Maîtriser la complexité des systèmes: apport de la fouille de données

Jean-François Mari, Amedeo Napoli

Maîtriser la complexité des systèmes: apport de la fouille de données

Equipe Orpailleur

LORIA (CNRS – INRIA Nancy Grand-Est – Université de Lorraine)
B.P. 239, 54506 Vandoeuvre les Nancy, France
http://orpailleur.loria.fr/

Comité de l'Administration Régionale (CAR)
Châlons en Champagne, 23 novembre 2016

Apport de l'Intelligence Artificielle en fouille de données

- Nettoyer les données
  - diminuer le bruit de mesure
  - cerner la variabilité dûe aux processus vivants par modélisation stochastique
- Agréger et classer les données :
  - élémentaire : binaires, nombres
  - ou complexes : ensembles, séquences, relations
- Décrire les ressemblances / différences entre familles (les classes)
- Visualiser: les résultats agrègés (la photo de famille !)

Les formes découvertes peuvent être interprétées comme des unités de connaissance pour un univers à explorer.

Nature des données d'utilisation du territoire

- Données temporelles car elles sont issues d'un calendrier de travail lié aux saisons
- Données spatiales car elles sont issues de territoires mis en valeur par les agriculteurs ;
- Données hétérogènes à différentes échelles

Enjeux : rendre visibles ces données par des moyens numériques

- Permettre leur réutilisation
- Rendre ces données inter-opérables
- Développer des nouveaux services

Dynamiques territoriales dans le bassin du Yar (analyse sur 12 ans à partir d'images satellite à l'aide du logiciel ARPENTAge)

