



BIG DATA

L'algorithme

qu'est-ce que c'est ?

De gros titres, tels que "Algorithmes - les vrais maîtres du monde ?" ou "Les algorithmes influencent-ils notre quotidien?", sont à la une des journaux, notamment dans le contexte du Big Data. Dans ce dossier, nous expliquerons ce que sont les algorithmes, comment ils fonctionnent et à quoi ils servent. Nous aborderons également les aspects éthiques.

Un algorithme est une instruction ou une règle logique et explicite sur la manière dont un problème est à résoudre, et se compose de différentes étapes exécutées l'une après l'autre ou simultanément. Pour la majorité des algorithmes, une entrée donnée conduit toujours à la même sortie.

Dans la plupart des cas, les algorithmes s'appliquent sous forme de logiciels informatiques, mais ils peuvent également être formulés en langage humain. L'exemple le plus illustre d'un algorithme de notre quotidien est une recette de cuisine.

D'où vient le nom ?

Le terme algorithme est dérivé du nom du mathématicien et géographe korasmien Muhammad Ibn Musa al-Khuwarizmi, originaire de la région de l'Iran d'aujourd'hui, qui a passé la majeure partie de sa vie à Bagdad. En 825, il écrit un manuel sur les chiffres indiens et y explique également le fonctionnement des algorithmes.

Le livre sur les chiffres, que nous connaissons aujourd'hui sous le nom de "chiffres arabes", a été traduit en latin au 12e siècle. La dernière partie de son nom (al-Khuwarizmi) a été traduite par "Algorismi". Les premiers mots étaient "*Dixit Algorismi*" (Algorismi a dit), d'où s'est finalement dérivé le terme algorithme.

Caractéristiques d'un algorithme

Tous les algorithmes partagent certaines caractéristiques qui les définissent.

Explicite : chaque algorithme doit être sans équivoque, c.à.d. il ne doit pas contenir des instructions qui se contredisent.

Exécutable : chaque étape doit être exécutable (elle ne doit donc pas p.ex. contenir une division par zéro)

Fini : l'algorithme doit avoir une fin définie et ne donc pas continuer à l'infini.

Terminé : un algorithme doit toujours se terminer après un nombre fini d'étapes et produire un résultat

Déterminé : pour les mêmes entrées, l'algorithme doit toujours donner les mêmes sorties, donc le même résultat

Défini : chaque étape est définie explicitement, il existe donc qu'une seule possibilité d'action après chaque étape.

En ce qui concerne la solution informatique de problèmes, il n'existe pas seulement des algorithmes. En effet, il existe aujourd'hui également d'autres outils qui se basent toutefois toujours d'une quelconque façon sur des algorithmes.

Heuristiques

Bien que nous disposions aujourd'hui d'une puissance de calcul énorme, certaines tâches ne peuvent pas être résolues dans un temps satisfaisant par les algorithmes. Un exemple serait le trajet optimal pour un service de livraison qui doit livrer des colis à plusieurs clients dans une même région. Pour un tel problème, il existe certes une solution optimale et parfaite, mais la trouver prend beaucoup de temps.

Une heuristique est appliquée lorsqu'un simple algorithme est trop lent. Une heuristique se base souvent sur des variables aléatoires. L'"algorithme glouton" (ou "greedy-algorithm" en anglais) est une heuristique qui, à chaque étape, choisit l'alternative la plus prometteuse. Raison pour laquelle cette heuristique trouve une solution rapidement, mais cette dernière n'est pas toujours la meilleure. Les heuristiques sont plus rapides que les algorithmes, et donc plus sujettes aux erreurs.

Intelligence artificielle

L'intelligence artificielle (en anglais *artificial intelligence*) est souvent abrégée par le sigle "IA". Des comportements intelligents sont simulés à l'aide de logiciels. Le terme est problématique car il n'existe pas de définition précise de "intelligence". En principe, on essaie d'imiter l'intelligence humaine - l'ordinateur doit être capable de gérer seul des problèmes. Ces dernières années, la recherche dans le

domaine de l'IA s'est orientée vers les "réseaux de neurones artificiels", qui reposent sur un modèle des cellules nerveuses du cerveau humain. Les termes "Deep Learning" ("apprentissage profond" en français) ou "Machine Learning" ("apprentissage automatique" en français) y sont étroitement liés.

