

Les modèles de fondation de l'IA générative

Les modèles de fondation de l'IA générative sont des modèles d'intelligence artificielle capables de générer du contenu à partir de données brutes, comme du texte, des images, des sons ou des vidéos. Ces modèles utilisent des techniques d'apprentissage profond pour apprendre les caractéristiques et les relations des données, et pour produire de nouveaux exemples qui respectent les mêmes règles. Les modèles de fondation de l'IA générative sont considérés comme une avancée majeure dans le domaine de l'IA, car ils permettent de créer des applications innovantes et créatives dans divers domaines.

Parmi les modèles de fondation de l'IA générative, les modèles de génération d'images sont ceux qui visent à créer des images réalistes et diversifiées à partir de données brutes ou conditionnées. Par exemple, un modèle de génération d'images peut prendre une description textuelle comme entrée et produire une image qui correspond au texte. Ou bien, il peut prendre une image existante et la modifier selon un style, une couleur ou un attribut donné. Les modèles de génération d'images peuvent être basés sur des réseaux de neurones convolutifs (CNN), qui sont capables de traiter et de reconnaître les caractéristiques visuelles des images, ou sur des réseaux antagonistes génératifs (GAN), qui sont des systèmes composés de deux réseaux concurrents : un générateur qui crée les images et un discriminateur qui les évalue. Les modèles de génération d'images offrent de nombreuses possibilités d'application, comme la synthèse de visages, le dessin assisté, la colorisation, la restauration ou la super-résolution d'images.

Les LLM, ou masters of laws, sont des diplômes de droit de niveau supérieur qui permettent aux étudiants et aux professionnels d'approfondir leurs connaissances juridiques dans un domaine spécifique. Les LLM sont généralement proposés par des universités prestigieuses, et sont reconnus internationalement comme des qualifications académiques de haut niveau. Les LLM sont souvent destinés aux personnes qui souhaitent se spécialiser dans un domaine du droit, comme le droit international, le droit des affaires, le droit fiscal, le droit de la propriété intellectuelle ou le droit des droits de l'homme. Les LLM peuvent également servir à acquérir une expérience juridique dans un pays ou une région différente, ou à se préparer à un doctorat en droit. Les LLM offrent de nombreux avantages aux étudiants et aux professionnels, comme l'amélioration de leurs compétences juridiques, la mise en réseau avec des pairs et des experts, l'ouverture de nouvelles opportunités de carrière ou de recherche, ou l'augmentation de leur crédibilité et de leur réputation.

Le service Azure OpenAI : un outil puissant pour l'intelligence artificielle

Un document qui présente le service Azure OpenAI, ses caractéristiques, ses bénéfices et ses applications.

Qu'est-ce que le service Azure OpenAI ?

Le service Azure OpenAI est un service cloud qui permet aux développeurs et aux chercheurs d'accéder à la plateforme OpenAI, une initiative qui vise à créer une intelligence artificielle générale (AGI) bénéfique pour l'humanité. Le service Azure OpenAI offre la possibilité d'utiliser des modèles d'IA pré-entraînés ou personnalisés, de les déployer sur le cloud ou en périphérie, et de les intégrer à des applications et des scénarios variés.

Comment fonctionne le service Azure OpenAI ?

Le service Azure OpenAI repose sur deux composants principaux : le portail Azure OpenAI et les API Azure OpenAI. Le portail Azure OpenAI est une interface web qui permet de créer, de gérer et de surveiller les ressources du service Azure OpenAI, telles que les abonnements, les clés, les modèles, les points de terminaison et les requêtes. Les API Azure OpenAI sont des interfaces de programmation qui permettent d'interagir avec les modèles d'IA du service Azure OpenAI, tels que GPT-3, Codex, DALL-E ou CLIP. Les API Azure OpenAI sont accessibles via des requêtes HTTP ou des SDK dans différents langages de programmation, tels que Python, Java, C#, Node.js ou Ruby.

Quels sont les avantages du service Azure OpenAI ?

