Deep and Shallow Learning at TAU

Marc Schoenauer and Michèle Sebag
TAU team, INRIA-CNRS-UPSud / Université Paris-Saclay

DATAIA-JST International Symposium on Data Science and AI
July 10., 2018

Non-standard uses of Deep Learning

Deep Learning At TAU
- Image Classification
- Population Genetics
- Power Grid Optimisation
- Combating Unemployment

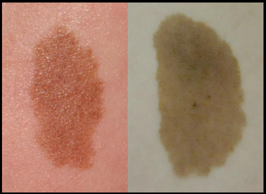
Not-so-Deep Learning at TAU
- Social Networks
- Causality

Wrap-up

(simple) Image Classification

G. Charpiat

- Dataset of skin pictures, from a hospital
- Classes: operate / don't operate
- Difficulties: small part of the image, detection, white balance...
- Methodology:
  - download pre-trained network and train with own dataset
  - OR build a 5+2-convolutional network and train – using standard libraries



Deep Learning At TAU

- Image Classification
- Population Genetics
- Power Grid Optimisation
- Combating Unemployment

Not-so-Deep Learning at TAU

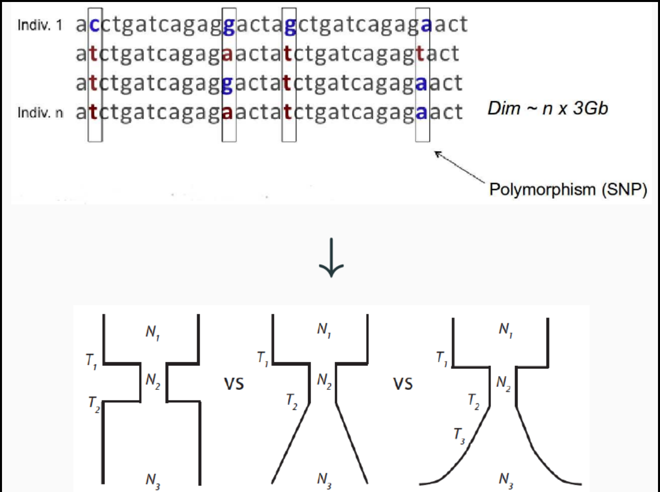
- Social Networks
- Causality

Wrap-up

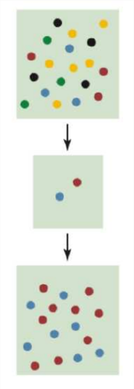
Deep Population Genetics

G. Charpiat & F. Jay

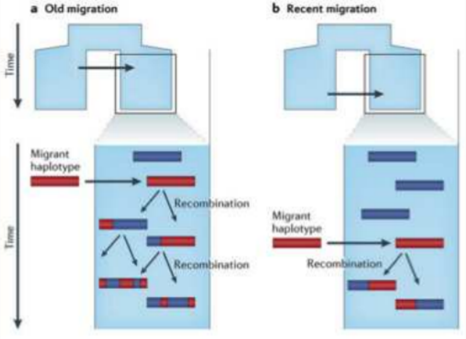
Reconstruct demographic history from today ADN sequences



