

Le Machine Learning

Prévoir des crimes avant qu'ils ne se produisent, proposer un diagnostic médical grâce à l'analyse de données de type Big Data et d'algorithmes sophistiqués, créer une voiture sans chauffeur totalement autonome... Le Machine Learning, ou apprentissage automatique, semble riche de promesses. Possible réalité ou pure fiction ? Voyons un peu ce qu'il en est.

Tentons tout d'abord une définition de ce concept en vogue. Le « Machine Learning », ou « apprentissage automatique » dans la langue de Molière, consiste à donner la faculté aux ordinateurs, au sens très large du terme, d'apprendre à partir des données qui leurs sont soumises, et non plus seulement d'exécuter des algorithmes séquentiellement. Cette théorie a pris son essor grâce aux travaux des mathématiciens Vapnik et

Chervonenkis (théorie VC) dans les années 60. Ces recherches se sont développées en marge de celles des probabilités et des statistiques. Le Machine Learning ne se concentre plus sur la façon de retrouver des objets abstraits, comme en suivant « à la lettre » une loi de probabilité. Il se concentre avant tout sur le côté opérationnel des choses, c'est-à-dire sur la prise de décision à partir d'une analyse empirique des données, en essayant en plus de déceler les erreurs potentielles. Or, la faculté d'analyse empirique est une des qualités qui manquent cruellement aux machines pour arriver à des conclusions similaires à celles du cerveau humain.

L'apprentissage automatique est un champ d'étude, parmi d'autres, de l'Intelligence artificielle. Il concerne à la fois la conception, l'analyse, le développement et l'implémentation de méthodes permettant à une machine d'évoluer par un processus systématique, ce afin de pouvoir réaliser des tâches difficiles voire impossibles via des moyens algorithmiques plus traditionnels. Des systèmes complexes peuvent être analysés, y compris pour des données associées à des valeurs symboliques. L'analyse peut concerner des graphes, des arbres ou des courbes. La classification est un exemple d'apprentissage automatique. Elle consiste à étiqueter chaque donnée en l'associant à une classe – un type de données complexe. La conduite autonome, sans humain, en est un autre, ainsi que le diagnostic.



Scikit-learn, la bibliothèque open source écrite en Python par des chercheurs de l'Inria.

Différents systèmes d'apprentissage existent. Nous les évoquons un peu plus loin dans cet article.

Principes du ML

Les algorithmes utilisés dans le Machine Learning permettent à un système piloté par ordinateur – un robot par exemple – ou assisté par ordinateur, d'adapter ses analyses et ses comportements en se fondant sur l'analyse de données empiriques issues d'une base de données ou de capteurs. La difficulté réside dans le fait que l'ensemble de tous les comportements possibles, compte tenu de toutes les entrées possibles, devient rapidement trop volumineux et de fait trop complexe à décrire – on parle d'explosion combinatoire – avec les langages de programmation classiques. Les développeurs du ML utilisent principalement le langage R, suivi de près par le langage Python, avec généralement la bibliothèque scikit-learn. Grâce à ces langages et bibliothèques open source adaptés aux principes du ML, ils codent des programmes capables d'ajuster un modèle permettant de simplifier cette complexité et de l'utiliser de manière opérationnelle. Ce modèle est adaptatif. Cela veut dire qu'il est capable de prendre en compte l'évolution des données. Le système d'analyse ou de réponse peut ainsi s'auto-améliorer. C'est une des formes de l'Intelligence

