

PYTHON FOR DATA ANALYSIS : HUMAN ACTIVITY RECOGNITION WITH SMARTPHONES

ALEXANDRE BROUTIN



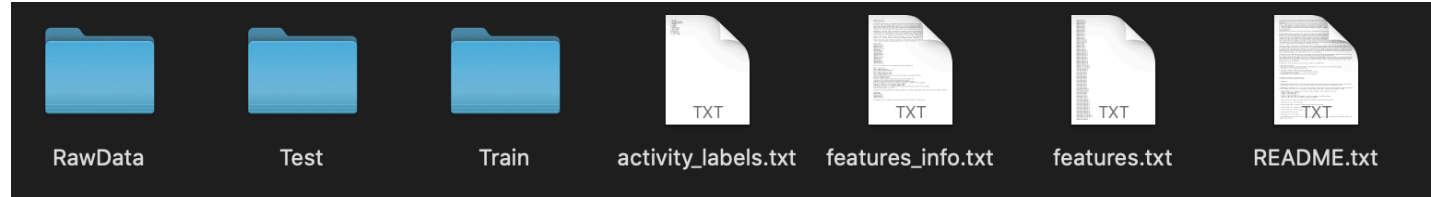
LES TENANTS ET ABOUTISSANT DU PROBLÈME



- Le but de ce projet était de réaliser une classification.
- Une base de données était fournie par l'UCI, mon sujet était "Human Activity Recognition with Smartphones". Il consistait à analyser les déplacements de 30 individus qui possédaient tous un téléphone, à l'aide d'un gyroscope et d'un accéléromètre. Ces 30 personnes réalisaient différentes actions dont 3 postures statiques (standing, sitting, laying) et 3 actions dynamiques (walking, walking_downstairs et walking_upstairs).
- Elle avait pour but de me fournir les données nécessaires pour classer l'activité de différentes personnes.



DESCRIPTION DATASET



- Le dossier téléchargé était composé de 3 autres dossiers et de 4 fichiers à sa base.
- On peut y voir les dossiers Train et Test, ce sont les données que j'ai utilisé. Nous avons aussi le dossier RawData, mais il contenait beaucoup de fichiers qui n'étaient pas ordonnés, normalisés et donc non plus concaténés...
- Les dossiers train et test étaient donc composés des valeurs de X, de y mais aussi de l'id des sujets. Au total, ils contenaient 563 colonnes et 10929 lignes.
- Nous avons ensuite le fichier activity_labels qui regroupe toutes les valeurs que peut prendre les différentes valeurs de y. y correspond à une action, une pose ou une transition de l'une à l'autre. Il y avait au total 12 possibilités. Vous allez pouvoir les voir dans le notebook. J'ai voulu les remplacer à l'intérieur de mes Dataframes pour que cela soit plus clair mais mon ordinateur n'arrivait pas à réaliser le calcul (L'algorithme est en commentaire dans le notebook).
- Les fichiers suivants étaient concernés en grande partie par les 563 colonnes dont je vous parlais ... features.txt et la liste de toutes ces colonnes, je vais m'en servir pour labelliser mon Dataframe. features_info.txt nous permet de mieux comprendre ces valeurs.
- Elles étaient nombreuses et pas évidentes à comprendre, j'ai mis réellement beaucoup de temps à savoir dans quel direction j'allais partir. Finalement, j'étudie X dans toute sa globalité, je n'étudie pas de colonnes en particulier. L'avantage était que je n'avais pas à en créer, en dehors de données temporelles.

REFLEXIONS

- Au niveau des données, j'ai pu réaliser quelque visuels intéressant qui nous permettent d'interpréter les données.
- A l'aide de Tnse, j'ai pu visualiser les activités et les déplacements des sujets avec des dimensions adaptés.
- J'ai aussi pu réaliser des histogrammes représentant d'une part si les sujet montent ou descendent plus d'escaliers, on a pu remarqué que les personnes avaient plus tendances à les descendre mais aussi à descendre plus rapidement que les monter. D'autre part, un autre histogramme m'a permit d'analyser le rapport des marches ascendantes/descendantes utilisées par les participants. Cela peut représenter leurs conditions physiques, ce qui pourrait être par leurs âges ou d'autres facteurs...
- Une fois que j'ai pu comprendre la base de données, j'ai pu avoir pas mal d'idées mais je n'ai pas pu toutes les réalisés même avec des recherches.
- J'ai ensuite pu voir qu'en appliquant nos modèles, j'obtenais de bonnes prédictions ce qui montre que ces données pouvaient être réellement utiles pour prédire leurs activités.

PROBLÈMES RENCONTRÉS

- Compréhension du Dataset – Comme dit précédemment, j'ai réellement mis du temps à trouver vers quelles directions me tournées. Je me suis bloqué avec la quantité de features et la compréhension du sujet.
- Réseaux de neurones et import de tensor flow – J'ai voulu appliquer un réseau de neurones que j'ai pu réaliser lors de l'un de mes projets cette année, mais je n'ai pas pu importer tensor flow... Ce n'est pas grand chose mais cela m'a bloquer pour justement « pas grand-chose ».
- De la RAM pour réussir à remplacer mes valeurs par les activity_labels – j'ai voulu clarifier mes visuels en remplaçant les entiers dans de mes valeur Y par leurs valeurs en string mais j'ai été bloqué par mon ordinateur ... J'ai voulu les remplacer directement dans le fichier, mais j'ai été confronté à un problème de type.

CONCLUSION

- Ce que j'ai réalisé – Ce projet m'a réellement permis d'avoir une réflexion sur une dataset en particulier sur ce sujet, j'ai pu réaliser des graphiques, diagrammes et même appliqués des modèle de Machine Learning/Deep Learning qui m'ont permis d'encore plus comprendre les possibilités et l'importance des données.
- Ce que j'ai appris – C'est la première fois que nous réalisons un projet aussi complet la manipulation de données et l'intelligence artificielle. Cela nous a réellement permis de nous mettre en situation et de se poser les bonnes questions. J'ai également prit énormément de compétences en Python et surtout appliqué au ML/DL, sans oublier toutes les librairies qui l'accompagne ... Cela m'a notamment encore plus encourager à travailler les données et les IA (Sans oublier le Python 😊).
- Ce que j'aurai encore pu réalisé – Je regrette quand même de ne pas avoir pu aller encore plus loin, en réalisant techniquement les autres idées que j'ai pu avoir sur cette base de donnée comme pouvoir afficher le parcours de tous les participants. Je l'ai commencé mais je n'ai pas pu le finaliser ... Si je peux donc donner une perspective d'amélioration pour ce projet, ce serait de choisir des datasets nécessitant moins de compréhension et plus de manipulation. Je regrette aussi de ne pas avoir pu réaliser l'API Django, même si c'est un domaine que je connaissais déjà mieux.