Demographic History Traces


Slatkin

Tightening

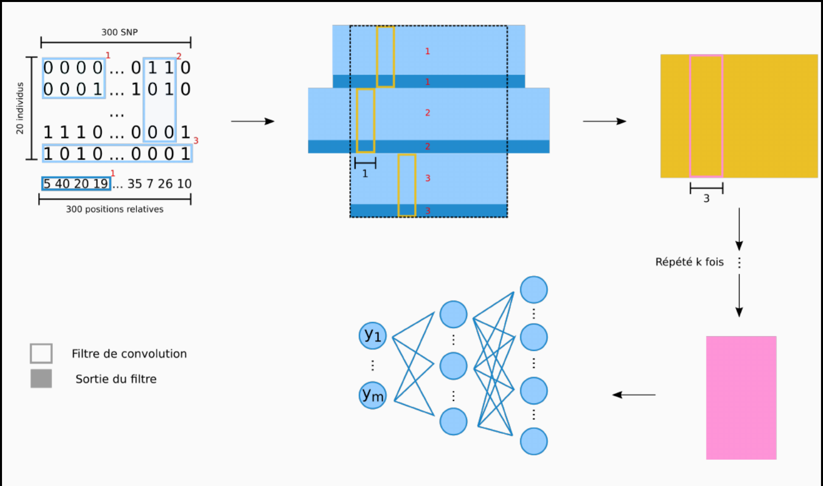


Sousa & Hey 2013

Interbreeding

The architecture

Calibration of a given model



The French Power Grid

B. Donnot, A. Marot (RTE), I. Guyon

Context

- Find curative actions — on/off switches
- to prevent "n-1" security loss

Network

- 7000 *consumers*, 3000 *producers* — including 300 renewable
- 10 000 electric lines, 30 000 topology switches
- 5mn time steps
- A physical simulator
  - expensive — .5s per simulation $\rightarrow$ 1 CPU day per week
  - fragile — 30% failures $\rightarrow$ find better initialization

Historical Data

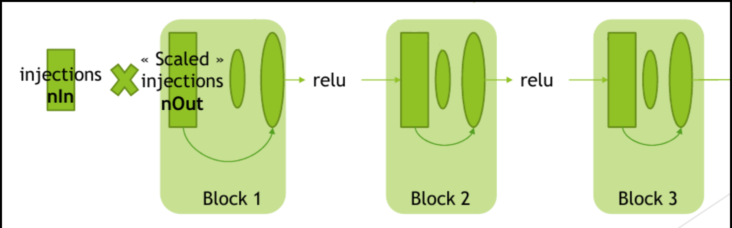
- 10 years record, $\sim$ 400 000 manoeuvres per years, but
- many possible causes, $<$ 20% curative actions — no way to tell

Deep Learning @ work

An ML/Optimization hybrid problem

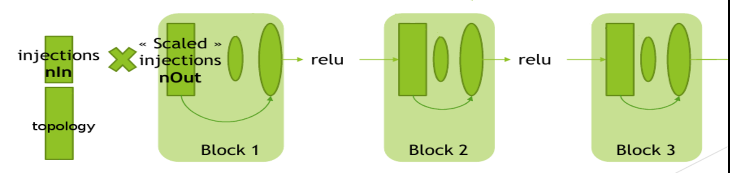
- Learn to distinguish curative from maintainance actions simulating "what-if" scenarii
- with greedy "n-1" constraint to be checked at all steps
- $\longrightarrow$ Deep surrogate of the simulator

Deep residual network

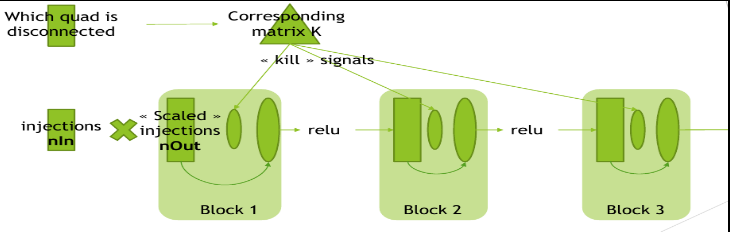


But how to include the topology?

One hot encoding one boolean input per line



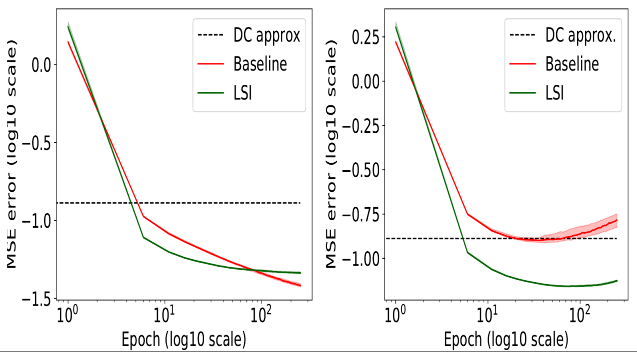
Guided dropout each missing line 'kills' some connections → additive adaptivity



Results on Artificial Data

118 nodes, 99 consumptions, 54 productions, 186 power lines
15000 simulated injections, 1-defect learning

