

FOUILLE DE DONNÉES DE CAPTEURS

2

Introduction

Motivations

3

- Environnement / Agronomie

Motivations

4

- Environnement / Agronomie
- Fouille de données de capteurs

Motivations

5

- Environnement / Agronomie
- Fouille de données de capteurs

Points communs ?

Motivations

6

- Environnement / Agronomie
- Fouille de données de capteurs

Points communs ?

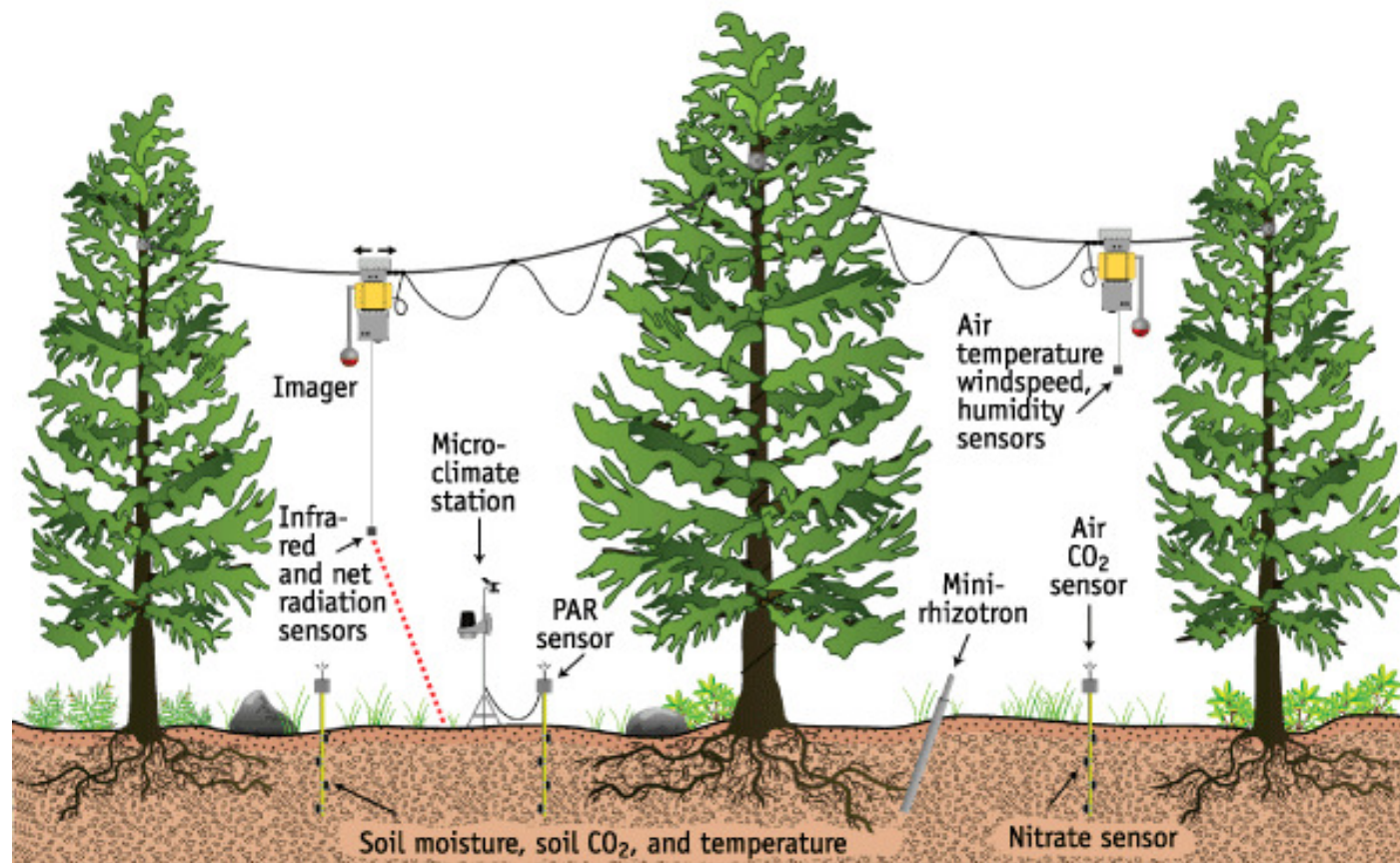


Les capteurs !!!

