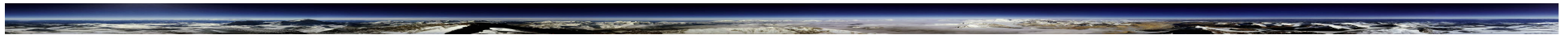


Apprentissage et Alimentation : l'apport du numérique



Thèse Sema Akkoyunlu

Et projet ANR « SHIFT » (2019-2022)

Antoine Cornuéjols

AgroParisTech – INRA MIA 518

antoine.cornuejols@agroparistech.fr

Où l'on parle d'« **avalanche** » de données

- Des données capturées à foison quand nous allons **sur Internet**
 - Sur quels sites
 - Combien de temps, les clics, les durées, les achats, ...
- **Smartphones**
 - Localisation même si on a dit non
 - Des tas d'applications pleines de curiosité
- **Bracelets** connectés
- Moyens de **paiement** (banques)
- Capteurs dans les **véhicules** (assurances)
- Compteurs Linky
- « Smart » cities

Tous les domaines sont touchés ...

- La **bio-informatique**

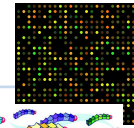
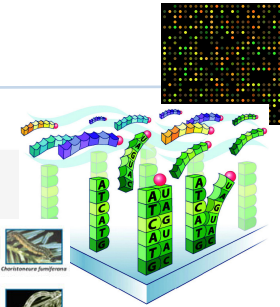
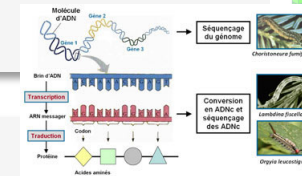
- La **sociologie**

- La **e-medecine (le me-data)**

- Le **domaine juridique**

- L' **industrie**

- L' **assurance**



Tous les domaines sont touchés ... même

- L'agriculture numérique

- Gros volume de données

- ◆ Capteurs
- ◆ Drones
- ◆ Réseaux sociaux et pro



Tous ... **sauf** ...

- **L'alimentation**

- Enquête **Nutrinet**

- ~ 277 000 internautes théoriquement sur des années
 - **Mais**
 - ◆ à 80% des femmes
 - ◆ Milieux socio-professionnels élevés
 - ◆ Abandonnent après quelques jours

**Manque de données
représentatives**

- **L'éducation**

- Peu de données sur ce qui se passe en classe ou devant un écran

Quelles **données** sur l'alimentation

Bases de données INCA

Étude Individuelle Nationale des Consommations Alimentaires

- Photographie des **habitudes de consommations alimentaires** de la population en France métropolitaine
- Études réalisées **tous les 7 ans**
 - INCA1 (1998-1999), **INCA2** (2006-2007) et INCA3 (2014-2015)
- Types de descripteurs
 - Caractéristiques des individus
 - Critères de choix des aliments
 - Attitudes et opinions en alimentation
 - Habitudes de vie / état de santé
 - Apports nutritionnels journaliers en 38 nutriments
 - Consommations alimentaires détaillées et quantifiées sur la semaine des individus
 - Descriptif des occasions de consommation : lieu, durée, ...

Bases de données INCA et ANSES

INCA2

- 4079 individus
- 541 526 lignes
- Table relatives aux données sur les ménages
- ...

Bases de données de l'ANSES sur la [composition des aliments](#)
(apports en substances bénéfiques)

Données d'entreprises privées

Ustensiles de cuisine connectés

« *Astuces, conseils ou recettes originales, nos **applications** nourrissent votre inspiration et vous accompagnent dans la **préparation de plats simples et savoureux** »*

- Avantages
 - Recueil automatique
- Limites
 - Ne recueillent pas toutes les informations voulues
 - Développement « agile » => données non cohérentes



Données d'entreprises privées

Chaînes de restauration

- *Les usagers qui l'acceptent*
- *Peuvent avoir leur consommation enregistrée*
 - Avantages
 - Enregistrement automatique et systématique
 - Limites
 - En contexte spécifique
 - Seulement à midi
 - Choix limités
 - Populations biaisées
 - Développement « agile » => données non cohérentes

Les données idéales

Données sur les **historiques de consommation** d'un (grand) ensemble d'individus

- Mieux si on peut suivre le comportement de **chaque individu**

On aimerait pouvoir faire du **clustering** pour identifier des **classes de consommateurs** en se basant sur :

