

Treillis de Galois et Analyse Formelle de concepts

Découverte de l'Analyse Relationnelle de Concepts avec RCAexplore

1 Présentation

Dans ces travaux, nous utiliserons comme dans le TP précédent :

- RCAexplore que vous trouverez à l'URL <http://dataqual.engees.unistra.fr/logiciels/rcaExplore>, pour construire des treillis ou des AOC-posets.
- Open office, Libre office ou Excel : Il est en effet recommandé de saisir vos données brutes dans une feuille de calcul afin de les manipuler avec facilité puis de les transformer dans les formats d'entrée du logiciel cible.

Vous allez construire un modèle pour l'Analyse Relationnelle de Concepts ainsi constitué :

- un contexte formel de restaurants (description : par ex. lieu, région, taille ... à choisir)
- un contexte formel de pizzas (description : par ex. prix, forme, ... à choisir)
- un contexte formel d'ingrédients (description : par ex. catégories, exotique ou non, sauce ou non, régime alimentaires acceptant / n'acceptant pas l'ingrédient, ... à choisir)
- un contexte relationnel connectant les restaurants aux pizzas qu'ils servent
- un contexte relationnel connectant les pizzas à leurs ingrédients

Vous rédigerez un compte-rendu d'expérience incluant toutes vos données et tous vos résultats. Il peut être rendu par groupes de 2 ou 3 étudiants, il doit alors indiquer tous les noms des participants du groupe. Donnez des noms compréhensibles à vos fichiers en les numérotant par étape de travail.

2 Collecte du jeu de données

Vous devez rassembler :

- votre fichier de données pour les deux restaurants de pizzas du précédent TP
- des fichiers de données récupérés sur Moodle afin d'avoir 5 restaurants de plus :
 - Gosset et al. et/ou Tornil et al. vous permettront d'avoir des données sur Pizza Hut et/ou sur l'Annexe et/ou sur Dominos (suivant votre fichier de départ)
 - Ramzi et al. vous permettra d'avoir des données sur Fred Pizza et sur Kosmos
 - Montet et al. vous permettra d'avoir des données sur Gepetto et Terrasse
- Vous constituez avec tout cela des données avec 7 restaurants et vous pouvez réduire à une trentaine de pizzas (en les répartissant entre les restaurants). Il se peut qu'il vous soit plus facile et surtout plus intéressant de tout garder, nous verrons en séance de TP. Commencez par tout garder.

Compte-rendu : Faites la liste des 7 restaurants et indiquez combien de pizzas par restaurant vous considérez. Indiquez si vous avez gardé ou non toutes les pizzas. Donnez les 4 fichiers rcft de départ avant toute modification et après modification si vous avez réduit quelque chose.

3 Alignement des jeux de données

Le premier problème à résoudre est que les pizzas et les ingrédients n'ont pas forcément été décrits de manière uniforme. Dans cette étape, vous allez aligner les dénominations et les descriptions en utilisant le maximum d'informations venant de tous les ensembles de données que vous avez récupérés.

Compte-rendu : Une section de votre compte-rendu décrira les difficultés que vous avez rencontrées et les choix d'alignement que vous avez effectués.

4 Constitution de la famille de contextes relationnels

A cette étape, vous produisez la famille de contextes relationnels :

- un contexte formel de restaurants (description : par ex. lieu, région, taille ... à choisir)
- un contexte formel de pizzas (description : par ex. prix, forme, ... à choisir)
- un contexte formel d'ingrédients (description : par ex. catégories, exotique ou non, sauce ou non, régime alimentaires acceptant / n'acceptant pas l'ingrédient, ... à choisir)
- un contexte relationnel connectant les restaurants aux pizzas qu'ils servent
- un contexte relationnel connectant les pizzas à leurs ingrédients

Compte-rendu : Une section de votre compte-rendu décrira les difficultés que vous avez rencontrées éventuellement, donnera des métriques (nombre de lignes et de colonnes de chaque table et densité (nombre de croix sur le nombre de lignes multiplié par le nombre de colonnes)). Vous rendez à côté le fichier RCFT produit contenant ses 5 contextes.

5 Construction des structures

A cette étape, vous explorez les structures construites avec plusieurs variantes d'algorithmes et d'opérateurs :

- Construisez les treillis (fca) ou les AOC-posets (ares).
- Utilisez les différents opérateurs existants.
- Produisez les métriques sur les nombres de concepts obtenus dans les différentes structures et éventuellement d'autres métriques si l'idée vous en vient.

Compte-rendu : Une section de votre compte-rendu décrira l'approche et commentera les résultats. Vous rendez à côté les dossiers contenant les structures produites en formats dot et pdf au moins.

6 Analyse qualitative des structures

Etudiez les questions suivantes :

1. Basez-vous sur les AOC-posets qui sont plus petits.
2. Indiquez dans les structures classant les restaurants et les pizzas quelques concepts apparaissant à l'étape 0, à l'étape 1, et à l'étape 2 du processus. Observez-les pour comprendre comment ils sont apparus.
3. Analysez dans les AOC-posets des restaurants, construits suivant différents opérateurs, quelques concepts apparus à l'étape 2 montrant que vous avez compris leur signification. Ecrivez un texte les décrivant en langage naturel et en logique de description (mais sans les intégrer dans Protégé pour des questions de temps).
4. Extrayez quelques règles d'implication entre attributs relationnels, entre attributs non relationnels et entre les deux (toujours avec différents opérateurs) dans les trois AOC-posets.
5. Imaginez un scénario d'utilisation et déroulez-le (ex. recherche d'un restaurant "semblable" quand un autre est fermé, recherche d'un restaurant offrant un grand choix pour des amis végétariens, etc.)

Compte-rendu : Une section de votre compte-rendu décrira l'approche et les réponses à ces questions.