Motivations

7

- Exemple d'installation de capteurs environnementaux

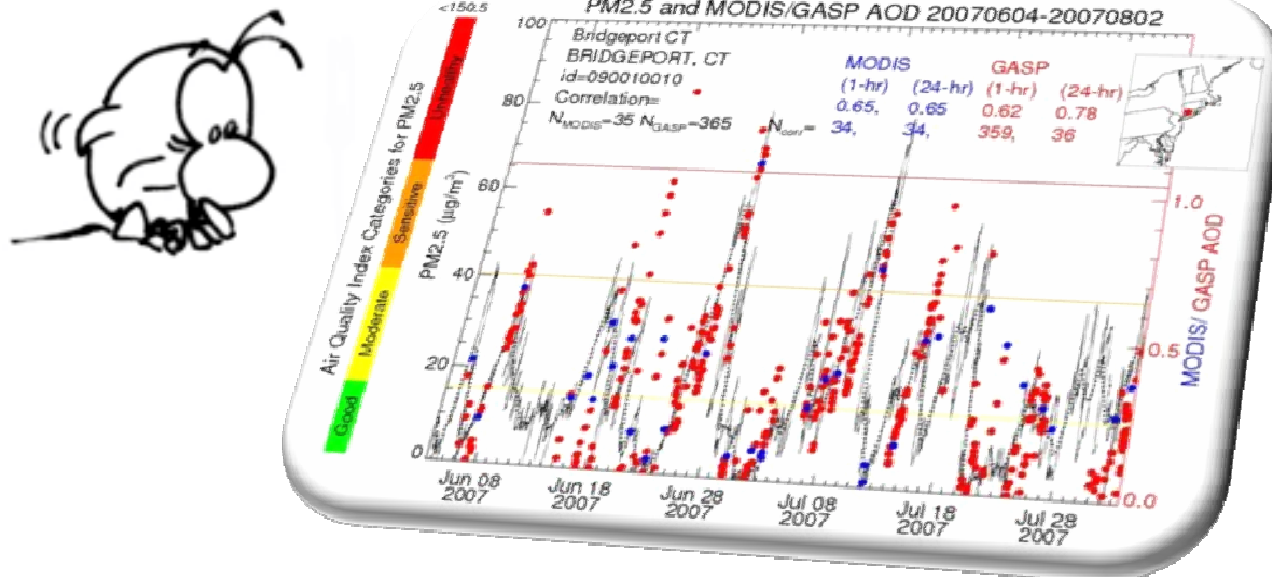


Pourquoi fouiller de telles données ?

8

□ Problèmes

- Des données en **grande quantité**
- **Difficiles** à analyser / interpréter
- **Peu** de connaissances *a priori*



Objectifs

9

- **Fournir** aux experts des **connaissances nouvelles** sur les données grâce aux techniques de fouille de données

- **Utiliser** ces connaissances pour **répondre à d'autres problématiques**
 - Détection d'anomalies
 - Prédiction