Le service Azure OpenAI présente plusieurs avantages, parmi lesquels :

- Il permet d'accéder à des modèles d'IA de pointe, capables de réaliser des tâches complexes et variées, telles que la génération de texte, la compréhension du langage naturel, la vision par ordinateur, la synthèse vocale ou la création de contenu multimédia.
- Il offre une grande flexibilité, en permettant de choisir entre des modèles pré-entraînés ou personnalisés, et de les adapter aux besoins spécifiques de chaque projet ou domaine.
- Il garantit une haute performance, en s'appuyant sur l'infrastructure cloud de Microsoft Azure, qui assure une disponibilité, une scalabilité et une sécurité optimales.
- Il facilite l'intégration, en proposant des API simples et des SDK dans différents langages de programmation, qui permettent de connecter les modèles d'IA du service Azure OpenAI à des applications et des scénarios existants ou nouveaux.

Quels sont les cas d'utilisation du service Azure OpenAI ?

Le service Azure OpenAI peut être utilisé pour de nombreux cas d'utilisation, dans différents domaines et secteurs d'activité. Voici quelques exemples :

- Dans le domaine de l'éducation, le service Azure OpenAI peut être utilisé pour créer des outils pédagogiques, tels que des assistants de rédaction, des générateurs de questions, des correcteurs automatiques ou des tuteurs virtuels.
- Dans le domaine de la santé, le service Azure OpenAI peut être utilisé pour créer des outils d'aide au diagnostic, à la prescription, à la recherche ou à la prévention, en exploitant les données médicales et les connaissances scientifiques.
- Dans le domaine du divertissement, le service Azure OpenAI peut être utilisé pour créer des outils de génération de contenu, tels que des scénarios, des dialogues, des personnages, des musiques ou des images.
- Dans le domaine du commerce, le service Azure OpenAI peut être utilisé pour créer des outils d'optimisation, tels que des recommandations personnalisées, des prévisions de ventes, des analyses de marché ou des chatbots.

Comment mettre en œuvre le service Azure OpenAI ?

Pour mettre en œuvre le service Azure OpenAI, il faut suivre les étapes suivantes :

- Créer un compte Microsoft Azure, si ce n'est pas déjà fait, et se connecter au portail Azure OpenAI.
- Souscrire à un abonnement au service Azure OpenAI, en choisissant le niveau de tarification adapté à ses besoins.
- Créer une clé d'API, qui permettra d'authentifier les requêtes au service Azure OpenAI.
- Choisir un modèle d'IA, parmi ceux proposés par le service Azure OpenAI, ou créer son propre modèle, en utilisant les outils de formation et de réglage fournis par le service Azure OpenAI.
- Créer un point de terminaison, qui permettra d'exposer le modèle d'IA sur le cloud ou en périphérie, et de le rendre accessible via une URL.
- Envoyer des requêtes au point de terminaison, en utilisant les API Azure OpenAI ou les SDK dans le langage de programmation de son choix.
- Analyser les résultats des requêtes, en utilisant les outils de suivi et de statistiques fournis par le portail Azure OpenAI.

Les services cognitifs Azure

Les services cognitifs Azure sont un ensemble de services cloud qui permettent aux développeurs de créer des applications intelligentes, en utilisant les capacités d'intelligence artificielle (IA) et de machine learning (ML) fournies par Microsoft. Les services cognitifs Azure couvrent différents domaines, tels que la vision, le langage, la recherche, la décision et la synthèse vocale. Ils offrent des API, des SDK et des outils faciles à utiliser, pour intégrer les fonctionnalités d'IA et de ML dans les applications web, mobiles ou de bureau.

La vision

Le domaine de la vision regroupe les services cognitifs Azure qui permettent aux applications de comprendre et d'analyser le contenu visuel, comme les images, les vidéos ou les flux en direct. Les services cognitifs Azure de la vision comprennent :

- Le service Vision par ordinateur, qui fournit des fonctions de détection, d'analyse, de description et de balisage des images, ainsi que de reconnaissance optique de caractères (OCR) et de génération de vignettes.
- Le service Reconnaissance faciale, qui permet d'identifier, de vérifier, de rechercher et de regrouper des visages dans des images ou des vidéos, ainsi que de détecter des attributs faciaux, comme l'âge, le genre, l'émotion ou les accessoires.
- Le service Analyseur vidéo, qui offre des capacités d'analyse avancée des vidéos, comme la détection et le suivi des visages, des objets, des logos, des scènes ou des activités, ainsi que la transcription, la traduction et la synthèse vocale des pistes audio.
- Le service Form Recognizer, qui permet d'extraire et de structurer les données clés contenues dans des formulaires papier ou numériques, comme des factures, des reçus, des chèques ou des cartes d'identité.
- Le service Custom Vision, qui permet de personnaliser les modèles de vision par ordinateur et de reconnaissance faciale, en les entraînant avec ses propres données, afin de répondre à des besoins spécifiques.

