



IMT Mines Alès
École Mines-Télécom



TRF RETAIL
PREDICTIVE AND PRESCRIPTIVE ANALYTICS



IMPACT DE MESURES DE SIMILARITÉ SÉMANTIQUE DANS UN ALGORITHME DE PARTITIONNEMENT

Jocelyn PONCELET, Pierre-Antoine JEAN,
Jacky MONTMAIN, François TROUSSET,
Sébastien HARISPE, Nicolas PECHEUR

Contexte

Objectifs

Méthodologie

Validation

Conclusion & perspectives

CONTEXTE

Regrouper des objets en fonction de leurs caractéristiques (clustering)

Segmentation de consommateurs

Informations essentielles → Stratégie commerciale/marketing

Approches traditionnelles :

Agrégation essentiellement monétaires (e.g. *RFM*, *CLV* ...)

Variété des produits [Kok et al. 2008]

Panier moyen du consommateur [Cachon et al. 2007a]

...

Limites :

Dimensions indépendantes

Représentation vectorielle → Fléau des dimensions [R. Bellman 1957]

Exemple

Consommateur 1

- Concombre 1€
- Cote de porc 6€

Consommateur 2

- Coca 1€
- Bière 6€

Consommateur 3

- Banane 1€
- Poulet 3€

Exemple

Comparaison vectoriel « simple »

Consommateur 1

- Concombre 1€
- Cote de porc 6€

Consommateur 2

- Coca 1€
- Bière 6€

Consommateur 3

- Banane 1€
- Poulet 3€

Exemple

Agrégation : Chiffre d'affaires

Consommateur 1

- Concombre 1€
- Cote de porc 6€

7€

Consommateur 2

- Coca 1€
- Bière 6€

7€

Consommateur 3

- Banane 1€
- Poulet 3€

4€

Exemple

Agrégation : Chiffre d'affaires

Consommateur 1

- Concombre 1€
- Cote de porc 6€

7€

Consommateur 2

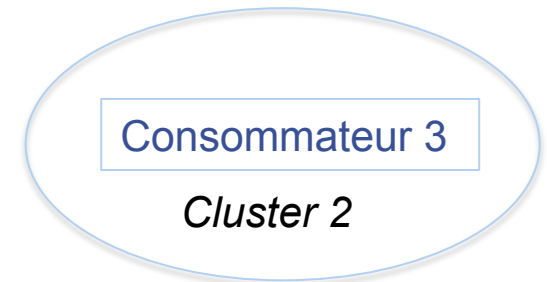
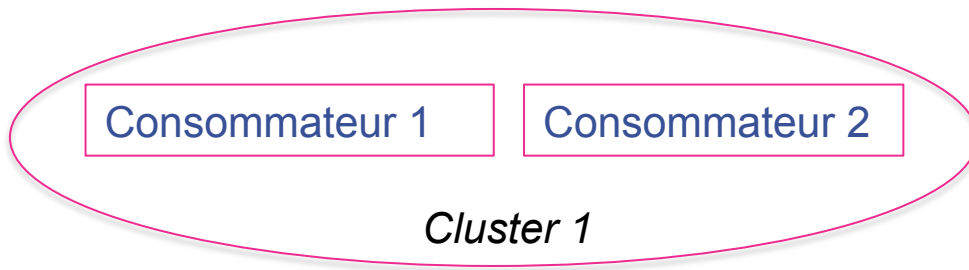
- Coca 1€
- Bière 6€