Infobox : test de Turing

Comment déterminer si une machine est considérée comme intelligente ? Pour répondre à cette question, le mathématicien britannique Alan Turing a développé un test qui porte son nom, le test de Turing. Sans contact visuel ni auditif (comme dans un chat), un interrogateur humain discute avec deux personnes, un être humain et une machine. Toutes les deux doivent persuader le testeur qu'elles sont humaines. Si le testeur n'est pas en mesure de dire avec certitude qui des deux

interlocuteurs est une machine, la machine a passé le test avec succès. Jusqu'à présent, aucune IA n'a réussi le test de Turing - cependant, de nombreux progrès ont été réalisés ces derniers temps. Ainsi, l'Université de Chicago a présenté en août 2017 une IA capable de rédiger des critiques (p.ex. de restaurants) dont il était impossible de distinguer de celles écrites par des êtres humains.

Machine Learning

Machine Learning, en français "apprentissage automatique", désigne une forme spéciale de l'IA : grâce à des exemples, un système apprend à généraliser ces exemples après une phase d'apprentissage. Le système reconnaît dans ces données d'apprentissage des schémas et des régularités et, dans le meilleur des cas, peut les utiliser pour de nouvelles données. L'apprentissage automatique est un algorithme qui s'adapte lui-même - si tout se

passé comme prévu, la machine s'améliore donc toute seule.

L'une des méthodes appliquées dans la réalisation d'une phase d'apprentissage s'appelle "Deep learning" ou "apprentissage profond". Pour cela, des réseaux de neurones artificiels sont utilisés qui disposent de nombreuses couches intermédiaires, appelées "hidden Layers", entre l'entrée et la sortie.

En gros, ces couches sont organisées de la même manière que les neurones du cerveau humain. Dans chaque couche, les informations de la couche d'entrée sont abstraites et généralisées. Cela permet à un ordinateur "d'apprendre" des concepts compliqués, tout en les composant de concepts plus simples. Alors qu'il serait très difficile de programmer manuellement les connaissances nécessaires (p.ex. reconnaissance d'une écriture manuscrite), cette difficulté peut être surmontée à l'aide de l'apprentissage automatique, sans qu'un être humain doive formaliser l'information.

Data-Mining

"Data-Mining" signifie littéralement "forage des données" (comme une mine). Le terme peut être trompeur, car il ne s'agit pas d'exploiter de nouvelles données, mais de filtrer de nouveaux liens, interrelations et tendances des "montagnes de données" existantes. Le Data-Mining est étroitement apparenté au Machine Learning, et les mêmes algorithmes sont souvent appliqués. L'objectif est tout de même de trouver de *nouveaux* schémas et régularités, tandis que, en ce qui concerne l'apprentissage automatique, les schémas donnés doivent être reconnus par l'ordinateur. Pour cela, des méthodes statistiques et des algorithmes sont souvent appliqués.

Algorithmes to go :

Que se cache-t-il derrière les applications de l'apprentissage automatique ?

L'apprentissage automatique, ou Machine Learning, a actuellement le vent en poupe - de nombreux éditeurs de sites Web et entreprises essaient d'évaluer leurs données à l'aide de ces nouvelles méthodes. Mais rares sont ceux qui disposent du savoir-faire et de la puissance informatique nécessaires pour développer leurs propres applications de Machine Learning. Comme de nombreux autres services, les applications d'apprentissage automatique sont également proposées dans le cloud. Il existe aujourd'hui quatre grands fournisseurs, connus de tous : Microsoft avec [Microsoft Azure](#), IBM avec [IBM Watson Machine Learning](#), Google avec sa [Google Cloud Platform](#) et Amazon avec [Amazon Machine Learning](#). Selon les fournisseurs, différentes connaissances préalables sont nécessaires ou différents aperçus dans les processus possibles. Le service de Machine Learning de Amazon est facile d'utilisation et nécessite très peu de connaissances préalables, mais fonctionne comme "black box" - en tant que personne extérieure à Amazon, il n'est pas possible de voir et de comprendre précisément quels algorithmes sont appliqués.

Des exemples d'applications réalisées à l'aide de l'apprentissage automatique sont les méthodes de diagnostic médicaux automatisées, reconnaissance de la fraude à la carte bancaire, analyses des marchés d'actions, analyse de l'ADN, reconnaissance vocale, reconnaissance de l'écriture, reconnaissance de texte, ainsi que des systèmes autonomes, tels que les voitures sans conducteurs.

Ainsi, lors de l'analyse des valeurs aberrantes, le but est de trouver des données qui ne vont pas avec le reste de la montagne de données - cela peut être utile lors de la reconnaissance de la fraude à la carte bancaire par exemple.