Le deep learning

Le deep learning est un algorithme de Machine Learning qui fait couler pas mal d'encre en ce moment. Les forts investissements réalisés par Google et Amazon dans ce domaine n'y sont pas étrangers. Le deep learning, ou apprentissage profond, et les réseaux de neurones ne datent pourtant pas d'hier, mais des années 80. À l'époque, ils n'avaient pas plus décollés que la plupart des projets d'IA. Les scientifiques recherchaient en premier lieu l'interprétabilité des modèles. Les besoins sont différents maintenant, et cela les a enfin remis au goût du jour. Le deep learning fournit par exemple d'excellents résultats dans l'analyse d'image, de son ou de texte. C'est un modèle paramétrique proposant de nombreux réglages. Il permet de décrire, par exemple, la structure d'un réseau de neurones en précisant le nombre de noeuds qu'il comportera et de combien de couches il sera constitué.

artificielle, et même un de ses principes élémentaires. Les programmes ainsi développés, selon bien entendu leur qualité et leur degré de perfectionnement, peuvent intégrer des capacités de traitement probabiliste des données, d'analyse de données issues de capteurs, de reconnaissance – de forme, d'écriture, vocale, etc. –, d'informatique théorique ou de data-mining et autres sciences tournant autour du big data.

Les applications du Machine Learning

Le Machine Learning est aujourd’hui en plein essor. Il est utilisé pour un spectre d'applications de plus en plus large, tel que les moteurs de recherche, comme Google, l'aide au diagnostic médical, Watson d'IBM, et technique de manière générale, la détection de données aberrantes ou manquantes, celle de fraudes à la carte de crédit, l'analyse des marchés financiers y compris l'analyse du marché boursier (logiciels-robots « traders »), la reconnaissance vocale, d'écriture manuscrite ou autre, l'analyse et indexation d'images et de vidéos, particulièrement pour la recherche d'image par le contenu, et la robotique – la locomotion particulièrement. L'apprentissage automatique est utilisé par exemple pour doter des ordinateurs ou des machines de systèmes de perception de leur environnement : vision, reconnaissance d'objets – visages, schémas, langages naturels, écriture, formes syntaxiques... La bio-informatique, la chémoinformatique (secteur de la chimie), les interfaces cerveau-machine, la classification des séquences d'ADN, le jeu, le génie logiciel, la conception de sites web adaptatifs et bien d'autres domaines y ont également recours. Des algorithmes de Machine Learning ont aussi été déployés à grande échelle pour le filtrage anti-spam des messageries électroniques, l'optimisation des stocks, la segmentation et le ciblage des clients ainsi que dans la maintenance industrielle. Il est

[Robot Learning Activities](#)
[Contact](#)
[Join Us](#)
[Back to Technical Committees](#)

Robot Learning

The robot hardware is progressively becoming more complex, which leads to growing interest in applying machine learning and statistics approaches within the robotics community. At the same time, there has been a growth within the machine learning community in using robots as motivating applications for new algorithms and formalisms. Considerable evidence of this exists in the use of robot learning approaches in high-profile competitions such as RoboCup and the DARPA Challenges, and the growing number of research programs funded by governments around the world. Additionally, the volume of research is increasing, as shown by the number of robot learning papers accepted to IROS and ICRA, and the corresponding number of learning sessions.



The primary goal of the Technical Committee on Robot Learning is to act as a focus point for wide distribution of technically rigorous results in the shared areas of interest around robot learning. Without being exclusive, such areas of research interest include:

L'éducation des robots risque de vite devenir un sujet de préoccupation aussi délicat que celle des enfants.

employé notamment pour la maintenance prédictive des matériels installés sur les plates-formes pétrolières ou les avions (moteurs). Le développement des objets connectés ne peut certainement pas se faire sans lui, et là encore le champ applicatif est presque infini.

Locomotion

Un système d'apprentissage automatique peut permettre à un robot ayant la capacité de bouger ses membres mais ne sachant initialement rien de la coordination des mouvements permettant la marche, d'apprendre tout seul à marcher ou à avancer sur un terrain inconnu (robots d'exploration spatiale). Le robot commencera par effectuer des mouvements aléatoires, puis, en sélectionnant et privilégiant les mouvements lui permettant d'avancer, mettra peu à peu en place une marche de plus en plus efficace. Il pourra aussi analyser le terrain grâce à ses différents capteurs et aux algorithmes de reconnaissance de forme, de volume et autres facteurs.