7€

Consommateur 3

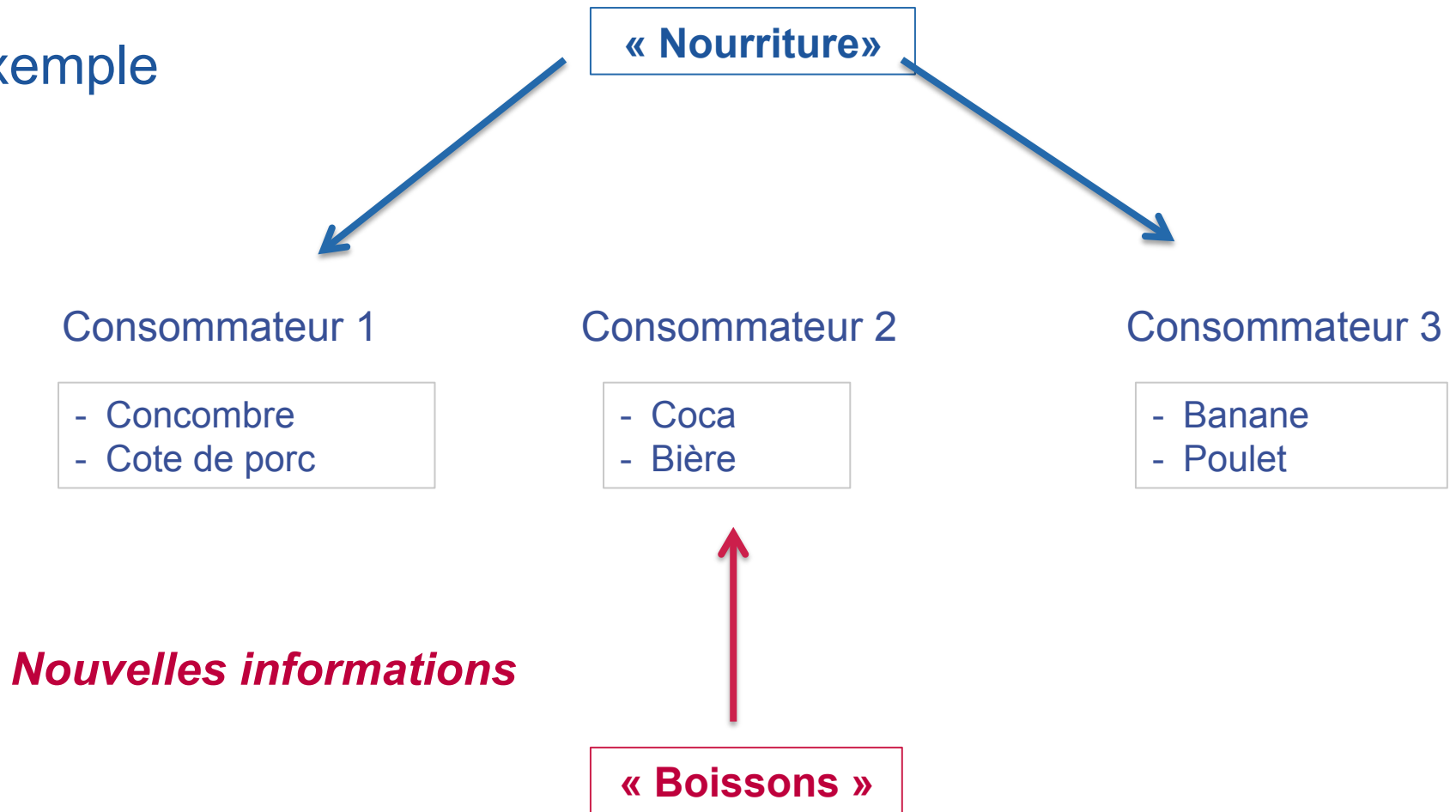
- Banane 1€
- Poulet 3€

4€



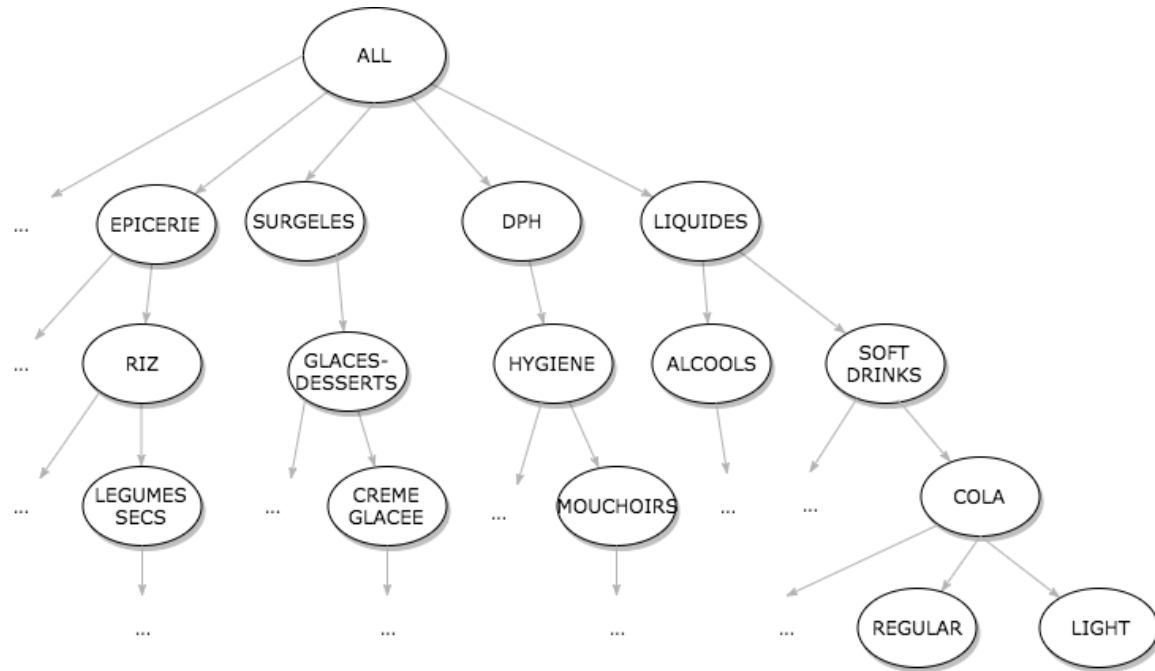
CONTEXTE

Exemple



OBJECTIFS

- Exploiter la typologie des produits
- Structure marchandise/
Taxonomie de produits
(connaissances *a priori*)
- Apporter de la sémantique dans les informations liées aux clients
- **Que viennent acheter mes clients dans mon magasin?**