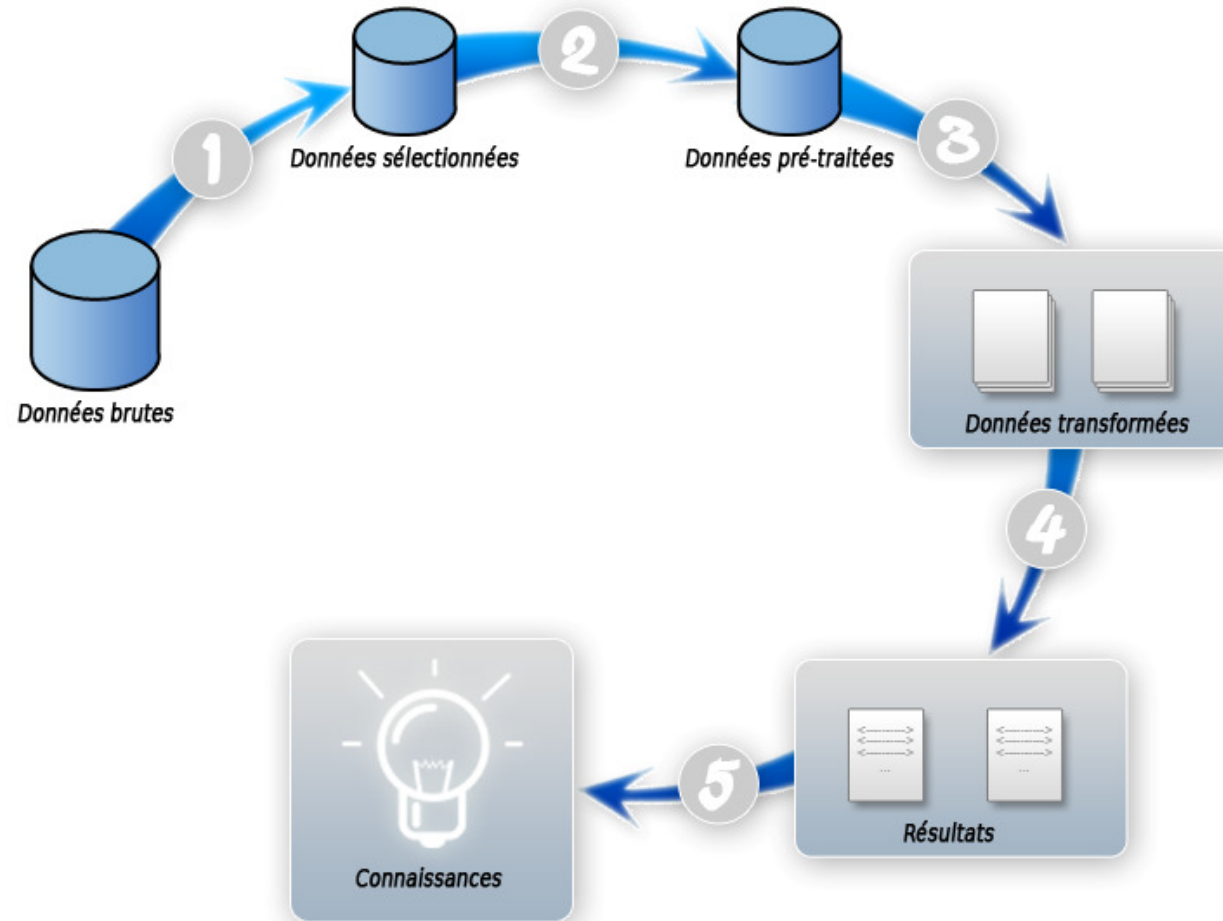


10

Extraire des connaissances dans les données de capteurs

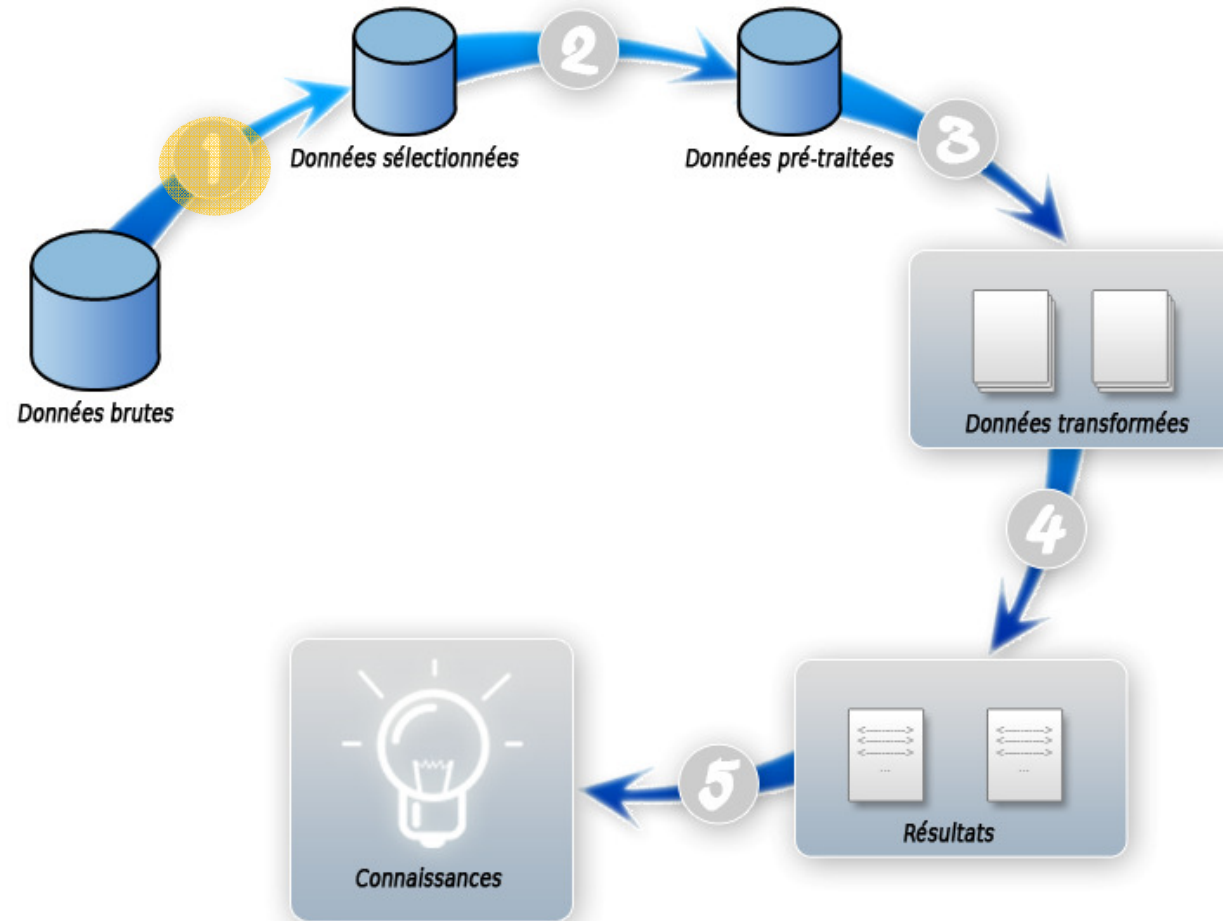
Le processus d'Extraction de Connaissances dans les Données