Le langage

Le domaine du langage regroupe les services cognitifs Azure qui permettent aux applications de comprendre et de manipuler le langage naturel, c'est-à-dire le langage parlé ou écrit par les humains. Les services cognitifs Azure du langage comprennent :

- Le service Analyse de texte, qui fournit des fonctions d'analyse syntaxique, sémantique et sentimentale des textes, ainsi que de détection des entités nommées, des mots-clés, des liens ou des langues.
- Le service Reconnaissance vocale, qui permet de transcrire le discours en texte, en reconnaissant les mots prononcés, le locuteur, l'intention ou les commandes.
- Le service Traduction vocale, qui permet de traduire le discours d'une langue à une autre, en conservant la voix et le ton du locuteur.

- Le service Génération de langage naturel, qui permet de produire du texte à partir de données structurées, comme des tableaux, des graphiques ou des images, en utilisant des modèles pré-entraînés ou personnalisés.
- Le service Linguistic Analysis, qui fournit des fonctions d'analyse morphologique, syntaxique et sémantique des textes, en utilisant des règles et des dictionnaires linguistiques.
- Le service QnA Maker, qui permet de créer des agents conversationnels capables de répondre aux questions des utilisateurs, en se basant sur des sources de connaissances existantes, comme des documents, des sites web ou des FAQ.

Les bases de données vectorielles

Les bases de données vectorielles sont des systèmes de gestion de données qui stockent et manipulent des vecteurs, c'est-à-dire des séquences de nombres réels qui représentent des entités ou des objets. Ces vecteurs peuvent être obtenus à partir de diverses sources, comme des images, des textes, des sons, des vidéos, ou des signaux biométriques. Les bases de données vectorielles permettent d'effectuer des opérations efficaces sur les vecteurs, comme la recherche par similarité, le calcul de distances, le regroupement, la classification, ou la recommandation.

Les bases de données vectorielles sont utilisées dans de nombreux domaines d'application, comme la vision par ordinateur, le traitement du langage naturel, la bioinformatique, ou le commerce électronique. Elles offrent plusieurs avantages par rapport aux bases de données traditionnelles, comme la réduction de la dimensionnalité, la généralisation, la robustesse au bruit, ou la facilité d'interprétation. Elles présentent aussi des défis spécifiques, comme le choix des méthodes de vectorisation, l'indexation des vecteurs de grande dimension, ou la garantie de la sécurité et de la confidentialité des données.

Dans cet article, nous allons présenter les principes et les caractéristiques des bases de données vectorielles, ainsi que les principales techniques et outils utilisés pour les mettre en œuvre. Nous allons également illustrer les applications et les bénéfices des bases de données vectorielles à travers quelques exemples concrets.

Principes et caractéristiques des bases de données vectorielles

Une base de données vectorielle est un système de gestion de données qui stocke et manipule des vecteurs, c'est-à-dire des séquences ordonnées de nombres réels qui représentent des entités ou des objets. Un vecteur peut avoir une dimension fixe ou variable, selon le type de données qu'il encode. Par exemple, un vecteur peut représenter les pixels d'une image, les mots d'un texte, les fréquences d'un son, ou les caractéristiques d'un produit.

Les vecteurs sont généralement obtenus à partir de données brutes, comme des fichiers multimédias, des documents, ou des flux de données, en utilisant des méthodes de vectorisation, qui transforment les données en vecteurs selon certains critères. Il existe différentes méthodes de vectorisation, selon le domaine et le but de l'application. Par exemple, on peut utiliser des méthodes basées sur des descripteurs, qui extraient des caractéristiques pertinentes des données, comme les couleurs, les formes, ou les textures pour les images, ou les mots, les phrases, ou les concepts pour les textes. On peut aussi utiliser des méthodes basées sur des modèles, qui apprennent des représentations vectorielles à partir de données massives, en utilisant des techniques d'apprentissage automatique, comme les réseaux de neurones, les modèles probabilistes, ou les modèles géométriques.