- Age
 - Taille de la famille
 - Éducation
 - ...
- Mieux si les **historiques** sont **longs** (plusieurs semaines / mois)
 - Mieux si les **détails des repas** sont connus

Bilan sur le recueil de données

1. Difficile de mesurer la consommation individuelle

- Pas de système de **mesure automatique**
- **Compliqué** (e.g. reconnaissance de photos de plats)

vs. montres connectées

2. Difficulté à recueillir des **historiques longs**

- **Coûteux** pour les individus
- **Pas d'incitation** à le faire

vs. médecine personnalisée

3. Populations et contextes **biaisés**

Vers des recommandations personnalisées

Projet ANR « SHIFT »: l'objectif

1. Identifier **ce qui détermine** les choix alimentaires des consommateurs
 - **Saisonniers**
 - **Sociétaux** (e.g. repas de famille, sur le lieu de travail, fête, ...)
 - Les repas pris **avant** et **après**
2. Réaliser un **système de recommandation personnalisé**
 - Fonctionnant **en-ligne**
 - Prenant en compte le **contexte** des repas
 - S'appuyant sur les facteurs **causaux**
Par contraste avec les facteurs qui sont corrélés mais non causaux

Tâches pour le projet ANR « SHIFT »

Task 1.1 Collecting a **large and comprehensive** enough dataset

- **Primary source of data:** INCA2 (and INCA3 if available on time)
 - But the temporal depth is very limited (a few days)
- **A new source of data is to be gathered**
 - A data collection campaign will be organized and performed
 - It is expected to gather data on 500 young adults for periods at least 4 months long over a year
 - Use of **apps on smartphones** (e.g. MyFitnessPal)

Tâches pour le projet ANR « SHIFT »

Task 1.2 **Preprocessing** the database and correcting the biases

— Identification of the systematic biases

- Biases that may be related to *gender, age, societal and cultural status, habits, ...*
- Exemples are almost entirely **positive**
- There might be systematic **missing information**
(E.g. *Side consumption outside of the main meals*)

Apprentissage **prescriptif**

- Apprentissage « **prescriptif** » (recherche de *causalités*)

1. J'observe que les gens qui mangent des glaces sont souvent en maillot de bain
2. Je voudrais vendre davantage de glaces

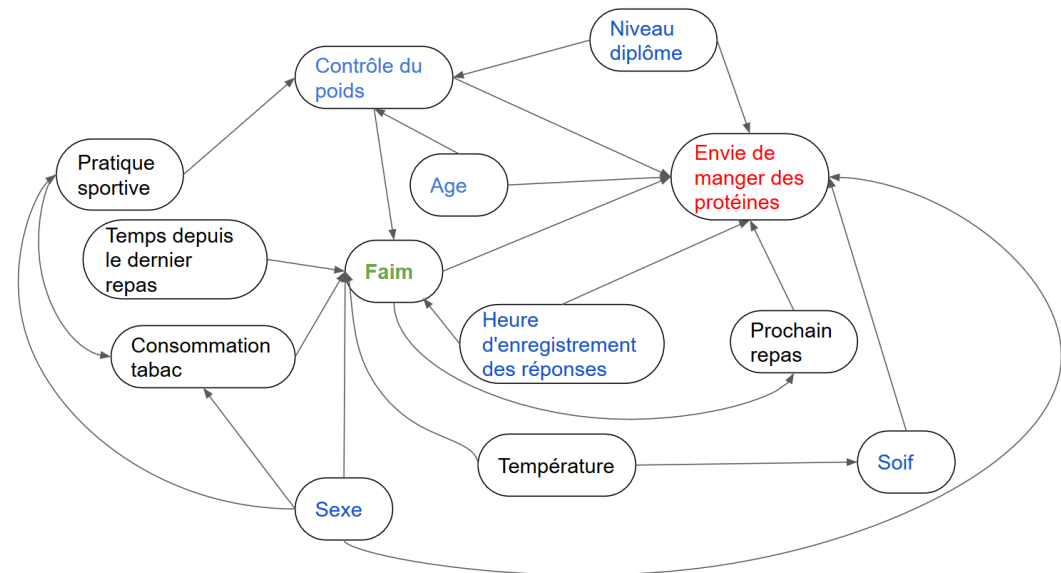
→ Je demande aux gens de se mettre en maillot de bain

- Quelles **recommandations** faire à un consommateur pour qu'il baisse sa consommation d'aliments carnés ?
- Quel impact **si on double le prix** de ... ?
- Quel rendement aurais-je eu l'année dernière **si j'avais** planté du ... au lieu de ...