11



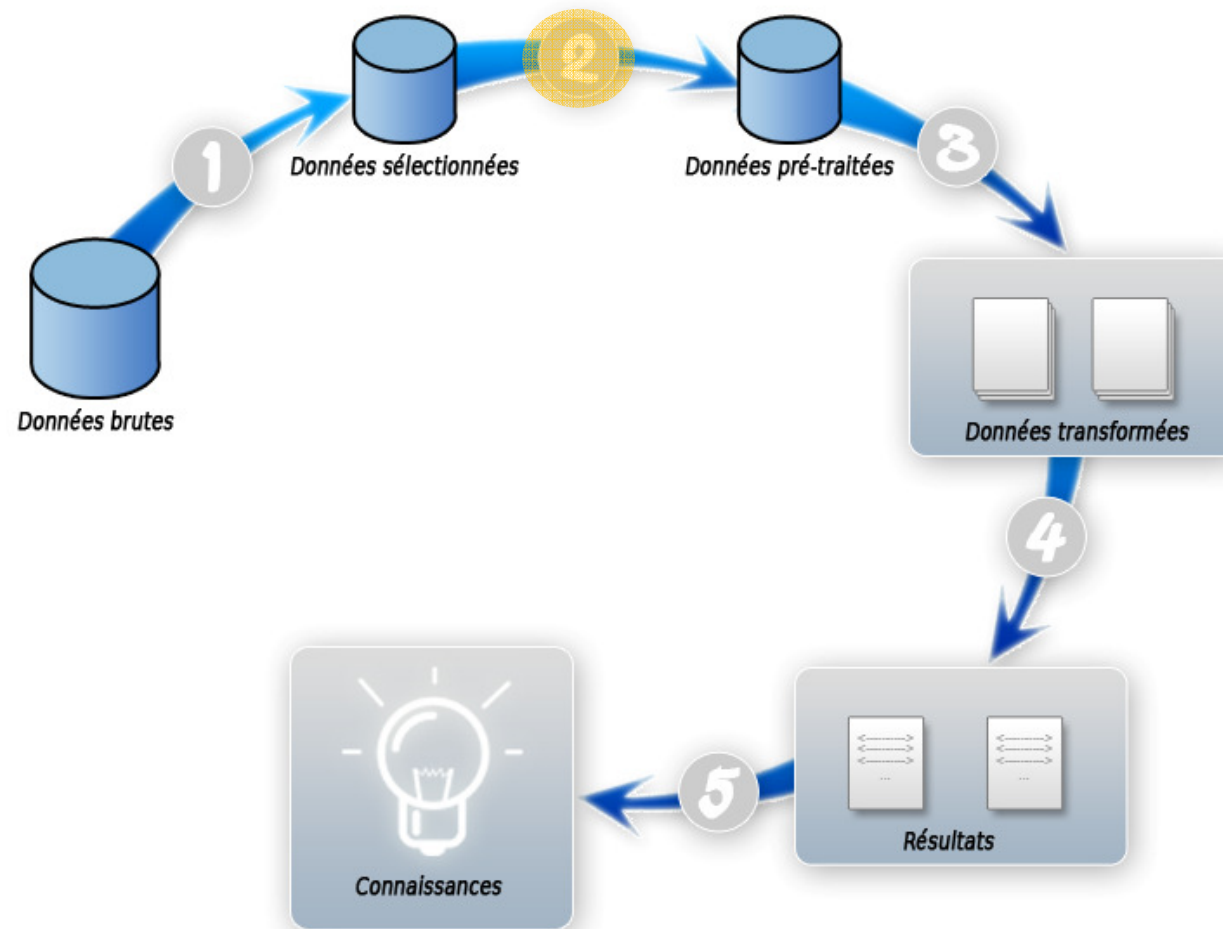
Le processus d'Extraction de Connaissances dans les Données