Le deep learning pour la classification d'images et la détection d'objets

Introduction

La classification d'images et la détection d'objets sont deux tâches importantes de la vision par ordinateur, qui consistent à identifier et à localiser les objets présents dans une image. Ces tâches ont de nombreuses applications pratiques, comme la reconnaissance faciale, la surveillance, la robotique, ou la médecine. Le deep learning, qui est une branche de l'apprentissage automatique basée sur des modèles profonds, a révolutionné le domaine de la vision par ordinateur, en offrant des performances supérieures aux méthodes classiques, basées sur des caractéristiques manuelles ou des modèles simples. Dans ce document, nous allons présenter les principes et les caractéristiques du deep learning pour la classification d'images et la détection d'objets, ainsi que les principaux modèles et techniques utilisés.

Principes du deep learning

Le deep learning est une méthode d'apprentissage automatique qui utilise des réseaux de neurones artificiels, qui sont des structures composées de couches de neurones interconnectés, qui reçoivent des données en entrée et produisent des sorties en fonction de paramètres appris. Un réseau de neurones peut être vu comme une fonction non linéaire complexe, qui approxime une relation entre les données d'entrée et les données de sortie. Le réseau de neurones apprend ses paramètres à partir d'un ensemble d'exemples, appelé jeu de données, en utilisant un algorithme d'optimisation, généralement basé sur la descente de gradient, qui minimise une fonction de coût, qui mesure l'erreur entre les sorties prédites et les sorties attendues. Le réseau de neurones est dit profond lorsque le nombre de couches est élevé, ce qui lui permet de capturer des niveaux de plus en plus abstraits et complexes de représentation des données. Le deep learning se distingue des autres méthodes d'apprentissage automatique par sa capacité à apprendre directement à partir des données brutes, sans avoir besoin de définir des caractéristiques spécifiques au domaine.

Classification d'images

La classification d'images est la tâche qui consiste à attribuer une étiquette à une image, parmi un ensemble de classes prédéfinies. Par exemple, on peut vouloir classer une image selon qu'elle contient un chat, un chien, ou un oiseau. La classification d'images peut être vue comme un problème de classification supervisée, où l'on dispose d'un jeu de données d'images annotées avec leurs classes respectives, et où l'on cherche à apprendre un modèle qui prédit la classe d'une image inconnue. Le deep learning a permis de réaliser des progrès significatifs dans la classification d'images, en utilisant des réseaux de neurones convolutifs, qui sont des modèles spécialisés pour le traitement des données visuelles.

Réseaux de neurones convolutifs

Un réseau de neurones convolutif (CNN) est un type de réseau de neurones qui utilise des opérations de convolution, qui consistent à appliquer des filtres linéaires sur les données d'entrée, pour extraire des caractéristiques locales et invariantes. Un CNN est composé de plusieurs couches, qui peuvent être de différents types, comme des couches de convolution, des couches d'activation, des couches de regroupement, ou des couches entièrement connectées. Les couches de convolution appliquent des filtres sur les données d'entrée, pour produire des cartes d'activation, qui représentent la réponse du filtre à une région de l'entrée. Les filtres sont appris par le réseau, et sont généralement de petite taille, comme 3x3 ou 5x5 pixels. Les couches d'activation appliquent une fonction non linéaire aux cartes d'activation, pour introduire de la non-linéarité dans le modèle. Les fonctions d'activation les plus courantes sont la fonction sigmoïde, la fonction tangente hyperbolique, ou la fonction ReLU. Les couches de regroupement réduisent la dimensionnalité des cartes d'activation, en appliquant une opération d'agrégation, comme le maximum, la moyenne, ou la norme, sur des régions voisines. Les couches de regroupement permettent de réduire le nombre de paramètres, de limiter le surapprentissage, et d'augmenter l'invariance à la translation. Les couches entièrement connectées sont des couches classiques de réseau de neurones, qui relient tous les neurones d'une couche à ceux de la couche suivante. Les couches entièrement connectées permettent de combiner les caractéristiques extraites par les couches précédentes, et de réaliser la tâche finale, comme la classification. Un CNN typique pour la classification d'images est composé de plusieurs blocs de couches de convolution, d'activation, et de regroupement, suivis de quelques couches entièrement connectées. La dernière couche entièrement connectée produit un vecteur de scores, qui représente la probabilité d'appartenance à chaque classe. La fonction de coût utilisée est généralement l'entropie croisée, qui mesure la divergence entre la distribution des scores et la distribution des classes réelles. Le réseau apprend ses paramètres en utilisant la rétropropagation du gradient, qui consiste à calculer le gradient de la fonction de coût par rapport aux paramètres, en utilisant la règle de la chaîne, et à mettre à jour les paramètres en suivant la direction opposée au gradient.