Reconnaissance manuscrite et vocale

La reconnaissance de caractères manuscrits est une tâche complexe car deux caractères similaires ne sont

jamais exactement égaux. Il est plus aisés de concevoir un système d'apprentissage automatique qui apprend à reconnaître des caractères en observant des exemples, c'est-à-dire des caractères déjà identifiés. Un procédé similaire est utilisé pour les dictées vocales.

La gestion des risques financiers

Autre secteur, lui aussi avide de nouveaux algorithmes et fort de capitaux à investir, le domaine de la gestion des risques financiers s'intéresse au Machine Learning depuis déjà quelques années. Des modèles se basant sur le Machine Learning sont capables de fournir une prédiction des risques pris par des établissements financiers dans le cadre de la délivrance de prêts ou de souscriptions de contrats d'assurance.

La lutte contre la criminalité

Le Machine Learning a aussi trouvé sa voie dans la lutte contre la criminalité. Plusieurs villes américaines font déjà appel à IBM et à son système Watson afin de prédire les zones d'une ville où sont susceptibles de survenir les prochaines agressions, d'éventuels cambriolages ou crimes divers. Ils peuvent ainsi mieux cibler la répartition et l'envoi d'effectifs

de polices sur les quartiers d'une ville, compensant ainsi un manque d'effectifs potentiel. Les résultats seraient assez convaincants d'après les municipalités concernées. L'Allemagne ainsi que d'autres pays s'intéressent de très près à ce nouveau concept imaginé, rappelons-le, par le grand Philippe K. Dick dans son livre *Minority Report*.

Le diagnostic médical

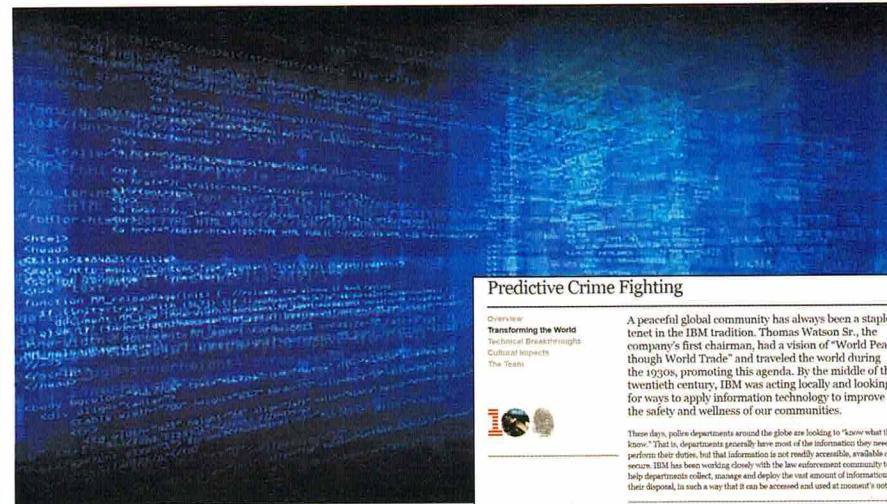
Un centre de cancérologie à New York, le Memorial Sloan Kettering, a recruté un spécialiste en diagnostic médical d'un genre peu banal : l'IA (intelligence artificielle) Watson, développée par IBM. Watson a ingurgité et étudié près de 2 millions de pages de journaux spécialisés et presqu'autant de rapports cliniques. Ses capacités d'analyse hors pair le mettraient d'ores et déjà en position de surpasser ses collègues humains, d'après ses créateurs et ses tuteurs. Quoi qu'il en soit, il possède un avantage incontestable sur ses homologues humains : le don d'ubiquité. IBM a annoncé que Watson sera disponible à terme sous forme de service cloud pour tous les hôpitaux de la Planète. Si son coût d'utilisation n'est pas trop élevé, cela pourrait rendre des services inestimables aux hôpitaux les moins bien lotis en médecins et spécialistes.

Les algorithmes

De nombreux types d'algorithmes, spécifiques à chaque domaine, sont utilisés dans le Machine Learning. En voici quelques exemples :

l'analyse discriminante linéaire ;
les algorithmes et la programmation génétiques ;
les arbres de décision ;
le boosting ;
les machines à vecteur de support (SVM) ;
les méthodes statistiques, telle que le modèle de mixture gaussienne ;
la régression logistique ;
les réseaux de neurones (avec apprentissage supervisé ou non) ;
Toutes ces méthodes sont combinées afin d'obtenir le plus de variantes possibles d'apprentissage. L'utilisation de tel

Memorial Sloan Kettering's Expertise Combined With the Power of IBM Watson is Poised to Help Doctors Make Better Treatment Choices



Après avoir gagné à Jeopardy, Watson d'IBM s'est trouvé un nouveau hobby : la lutte contre le crime.

ou tel algorithme dépendra de fait fortement de son adaptabilité avec la tâche à effectuer (classification, diagnostic, estimation de valeurs, parcours de graphe). La qualité du travail obtenu dépendra fortement de facteurs initiaux dits contraignants, liés bien évidemment aux données en soi, dont :

le nombre d'exemples – moins il y en a, plus l'analyse sera difficile ; plus il y en a, plus le besoin en ressources informatiques sera élevé et plus longue sera l'analyse. La fiabilité sera accrue si les erreurs sont correctement identifiées ; le nombre et la qualité des attributs décrivant ces exemples ; le pourcentage de données renseignées et manquantes ; le bruit : le nombre et la localisation des valeurs douteuses (erreurs potentielles et autres valeurs aberrantes) ou naturellement non-conformes au pattern de distribution générale.

La tentation de s'inspirer des êtres vivants pour concevoir des machines capables d'apprendre, comme avec les réseaux de neurone, a toujours été forte. Bien que l'apprentissage automatique soit avant tout un sous-domaine

Predictive Crime Fighting

Overview
Transforming the World
Technical Breakthroughs
Cultural Insights
The Teams



A peaceful global community has always been a staple tenet in the IBM tradition. Thomas Watson Sr., the company's first chairman, had a vision of "World Peace through World Trade" and tried to make it a reality during the 1920s, promoting this agenda. By the middle of the twentieth century, IBM was acting locally and looking for ways to apply information technology to improve the safety and wellness of our communities.

These days, police departments around the globe are looking to "know what they know" about their communities. They want to perform their duties, but that information is not readily accessible, available or secure. IBM has been working closely with the law enforcement community to help departments collect, manage and deploy the vast amount of information at their disposal, in such a way that it can be accessed and used at moment's notice.



Le « docteur » Watson, officiel et diagnostique à tout va au Memorial Sloan Kettering de New York et dans d'autres établissements.

de l'informatique, il est également très intimement lié aux sciences cognitives, aux neurosciences, à la biologie ou encore à la psychologie et pourrait, en croisant les domaines précités aux biotechnologies et aux nanotechnologies, aboutir à des systèmes d'intelligence artificielle extrêmement performants. Ceci représente d'ailleurs un sujet d'inquiétude – une IA trop intelligente – pour quelques membres de la communauté scientifique tels que l'astrophysicien Stephen Hawking.

Les différents types d'apprentissage

Les algorithmes d'apprentissage peuvent se catégoriser selon le type d'apprentissage qu'ils emploient : supervisé ou non, partiellement ou semi-supervisé, par renforcement ou encore par transfert.

L'apprentissage supervisé

Lorsque les classes sont prédéterminées et les exemples connus, le système apprend à classer selon un certain modèle. C'est en simplifié ce que l'on appelle l'apprentissage supervisé – ou d'analyse discriminante. Un expert, appelé aussi oracle, doit au préalable étiqueter des exemples. Le processus se déroule en deux phases. Lors de la première phase, dite d'apprentissage ou hors ligne, il faut déterminer un modèle des données étiquetées. Dans la seconde phase, dite de test ou en ligne, l'étiquette d'une nouvelle donnée doit être prédite à partir d'un modèle connu. Il est parfois préférable d'associer une donnée non pas à une classe unique mais à une probabilité d'appartenance à chacune des classes prédéterminées. C'est ce qui s'appelle l'apprentissage supervisé probabiliste.

Segmentation par la méthode des *k*-means

La méthode des *k*-means est un outil de classification classique qui permet de répartir un ensemble de données en *k* classes homogènes. La plupart des images (photos, dessins vectoriels 2D, synthèses 3D, ...) vérifient localement des propriétés d'homogénéité, notamment en terme d'intensité lumineuse. L'algorithme des *k*-means permet donc d'apporter une solution à la segmentation d'images.

Présentation

La méthode des *k*-means a été introduite par J. McQueen [10] en 1971 et mise en œuvre sous sa forme actuelle par E. Forgy [5]. De nombreuses variantes se sont succédées depuis afin d'étendre ses capacités de classification (séparations non linéaires) : *kernel k-means* [13] (*k*-means basée sur des méthodes à noyaux), améliorer ses performances : *global k-means* [8], *k-Harmonic means* [14], automatiser le choix du nombre de clusters : *Gaussian-means* [6], *X-means* [11].

Dans le cadre de la classification non supervisée, on cherche généralement à partitionner l'espace en classes concentrées et isolées les unes des autres. Dans cette optique, l'algorithme des *k*-means vise à minimiser la variance intra-classe, qui se traduit par la minimisation de l'énergie suivante :

$$E = \frac{1}{2} \sum_{c \in \mathcal{C}} \sum_{x \in c} \|x - m_c\|^2 \quad (1)$$

$$= \frac{1}{2} \sum_{c \in \mathcal{C}} \#c \times V(c) \quad (2)$$

$$E = \frac{1}{2} \sum_{x \in \mathcal{D}} \min_{c \in \mathcal{C}} \|x - m_c\|^2 \quad (3)$$

Principe de la segmentation par la méthode des *k*-means sur le site de Telecom-ParisTech.

avec \mathcal{C} l'ensemble des clusters et pour chaque cluster c , m_c son centre (appelé noyau), $V(c)$ sa variance, $\#c$ le nombre de ses éléments et $\mathcal{D} = \bigcup_{c \in \mathcal{C}}$ l'ensemble des données que l'on cherche à classer.

L'analyse discriminante linéaire ou les SVM (Support Vector Machine, machines à vecteurs de support ou séparateurs à vaste marge) en sont des exemples typiques. En médecine, le système peut,

en fonction de points communs détectés avec les symptômes d'autres patients connus (les exemples), catégoriser de nouveaux patients au vu de leurs analyses médicales en risque estimé (probabilité) de développer telle ou telle maladie. N'est-il pas mon cher Watson ?

Open source et Machine Learning

Ce ne sont, heureusement, pas les éditeurs de logiciels, grands ou non, qui ont les clés du marché du Machine Learning. Ils possèdent néanmoins un avantage certain : ils ont les moyens matériels nécessaires pour l'intégration des données, notamment pour les grands groupes où il faut contrôler les flux de données, leur hébergement et leur fiabilité – c'est ce que l'on appelle la véracité des données. La véritable innovation réside dans les algorithmes, ce sont eux les futurs « rois » du Machine Learning, et sur ce plan les éditeurs commerciaux sont plutôt à la traîne par rapport aux communautés open source. Il ne faut pas oublier que les packages open source très performants ont fortement contribué à populariser le Machine Learning auprès des chercheurs. Des chercheurs de l'Inria utilisent l'outil qu'ils ont développé, Scikit-Learn, une toolbox Python de Machine Learning robuste et fiable. Ils l'ont mise à disposition de tous. Chercheurs, ingénieurs et industriels l'utilisent pour leurs projets. D'autres projets open source de ML existent, certains plus adaptés aux données massives comme Mahout et Spark. Bien que les gros éditeurs propriétaires – nous parlons du code source, pas de la mise à disposition, toujours sous caution – tels que IBM et SAS intègrent des algorithmes puissants à leurs solutions, ils ne séduisent pas pour autant les chercheurs. Il est par conséquent fort probable qu'à terme le marché du ML sera dominé par des packages open source, complétés par des services cloud, peut-être tels que Watson, l'un n'empêche pas l'autre. Plutôt que de se « maquer » avec un grand éditeur de A à Z – du ML au Cloud, disons –, il est toujours possible pour ceux ne disposant pas de ressources de calcul suffisantes de louer les services de Cloud tels que celui d'Amazon Web Services et d'y faire tourner des algorithmes open source.



L'apprentissage non supervisé

L'apprentissage non supervisé, appelé également classification automatique ou encore clustering, désigne le cas de figure où le système ou l'opérateur ne disposent que d'exemples, mais pas d'étiquettes, et que le nombre de classes ainsi que leur nature n'ont pas été prédéterminés. Aucun oracle n'est alors requis. C'est à l'algorithme de découvrir par lui-même la structure plus ou moins cachée des informations. Le partitionnement de données ou data clustering est un exemple typique d'algorithme d'apprentissage non supervisé. Les outils mathématiques, méthodes et logiciels utilisés peuvent être de simples tableurs comme Excel, les langages R, Python ou la méthode des *k*-means.

L'apprentissage semi-supervisé

Pouvant être réalisé de manière probabiliste ou non, l'apprentissage semi-supervisé vise à faire apparaître la

distribution sous-jacente des exemples dans leur espace de description. Il est mis en œuvre lorsque des données – ou étiquettes – sont manquantes. Le modèle doit utiliser des exemples non étiquetés mais néanmoins capables de fournir des renseignements. Il est employé en médecine afin de proposer une aide au diagnostic ou au choix des tests les moins onéreux.

L'apprentissage partiellement supervisé

Probabiliste ou non, l'apprentissage partiellement supervisé est utilisé lorsque l'étiquetage des données est partiel. C'est le cas quand un modèle énonce qu'une donnée n'appartient pas à une classe A, mais appartient peut-être à une classe B ou C – par exemple trois classes de maladie évoquées dans le cadre d'un diagnostic différentiel.

L'apprentissage par renforcement

Dans l'apprentissage par renforcement, l'algorithme apprend un comportement à partir d'une observation. L'action de l'algorithme sur l'environnement produira une valeur de retour qui guidera l'algorithme d'apprentissage. Un exemple classique en est l'algorithme de Q-learning, une classe

Microsoft Cortana Analytics Suite

Cortana Analytics

A fully managed big data and advanced analytics suite that enables you to transform your data into intelligent action.

WHY CORTANA ANALYTICS

Analytics that enables action	Perceptual intelligence	Fast and flexible	Secure and scalable	Personal digital assistant
Take action ahead of your competitors by going beyond looking in the rear-view mirror to predicting what's next.	Get closer to your customers. Infer their needs through their interaction with natural user interfaces.	Build with partner solutions for your industry to get going quickly and tailor to your specific needs.	Get value from your data in a secure and scalable way regardless of its size and complexity.	Get things done with Cortana in more helpful, proactive, and natural ways.

Cortana, de Microsoft, n'est pas qu'une voix féminine agréable mais aussi une suite analytique de ML.

d'apprentissage par renforcement ne nécessitant aucun modèle initial de l'environnement.

L'apprentissage par transfert

L'apprentissage par transfert désigne la capacité d'un système à reconnaître et appliquer des connaissances et des compétences apprises à partir de tâches antérieures sur de nouvelles tâches ou domaines partageant des

similitudes. Deux questions cruciales se posent avec cette technique : comment identifier les similitudes entre les tâches cibles et sources et comment transférer la connaissance des tâches sources vers les tâches cibles ? Les réponses ne sont pas toujours évidentes, les questions non plus d'ailleurs.

Big Data et Machine Learning

Le Big Data est lui aussi en pleine expansion. L'essor du Machine Learning s'explique notamment par la disponibilité de données en très grands volumes et le désir de les interpréter à diverses fins. Des moteurs de recherche, sites et logiciels divers – même quand ils n'en ont pas l'air – collectent des quantités phénoménales d'informations sur tout et n'importe quoi, sur toute la Planète, votre parcours de navigation sur Internet, vos préférences, votre configuration, des statistiques météo, la fréquence des pluies de sauterelles, l'historique des crises de toutes sortes, les réseaux sociaux, les objets connectés et bien d'autres choses encore. Toutes ces données sont passées à la moulinette de puissants algorithmes « magiques » afin d'obtenir des

Amazon Machine Learning

Amazon Machine Learning, au dire même de son créateur, est un service géré de conception de modèles d'apprentissage-machine et de génération de prédictions permettant de développer des applications robustes, évolutives et intelligentes. AML permet d'utiliser une technologie d'apprentissage-machine puissante sans, disons, mettre les mains dans le cambouis. Le processus de conception de modèles d'apprentissage-machine se décompose en trois opérations : analyse des données, formation des modèles et évaluation. L'étape d'analyse des données calcule et visualise la distribution des données et suggère des transformations afin d'optimiser le processus de formation du modèle. L'étape de formation du modèle recherche et stocke les schémas prédictifs au sein des données transformées. Dans l'étape finale, facultative, la précision du modèle est évaluée. Amazon Machine Learning combine des algorithmes d'apprentissage-machine à des outils visuels interactifs destinés à faciliter la création, l'évaluation et le déploiement des modèles de ML. Une fois qu'un modèle a été conçu, la console de paramétrage et l'outil d'évaluation de modèle intuitifs aident à en comprendre les atouts et les faiblesses et à ajuster ses performances.

The screenshot shows the Dataiku platform interface with sections for Machine Learning, Data Mining, Data Visualization, and Big Data. It includes examples like 'A Guided Path Through The Machine Learning Maze' and 'Automatic Feature Engineering'.

Dataiku propose divers services d'analyse dans le domaine du Machine Learning.

indicateurs et des recommandations sur les comportements, les profils et les tendances des personnes, des groupes et même des États, si tant est que ces conclusions soient véritables. Ces traitements se font en principe sans le moindre jugement arbitraire humain, si l'on fait abstraction des choix de tel ou tel algorithme, et donc de manière beaucoup plus fiable et efficace d'après ses aficionados. Big Data et Machine Learning vont donc de pair. Les entreprises vont aussi devoir intégrer dans leurs modèles prédictifs des données très hétérogènes, voire incomplètes. C'est le cas dans la maintenance prédictive, le monitoring d'infrastructures complexes ou de réseaux de transport.

D'après la récente légende, ces programmes d'Intelligence artificielle peuvent prédire l'efficacité de médicaments avant même d'avoir effectué le moindre essai clinique, prévoir des révoltes ou bouleversements politiques à venir ou déterminer le vainqueur d'une Coupe du monde avant le premier match. Les voyantes et autres pythons n'ont qu'à bien se tenir, ou conserver des tarifs raisonnables, car les IA ne seront peut-être pas à la portée de toutes les bourses... à moins qu'elles ne soient accessibles librement à tous ?

Lokad travaille plutôt dans la classification, l'optimisation de stocks et d'inventaire et consorts.