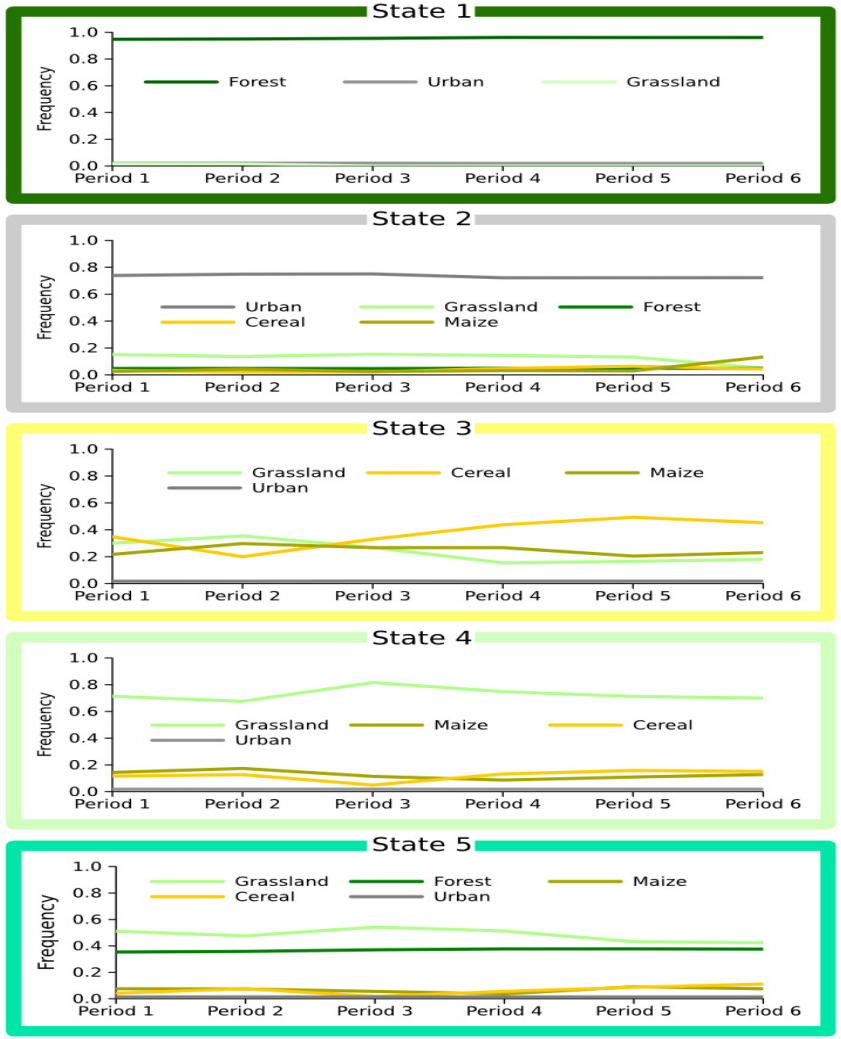
region 1 : forêts et zones pérennes

region 2 : urbain

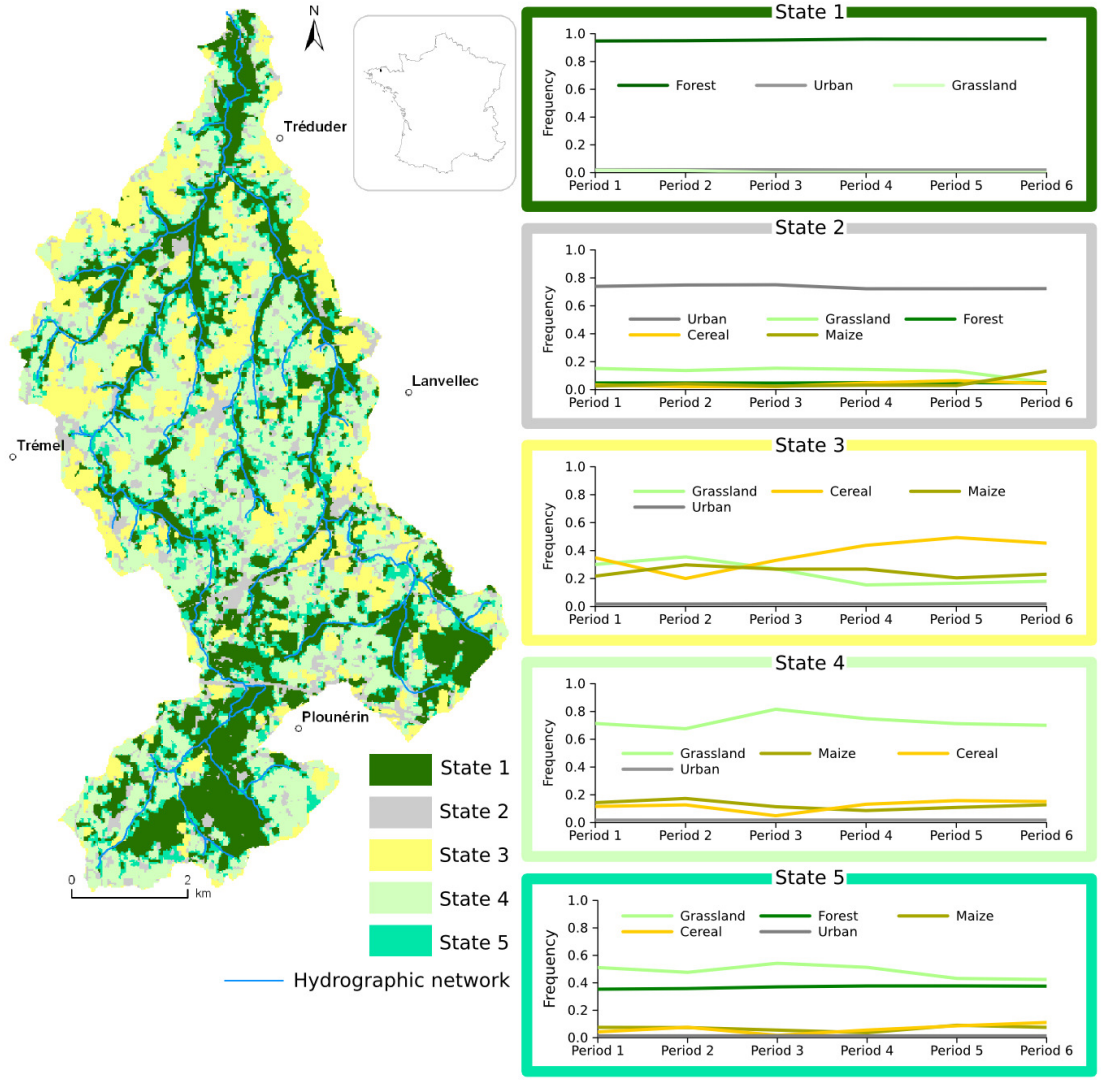
region 3 : disparition des prairies

region 4 : maintien des prairies

region 5 : sans dynamique

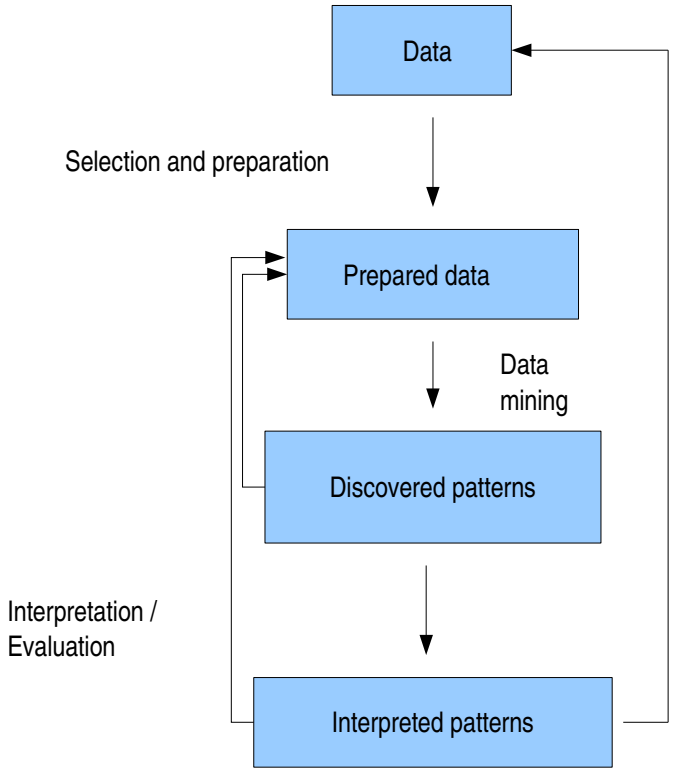


Disparition des prairies dans le bassin du Yar : bassin à algues vertes



Knowledge Discovery in Databases (KDD)

- KDD is applied to large volumes of complex data for discovering patterns which can be significant and reusable.
- KDD is based on three main steps: data preparation, data mining, and interpretation of the discovered units.
- KDD is iterative and interactive, i.e. it can be replayed and it is guided by an analyst.



Research tracks in the Orpailleur Team

- Knowledge Discovery:
  - pattern mining, FCA and extensions, association rules, mining big data, visualization
  - mining complex data: sequences, trees, graphs, linked open data
  - text mining, information retrieval, recommendation
  - preferences in KD: skylines, skycubes, aggregation measures
  - privacy and reputation in KD
- KD in Life sciences:

  data integration, mining complex data in biology and pharmacology
- Knowledge engineering:
  - knowledge mining, ontology engineering
  - Inductive Logic Programming
  - decision making, dimensionality reduction
  - vulnerability management: detection and representation

Towards "Exploratory Knowledge Discovery"

- Knowledge Discovery and Knowledge Engineering are complementary.
- Use a declarative approach for problem solving: i.e. "describe the problem and the solver will take care of the solution".
- Define interactive mechanisms to identify seeds for pattern space exploration consistent w.r.t. domain knowledge (interaction, constraints, dimensionality reduction).
- Propose mechanisms to explore the neighborhood of the seeds until initial constraints are satisfied by query reformulation (redescription).
- Address threshold issues w.r.t. analyst queries thanks to the skyline analysis of the pattern space (preferences).

An Ordinal Approach to EKD: Small is Beautiful...

- Classification is a polymorphic process,
- and a good candidate for bridging discovery and representation of patterns.
- Partial Orders and their properties can be used for revisiting classification:
  - Discovery of classes for understanding data.
  - Organization of classes into a partial order.
  - Representation and Reasoning: instantiation, class definition, information retrieval...
- Formal Concept Analysis can play some of these roles...