Chapitre 1

Introduction

L'intelligence artificielle fait la une des médias en ce début de XXI^e siècle. Ceci est du à de nombreuses réalisations récentes déjà disponibles ou en cours de développement et à l'investissement sur ce sujet des grands acteurs industriels, particulièrement ceux du Web et des réseaux sociaux. Des exemples, parmi les plus marquants, sont le développement de véhicules autonomes, de robots pour l'aide à la personne ou pour la médecine, les premières interfaces cerveau-machine, la reconnaissance de la parole et la traduction automatique. Ces services numériques omniprésents améliorent la vie quotidienne comme dans la vie professionnelle. Ils apportent également de nouveaux risques pour la société avec la diminution du nombre d'emplois et la très grande consommation énergétique induite par ces services. Ils engendrent aussi des dangers d'atteinte à la vie privée et la délégation de la prise de décision à des entités numériques, comme des véhicules autonomes, pose également des questions complexes et essentielles de droit.

Ces réalisations en intelligence artificielle voient le jour grâce à l'avènement du monde numérique et à des avancées scientifiques importantes, en particulier, en apprentissage machine. En effet, la seconde moitié du XX^e siècle a été consacrée au développement de l'informatique avec des ordinateurs qui doublent leur capacité de calcul tous les deux ans, la miniaturisation des ordinateurs, le déploiement de capteurs (caméras, montres connectées, ...) et le déploiement de réseaux permettant une interconnexion des objets et des données. Une des conséquences est la disponibilité de données numériques dans de nombreux domaines d'application. Ces grandes quantités de données peuvent alors être utilisées par des programmes d'apprentissage machine pour résoudre des tâches. Pour illustrer l'usage des données et de l'apprentissage machine, nous considérons l'exemple de la conduite autonome et de la traduction automatique.

Pour la conduite autonome, des images animées en très grand nombre vont être générées par des capteurs vidéos sur des voitures. Des programmes vont être entraînés sur ces images pour apprendre à reconnaître les objets dans les images. D'autres programmes vont apprendre, à partir des actions réalisées par un chauffeur (humain ou machine), à prendre une décision d'action de conduite pour le véhicule. Pour la traduction automatique, on dispose, par exemple avec la communauté européenne, de grands corpus de textes écrits dans plusieurs langues. Il est alors possible en utilisant ces textes et leurs traductions d'entraîner des programmes pour traduire d'une langue vers l'autre.

Les fondements, modèles et algorithmes de l'apprentissage machine ont eux aussi été développés dans la seconde moitié du XX^e siècle. Une théorie a été développée pour définir ce que signifie apprendre par programme et a permis de poser les limites de ce qui est apprenable en un temps de calcul raisonnable. En parallèle, de nombreux systèmes d'apprentissage ont été définis comme, en particulier, les séparateurs à vastes marges, les méthodes d'ensemble et les réseaux de neurones. Ces méthodes ont été appliquées avec succès sur des tâches comme la reconnaissance de caractères manuscrits, la prédiction de scores clients ou le diagnostic médical.

Si les modèles et méthodes étaient définis, la mise à disposition de grands jeux de données, le développement de la capacité des machines, la mise à disposition de bibliothèques d'apprentissage ont permis de nouvelles avancées au XXI^e siècle. En particulier, les réseaux de neurones ont pu être appliqués avec succès dans de nombreux domaines. Ces succès entraînent des changements dans de nombreux domaines scientifiques ce que nous illustrons par deux exemples.

En traitement d'images, de très grands jeux de données sont disponibles. Des architectures spécifiques de réseaux de neurones ont été développées. Ces réseaux ont pu être entraînés sur ces jeux de données. Tout ceci a permis de résoudre des tâches de classement d'images (classer une image dans un petit nombre de catégories) et des tâches de reconnaissance d'objets dans des images avec des performances jamais atteintes auparavant.

