

Utilisation d'un journal alimentaire numérique à champ libre : évaluation préliminaire de l'exploitabilité des données récoltées avec une approche semi-supervisée.

Sema Akkoyunlu¹, Nicolas Darcel², Cristina Manfredotti¹, Antoine Cornuéjols¹, Fabien Delaere³

[1]UMR MIA-Paris, AgroParisTech, Inra, Université Paris-Saclay; [2]UMR PNCA, AgroParisTech, Inra, Université Paris-Saclay; [3]Danone Nutricia Research, Palaiseau

Un carnet alimentaire permet une analyse fine des consommations et de leur dynamique mais constitue un processus long et fastidieux pour les sujets comme pour l'expérimentateur. De plus en plus d'outils numériques ergonomiques facilitent l'acquisition des données par le sujet, mais sont souvent dénués d'une classification d'aliments hiérarchisée, ou sont assimilables à un carnet ouvert. Une analyse des aliments ou des nutriments ingérés requiert alors une phase lourde de codage des aliments par l'expérimentateur. L'objectif de cette étude est d'évaluer la performance d'un algorithme de codage automatique des aliments issus d'un carnet alimentaire électronique ouvert.

Les données de consommation alimentaires sont récoltées à l'aide d'une application mobile, *MyFitnessPal*. L'utilisateur note ses consommations librement ou peut scanner le code barre de l'aliment pour obtenir son nom commercial. La classification CIQUAL (Anses) est organisée suivant une hiérarchie à 4 niveaux : groupe, sous-groupe, sous-sous-groupe et aliment. Notre objectif est d'assigner un code à chaque acte de consommation - en priorité, un code d'aliment puis de sous-sous-groupe, de sous-groupe et enfin de groupe.

L'algorithme consiste à utiliser le nom du produit (exemple : "Danone Activia") comme requête pour trouver les aliments similaires selon leur dénomination dans la base CIQUAL. Les résultats de la requête sont classés selon un score de pertinence basé sur les occurrences des mots. Si la majorité des résultats font partie d'un même sous-sous-groupe, alors nous assignons le code correspondant à l'aliment. Sinon, nous passons au sous-groupe puis au groupe. Si aucun consensus n'est obtenu, l'aliment ne peut être catégorisé. Dans ce cas, la base OpenFoodFacts, base *open source* de produits alimentaires non hiérarchisée, est utilisée. Si les informations sur la catégorie existent (exemple : {"produits laitiers", "yaourt"}) alors nous les utilisons pour trouver la catégorie correspondante dans la base CIQUAL. Les catégories proposées par l'algorithme ont ensuite été vérifiées manuellement pour en estimer la performance.

Un journal tenu durant un mois contenant 1474 actes de consommation comptant 442 aliments différents est utilisé. En combinant les résultats sur les deux bases, nous obtenons 431 correspondances (97%) dont 337 (78%) ont été jugées bonnes a posteriori. Parmi les bonnes correspondances, 46 (14%) sont des codes aliment, 156 (46%) des codes sous-sous-groupe, 120 (36%) des codes sous-groupe et enfin 15 (4%) des codes groupe. Les 11 (2%) non-correspondances correspondent à des aliments mal orthographiés ou inexistant dans les deux bases.

En limitant les actions de recherche par le sujet, l'utilisation d'un carnet alimentaire électronique à champ libre est une alternative intéressante aux outils actuels de recueil alimentaire. Le problème du codage des aliments peut être allégé en utilisant un algorithme

de codage automatique. Ce dernier peut être amélioré en tenant compte des fautes d'orthographe.

1. <https://ciqua.anses.fr/>
2. <https://fr.openfoodfacts.org/>