Analyser, comprendre le monde
Complémentarité entre apprentissage et visualisation

Michele Sebag, Jean-Daniel Fekete

Coll. Philippe Caillou

Journée AFIA-IHM — 11 mars 2021

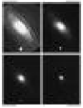
Machine Learning in general

Our team: AO/TAU, LISN, CNRS, INRIA, UPSaclay

ML-IHM complementarity, 2 case studies

Cartolabe: today, to-morrow

Learning, XVIII-XIX^e centuries

Ast. series 

Pierre de Rosette 

World

Natural phenomenons

Human–related phenomenons

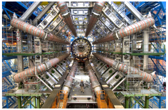
Data / Principles

Maths. Modelling

Common Sense

You are here

Learning, XXI^{e} century

Sc. data 

Google

World

Natural phenomenons

Human–related phenomenons

Data / Principles

Maths. Modelling

Common Sense

You are here

Types of Machine Learning problems

WORLD – DATA – USER

Observations	+ **Target**	+ Rewards
Understand **Code**	**Predict** **Classification/Regression**	Decide Action Policy/Strategy
Unsupervised **LEARNING**	**Supervised** **LEARNING**	Reinforcement LEARNING

Machine Learning in general

Our team: AO/TAU, LISN, CNRS, INRIA, UPSaclay

ML-IHM complementarity, 2 case studies

Cartolabe: today, to-morrow

New Artificial Intelligence

A Case of Irrational Scientific Exuberance

- Underspecified goals — Big Data cures everything
- Underspecified limitations — Big Data can do anything (if big enough)
- Underspecified caveats — Big Data and Big Brother

TAU goals

TAckling the Underspecified

- Learn trustable models
- Mix data **and** knowledge in basic sciences — principles, models, softwares
- Characterize the learning landscape

an end and a mean to knowledge transfer

TAU

Three scientific pillars

- Enforcing fairness and robustness — Fair AI
- ML with/for scientific modelling
  From data assimilation to NN-based Numerical Analysis
- Learning to learn

Main three application domains

- Energy
- Human and Social Sciences
- Numerical Engineering

Organization of challenges

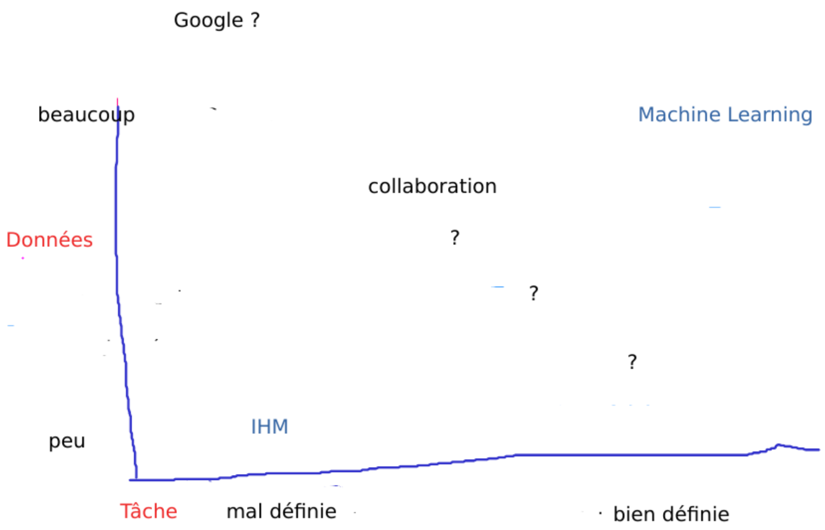
Machine Learning in general

Our team: AO/TAU, LISN, CNRS, INRIA, UPSaclay

ML-IHM complementarity, 2 case studies

Cartolabe: today, to-morrow

ML - IHM : Complémentarité



Exploratory Data Analysis

Clustering

- Input: data
- Desired output: making sense of it...
- How:
  - Divide and conquer
  - Visualize = project in 2-3 D.

Issues

- Which distance ?
- Which projection criterion ?