12



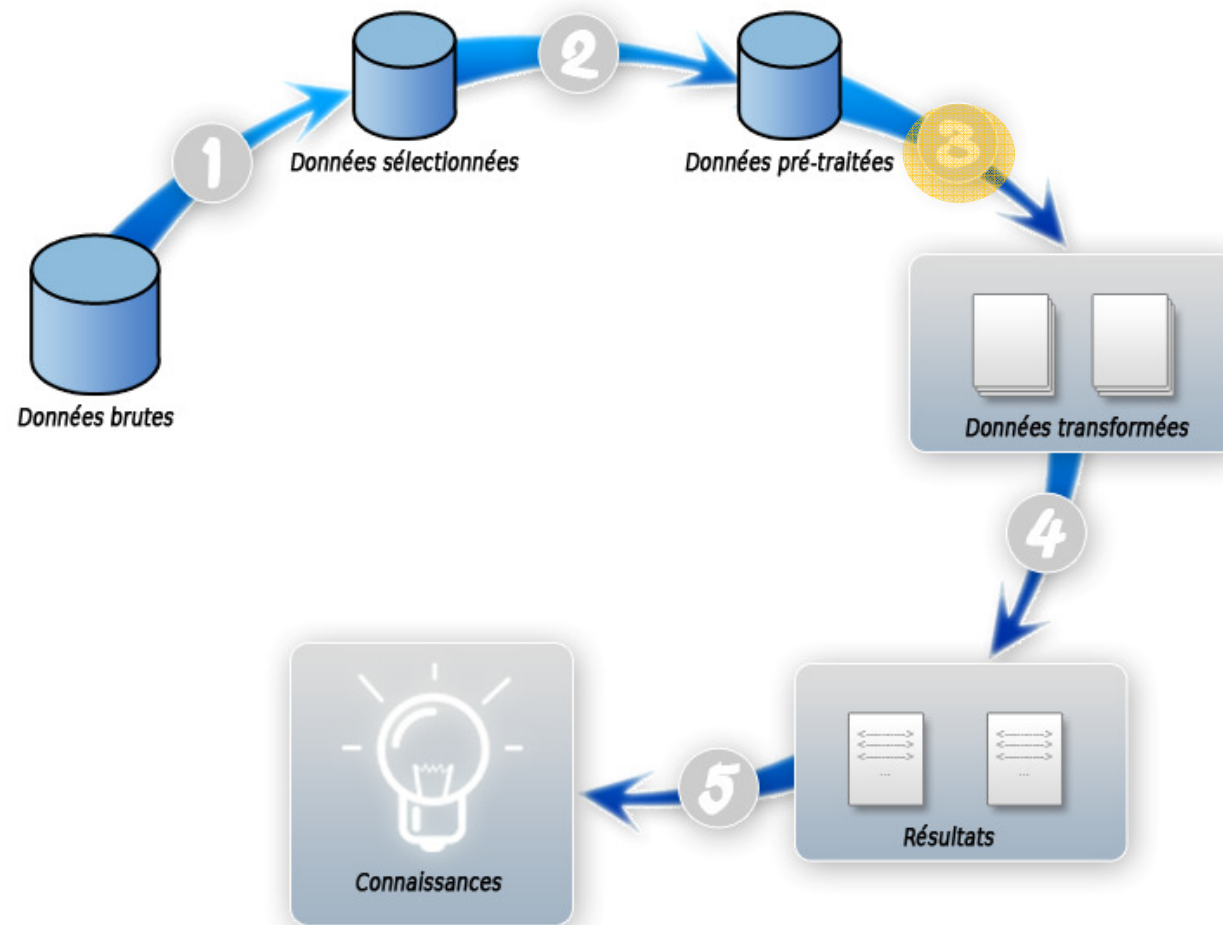
Le processus d'Extraction de Connaissances dans les Données