One-hot encoding vs Guided dropout

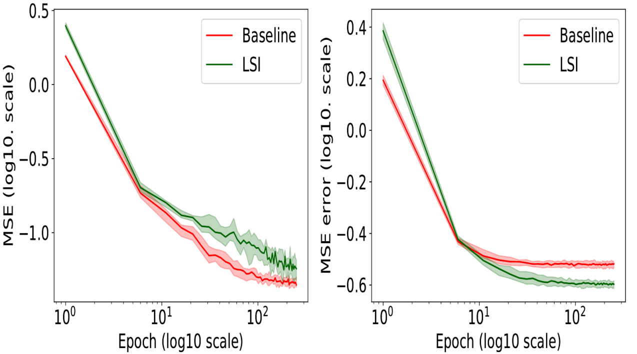


1-defect generalization 2-defects generalization

Results on Real Data

368 nodes, 246 consumptions, 122 productions, 387 power lines
Real data from 2012 to May 2017 for training,
and from June and July 2017 for test, 1-defect learning

One-hot encoding vs Guided dropout



1-defect generalization

2-defects generalization

Combating Unemployment with Big Data

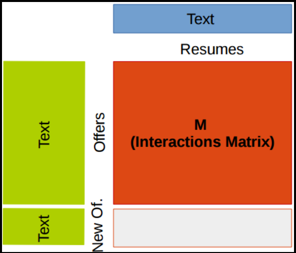
Coll. Qapa

Combating Unemployment

- How to deal with *Market friction*?

  unfilled job offers **and** unemployed people

- A collaborative filtering problem



Goal Matching job offers and CVs

From descriptions + history of interactions

How Extending Collaborative Filtering (recommending job seekers to recruiters)

The MAJORE project (MAtching JObs and REsumes)

A Collaborative Filtering problem

The Qapa Data

2012-2016

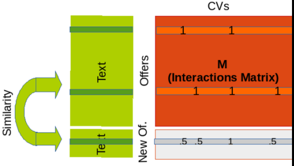
- 2 millions of jobs offers
- 1.5 millions of CVs
- 13 millions of interactions

2 months period

- 11,000 recruiters (users)
- 7,000 CVs (items)
- 80,000 interactions

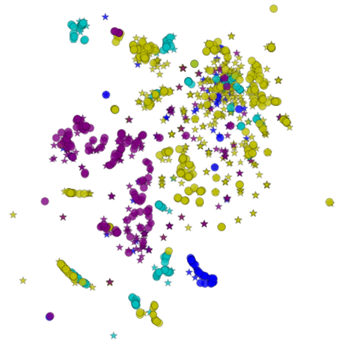
The Cold Start problem

- Recommend new offers (jobs)
- Using jobs (offers)similarities
- direct approaches fail (tf-idf or LSA)

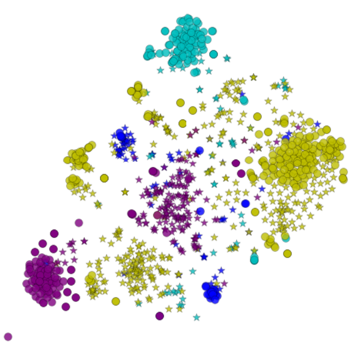


Representation of CVs and Job Offers

Offers (circle) and CVs (star)



Rows/col of $\mathcal{M} \rightarrow$ SVD(500) $\rightarrow$ t-SNE

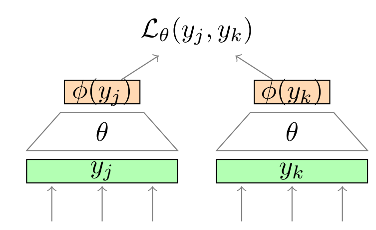


Bags of words $\rightarrow$ LSA(500) $\rightarrow$ t-SNE

Seekers/recruiters understand each other (left);
but they don't speak the same language (right)
$\longrightarrow$ **DNNs to learn an ad hoc representation**

Learning a Representation

Siamese Networks



with loss $\mathcal{L}_\theta(\mathbf{y}_j, \mathbf{y}_k) = \mathbf{1}_{\{sim^*(j,k)=1\}} \|\Phi_\theta(\mathbf{y}_j) - \Phi_\theta(\mathbf{y}_k)\|_2^2 + \mathbf{1}_{\{sim^*(j,k)=0\}} \left(m - \|\Phi_\theta(\mathbf{y}_j) - \Phi_\theta(\mathbf{y}_k)\|_2\right)_+^2$

where $sim^*(j,k) = \begin{cases} 1 & \text{if } \langle \mathcal{M}_{\cdot,j}, \mathcal{M}_{\cdot,k} \rangle > 0 \\ 0 & \text{otherwise} \end{cases}$

Seeded Influencer Identification

Ph. Caillou & M. Sebag, coll SME Augure

Goal

- Find the influencers in a given social network
- from public data sources — tweets, blogs, articles, ...
- Commercial goal: so as to bribe them :-)
- Scientific goal: with little user input

What is an Influencer?

