

Organigramme

Sonny Klotz - Jean-Didier Pailleux - Malek Zemni

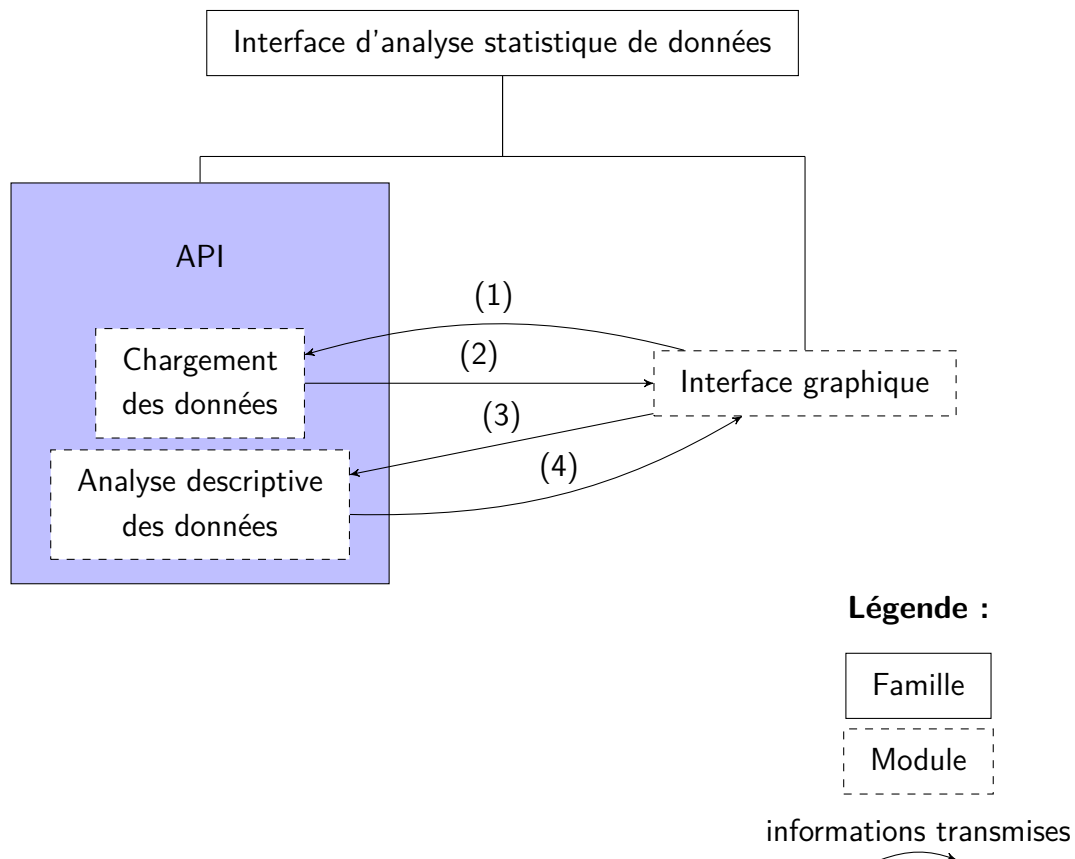
*Interface de chargement, de contrôle
et d'analyse statistique des données
pour la constitution d'un graphe de flux*

08/03/2017

1 Présentation de l'organigramme

Le travail peut être décomposé en 3 modules :

- Une interface graphique : va faire transiter les informations entre les différentes APIs de l'application et fournir des rendus visuels des analyses
- Une API de chargement des données : va effectuer des vérifications sur les données fournies (.csv) puis retourne le résultat à l'interface graphique
- Une API d'analyse descriptive des données : va permettre de calculer les informations statistiques nécessaires



Notes :

- (1) Fichier .csv formaté
- (2) Structure décrivant les données du fichier .csv après vérification : nombre de lignes et de colonnes, entrées erronées, noms des colonnes...
- (3) Colonne de données d'un type attendu
- (4) Résultats statistiques (moyenne, quantiles, écart-type, ...)

