

TD 1 - Fusion de données

Sébastien Bindel et Mustafa Al Samara

Semestre 2

I. Synopsis

Il vous ait proposé dans cette activité de découvrir une structure de données complexe mais très utilisée à savoir l'objet DataFrame. Cette structure est disponible avec la bibliothèque Pandas et sera utilisée pour stocker les données relatives à des mesures de consommation énergétique d'un dispositif de transmission Lora qui est une technologie similaire au Wi-Fi mais avec une portée de communication supérieure, un débit plus faible. L'objectif sera de concevoir un module dont le but est de pouvoir agréger les fichiers et de modifier les jeux de données.

II. Présentation du jeux de données

Les données sont réparties dans huit fichiers, chacun est associé à un ensemble exclusif de paramètres de transmission qui sont les suivants :

- le facteur d'étalement (valeur à 7 ou 12),
- la bande passante, fixée à 125 kHz,
- le taux de bits redondants : 4/5, 4/6, 4/7, 4/8.

Le choix de ces paramètres permet de voir leur influence sur le temps d'une transmission LoRa et donc la consommation du dispositif. Chaque fichier contient des données dont la structure est identique et décrite dans le tableau 1.

Champs	Type	Description
Timestamp (S)	Décimaux	instant de la prise de mesure
File Main Current (A)	Décimaux	courant consommé
File Main Voltage (V)	Décimaux	tension appliquée
File Main Energy (J)	Décimaux	énergie consommée
UART	Chaîne de caractères	information générée par le dispositif

TABEAU 1 – Description de la structure des données

Exercice 1

Comment sont nommés les fichiers contenant les jeux de données ?

III. Agrégation des jeux de données

Dans cette partie, il vous est présenté comment agréger des données issues de plusieurs fichiers. La première étape consiste à extraire les informations dans le nom des fichiers afin de nommer le fichier final. La seconde consiste à lire les données, à déterminer le type de chaque variable chargée et d'ajouter le taux de redondance des bits pour enfin fusionner les données.

III.1. Extraction des informations dans le nom des fichiers

Exercice 2

Créer la fonction `get_sf(string)` qui prend en paramètre le nom du fichier et qui renvoie la valeur du facteur d'étalement contenue dans le nom de fichier.

Exercice 3

Créer la fonction `get_bw(string)` qui prend en paramètre le nom du fichier et qui renvoie la valeur de la bande passante contenue dans le nom de fichier.

Exercice 4

Créer la fonction `get_cr(string)` qui prend en paramètre le nom du fichier et qui renvoie la valeur de la bande passante contenue dans le nom de fichier.

III.2. Lecture du jeux de données

Exercice 5

Créer la fonction `load_data(string)` qui récupère les données d'un fichier csv dont le nom est donné en paramètre. La lecture du fichier se fera avec la commande `read_csv()` de la bibliothèque Pandas.

III.3. Modification des jeux de données

Exercice 6

En appelant l'attribut `dtypes` de votre DataFrame regarder le type de chaque variable (UART object). Modifier la variable UART en chaîne de caractères.

Exercice 7

Renommer l'entête de votre DataFrame en `Time(S)`, `Current(A)`, `Voltage(V)`, `Energy(J)`, `UART`.

Exercice 8

Ajouter une variable CR dans votre Dataframe dont la valeur a été extraite du nom du fichier.

III.4. Sauvegarde des jeux de données

Exercice 9

Sauvegarder vos données dans un fichier csv où le format du nom est le suivant : dataset-bw-sf.csv

III.5. Agrégation automatique des jeux de données

Exercice 10

Écrire un programme qui lit l'ensemble des fichiers pour les concaténer en deux fichiers, chacun relative à un facteur d'étalement.