- Highly retweeted? — but # retweets disagrees with # followers
- Sources of information cascades?
- # invitations to join?
- Topic-specific PageRanking?
- Presence on Wikipedia?

Seeded Influencer Ranking

Input

- A social network
- (Big) Data of user interactions — Tweets, blogs, messages, . . .
- Some identified influencers

The global picture

- Derive features representing the users — using content and traces
- Optimize scoring function — giving highest scores to known influencers
- return best-k scoring users — the most (known + new) influencers

Data

Sources

- random 10% of all retweets of November 2014
- 100M tweets, 45M unique tweets,
- with origin-destination pairs
- 43M nodes graph
- only words in at least 0.001% and at most 10% tweets considered
- leaving 10.5M candidates

Features

100 Content-based Features

foreach medium eventually

- Foreach user x, identify $\mathcal{W}(x)$, the N words with max tf-idf

  term frequency - inverse document frequency
- 50 words that appear most often in all $\mathcal{W}(\mathit{influencer})$
- 50 words with max sum of tf-idf over $\bigcup \mathcal{W}(\mathit{influencer})$
- each selected word is a feature for x: 0 if not present in $\mathcal{W}(x)$, ti-idf otherwise

  Many are null for most candidates

6 Network Features

- from the weighted graph of retweets

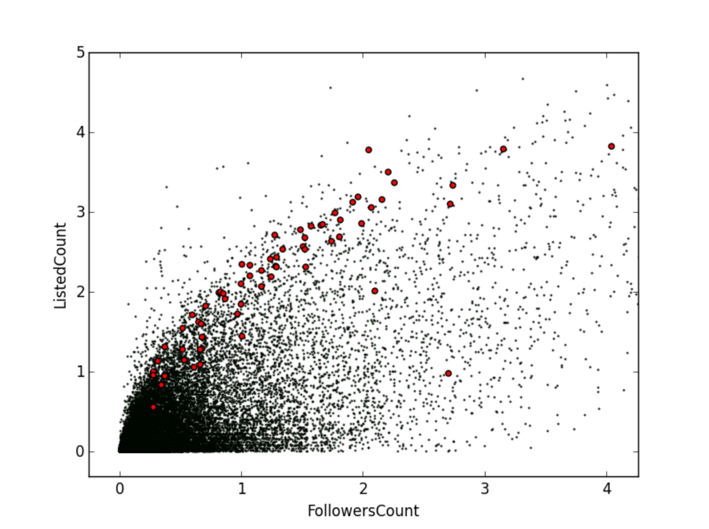
  centrality, PageRank, ...

5 Social Features

- from tweeter user profiles

  # tweets, # followers, ...

Better than Basic Social Features



Transfer to Augure Company

TOUTE L'ACTUALITÉ / BUSINESS / **FINANCEMENT**

Augure lève 15 M€ pour identifier les influenceurs grâce au big data

Maryse Gros, publié le 31 Mars 2015

. . . in the blog *Le monde informatique* . . .

marketing de la société et directeur général de la filiale espagnole.

Un programme de recherche avec l'Inria

Depuis deux ans, Augure a lancé avec l'Inria un programme de recherche destiné à exploiter les technologies d'apprentissage machine pour aider les marques à exécuter et mesurer leurs campagnes de relations avec les influenceurs. L'Augure Lab est dirigé par Fabien Barzic et le comité scientifique constitué avec l'Inria est supervisé par Michèle Sebag, directrice de recherche au CNRS. L'objectif est de tirer profit des big data pour « optimiser les processus d'identification des influenceurs », explique la société. L'investissement de 15 M€ engrangé par Augure va en partie contribuer à renforcer ce programme d'innovation technologique.