Modèles et techniques

Les CNN ont été popularisés par le modèle AlexNet, qui a remporté le défi ImageNet en 2012, en battant les méthodes classiques basées sur des caractéristiques manuelles. ImageNet est un jeu de données d'images de grande taille, qui contient plus de 14 millions d'images, réparties en plus de 20 000 classes. Le modèle AlexNet est composé de huit couches, dont cinq couches de convolution et trois couches entièrement connectées, et utilise la fonction ReLU comme fonction d'activation, ainsi que des techniques de régularisation, comme le dropout et la normalisation des lots. Le modèle AlexNet a permis de réduire l'erreur de classification sur ImageNet de plus de 10 points.

Depuis, de nombreux modèles de CNN ont été proposés, en augmentant la profondeur, la largeur, ou la complexité du réseau, pour améliorer les performances sur la classification d'images. Parmi ces modèles, on peut citer VGG, qui utilise des filtres de taille 3x3 et des couches de convolution successives, Inception, qui utilise des blocs modulaires avec des branches parallèles de différentes tailles, ResNet, qui utilise des connexions résiduelles qui permettent de sauter des couches, DenseNet, qui utilise des connexions denses qui relient toutes les couches entre elles, ou encore EfficientNet, qui utilise une recherche automatique de l'architecture optimale du réseau. Ces modèles ont atteint des niveaux de performance proches ou supérieurs à ceux des humains sur la classification d'images.

Outre les modèles, plusieurs techniques ont été développées pour améliorer le deep learning pour la classification d'images. Parmi ces techniques, on peut citer l'augmentation des données, qui consiste à appliquer des transformations aléatoires sur les images d'entraînement, comme le recadrage, le retournement, la rotation, le bruit, ou la variation des couleurs, pour augmenter la diversité et la robustesse du modèle. On peut aussi citer le transfert d'apprentissage, qui consiste à utiliser un modèle pré-entraîné sur un jeu de données de grande taille, comme ImageNet, et à l'adapter à un jeu de données de plus petite taille, en conservant les couches de convolution, qui captent des caractéristiques génériques, et en ré-entraînant les couches entièrement connectées, qui captent des caractéristiques spécifiques. On peut également citer l'apprentissage par auto-étiquetage, qui consiste à utiliser un modèle pré-entraîné pour générer des étiquettes pour des images non annotées, et à les utiliser comme données d'entraînement supplémentaires.

Détection d'objets

La détection d'objets est la tâche qui consiste à identifier et à localiser les objets présents dans une image, en leur attribuant une classe et une boîte englobante. Par exemple, on peut vouloir détecter les personnes, les voitures, ou les animaux dans une image. La détection d'objets peut être vue comme un problème de régression supervisée, où l'on dispose d'un jeu de données d'images annotées avec les classes et les boîtes englobantes des objets, et où l'on cherche à apprendre un modèle qui prédit les classes et les boîtes englobantes des objets dans une image inconnue. Le deep learning a permis de réaliser des progrès significatifs dans la détection d'objets, en utilisant des modèles basés sur des CNN, qui peuvent être classés en deux catégories : les modèles à deux étapes, qui utilisent une région d'intérêt pour localiser les objets, et les modèles à une étape, qui détectent les objets directement à partir de l'image entière.

Modèles à deux étapes

Les modèles à deux étapes sont des modèles qui utilisent une région d'intérêt (RoI), qui est une sous-région de l'image qui contient potentiellement un objet, pour localiser les objets. Les modèles à deux étapes sont composés de deux parties : un extracteur de caractéristiques, qui utilise un CNN pour extraire des caractéristiques de l'image entière, et un détecteur de région, qui utilise un autre CNN pour détecter les objets à l'intérieur de chaque RoI. Les modèles à deux étapes sont généralement plus précis que les modèles à une étape, mais aussi plus lents.

Le modèle R-CNN est le premier modèle à deux étapes, qui utilise un algorithme de segmentation, comme le recherche sélective, pour générer environ 2000 RoI par image, et qui utilise un CNN pré-entraîné, comme VGG ou AlexNet, pour extraire des caractéristiques de chaque RoI. Le modèle R-CNN utilise ensuite un classifieur SVM, qui utilise les caractéristiques extraites comme entrée, pour prédire la classe de chaque RoI, et un régresseur linéaire, qui utilise également les caractéristiques extraites comme entrée, pour affiner les boîtes englobantes de chaque RoI. Le modèle R-CNN a permis de réduire l'erreur de détection sur le jeu de données PASCAL VOC, qui est un jeu de données de référence pour la détection d'objets, qui contient environ 20 000 images, réparties en 20 classes. Cependant, le modèle R-CNN est très lent, car il nécessite d'exécuter le CNN sur chaque RoI individuellement, et utilise plusieurs composants distincts, qui doivent être entraînés séparément.

Le modèle Fast R-CNN améliore le modèle R-CNN, en utilisant une couche de regroupement de région (RoIPool), qui permet de réduire le nombre d'exécutions du CNN, en extrayant les caractéristiques de

l'image entière une seule fois, et en les partageant entre les différentes RoI. Le modèle Fast R-CNN utilise également un seul réseau, qui combine le classifieur et le régresseur, et qui peut être entraîné de bout en bout. Le modèle Fast R-CNN est plus rapide et plus précis que le modèle R-CNN, mais il dépend toujours d'un algorithme externe pour générer les RoI, ce qui limite sa performance et sa flexibilité.

Le modèle Faster R-CNN résout le problème du modèle Fast R-CNN, en utilisant un réseau de proposition de région (RPN), qui est un autre CNN, qui prend les caractéristiques extraites par l'extracteur de caractéristiques comme entrée, et qui prédit les RoI à partir de plusieurs ancres, qui sont des boîtes englobantes de tailles et de ratios variés, situées à différentes positions de l'image. Le modèle Faster R-CNN utilise ensuite la même couche RoIPool et le même détecteur de région que le modèle Fast R-CNN, pour prédire la classe et la boîte englobante de chaque RoI. Le modèle Faster R-CNN est plus rapide et plus précis que le modèle Fast R-CNN, car il utilise un seul réseau, qui peut être entraîné de bout en bout, et qui génère les RoI de manière adaptative.

Modèles à une étape

Les modèles à une étape sont des modèles qui détectent les objets directement à partir de l'image entière, sans utiliser de RoI. Les modèles à une étape sont composés d'un seul réseau, qui utilise un CNN pour extraire des caractéristiques de l'image, et qui prédit la classe et la boîte englobante de chaque objet à partir de plusieurs ancres, comme le modèle Faster R-CNN. Les modèles à une étape sont généralement plus rapides que les modèles à deux étapes, mais aussi moins précis.

Le modèle YOLO est le premier modèle à une étape, qui divise l'image en une grille de cellules, et qui prédit la classe et la boîte englobante de chaque objet à partir de plusieurs ancres par cellule. Le modèle YOLO utilise un seul réseau, qui combine l'extracteur de caractéristiques et le détecteur d'objets, et qui peut être entraîné de bout en bout. Le modèle YOLO est très rapide, car il traite l'image entière en une seule fois, mais il a tendance à manquer les petits objets, ou les objets proches les uns des autres, car il limite le nombre d'objets par cellule.

Le modèle SSD améliore le modèle YOLO, en utilisant plusieurs cartes de caractéristiques à différentes échelles, pour détecter les objets de différentes tailles, et en utilisant plus d'ancres par cellule, pour détecter les objets de différents ratios. Le modèle SSD utilise également un seul réseau, qui combine l'extracteur de caractéristiques et le détecteur d'objets, et qui peut être entraîné de bout en bout. Le modèle SSD est plus précis que le modèle YOLO, car il capture mieux les objets de différentes tailles et formes, mais il est aussi plus lent.

Conclusion

Le deep learning pour la classification d'images et la détection d'objets est un domaine en constante évolution, qui continue de proposer des modèles et des techniques innovants, pour relever les défis liés à la complexité et à la diversité des données visuelles.