D'autres méthodes du Data-Mining regroupent certaines données (p.ex. lors de l'analyse cluster et de classification), ou cherchent des liens entre les données (p.ex. analyse de régression et d'association). C'est ainsi que les annonces du type "Les clients qui ont acheté ce produit ont également acheté..." dans les boutiques en ligne voient le jour.

Applications

Comme mentionné plus haut, les algorithmes sont centenaires. Cela fait donc longtemps qu'ils font partie de notre vie et que nous ne les remarquons plus, car p.ex. les recettes de cuisine et les lois sont elles aussi des algorithmes. Nous avons dressé une liste (non exhaustive) avec des exemples de domaines dans lesquels les techniques mentionnées plus haut sont appliquées :

Secteur financier : pour reconnaître les tentatives de fraude et lors du credit scoring, pour évaluer la solvabilité des clients et les risques.

Marketing : sélection de groupes-cibles ("Quel concept de publicité pour qu'elle atteigne nos clients?"), analyse du panier ("Quels produits sont achetés ensemble?" - les supermarchés sont aménagés selon ces connaissances), profils clients ("Qu'est-ce qu'achète ce client? Qu'est-ce que pourrait l'intéresser?")

Médecine : analyser de nouveaux principes actifs pour les médicaments, surveillance des effets secondaires

Industrie : les processus automatisés sont contrôlés par des algorithmes, mais le Data-Mining et le Machine Learning permettent de les améliorer sans cesse, p.ex. dans l'industrie chimique et la plasturgie

Internet : sur Internet notamment, les algorithmes, l'IA, le Machine Learning et le Data-Mining jouent un grand rôle. Il existe par exemple les applications suivantes :

Sécurité informatique : reconnaissance d'attaque (firewall intelligent)

Expérience utilisateur sur les réseaux sociaux : se base en grande partie sur les analyses de réseau sur les réseaux sociaux, permet p.ex. la recommandation de nouveaux amis sur Facebook

Analyse du comportement utilisateur : p.ex. à l'aide de log files, cookies et d'autres données que l'utilisateur laisse sur les sites Web

Système de recommandation de produits, musique ou films, p.ex. service de streaming Netflix.

Technique de sécurité : reconnaissance faciale et reconnaissance de "comportements suspects" pour les surveillances vidéo, soupçon de terrorisme

Il n'y a quasiment pas de limites pour les domaines d'application où des algorithmes et des techniques comme le Machine Learning peuvent être utilisés. L'entreprise [hudl](#) affirme par exemple qu'elle aide des sportifs avec des analyses vidéo et du Machine Learning à se préparer aux compétitions, tandis que la start-up "[Upserve](#)" vise à aider les restaurants à prédire le nombre de clients - et même quels plats inciteraient les clients à revenir manger dans le même établissement.

Aspects éthiques et problèmes

Depuis un certain temps, des inquiétudes ont été soulevées quant à l'utilisation d'algorithmes. Des titres à la une, tels que "Algorithmes - les vrais maîtres du monde ?" ou aussi "Deux chatbots ont développé une propre langue" pourraient faire croire que les algorithmes seraient "mauvais" - alors qu'il ne s'agit en fait que d'outils. Les problèmes et les aspects éthiques qui en résultent dépendent notamment des fins, de la manière, des personnes qui programment ces techniques et de leur utilisation.

Des tâches que la plupart des personnes peuvent faire sans problème - par exemple reconnaître un visage ou désigner un animal ou un objet sur une photo - posent de grandes difficultés à l'ordinateur. Souvent aussi lorsque les algorithmes ont subi un processus de "Deep Learning" et ont été entraînés à l'aide de milliers d'images pour reconnaître des objets sur des photos. Dans un test du réseau de neurones de Google, [l'informaticienne Julia Evans a découvert par exemple que l'IA pensait que la reine d'Angleterre portait un bonnet de douche \(au lieu d'une couronne\) sur la tête.](#)

L'informaticienne a développé une méthode qui permet de modifier les images pour qu'elles semblent identiques pour les êtres humains, mais que l'IA puisse y "voir" des choses complètement différentes - par exemple un vautour au lieu d'un panda, ou une serviette au lieu d'un chat. Il lui suffisait de modifier quelques pixels. En revanche, elle n'a pas réussi à tromper l'IA avec des images de chiens manipulées - l'IA a toujours reconnu les chiens dans les photos. L'une des raisons est sûrement parce que l'IA a été entraînée avec de nombreuses photos de chien et qu'elle excelle donc dans la reconnaissance de chiens. Contrairement aux photos de pandas et de couronne.