De la même façon, des architectures de réseaux de neurones ont été définies pour traiter des textes en langage naturel. Avec ces architectures et les masses de données de documents multi-lingues, des réseaux ont pu être définis et entraînés pour la traduction automatique. En quelques années, les réseaux de neurones ont supplanté des méthodes qui avaient plus d'un demi-siècle d'expérience. En effet, ces réseaux obtiennent des performances inégalées en traduction automatique dépassant les méthodes d'alignement statistique des années 2000 et les systèmes de traduction à base de règles développés et améliorés depuis les années 1970.

L'apprentissage machine a donc contribué aux développements récents des progrès de l'intelligence artificielle. Mais que veut dire apprendre pour une machine ! Une définition attribuée à Mitchell [1] peut s'écrire :

> Apprendre, c'est améliorer les *performances* sur une *tâche* par l'*expérience*.

Cette définition est très générale et s'applique à de nombreux cadres : apprendre à jouer aux échecs consiste à améliorer son niveau de jeu au fur et à mesure des parties jouées ; apprendre à conduire consiste à améliorer son niveau de maîtrise du véhicule au fur et à mesure des kilomètres parcourus ; apprendre à traduire consiste à traduire des textes et à améliorer la qualité des traductions au fil des textes traduits. La définition précédente doit être davantage spécifiée pour se traduire en un problème d'apprentissage machine. En particulier, une tâche complexe comme la conduite autonome d'un véhicule se décompose en de nombreuses tâches : apprendre à reconnaître des objets dans les images envoyées par les capteurs vidéo ; apprendre à représenter la situation actuelle à partir des images et des informations, comme la vitesse, envoyées par les capteurs physiques ; apprendre à prendre une décision d'action de conduite dans une situation. Nous allons donc préciser la définition de l'apprentissage machine considérée dans cet ouvrage en précisant les notions de tâche, d'expérience et de performance.

Nous supposons que *la tâche* peut être modélisée comme une transformation de données d'entrées en des données de sortie. Par exemple, classer une image ou une partie d'image consiste à associer, à une représentation numérique d'une image, une classe ou catégorie. Dans le cadre de conduite autonome, cela pourrait être classer une partie de l'image comme personnage animé ou non, classer comme panneau de signalisation avec la définition du panneau. Traduire un texte, c'est prendre la représentation numérique d'un texte dans une langue et trouver la représentation dans la langue de traduction. La nature des données d'entrée et de sortie, la forme de la transformation mènent à des problèmes différents de difficultés différentes.

Nous supposons que *l'expérience* est constituée de données, en général de données historiques. Pour une tâche de transformation, nous supposons que l'expérience est donnée sous forme d'exemples qui sont des couples constitués de la représentation d'une donnée d'entrée et de la représentation de la sortie attendue pour cette entrée. Par exemple, en reconnaissance d'images, on suppose disposer d'un grand nombre d'images dont on connaît la classe. En traduction automatique, on suppose disposer d'un grand corpus de textes dans deux langues.

1. T. M. Mitchell, *Machine Learning*, McGraw-Hill, (1997).

La performance mesure la capacité à résoudre la tâche. Mais, si un programme classe correctement les images (ou traduit les textes) sur lesquelles il a appris, est-il capable de classer de nouvelles images (de traduire de nouveaux textes) ? La performance correspond donc à la capacité de bien généraliser, c'est-à-dire de bien résoudre la tâche sur de nouvelles données. Ceci fait la particularité et la difficulté de l'apprentissage machine. En effet, la performance dépend d'un avenir inconnu : quelles images nouvelles faudra-t-il classer ? Quels textes nouveaux faudra-t-il traduire ? Comme cet avenir est inconnu, des hypothèses devront être faites par le système d'apprentissage machine. Ceci fait qu'il existe différents systèmes d'apprentissage machine avec leurs avantages respectifs.