Exemple de structure marchandise

MÉTHODOLOGIE

Exemple

Consommateur 1

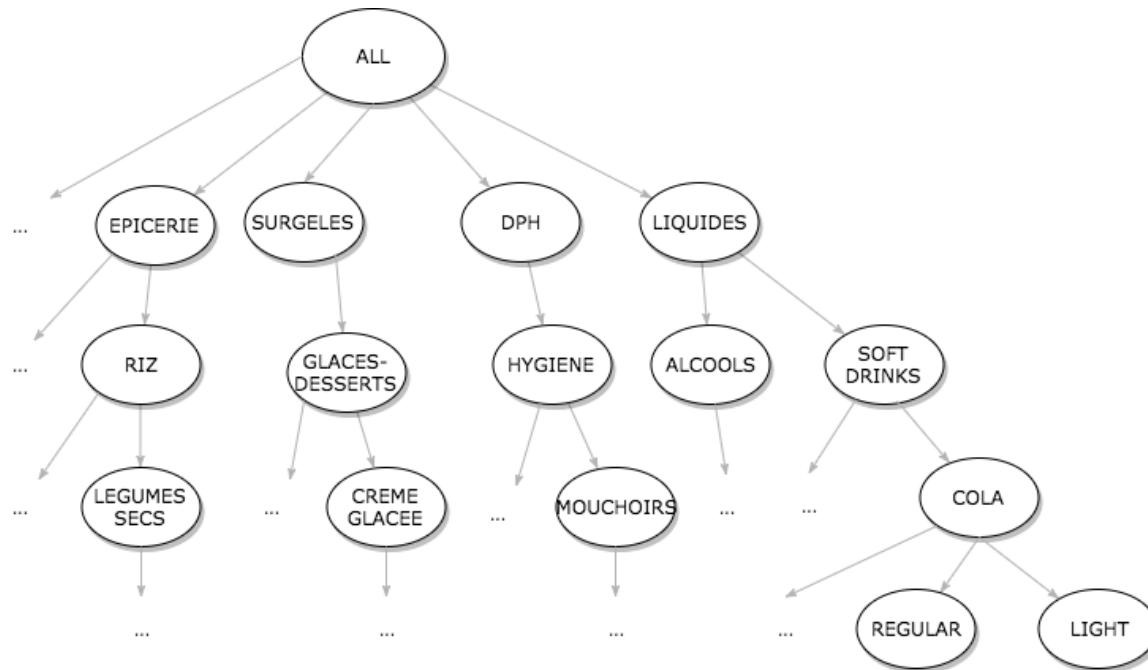
- Concombre
- Cote de porc

Consommateur 2

- Coca
- Bière

Consommateur 3

- Banane
- Poulet



Exemple de structure marchandise

MÉTHODOLOGIE

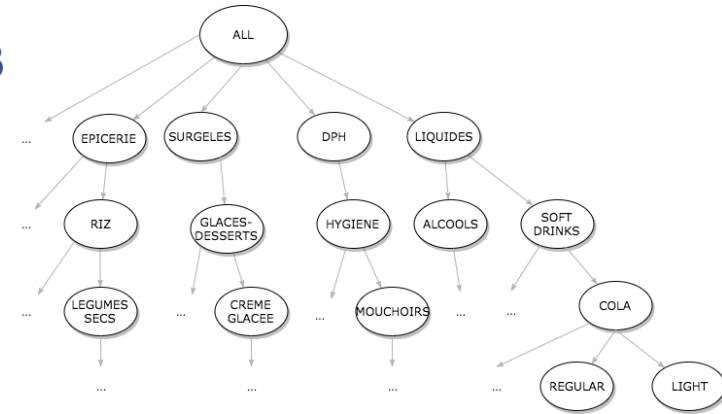
Exemple

Consommateur 1 Consommateur 2 Consommateur 3

- Concombre
- Cote de porc

- Coca
- Bière

- Banane
- Poulet



Exemple de structure marchandise

Consommateur 1

- Légumes
- **Viandes**

Consommateur 2

- Boisson
- Alcool

Consommateur 3

- Fruits
- **Viandes**

MÉTHODOLOGIE

Exemple

Consommateur 1

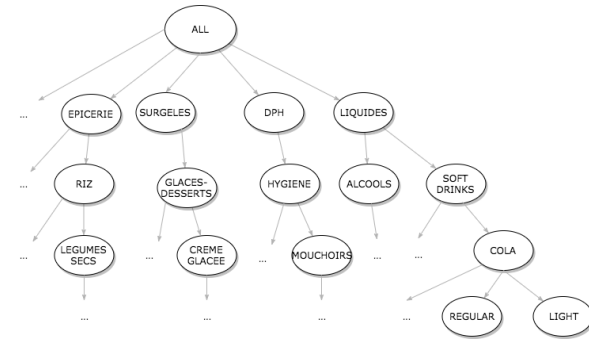
- Concombre
- Cote de porc

Consommateur 2

- Coca
- Bière

Consommateur 3

- Banane
- Poulet



Consommateur 1

- Légumes
- **Viandes**

Consommateur 2

- Boisson
- Alcool

Consommateur 3

- Fruits
- **Viandes**

Consommateur 1

- **Alimentaire**

Consommateur 3

- **Alimentaire**

Cluster 1

Consommateur 2

- **Liquide**

Cluster 2