FIGURE 1 – Organigramme des différents modules du logiciel

2 Fonctionnalités des modules

2.1 API

Application Programming Interface constituent les paquets utilisables par les développeurs (intégrée), qu'on va livrer au client en plus de l'application elle-même.

Parmi les modules du logiciel, on a 2 APIs :

- Une API de chargement des données
- Une API d'analyse statistique des données

2.1.1 Chargement des données

Le module de chargement des données s'occupe de la préparation et de la vérification du fichier fourni en entrée. Ces principales fonctionnalités sont :

- Ouverture du fichier en vérifiant qu'il a la bonne extension `.csv` et qu'il est accessible en lecture.
- Donner le nombre de lignes et de colonnes
- Détection des type de colonnes : ça sera plutôt une détection de la cohérence des types, on parcourt les entrées de chaque colonne et on vérifie si leur type correspond bien au type attendu.

Dans le même parcours, on pourra aussi gérer les erreurs :

- détecter les valeurs erronées (valeurs qui ne correspondent pas au type attendu)
- détecter les cases vides
- Nommer les colonnes.

En sortie, on aura donc une structure décrivant les données du fichier `.csv` après vérification : nombre de lignes et de colonnes, entrées erronées, noms des colonnes...

2.1.2 Analyse descriptive des données

Ce module est utilisé pour fournir des informations de statistiques descriptives sur les colonnes de données qui lui seront fournies. Ses fonctionnalités se décomposent en trois pour les différents types de données à traiter :

1. qualitatif : calcul des effectifs et fréquences d'apparition
2. quantitatif discret :
 - effectifs, effectifs cumulés, fréquence et fréquence cumulée
 - moyenne
 - médiane et autres quantiles
 - variance et écart-type

- anomalies : boîte à moutaches de Tukey
 - symétrie : coeff de Pearson ou coeff de Yule
 - aplatissement : coeff de Fisher
3. quantitatif continu : on regroupe les valeurs en classe d'intervalles, on peut ensuite appliquer les mêmes techniques que pour les variables discrètes

2.2 Interface graphique

Le module de l'interface graphique s'occupera de la manière dont l'application sera représentée à l'écran pour l'utilisateur, ce qui correspond au positionnement des éléments textuels, boutons et des fonctionnalités disponible dans une fenêtre. Voici une liste des principales fenêtres composant ce module avec leurs fonctionnalités :

1. Fenêtre de chargement pour récupérer un fichier `.csv` avec validation du choix pour passer à la prochaine fenêtre (En renseignant son chemin dans le système de fichier, ou de la manière d'un Drag & Drop).
2. Fenêtre de contrôle préliminaire avec visualisation d'un échantillon du `.csv`.
 - Affichera le nombre de lignes/colonnes contenu dans le `.csv`.
 - Affichera le titre du fichier.
 - Affichera un échantillon du contenu du `.csv` (environ les 1000 premières lignes) avec un système de scroll.
 - Affichage des lignes erronés (numéro de la ligne + contenu + type d'erreur).
 - Mise en place d'un système de navigation sous forme d'onglet (Onglet erreurs, onglet échantillon,...). Cela permettra d'éviter que la fenêtre contienne trop d'informations.
 - Affichage d'un bouton pour passer à l'étape de l'analyse de données.
3. Fenêtre correspondant à l'étape de l'analyse de données.
 - L'Utilisateur devra sélectionner la colonne avec un clic, puis pourra lancer l'analyse sur celle-ci.
 - La fenêtre affichera les résultats de l'étude qualitative (Médiane, Quantile et anomalie) d'une part et de l'étude quantitative (Histogramme et Diagramme de secteur) d'autre part.
 - Un bouton pour lancer l'exportation des résultats sera disponible (Écriture dans un nouveau fichiers).

3 Sources

- <http://www.math.univ-toulouse.fr/~baccini/zpedago/asde.pdf>
- <http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-l-des-uni.pdf>
- <http://iml.univ-mrs.fr/~reboul/cours2.pdf>
- <http://commons.apache.org/proper/commons-math/userguide/stat.html>
- <https://hal.archives-ouvertes.fr/halshs-00287751/document>
- Wikipédia : Données abérantes.