La recherche de relations causales

Qu'est-ce qui **cause** l'appétence pour des plats protéinés ?

- La **faim** ?
- L'**heure** dans la journée ?
- Le **genre** ?
- L'**aspect visuel** ?
- L'**aspect olfactif** ?
- La richesse en **protéines des repas précédents** ?
- ...



Un coach nutritionnel

Étant donné :

- Un **individu** (et des données sur ses habitudes de consommation)
- Des données sur les **comportements alimentaires** d'un ensemble représentatif d'individus
- Des connaissances sur **ce qui détermine les choix** de consommation alimentaire

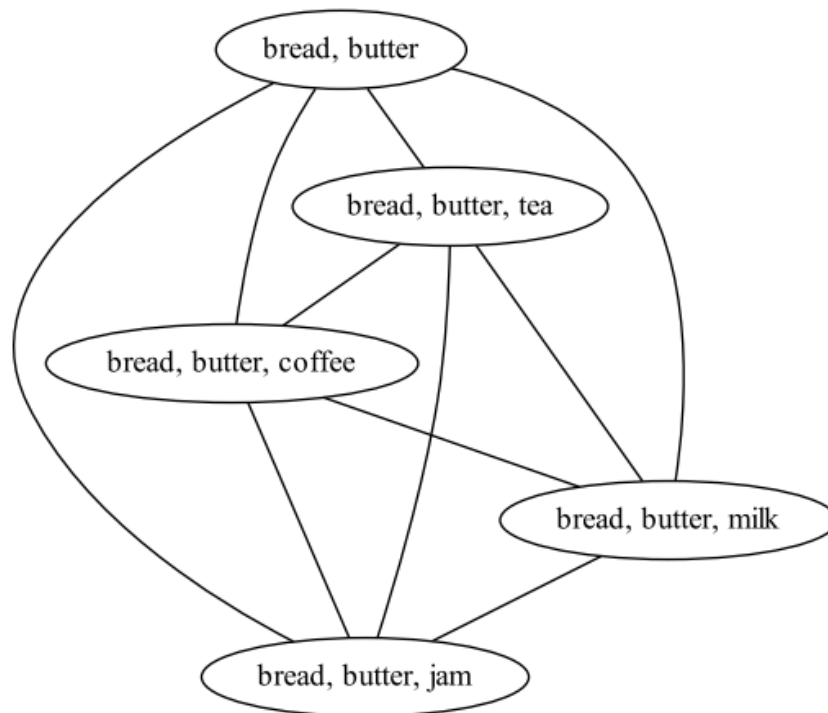
Suggérer des opérateurs

- Qui induisent des **modifications désirables d'habitudes** de consommation
- Qui sont **personnalisés**
- En **temps réel**