Mesures de Similarité Sémantique (SSM)

Similarité Groupwise (GW)

Similarité entre 2 ensembles de concepts/objets

Méthode ensembliste

Méthode utilisant les similarité Pairwise

Similarité Pairwise (PW)

Similarité entre 2 concepts dans une taxonomie

Plus court chemin

Contenu Informationnel & MICA (Most Informative Common Ancestor)

Contenu Informationnel (IC)

Quantité d'information associée à un concept

Utiliser SSM pour exploiter des connaissances *a priori*

Prérequis :

- Ensemble d'objets (*e.g. consommateurs*)
- Ensemble de concepts hiérarchisés au sein d'un **ordre partiel** (*e.g. produits & taxonomie de produits*)
- Une « étiquette » associée à chacun des objets pour évaluer l'approche (*e.g. Cluster finaux*)

Jeu de données public issu de la grande distribution **indisponible**

Analogie pour valider l'approche sémantique

VALIDATION DE LA MÉTHODOLOGIE

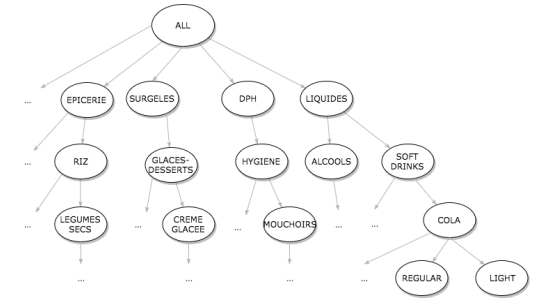
Analogie afin de démontrer l'intérêt d'approches sémantiques