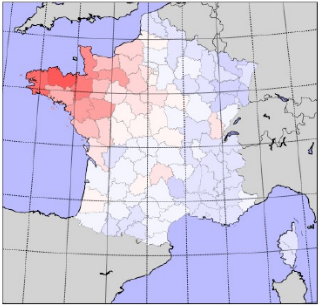
Case study 1: Nutriperso

Coll. O. Allais, P. Caillou, A. Constantinescu, K. Gasnikova

The Kantar data

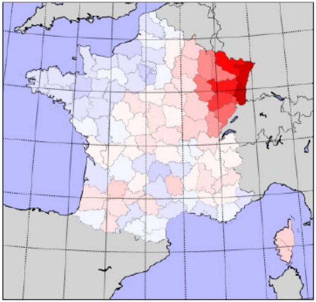
- A wealth of data: 20,000 households, 180,000 food items
- Goal: assess impact of food on health.

Identify (elements of) diets: some have a geographical support



Thym frais,Fromage camenbert, Beurre, Cidre, Café, Produit sucrant, Lait, Crème fraîche, Vinaigre, Rillettes

PRESIDENT, PAYSAN BRETON, COIC.ESCALE BRETONNE, GRAND FERMAGE, SAN MARCO,



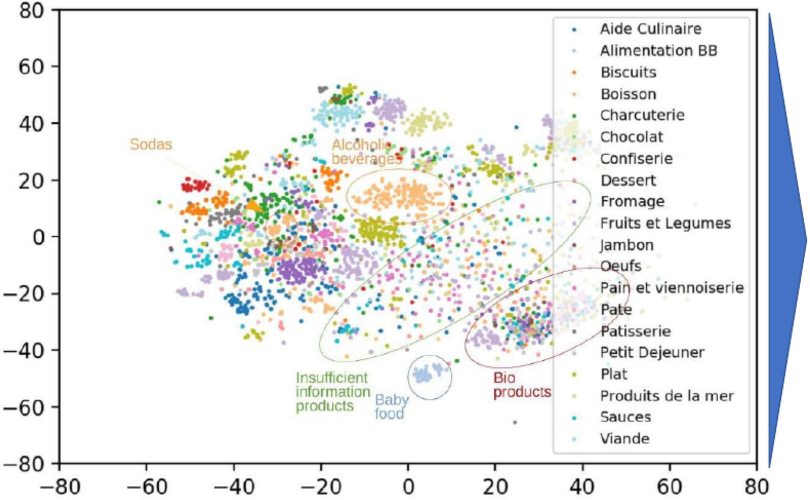
Gelée A Préparer, Fécule, Aide A La Pâtisserie,Jambon surgelé, Chapelure, Farine, Bouillon Et Court-Bouillon, Autre aide culinaire

SCHMIDT, GEO, ADAM, TEMPE, KAUFFERS, AGNESI,CUSENIER

Case study 1: Nutriperso, foll'd

Coll. O. Allais, P. Caillou, A. Constantinescu, K. Gasnikova

Localize types of food

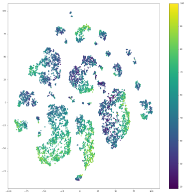


Case study 2: Eurostate

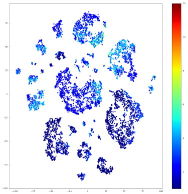
Coll. N. Schwencke, E. Ollion

The SRCV data

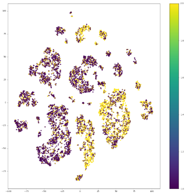
- A wealth of data: 10,000 households, 1,000 +features
- Goal: describe population.



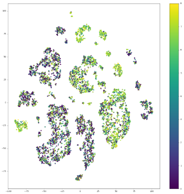
Age



N. persons



Gender



Size urban unit

Machine Learning in general

Our team: AO/TAU, LISN, CNRS, INRIA, UPSaclay

ML-IHM complementarity, 2 case studies

Cartolabe: today, to-morrow

Current ML pipeline: CARTOLABE-DATA

From words/vectors to dimensionality reduction

- Each entity (article, author, lab) $\rightarrow$ a vector in $\mathbb{R}^d$
- Method choice ?
  - Singular Value Decomposition
  - Latent Dirichlet Allocation
  - BERT
- Hyper-parameter: number of dimensions d

Naming regions

- Most frequent & discriminant terms
- Manually tuned to avoid nonsense

Next

The need

- There is no such thing as a good distance for all
- Learn the distance
- More generally: Interactive intent modelling Ruotsalo et al. 18

From usages to expectations

- How to get feedback ?
- Asking scores ? :-(
- Providing summaries