit à partir de zéro. Grâce au modèle pré-entraînés nous avons une base de connaissances. Ainsi des techniques comme le prompt engineering peut être appliquées avec peu d'effort.



Le prompt engineering[5] implique de créer des instructions (prompt) pour ces modèles pré-entraînés. Le but étant de les guider pour produire la sortie souhaitée. Ainsi, la création d'un modèle spécifique à un problème modèle est remplacé par la définition d'un prompt. Nous évitons donc les tâches difficiles tel que le réglage des hyperparamètres ou le calcul coûteux de l'entraînement. Cela est remplacé par la définition méticuleuse du prompt en utilisant les bons exemples, contextes et directives. Néanmoins, cette tâche nécessite une connaissance approfondie du fonctionnement interne du modèle pré-entraîné et du problème à résoudre. Il faut garantir que les exemples utilisés dans le prompt soient aussi représentatifs et diversifiés que possible, reflétant au plus près la distribution réelle des paires entrée-sortie qui définissent le problème.

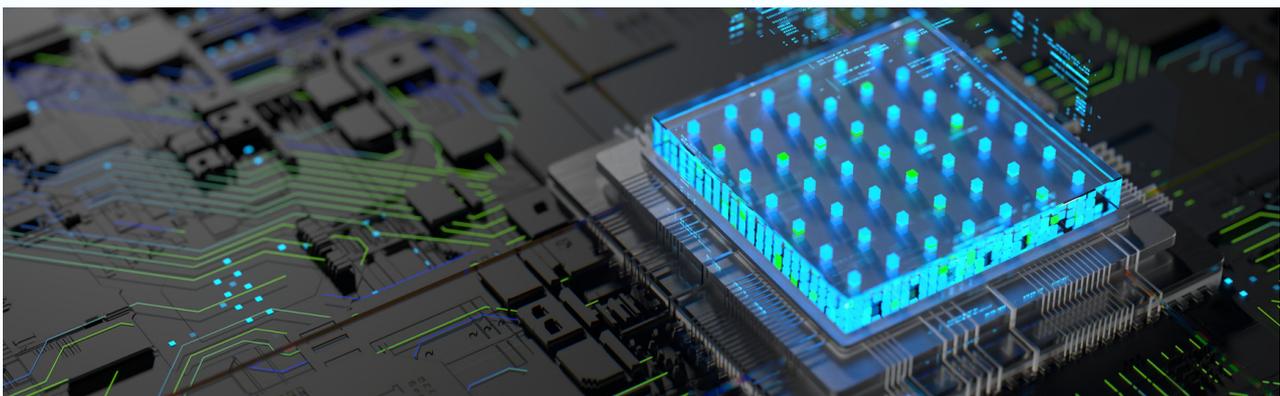
Dans ce cas-ci, la tâche de traduire du texte de l'anglais au français est bien plus simple à réaliser. Il nous suffit de nous munir d'un modèle pré-entraîné tel que GPT-4 et de lui fournir un prompt. Ce prompt contient 1) la définition de la tâche, 2) quelques exemples de phrases ainsi que leurs traductions, et enfin 3) la phrase actuelle à traduire (voir dans l'illustration ci-dessus). Et c'est tout ! GPT-4 comprends le concept de la traduction grâce à son pré-entraînement. Il suffit de lui demander correctement avec un prompt.

Ainsi, le prompt engineering tire parti de l'apprentissage étendu que ces modèles ont déjà subi. Ils rendent possible leur application à une variété de tâches sans nécessiter toutes les complexités du ML traditionnel. De la génération créative à la résolution de tâches complexes, la polyvalence de du prompt engineering est impressionnante.

Grâce à cette approche, la charge de la collecte de données, du prétraitement et de l'annotation est considérablement réduite. Ces étapes constituent généralement une partie significative de la tâche de modélisation en termes de temps et de coût. L'utilisation de modèles pré-entraînés allège considérablement ces exigences, simplifiant le processus de déploiement de solutions de ML.