Opérateurs

Suggestions de substitution d'un plat par un autre

- Remplacer des frites par du riz

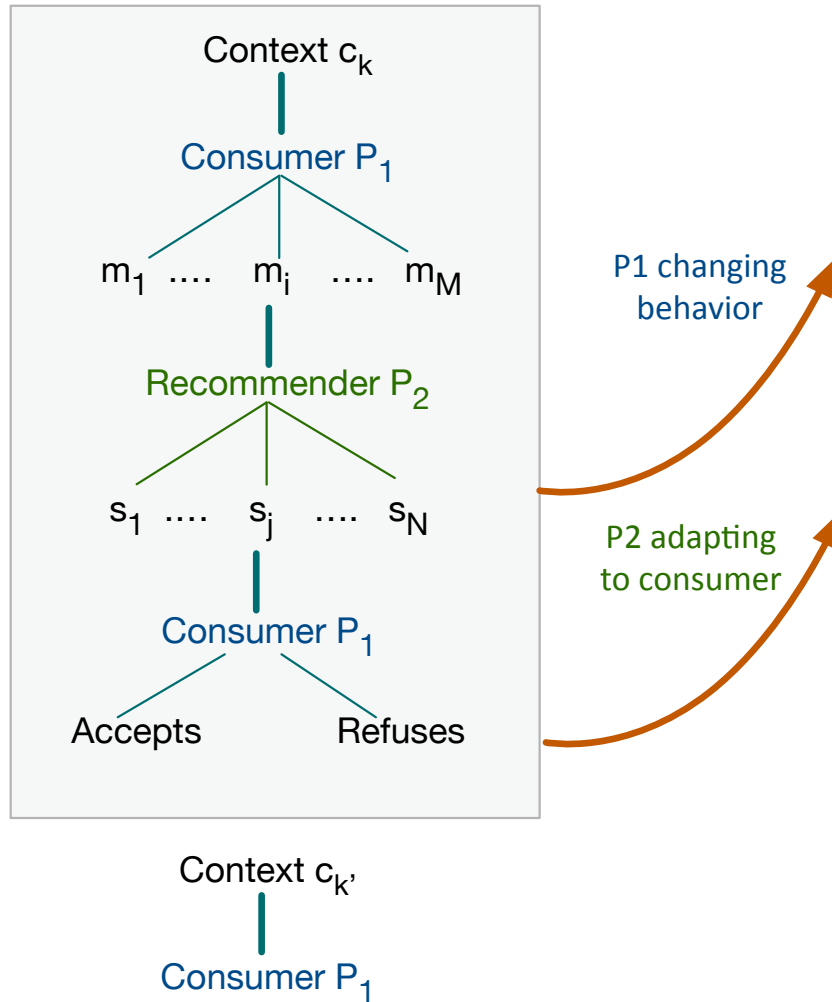


Breakfast and lunch		
Food Item	Substitute item (ordered by score)	Score
Bread	Rusk	0.2234
	Viennoiserie	0.1359
	Cakes	0.0745
Coffee	Tea	0.2799
	Cocoa	0.1729
	Chicory	0.1486
Tea	Coffee	0.2799
	Cocoa	0.1721
	Chicory	0.1289

Score de substituabilité :

$$f(x, y) = \frac{|C_x \cap C_y|}{|C_x \cup C_y| + |A_{x:y}| + |A_{y:x}|}$$

Un modèle



A model

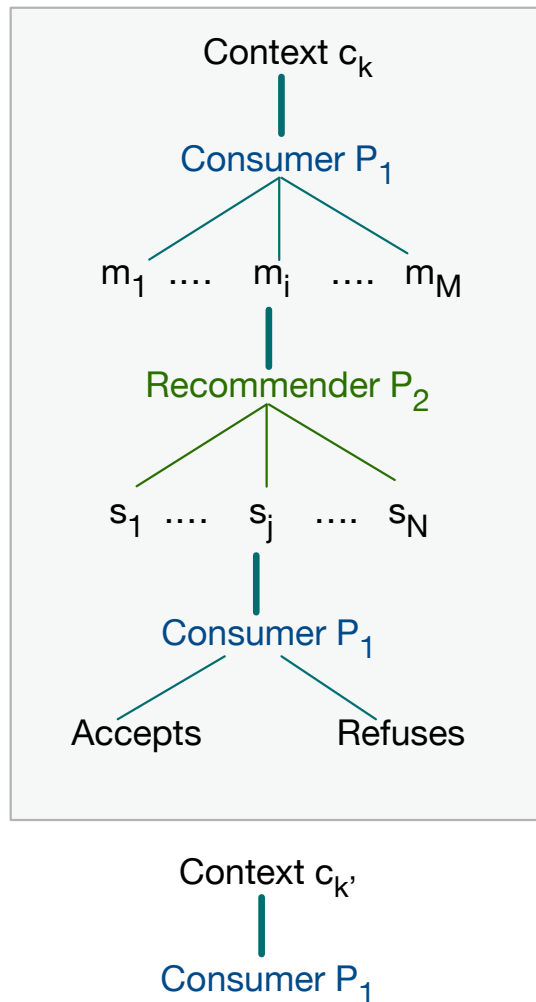
- Two “players”
 - P_1 : the **consumer**
 - P_2 : the **recommending system**
- P_2 acts as a “coach” to P_1
 - **Suggests** changes to P_1 's choices when appropriate
 - **Adapts** to the specificities of P_1 (probability of following suggestions)
- P_1 **may follow** or **not** the suggestions of P_2
 - If yes: **modifies** (slightly) its rules for future choices

Les objectifs

- **But :**
 - Que P_1 **atteigne** le plus haut **score Pandiet** possible étant données ses caractéristiques
 - Aussi **rapidement** que possible