13



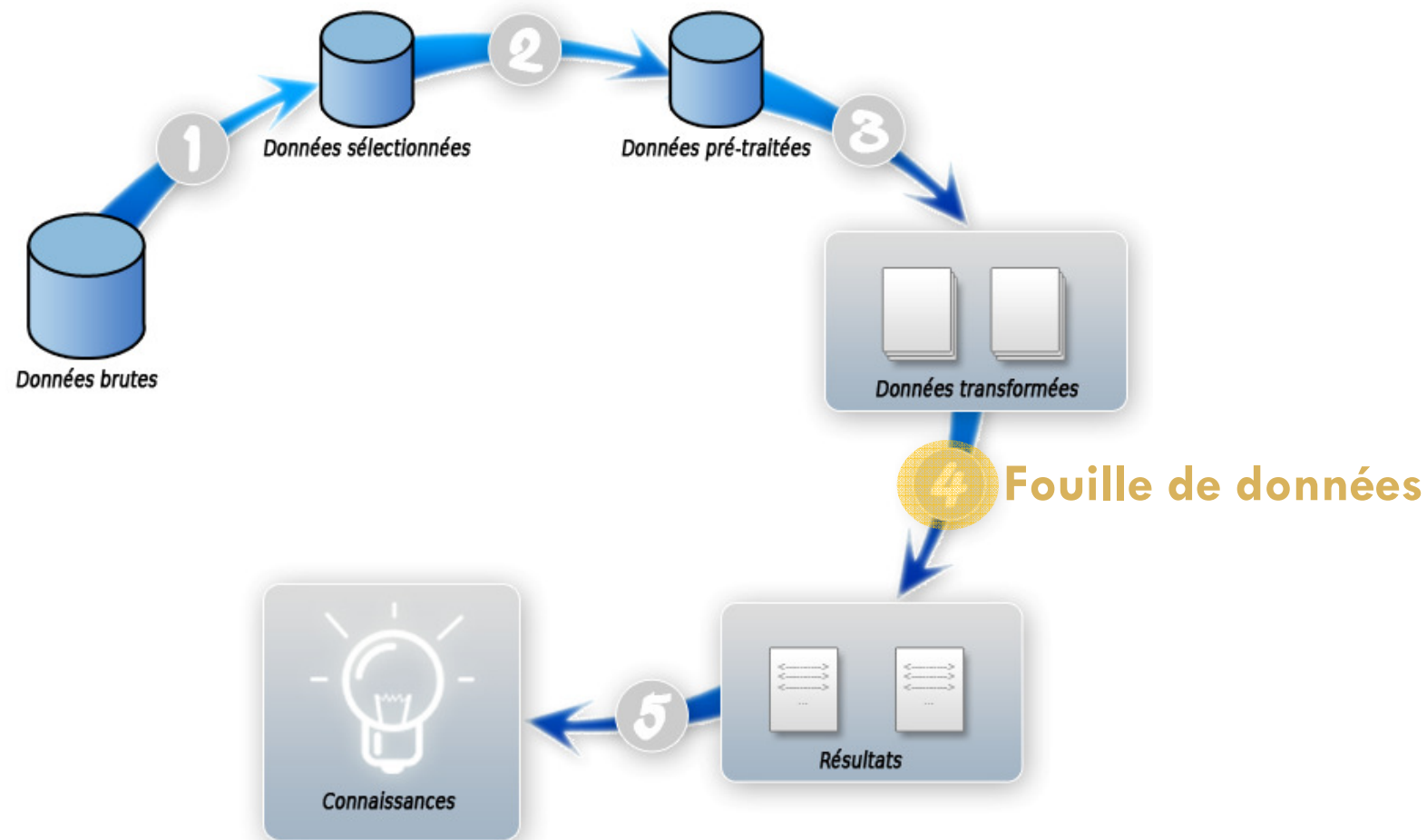
Le processus d'Extraction de Connaissances dans les Données

14



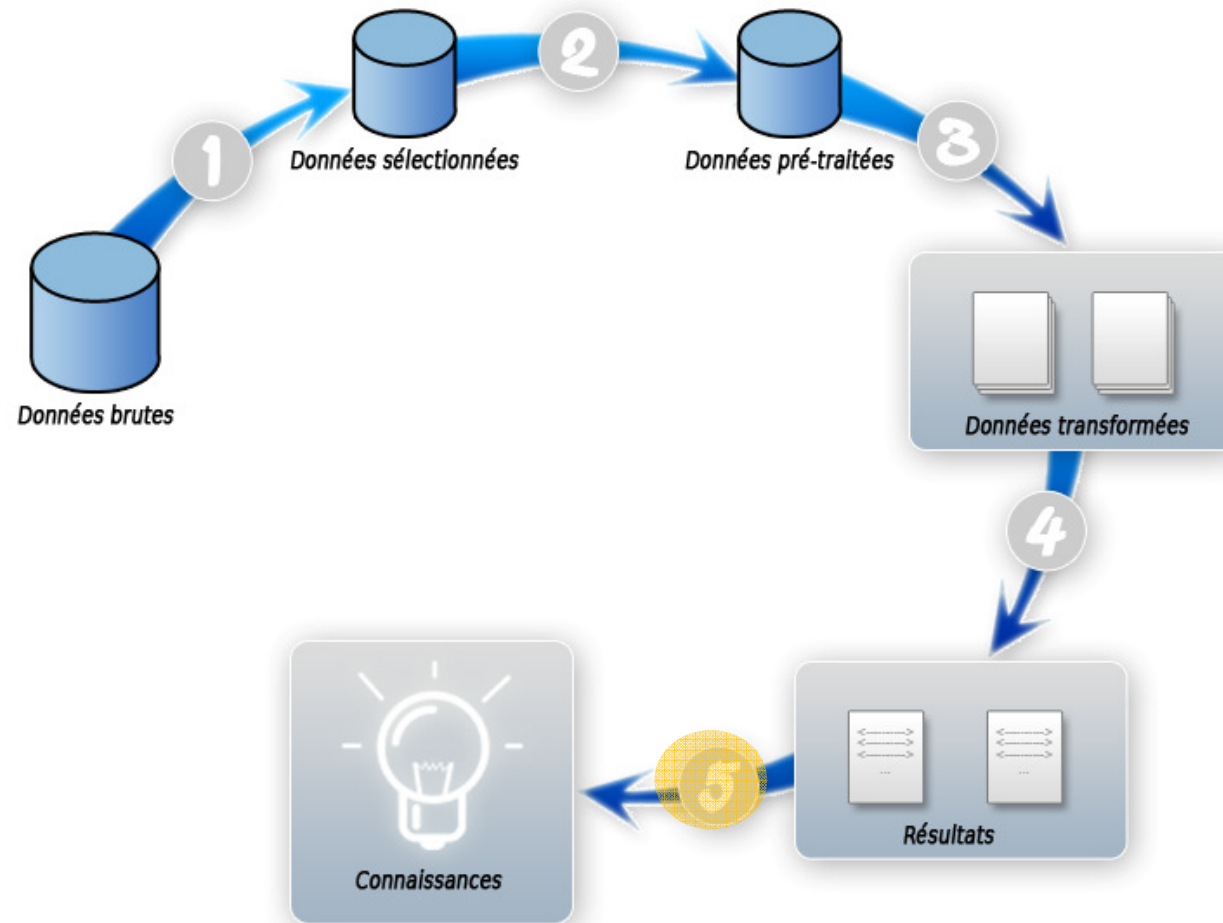
Le processus d'Extraction de Connaissances dans les Données