Nous considérons, dans cet ouvrage, l'apprentissage à partir d'exemples pour des tâches de transformation avec l'objectif de généraliser sur de nouvelles données. Les exemples sont des couples constitués d'une donnée d'entrée et de la sortie attendue. Un tel problème d'apprentissage est dit *supervisé* car, pour chaque entrée, la sortie attendue est connue. Il existe d'autres modes d'apprentissage comme, par exemple, l'apprentissage non supervisé pour lequel on ne dispose que des données d'entrée avec l'objectif de trouver des groupes cohérents dans les données. Nous supposons également un mode d'apprentissage pour lequel on apprend à partir d'un échantillon. Il existe d'autres formes d'apprentissage comme l'apprentissage en ligne pour lequel l'apprenant reçoit un exemple à la fois et doit prédire une sortie à chaque présentation d'exemple. Signalons également l'apprentissage par renforcement pour lequel l'apprenant choisit des actions, reçoit des réponses de l'environnement et doit apprendre une stratégie pour résoudre la tâche. C'est le cas d'un robot qui apprend à naviguer ou d'un programme informatique qui apprend à jouer.

Les données d'entrée sont des représentations des objets du monde comme un patient, une image ou un texte. La représentation la plus fréquente des objets est de considérer des propriétés de l'objet, appelées attributs ou variables, et de les ranger dans des vecteurs (des tableaux de valeurs à une dimension). Par exemple, on peut représenter un patient par un vecteur contenant un certain nombre d'attributs comme l'âge et différents résultats d'examens médicaux. La construction de ces représentations est dite manuelle car les attributs sont souvent choisis par un ou des experts humains.

La représentation peut également provenir du signal numérique. On peut, par exemple, représenter une image par un vecteur où chaque attribut correspond à un pixel [2] de l'image et la valeur correspondante est la valeur d'une propriété du pixel comme le niveau de gris. Nous nous limitons ici

2. Un pixel est la partie élémentaire d'une image et le nombre de pixels permet de mesure la qualité, parfois appelée résolution, d'une image.

aux représentations vectorielles mais notez qu'il est possible de dépasser les représentations vectorielles. Par exemple, une image peut être vue comme une grille de pixels, un texte comme une suite de lettres ou de mots ou de phrases, un réseau social peut être considéré comme un graphe avec des individus connectés par des relations.

Le *choix de représentation* est souvent essentiel en apprentissage machine. En effet, il est facile de comprendre que la richesse du langage de représentation influe sur ce qu'il est possible d'apprendre par programme. Par exemple, si un patient est décrit par son âge et son poids, il sera impossible d'apprendre un programme capable d'émettre un diagnostic. De même, la résolution d'une image, c'est-à-dire le nombre de pixels par image, va influer sur ce qui peut être prédit sur l'image. Ce choix essentiel peut être fait par expertise ce qui est souvent le cas dans des domaines d'application comme le marketing et la médecine.

Cependant, dans de nombreux domaines comme l'image, la construction manuelle de représentations est désormais abandonnée. Le système d'apprentissage apprend à partir des représentations numériques des images. C'est le cas, tout particulièrement, pour les réseaux de neurones (profonds) qui prennent en entrée des représentations brutes des images et apprennent à résoudre la tâche tout en construisant des nouvelles représentations des images dans les couches internes du réseau. La même constatation peut être faite pour les textes en langage naturel où il est remarquable que des représentations numériques des textes apprises dans des réseaux donnent de meilleurs résultats que des représentations construites à partir de l'expertise des linguistes. Beaucoup des progrès récents en apprentissage machine, et donc en intelligence artificielle, sont une conséquence de *l'apprentissage de représentations*.

Dans l'apprentissage supervisé à partir d'exemples, les sorties peuvent être de différente nature. On parle de *classification supervisée* lorsque la sortie correspond à un nombre fini de classes. Par exemple, classer une image comme contenant un visage ou non, classer une image dans un nombre fini de catégories prédéfinies. On parle de *régression* lorsque la sortie attendue est un nombre réel. Par exemple, attribuer à une image une probabilité de contenir un visage ou attribuer un score de pertinence à une image pour une requête dans un système de recherche d'images.