- **Meta but**
 - Trouver la **politique optimale** pour le **coach** (P_2)
 - Personnalisée
 - Sensible au contexte

A model: learning phase



P1 changing behavior

P2 adapting to consumer

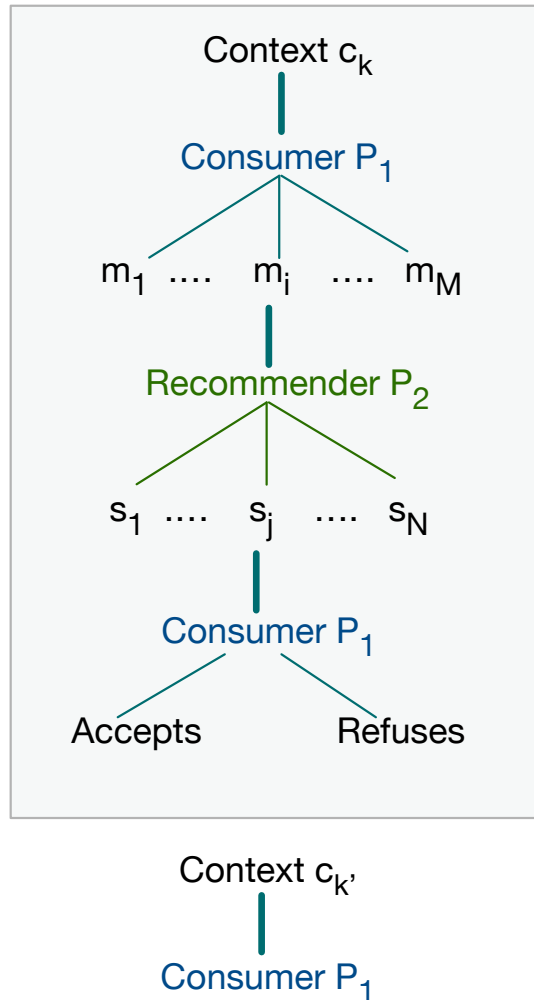
- P_1 : **speed of convergence** to optimized Pandiet score
- P_2 : **adaptability** to player P_1 and to context

Performance of learning depends on similarities

- between contexts
- between consumers

So as to generalize from specific experiences

A model: learning phase



- P_1 selects m_i according to

$$m_i = \underset{m}{\text{ArgMax}} [Q(m, c_k) + \text{noise}]$$

$$Q(s_j(m_i), c_k) \leftarrow Q(s_j(m_i), c_k) + \varepsilon$$

- P_2 selects s_j according to

$$s_j = \underset{s}{\text{ArgMax}} \text{score}(s, c_k)$$

$$\text{score}(s, c_k) = \text{Pandiet_gain} [s_j(m_i) - m_i] \times \underbrace{p(s_j(m_i) \text{ accepted by } P_1 \text{ in context } c_k)}_{\text{Estimated on line}}$$

Estimated on line

$$\tilde{p}(s_j(m_i) | c_k) = \frac{s_j(m_i) \text{ accepted in } c_k}{1 + \# \text{ tests}}$$

- P_1 selects 'accepts' according to $p(s_j(m_i) | c_k)$

Un problème de planification (en IA)

Mais :

- Les **opérateurs de changement** peuvent seulement être **suggérés**
 - Il faut qu'ils soient **acceptables**
- Leurs effets prennent du **temps pour être mesurables**
- Les opérateurs peuvent **interagir** de manières diverses
 - E.g. Changer le contexte du repas
peut rendre une substitution de plat impossible

Comment **mesurer l'efficacité** du système de recommandation ?

Spécificités

- On veut modifier **des habitudes**
 - **Renouvellement sans fin** des choix.
 - Dans des **contextes répétés**
- Modification **graduelle** de la fonction de décision du consommateur
 - Accepter des « **reculs** » parfois : stratégie de recommandation **non myope**
 - **Suggestions acceptables** ... et suffisamment variées
- **Non indépendance des choix** proches dans le temps

Conclusions

1. Le **recueil de données** en alimentation est un défi qui reste à relever
2. La **recommandation personnalisée** en alimentation
 - Est un problème **intéressant**
 - Voir aussi le concept de **nudging**
 - Demandant des développements **méthodologiques**
 - Avec un **fort impact** potentiel
 - Sur la **santé** publique
 - Sur la **durabilité** des systèmes de productions