15



Le processus d'Extraction de Connaissances dans les Données

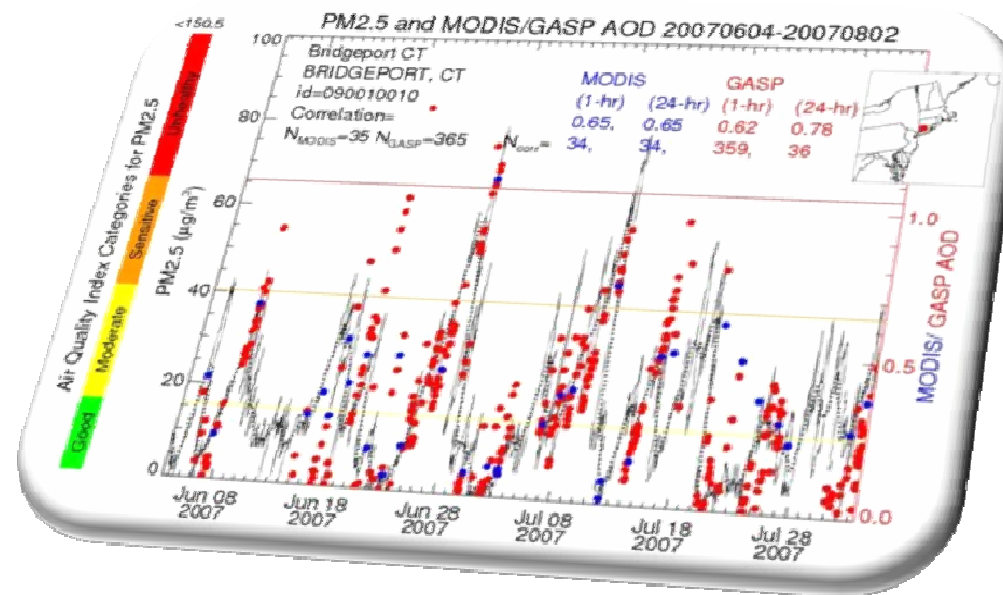
16



Description des données de capteurs

17

- Comment **décrire** les données relevées par des capteurs ?



Description des données de capteurs

18

- Comment **décrire** les données relevées par des capteurs ?
- Nous allons **décrire les fragments de données fréquents**

Description des données de capteurs

19

- Comment décrire les données relevées par des capteurs ?
- Nous allons décrire les fragments de données fréquents
- **Recherche des motifs fréquents**
 - ▣ Règles d'association
 - ▣ Motifs séquentiels

Description des données de capteurs

20

- Comment décrire les données relevées par des capteurs ?
- Nous allons décrire les fragments de données fréquents
- **Recherche des motifs fréquents**
 - ▣ **Règles d'association**
 - ▣ Motifs séquentiels

Règles d'association

21

- **1993** – Agrawal et al
- Rechercher des **corrélations entre attributs** (items) au sein de grandes masses de données

Règles d'association

22

Exemple

Pastis → Bière [30%, 80%]

- ▣ Support : 30% des clients ont acheté **du caviar et du champagne**
- ▣ Confiance : 90% des gens qui ont acheté du **caviar** ont aussi acheté du **champagne**

Règles d'association

23

- La recherche de règles d'association consiste à trouver dans un ensemble de données, **toutes les règles** dont le support et la confiance sont supérieurs à des **paramètres fixés par l'utilisateur**
- De (très) nombreux algorithmes existent pour extraire des règles d'association dans des données, mais...