Nous considérons, dans cet ouvrage, principalement la classification supervisée et la régression. Mais l'apprentissage machine peut s'étendre à des tâches de prédiction plus complexes. Par exemple, la prédiction d'ordres pour ordonner les réponses d'un système de recommandation ou la prédiction de sorties structurées pour prédire un texte de sortie à partir d'un texte d'entrée en traduction automatique.

Pour terminer cette introduction à l'apprentissage machine, rappelons que l'objectif est de bien prédire pour de nouvelles données. C'est un problème difficile et de nombreuses solutions peuvent être proposées selon les procédés de prédiction choisis et la façon de les apprendre. C'est pour cette raison qu'il existe différents systèmes d'apprentissage avec leurs avantages respectifs. Leur présentation, leur compréhension et la présentation des éléments de choix d'un système d'apprentissage adapté à une tâche sont les objectifs de cet ouvrage.

Après avoir introduit l'apprentissage machine, il est important de comprendre les différences entre l'approche classique par programmation et l'approche par apprentissage machine pour la résolution d'une tâche. Dans l'approche classique, l'informaticien écrit un programme dans un langage de programmation pour résoudre la tâche cible. Cette approche est la plus répandue mais a montré ses limites pour des tâches comme, par exemple, le classement d'images et la reconnaissance de la parole. En effet, il s'est avéré impossible à l'expert programmeur d'exprimer tous les choix possibles à cause de la difficulté de la tâche, du bruit dans les données, de l'incertitude dans les réponses.

En apprentissage machine, on suppose avoir des données en grande quantité et le système d'apprentissage est un programme qui va générer un programme pour résoudre la tâche. Par exemple, à partir de productions orales et de leurs transcriptions écrites, le système d'apprentissage génère un programme capable de prendre en entrée une production orale et de sortir sa transcription textuelle. C'est donc un programme (le système d'apprentissage) qui génère un programme plutôt qu'un programmeur qui écrit un programme. Les deux approches sont complémentaires. L'approche par apprentissage est réservée à des tâches pour lesquelles la programmation est difficile et pour lesquelles on dispose de beaucoup de données pour que l'apprentissage soit possible.

Pour finir, nous discutons la position de l'apprentissage dans le paysage scientifique et dans la société. L'apprentissage machine est principalement lié aux mathématiques et à l'informatique : les probabilités pour la modélisation, les statistiques pour l'évaluation et certaines méthodes inférentielles, l'optimisation pour la recherche de solutions, l'algorithmique, la complexité, le calcul scientifique et les mathématiques appliquées pour des implantations efficaces. La communauté apprentissage machine est concernée par les questions allant de la théorie de l'apprentissage (définir ce que veut dire apprendre) aux applications de l'apprentissage. Au sein de cette communauté ont été développées les méthodes les plus performantes comme les méthodes d'ensemble, les séparateurs à vastes marges et les réseaux de neurones. D'autres communautés ont contribué au développement de l'apprentissage machine comme la sociologie et la physique dans le domaine de l'apprentissage pour les réseaux, l'économie et la théorie des jeux pour l'apprentissage en ligne.

Toutes les disciplines scientifiques et tous les domaines d'application sont désormais concernés par l'apprentissage machine dans l'utilisation de systèmes d'apprentissage pour résoudre des tâches de la discipline ou du domaine d'application. L'apprentissage machine a une place essentielle dans la science des données et dans l'intelligence artificielle. Ses applications font désormais partie de notre quotidien dans les applications que nous utilisons. Tous les grands groupes industriels en ont fait un domaine de recherche prioritaire pour des applications comme les assistants personnels, la maison intelligente, la ville intelligente, la médecine assistée et les véhicules autonomes, entre autres applications. Ceci génère des problèmes sociétaux sur lesquels nous revenons dans la conclusion comme l'égalité de traitement entre les utilisateurs, le respect de la vie privée et les questions de droit liées à l'utilisation de systèmes autonomes. Enfin, ces systèmes engendrent des problèmes énergétiques liés à la consommation des machines comme, par exemple, les fermes de calcul nécessaires à un moteur de recherche d'information.