L'apport financier servira aussi à soutenir la croissance organique de la société en Europe et à

Causality

D.Kalainathan, O.Goudet, P.Caillou, M.Sebag, P.Tubaro, I.Guyon

Quality of life at work and firm performance: A controversial issue

- quality of life at work ! a mean of enhancing productivity and competitiveness
  La fabrique de l'industrie, Bourdu, E. et al., 2016
- Literature review and analytical study of relationship between social and financial performance
  Allouche et Laroche, 2005

Off-the-shelf Data Analysis

- 408 firms
- 222 "social" variables grouped into 8 domains: Employment structure, Employment dynamic, Job security, Renumeration levels, Training, Health and safety, Professional equality, Social relations
- 21 financial ratios: Global business results (6), Productivity and capital intensity (4), Balance sheet ratios (5), Investment efforts (3), Profitability (3)

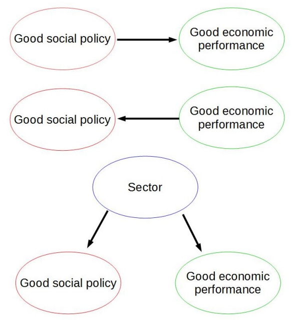
Need for Causality Analysis

(very partial) Conclusions

- The firms in the cluster with the best social policy $\rightarrow$ best productivity
- The firms in the cluster with the worst social policy $\rightarrow$ worst productivity per employee . . . but good financial profitability

Can we make recommendations for managers?

- Correlation doesn't mean causality
- Different causal models for the same observed correlation



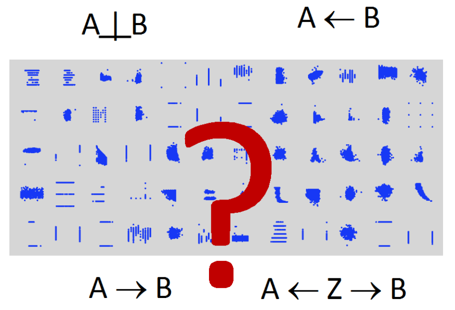
Possible causal models

A Machine Learning approach

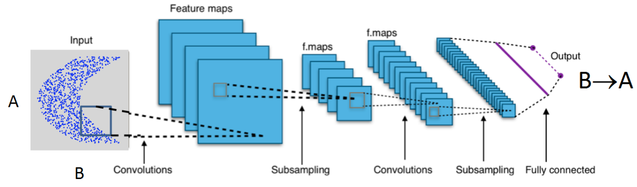
- From set of labeled pairs ($X \rightarrow Y$, $Y \rightarrow X$, $X \perp Y$ and $X \parallel Y$)
- Learn a causation score
- Derive causation graph . . . and prune it

Causality Challenge

I. Guyon@TAO

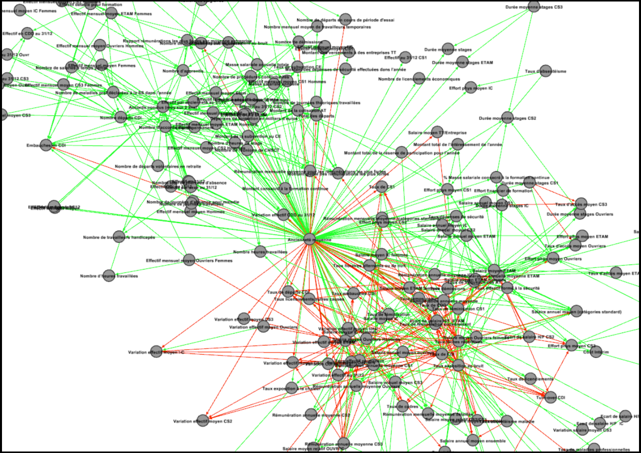


Training Data: labeled densities



Similar to images

Full causal graph



Amiqap data: positive and negative causal effects

Need for tools to prune/analyze causality graphs

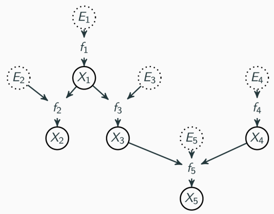
Causal Modeling

Relevance: From models to causal models

- Decreased sensitivity wrt data distribution
- Support interventions — clamping variable value
- Hopes of explanations

Formalism: Functional Causal Models $(X_1, \ldots, X_d)$ — Pearl 09

- $Pa(X_i)$: Direct causes for X_i
- All unobserved influences: noise variables E_i