Malheureusement, il existe également de nombreux exemples moins drôles. Joy Buolamwini, docteure au MIT Media Lab, explique qu'elle rencontrait toujours le même problème : les logiciels de reconnaissance facile ne fonctionnaient pas sur elle. Joy Buolamwini est noire. Elle réalisa vite que le

logiciel reconnaissait certes un masque blanc en tant que "visage", mais pas son visage à elle. Selon la reconnaissance d'image de Google, certaines personnes de couleur noire étaient des gorilles, tandis qu'un appareil photo de Nikon demandait aux personnes aux traits asiatiques si elles avaient cligné des yeux et proposait de refaire la photo sans ce "défaut". Les algorithmes sont-ils racistes ?

Joy Buolamwini appelle ce phénomène "algorithmic bias" (biais algorithmiques). Les algorithmes adoptent souvent les façons de penser ou les préjugés (souvent inconscients ou systématiques) de leurs développeurs. Dans le cas de la reconnaissance facile par exemple, la raison était que les données d'entraînement ne contenaient pas suffisamment de "People Of Color non blancs". Et parce que l'informatique est un secteur dans lequel travaillent avant tout des hommes blancs, personne ne s'est rendu compte "par hasard" de ce problème.

Reconnaître correctement des visages et des objets sur des photos est une chose, mais pour d'autres applications (possibles) d'algorithmes la situation peut s'avérer plus compliquée : lorsque des algorithmes, qui doivent déterminer la solvabilité des clients, contiennent également des biais (racistes), certaines personnes peuvent être discriminées sans s'en rendre compte. Il va de même pour les IA qui doivent classer des candidatures pour des postes de travail ou les universités. Pour ce qui est de la surveillance vidéo, qui doit reconnaître de manière "intelligente" un comportement prétendu suspect, il pourrait y avoir - sans intention - du "racial profiling", à savoir que des actions de People of Color soient plutôt considérées comme "suspectes" que celles des personnes blanches.

En principe, pour chaque algorithme qui prend des décisions importantes se pose la question de la transparence. Qui vérifie si une erreur ou un "algorithmic bias" s'est introduit dans le code ? Comment s'assurer que des données nourries aux algorithmes de Machine Learning sont représentatives ?

Éventuelles solutions

L'une des solutions possibles est que le code soit ouvert, à savoir disponible en tant que code open source. De nombreux projets sont organisés de la sorte, tels que le navigateur Firefox ou le système d'exploitation Linux. Non seulement les logiciels sont libres, mais le code qui se cache derrière est également public et toutes les personnes intéressées sont invitées à participer à son développement. Cela présente l'avantage que d'éventuels bugs et vulnérabilités puissent être trouvés plus rapidement et ensuite supprimés. Certains problèmes d'algorithmes pourraient ainsi être éliminés - des bénévoles ou des ONG pourraient examiner le code et améliorer les vulnérabilités qui pourraient poser des problèmes éthiques. En revanche, pour les algorithmes qui se basent sur le Deep Learning, cette démarche serait uniquement utile si les données d'entraînement (à savoir les photos qui sont utilisées pour entraîner la reconnaissance faciale) sont publiques et peuvent être complétées.

Si nous essayions d'éliminer les problèmes éthiques d'algorithmes par ce que l'on appelle le crowdsourcing - à savoir l'implication de nombreux bénévoles qui recherchent des

erreurs - cela supposerait que nous serions plus nombreux à être en mesure de lire des codes et de les comprendre. Cette compétence devient de plus en plus importante - une raison de plus pour enseigner à l'école le codage et la compréhension des contextes algorithmiques.

Une autre solution serait la possibilité d'introduire un audit indépendant pour les algorithmes qui interviendrait dans les domaines sensibles de la vie. Avant qu'un algorithme puisse p.ex. être utilisé pour la détermination de la solvabilité, il devrait être contrôlé par un organisme indépendant - comme il est p.ex. le cas pour les émissions polluantes des véhicules.

Les algorithmes sont souvent compliqués - et les problèmes éthiques qui s'en accompagnent le sont aussi. De même qu'il est impossible de repousser le dentifrice dans son tube, il n'est pas possible d'arrêter le progrès. En revanche, il est important de mener le débat sur les avantages et les inconvénients, et de créer le progrès. Il est donc d'autant plus important d'être informé et de connaître les faits.

Pour toute question au sujet de l'arnaque en ligne ou sur l'utilisation d'Internet en général, contactez la BEE SECURE Helpline:

