

Détection de la maladie de Parkinson

La maladie de Parkinson-décrite par James Parkinson en 1817-est une maladie neurodégénérative caractérisée par la dégénérescence des neurones dopaminergiques. Puisque c'est une maladie idiopathique, ses causes explicites demeurent inconnues, d'où la nécessité de trouver une corrélation mathématique entre ses attributs et il se trouve que le «Machine Learning» excelle dans ce type de tâche.

Je propose de construire des modèles basés sur le Machine Learning pour prédire à postériori la probabilité qu'un individu soit atteint afin de commencer au plus tôt le suivi clinique et cela s'inscrit parfaitement dans le thème de cette année intitulé : Santé et Prévention.

Positionnement thématique (ETAPE 1)

INFORMATIQUE (Informatique Théorique), MATHÉMATIQUES (Mathématiques Appliquées), MATHÉMATIQUES (Analyse).

Mots-clés (ETAPE 1)

Mots-Clés (en français)	Mots-Clés (en anglais)
<i>Apprentissage automatique</i>	<i>Machine Learning</i>
<i>La maladie de Parkinson</i>	<i>Parkinson's disease</i>
<i>Classification</i>	<i>Classification</i>
<i>Naïve Bayes</i>	<i>Naïve Bayes</i>
<i>K plus proches voisins</i>	<i>K-Nearest Neighbors</i>

Bibliographie commentée

Le Machine Learning (ou apprentissage automatique mais le premier terme est largement plus utilisé) est une sous-partie de l'intelligence artificielle qui s'intéresse à créer des algorithmes qui se comportent et opèrent de manière intelligente ou qui simulent cette intelligence, la première définition a été donnée par Arthur Samuel en 1959 : « The field of study that gives computers the ability to learn without being explicitly programmed » Ce dernier a créé en 1952 un programme capable de jouer aux dames et de s'améliorer continuellement après chaque partie jouée[1]. En 1958 Frank Rosenblatt a conçu le premier réseau neuronal pour ordinateurs (le Perceptron), qui simule les processus de pensée du cerveau humain[2]. C'est à partir de là que l'ère du Machine Learning commença et que les modèles utilisés aujourd'hui ont été fondés, en commençant avec Joseph Berkson en 1944 qui a introduit le modèle de régression logistique[3], par la suite, Vladimir N. Vapnik and Alexey Ya. Chervonenkis en 1963 ont conçu le modèle SVM (Machine à vecteurs de support)[4]. Avec l'arrivée d'internet et la quantité phénoménale de données qui ont été générées (Big Data), le Machine Learning a connu son plus grand essor, c'est ainsi qu'en 1967 Thomas Cover and Peter E. Hart de Stanford ont publié un article concernant l'algorithme du plus proches voisins « KNN »[5], suivi par Robert Schapire qui a publié en 1990 un article intitulé « The Strength of Weak Learnability » pour introduire le concept de « Boosting » qui consiste à agglomérer plusieurs algorithmes faibles afin de donner naissance à un algorithme plus performant[6]. En se basant sur

ce principe, Tin Kam Ho a présenté dans une conférence en 1995, les Forêts de décisions aléatoires (Random decision forests)[7]. Deux ans après, L'ordinateur d'échecs d'IBM, «Deep Blue », bat le champion du monde Garry Kasparov aux échecs, une panique générale et un signe que l'intelligence artificielle rattrape à grande vitesse l'intelligence humaine[8]. En 2016, Lee Sedol l'un des meilleurs joueurs des années 2000 du Go -un jeu de stratégie abstraite bien plus complexe que les échecs avec 2.08×10^{170} positions légales- a annoncé sa retraite après être battu 4 fois sur 5 par le programme de Google « AlphaGo » [9], qui a ensuite enchaîné les victoires notamment contre Ke Jie (meilleur joueur en 2017), Mais le champ d'application du Machine Learning s'étend bien sûr au-delà des jeux. Grâce aux progrès technologiques, il s'est petit à petit fait une place dans de nombreux domaines. Aujourd'hui, on peut le retrouver dans les moteurs de recherche, la vision artificielle, les chatbots, la médecine, les transports autonomes...

En ce qui concerne mon étude, je vais essayer de prédire si un individu souffre ou va souffrir de la maladie de Parkinson à partir des données d'individus sains et malades. Grâce au modèle de Naïve Bayes qui s'appuie sur le théorème de Bayes et sur l'indépendance des données et d'après le modèle des K plus proches voisins (KNN) qui se base sur le calcul des distances entre les données, je vais générer deux algorithmes différents qui vont classer les patients, dans le but de comparer la précision des deux modèles.

Problématique retenue

Connaître les paramètres qui influencent la précision des modèles utilisés connue sous le nom d'Hyperparamètres, en étudiant les fondements mathématiques des algorithmes et les appliquer dans le cas de la maladie de Parkinson.

Objectifs du TIPE

- Déterminer les facteurs de risques les plus probables du Parkinson.
- Classifier les patients malade ou non malade avec grande exactitude.
- Réaliser un comparatif entre le modèle de Naïve Bayes et le modèle de KNN.
- Construire un troisième modèle plus performant à partir des deux autres modèles pour répondre à la problématique.

Références bibliographiques (ETAPE 1)

- [1] ARTHUR SAMUEL : Some Studies in Machine Learning Using the Game of Checkers : *IBM Journal of Research and Development*, Vol.3, Issue : 3, Pages : 210–229, Juillet 1959
- [2] FRANK ROSENBLATT : The Perceptron : A probabilistic model for information storage and organization in the brain : *Psychological Review*, Vol. 65, No. 6, 1958.
- [3] JOSEPH BERKSON : Application of the Logistic Function to Bio-Assay : *Journal of the American Statistical Association*, Vol. 39, No.227, Pages : 357–365, Septembre 1944
- [4] BERNHARD SCHOLKOPF, ZHIYUAN LUO, VLADIMIR VOVK : Empirical Inference : Festschrift in Honor of Vladimir N. Vapnik : *Berlin Springer*, Chapitre : *Early History of Support Vector Machines*, Pages : 13–20, 2014
- [5] THOMAS COVER, PETER E. HART : Nearest neighbor pattern classification : *IEEE*

Transactions on Information Theory, Vol.13, Issue : 1, Pages : 21–27, Janvier 1967

[6] ROBERT SCHAPIRE : The Strength of Weak Learnability : *Machine Learning, Vol.5, Issue : 2, Pages : 197–227, Juin 1990.*

[7] TIN KAM HO : Random decision forests : *Proceedings of 3rd International Conference on Document Analysis and Recognition, 14–16 Août 1995, Montréal, QC, Canada*

[8] HISTORY.COM EDITORS : Deep Blue defeats Garry Kasparov in chess match :
<https://www.history.com/this-day-in-history/deep-blue-defeats-garry-kasparov-in-chess-match>,
dernière consultation en février 2022.

[9] THE ATLANTIC : How Google's AlphaGo Beat a Go World Champion :
<https://www.theatlantic.com/technology/archive/2016/03/the-invisible-opponent/475611/>,
dernière consultation en février 2022.