

Extraction de connaissances a partir de données relationnelles sur la qualité de l'eau

Xavier Dolques¹ Florence Le Ber¹ Marianne Huchard²
¹ICUBE, Université de Strasbourg/ENGEEES, CNRS, Strasbourg
²LIRMM, Université Montpellier 2, CNRS, Montpellier

Journée scientifique et technique 20 mars 2014

Problématique

Problématique

Peut-on établir des liens entre des caractères physico-chimiques d'un cours d'eau et des modalités de traits biologiques des taxons qu'il contient ?

- ▶ on essaie ici de se détacher des taxons, spécifiques à certaines régions, pour obtenir des résultats sur des caractéristiques comparables à une plus large échelle.



Base de données Fresqueau

Base de données de qualité de l'eau

Quelques chiffres sur la partie de la base qui nous intéresse :

- ▶ stations de mesure (11329)
- ▶ prélèvements hydrobiologiques (29636 prélèvements composés de 70929 valeurs)
- ▶ prélèvements physicochimiques (125318 composés de 14603913 valeurs)



Base de données relationnelles

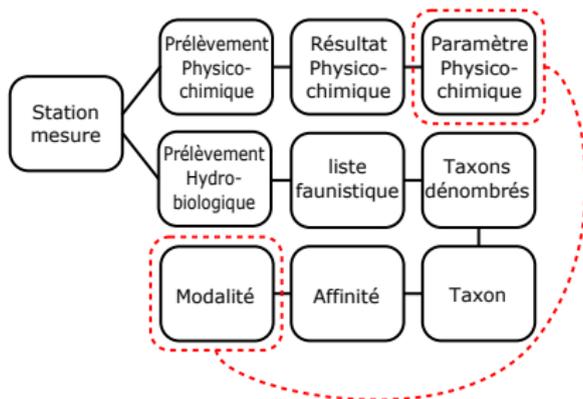
Aspect relationnel

- ▶ information répartie sur plusieurs tables
- ▶ relations entre les tables

Base de données relationnelles

Exemple

Dans la base Fresqueau, quels liens entre paramètres physico-chimiques et modalité de traits biologiques : 8 relations à parcourir pour passer de l'un à l'autre.



Extraction de connaissance

Répondre à la problématique

Comment extraire des connaissances à partir

- ▶ d'une grande quantité de données
- ▶ de données relationnelles



Analyse Relationnelle de Concepts

Méthode d'extraction de connaissance

- ▶ méthode automatique
- ▶ gestion des données relationnelles

Analyse Relationnelle de Concepts

Concept

Un concept est un ensemble de caractéristiques propres à un ensemble d'objets et permettant de le définir.

TABLE : Exemple de concept

identifiant	concept 93
caractéristiques	aff1(respiration : stigate); aff3(stade aquatique : larve); aff3(salinité : eau douce); aff3(pH : > 6)
objets	Acentria ; Aphelocheirus ; Epi- didae ; Micronecta ; Sialis

Concepts Relationnels

Relation avec d'autres concepts

On peut caractériser un ensemble d'objets par des concepts calculés. Par exemple il peut exister le concept des stations pour lesquelles il existe un taxon à forte abondance décrit par le concept 93

TABLE : Exemple de concept relationnel

identifiant	concept 156
caractéristiques	\exists abondance5(Concept93) ; \exists abondance3(Concept42)
objets	station3 ; station8 ; station24

Concepts Relationnels

Analyse relationnelle de concepts

Méthode de construction automatique des concepts relationnels à partir de données relationnelles.

Expressivité

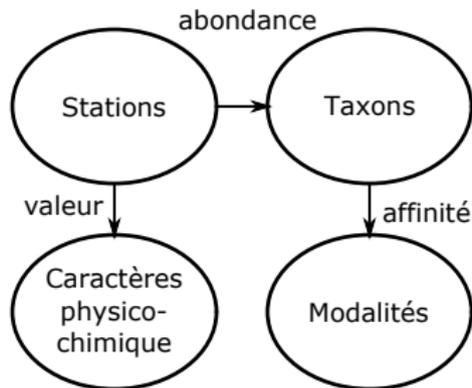
On peut faire varier l'expressivité des concepts relationnels obtenus en jouant sur les quantificateurs ($\exists, \forall \dots$)



Application

Données Alsace

- ▶ sous-ensemble de la base, jeu de données plus petit pour des tests
- ▶ 3 tables



Préparation des données

- ▶ discretisation des données
- ▶ transformation en relations binaires
- ▶ modélisation du schéma relationnel de l'analyse
 - ▶ choix des relations et de leur sens
 - ▶ choix des quantificateurs utilisés



Règles

PourMajoriteforte_abondance(
 \exists forte_affinite(distributiontransversale : berges et bras))
→ \exists bon_etat(berges)

Interprétation

On trouve en forte abondance des taxons dont la distribution transversale se situe sur les berges et dans les bras seulement si les berges sont en bon état.

Règles

*Pour*Majoriteforte_abondance(
 \exists forte_affinite(courant lent))
→ \exists mauvais_etat(hydrologie)

Interprétation

On trouve en forte abondance des taxons adaptés au courant faible seulement si l'hydrologie est altérée.

Conclusion

- ▶ les données relationnelles nécessitent des approches spécifiques
- ▶ l'analyse relationnelle de concepts permet de considérer ce type de donnée
- ▶ les concepts permettent de faire émerger des règles

Perspectives

Passage à un plus grand jeu de donnée

- ▶ consider des régions plus grandes, voire la base entière
- ▶ problèmes d'explosion combinatoire du nombre des résultats

Applications sur d'autres problèmes

- ▶ relations entre contexte (par ex. activité agricole) et résultats
- ▶ relations spatiales entre station (amont-aval)



Merci de votre attention