Figure 2: Exemple de prompt



Avantages et Inconvénients

Machine Learning Traditionnel

Avantages : Le ML traditionnel est hautement personnalisable. Il peut être adapté à des tâches spécifiques avec un haut degré de précision. De plus, nous collectons et prétraitons les données nous-mêmes, ce qui apporte plus de contrôle et de validation sur ce qui est ingéré par le modèle.

Inconvénients : Toutefois, entraîner un modèle est un processus gourmand en ressources qui nécessite des investissements importants. La collecte, le prétraitement et l'annotation des données sont des étapes coûteuses et représentent généralement la majorité du temps consacré aux projets de ML. Enfin, la formation d'un modèle peut également être extrêmement coûteuse et prends aussi beaucoup de temps.

Modèles Pré-entraînés

Avantages : Le Prompt Engineering sur des modèles pré-entraînés limite la nécessité de traiter des données. Cela conduit à un déploiement de modèle moins coûteux et plus rapide. Cela signifie également que ces modèles peuvent s'adapter plus rapidement à des changements dans la spécification du problème.

Inconvénients : Le principal problème avec les modèles pré-entraînés est que nous avons peu de contrôle sur les sources de données et les méthodes de prétraitement utilisées pour leur création. Ainsi, pour des tâches sensibles telles que les problèmes médicaux ou financiers, ces modèles peuvent être moins appropriés.

Conclusion

Dans le futur, le Prompt Engineering pourrait devenir omniprésent dans les applications industrielles, compte tenu de sa rapidité de déploiement et de son coût réduit. Nous pourrions assister à l'émergence de techniques de prompt engineering[5] plus sophistiquées, rendant les modèles pré-entraînés encore plus polyvalents. Parallèlement, le ML traditionnel continuera de progresser. En particulier dans les domaines nécessitant une extrême précision et une supervision des données.

En conclusion, les deux méthodes sur des modèles pré-entraînés offrent tous deux des avantages et des inconvénients uniques. Le choix entre ces approches doit être guidé par les exigences spécifiques de la tâche à accomplir. Il faut tenir compte de facteurs tels que la disponibilité des données et les contraintes de ressources. À mesure que le domaine évolue, la synergie de ces deux approches pourrait ouvrir la voie à des solutions plus innovantes et efficaces dans le domaine de l'intelligence artificielle.

Bibliographie

- 1.Poumay (Ph.D.) J. Comprendre l'intelligence artificielle. The Thought Process. Publié le 10 Février, 2022. <https://thethoughtprocess.xyz/science/informatique/2022/02/10/comprendre-lintelligence-artificielle/>
- 2.Poumay (Ph.D.) J. L'intelligence artificielle : ses développements passés et son impact futur sur le travail. The Thought Process. Publié le 21 Septembre, 2023. <https://thethoughtprocess.xyz/non-classe-fr/2023/09/21/lintelligence-artificielle-ses-developpements-passes-et-son-impact-futur-sur-le-travail/>
- 3.Poumay (Ph.D.) J. Embeddings : lier les mots aux nombres. The Thought Process. Publié le 19 Juin, 2020. <https://thethoughtprocess.xyz/science/informatique/2020/06/19/embeddings-lier-les-mots-aux-nombres/>
- 4.Poumay (Ph.D.) J. Notes de lecture #6 : Modèles de langage. The Thought Process. Publié le 18 Août, 2022. <https://thethoughtprocess.xyz/science/informatique/2022/08/18/notes-de-lecture-6-modeles-de-langage/>
- 5.Poumay (Ph.D.) J. Le Prompt Engineering : Comment parler avec les IA. The Thought Process. Publié le 22 Novembre, 2023. <https://thethoughtprocess.xyz/non-classe-fr/2023/11/22/le-prompt-engineering-comment-parler-aux-ia-comme-chatgpt/>