Consommateurs



Produits



Structure marchandise

VALIDATION DE LA MÉTHODOLOGIE

Analogie afin de démontrer l'intérêt d'approches sémantiques



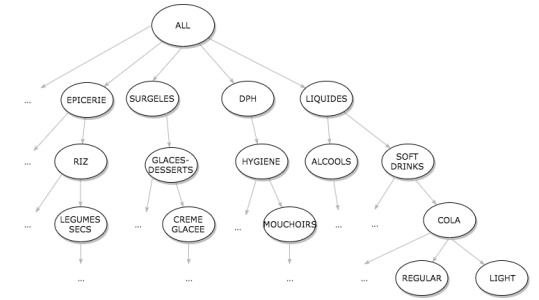
Consommateurs



Maladies



Produits



Structure marchandise

VALIDATION DE LA MÉTHODOLOGIE

Analogie afin de démontrer l'intérêt d'approches sémantiques



Consommateurs



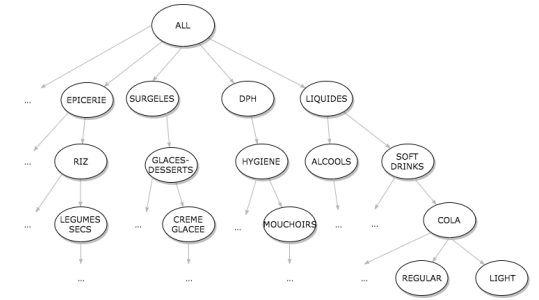
Maladies



Produits



Symptômes



Structure marchandise

VALIDATION DE LA MÉTHODOLOGIE

Analogie afin de démontrer l'intérêt d'approches sémantiques



Consommateurs



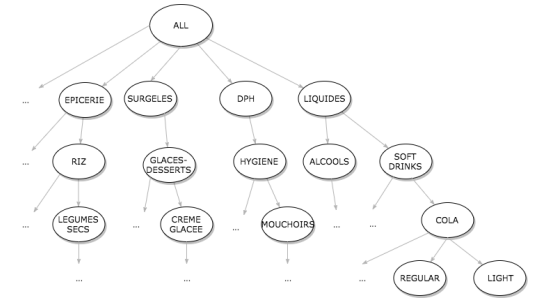
Maladies



Produits



Symptômes



Structure marchandise



MeSH
(*Medical Subject Headings*)

VALIDATION DE LA MÉTHODOLOGIE

*Zhou, X., J. Menche, A. L. Barabási, et A. Sharma (2014).
Human symptoms–disease network. Nature communications 5, 4212.*

Liste de maladies/symptômes désambiguïsés dans la taxonomie du MeSH

Un symptôme est à la maladie ce qu'un produit est au consommateur

1517 Maladies pour **223** symptômes

Vecteur TF-IDF [*Sparck J. 1972*]
(force d'association maladie/symptôme)

Vecteur binaire
(il existe une force d'association entre la maladie et le symptôme ou non)

Classification Ascendante Hiérarchique (CAH)

VALIDATION DE LA MÉTHODOLOGIE

Baseline de comparaison

Mesures vectorielles - distance de cosinus

Mesures vectorielles - distance de euclidienne

Mesures de similarité sémantique (*Semantic Measures Library**)

Mesure GW ensembliste – *Jaccard*

GW et PW sans IC – *BMA et Wu & Palmer*

[*Pesquita et al. 2007*] [*Wu et Palmer 1994*]

GW, PW avec IC Resnik - *BMA et PW de Resnik*

[*Pesquita et al. 2007*] [*Resnik P. 1995*]

GW, PW avec IC Seco - *BMA et PW de Resnik*

[*Pesquita et al. 2007*] [*Resnik P. 1995*]

Évaluation F1-mesure

Normalisation des vecteurs

*semantic-measures-library.org



Résultats

Performance des Mesures de Similarité Sémantique

16 pts d'amélioration de la F1-mesure

Combinaisons « complètes » des SSM (*GW & PW & IC*)

Analogie avec le domaine Biomédical

Amorcer la validation de partitionnement sémantique

Autres combinaisons de SSM

Limite identifiée : Complexité des SSM

Méthodes de résumé sémantique

MERCI



IMT Mines Alès
École Mines-Télécom