The screenshot shows the Lokad homepage with sections for 'Getting started', 'Automation', 'Scripting with Envision', and 'Envision user guide'. It also lists 'Data import' and 'Inventory optimization' services.

Les principaux acteurs du Machine Learning

Dans le domaine du ML, les trois grands acteurs qui se partagent le fromage sont sans conteste IBM, Google et Microsoft. Ils bénéficient tous trois de leur statut d'éditeur mondial ainsi que d'une forte empreinte dans les domaines du Cloud Computing et du Big Data. Le positionnement de Microsoft sur le sujet se fait sur divers plans : sur la partie système d'exploitation avec son assistante Cortana et ses capacités d'apprentissage, dans le

R&D

The Watt-Sun Program Is Poised to Improve Weather Forecasting for Solar Energy

August 2014 | by Jim Utley



A 20 megawatt solar farm in Tucson, Arizona, is one of the test sites for Watt-Sun. Solar-power panels seem to be popping up on both residential and commercial rooftops like Monet's mushrooms in the spring. And massive solar farms are being built in sun-blessed areas of the world. This is great news for renewable-energy proponents, who see solar—along with wind, waves and tides—as a way to wean the world off fossil fuels.

Le projet Watt-Sun d'IBM, censé améliorer le rendu des systèmes de production d'énergie solaire.

Cloud avec les API d'Azure qui donnent accès à des algorithmes prédictifs et avec la suite Office 365 et son nouvel outil Delve. Google est lui aussi présent sur ce marché avec ses API cloud.

IBM est bien connu pour ses recherches en matière d'Intelligence artificielle, recherches qui ont notamment engendré la technologie Watson. Là encore, IBM propose tout un jeu d'API donnant accès à de multiples applications de Machine Learning. Les fonctionnalités accessibles concernent par exemple le travail sur les langues (détection de langues, traduction contextuelle...) ou la modélisation de l'utilisateur avec la prédition de ses caractéristiques sociales. Le domaine médical représente une autre application de ML chez IBM, où l'infatigable Watson délivre une aide au diagnostic du cancer pour différents hôpitaux américains. Big Blue, qui ne dort jamais, a aussi construit un système complet de ML permettant d'améliorer le rendu des systèmes de production d'énergie solaire, le programme Watt-Sun. Le grand Bleu souhaite le mettre à disposition des acteurs du secteur.

Bien que plus modestes que ces géants de l'informatique, deux acteurs français œuvrent dans le domaine du ML : Dataiku et Lokad. ☀

THIERRY THAUREAUX

Du Machine Learning pour Wikipedia

La fondation Wikimedia a décidé d'installer une intelligence artificielle afin de lutter contre le vandalisme. Dans le but notamment de garder une haute qualité des articles publiés sur Wikipedia, la Wikimedia Fondation déploie un tout nouveau service baptisé ORES, pour Objective Revision Evaluation Service. Ce système à base d'open data et de Machine Learning open source devrait permettre d'évaluer les nouvelles contributions en vue d'y détecter d'éventuelles tentatives de vandalisme littéraire. Il aidera de plus les nouveaux contributeurs afin de leur éviter de réaliser des modifications parmi les plus dommageables. Ce service n'est disponible pour le moment que pour les versions anglaise, portugaise, turque et perse mais devrait rapidement s'étendre aux autres langues.

The screenshot shows the ORES homepage with sections for 'Content page', 'Discussion', 'Recent changes', 'Recent changes talk', 'Help', 'Batch', 'Comments', 'Wikidata', 'Feedback', 'Report bugs', 'Research', and 'Phabricator'. It also includes a 'Powered by Wikia' footer.

ORES, un tout nouveau service anti-vandalisme et d'aide aux contributeurs novices pour Wikipédia et les autres projets de la fondation Wikimedia.

À lire dans le prochain numéro, les meilleurs outils open source pour le Machine Learning.