Plan

Cet ouvrage d'introduction étudie donc principalement l'apprentissage supervisé à partir d'exemples pour des données vectorielles. Le contenu des différents chapitres est le suivant :

Dans le chapitre 2, nous présentons la problématique de l'apprentissage supervisé et précisons certains des points abordés dans cette introduction. En particulier, nous discutons du langage de représentation, nous revenons sur la question de l'objectif de bien apprendre sur des données non encore rencontrés et montrons pourquoi des biais d'apprentissage sont nécessaires, nous expliquons pourquoi il n'existe pas de meilleur système d'apprentissage et nous voyons comment évaluer les performances lorsqu'on ne dispose que d'un échantillon de données.

Dans le chapitre 3, nous proposons un premier système d'apprentissage de règles de classement représentées sous forme d'un arbre de décision. Ce système est basé sur un algorithme basé sur des critères de choix issus de la théorie de l'information. Cet algorithme est dit glouton dans le sens où, à chaque étape, un choix est fait sans qu'il puisse être remis en question par la suite. Nous présentons également les méthodes d'ensemble qui permettent de combiner des procédures de classement à l'aide de méthodes de votes. Nous présentons alors les forêts aléatoires et le « gradient boosting » qui sont parmi les meilleurs systèmes d'apprentissage pour la classification et la régression lorsque les représentations vectorielles des données sont suffisamment riches.

Dans le chapitre 4, nous considérons l'apprentissage de fonctions de classement appelées séparateurs linéaires et définies par la séparation de l'espace des données d'entrée par des hyperplans. Nous proposons différents algorithmes d'apprentissage. En particulier, nous introduisons la descente de gradient qui est une méthode d'optimisation permettant de trouver un « meilleur » séparateur. Nous présentons les séparateurs à vastes marges (SVMs) pour la séparation linéaire. Nous présentons enfin le « truc du noyau » qui permet de généraliser les SVMs à des fonctions de séparation non linéaires. Les SVMs sont également un des meilleurs systèmes d'apprentissage mais ont surtout permis de fixer les bases de l'apprentissage statistique et d'étudier les fonctions noyaux en apprentissage machine.

Dans le chapitre 5, nous introduisons les réseaux de neurones. Nous étudions particulièrement le perceptron multi-couches (PMC) qui est une architecture de réseaux adaptée pour les données vectorielles. Nous discutons l'influence du choix de l'architecture, c'est-à-dire du nombre de couches et du nombre de neurones par couche. Nous introduisons les principes généraux de l'algorithme de rétro-propagation du gradient qui permet d'apprendre les meilleures valeurs pour les paramètres d'un réseau. Nous présentons succinctement d'autres architectures adaptées aux images et aux textes. Les réseaux de neurones sont utilisés pour de nombreuses tâches avec des performances inégalées, principalement pour des problèmes sur des données telles que des images, des textes ou de la parole.

Dans le chapitre 6, nous donnons des éléments de méthodologie pour résoudre des problèmes en utilisant des systèmes d'apprentissage machine. Nous discutons des données et des choix de représentation. Nous présentons les méthodes de base pour évaluer les performances d'un système. Nous donnons des éléments pour le choix d'un système d'apprentissage en fonction des données et de la tâche cible. Enfin, nous donnons quelques pointeurs sur des bibliothèques logicielles d'apprentissage machine.

La conclusion synthétise les principales connaissances introduites. Elle décrit rapidement d'autres types d'apprentissage et présente quelques-uns des nombreux challenges qui restent à résoudre en apprentissage machine.

Enfin, le chapitre 8 contient des exercices simples permettant d'assurer la compréhension des notions introduites dans l'ouvrage.