Règles d'association

24

- La recherche de règles d'association consiste à trouver dans un ensemble de données, **toutes les règles** dont le support et la confiance sont supérieurs à des **paramètres fixés par l'utilisateur**
- De nombreux algorithmes existent pour extraire des règles d'association dans des données, **mais...**

Motifs séquentiels

25

...mais pas de notion de **temporalité** dans les règles d'association !

Pastis et bière sont corrélés... et après ?



Motifs séquentiels

26

...mais pas de notion de **temporalité** dans les règles d'association !

Pastis et bière sont corrélés... mais y a-t-il d'autres corrélations **dans le temps** ?



PUIS



Motifs séquentiels

27



PUIS



Ceci est un **motif séquentiel** !

< (Pastis Bière)(Aspirine) > [**20%**]

Motifs séquentiels

28



PUIS



Ceci est un **motif séquentiel** !

< (Pastis Bière)(Aspirine) > [20%]

20% des clients ont acheté du **pastis et de la bière**,
puis sont revenus pour acheter de l'**aspirine**.

Et mes capteurs ?...

30



- Comment fouiller des données de capteurs ?
 - ▣ Dans les exemples, un item était un **produit** (bière, pastis, ...)

 - ▣ Dans les données de capteurs, un item est la **valeur associée à un capteur** à un moment donné
 - **Temp_{bas}** le capteur de température mesure une **valeur basse**
 - **Temp_↓** le capteur de température mesure une **valeur en baisse**

Et mes capteurs ?...

31

- Comment fouiller des données de capteurs ?
 - **Temp_{bas}** le capteur de température mesure une **valeur basse**
 - **Temp_↓** le capteur de température mesure une **valeur en baisse**

 - **Exemple de motif séquentiel**
 - **< (Temp_↑) (CO2_↑) > [95%]**
 - Dans **95%** des jours étudiés, on a observé une **hausse de température** suivie d'une **hausse de CO2**.

33

Exploiter les résultats

Que faire de ces résultats ?

34



- **A quoi ça sert ?**

- Fournir une **meilleure compréhension** des données étudiées
- Ils sont **facilement interprétables** par des non-informaticiens
- Ils peuvent répondre à **d'autres problématiques** : détection d'anomalies ou prédiction de comportements

Que faire de ces résultats ?

35



- **A quoi ça sert ?**

- Fournir une **meilleure compréhension** des données étudiées
- Ils sont **facilement interprétables** par des non-informaticiens
- Ils peuvent répondre à **d'autres problématiques** : détection d'anomalies ou prédiction de comportements

Détection d'anomalies : intuition

36

- Les **motifs séquentiels** extraits décrivent ce qui arrive **fréquemment** dans le jeu de données étudié
- Par exemple, il est **très fréquent** de trouver une hausse de température suivie d'une hausse de CO2
 $\langle (\text{Temp}_{\uparrow}) (\text{CO2}_{\uparrow}) \rangle [95\%]$

Détection d'anomalies : intuition

37

- Les **motifs séquentiels** extraits décrivent ce qui arrive **fréquemment** dans le jeu de données étudié
- Par exemple, il est **très fréquent** de trouver une hausse de température suivie d'une hausse de CO2

$\langle (\text{Temp}_{\uparrow}) (\text{CO2}_{\uparrow}) \rangle [95\%]$



Dans ce cas, une chute de CO2 après une forte hausse de température est un **comportement discordant** par rapport à ce motif, et donc peut-être d'une anomalie

38

Conclusion

Conclusion

39

- **Pourquoi utiliser les motifs séquentiels?**
 - ▣ Fournir une **description compréhensible** des données
 - ▣ Répondre à d'autres problèmes : **détection d'anomalies** ou **prédiction**, par exemple



Questions ?



Où trouver ces outils ?

41

□ Règles d'association

- **Weka** - collection d'algorithmes

<http://www.cs.waikato.ac.nz/ml/weka>

- **aRules** – package R <http://r-forge.r-project.org/projects/arules>

- **Implémentation C++** - <http://www.cs.bme.hu/~bodon/en/apriori>

□ Motifs séquentiels

- **arulesSequences** - package R

<http://cran.r-project.org/web/packages/arulesSequences>

- **SPMF** (Java) <http://www.philippe-fournier-viger.com/spmf>

- **PrefixSpan** (C++) <http://www.cb.k.u->

tokyo.ac.jp/asailab/tabai/prefixspan/prefixspan.html
rabatel@lirmm.fr