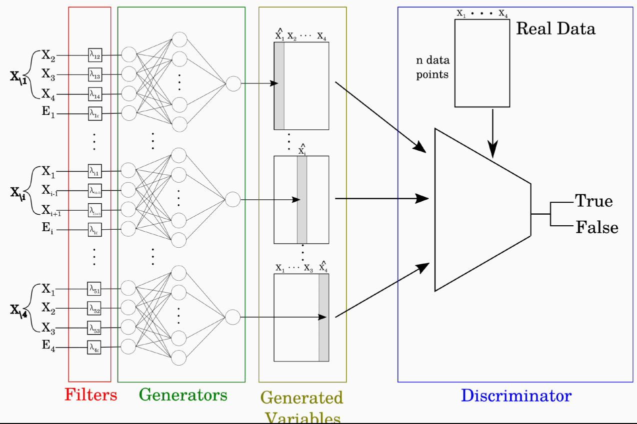
$$
\begin{cases}
X_1 = f_1(E_1) \\
X_2 = f_2(X_1, E_2) \\
X_3 = f_3(X_1, E_3) \\
X_4 = f_4(E_4) \\
X_5 = f_5(X_3, X_4, E_5)
\end{cases}
$$

Structure Agnostic Modeling

Goudet et al. 18

A dynamic process without self-loops

$$Xi = f_i(X_{\setminus i}, E_i)$$



Structure Agnostic Modeling (2)

Ingredients for $f_i(X_{\imath}, E_i)$

- Scaling factors $a_{i,j}$ impact of X_j on X_i
- Dense layer with non-linear activation function
- Linear readout

$$\hat{X}_i = m_i^t \tanh(\bar{W}_i^t(a_i \odot X) + n_i E_i + b_i) + \beta_i$$

Structure Agnostic Modeling (3)

Loss function

- Adversarial learning

$$L_i = \mathrm{E}_{x_i, x_{\setminus i}}[\log D(x_i, x_{\setminus i})] + \mathrm{E}_{e_i, x_{\setminus i}}[\log(1 - D(\hat{f}(e_i, x_{\setminus i}), x_{\setminus i}))]$$

- $+$ Regularization — enforcing graph sparsity

$$L_\lambda = \sum_{i=1}^{d} L_i + \lambda \sum_{i=1}^{d} ||a_i||_1, \lambda \geq 0$$

A competition between d sparse causal mechanisms $\hat{f}_i$ and a shared discriminator D.

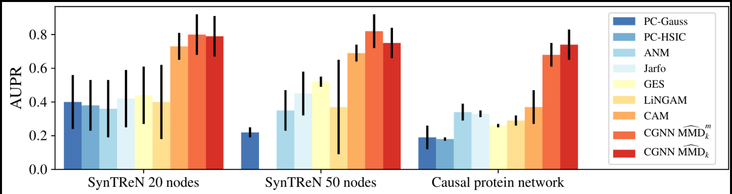
Discussion

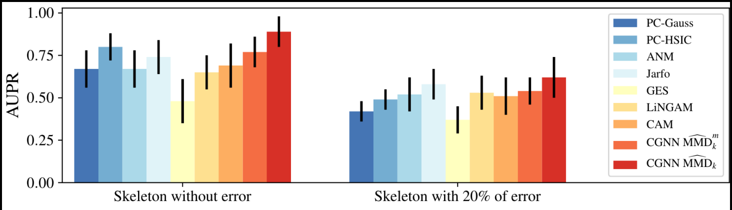
- No combinatorial search — scalability
- Cycles are possible: either genuine; or indicate non-identifiability

Results

on-going work

. . . on standard benchmarks





Deep Learning At TAU
- Image Classification
- Population Genetics
- Power Grid Optimisation
- Combating Unemployment

Not-so-Deep Learning at TAU
- Social Networks
- Causality

Wrap-up

AI@TAU

Fundamental

- Optimisation (Riemanian Geometry, SGD) — Y. Ollivier, G. Charpiat, MS^2
- DNNs structure learning — G. Charpiat, I. Guyon
- Causality — I. Guyon, M. Sebag

Energy and Safety

- Power Grid Optimisation — RTE
- Simulator Calibration and Algorithm Configuration — ADEME
- Intrusion detection / Reinforcement Learning — Thalès Theresis
- Validation scenarios for the autonomous vehicle — Renault

Computational Social Sciences

- Matching jobs and Resumes — Qapa
- Quality of life at work and firm performance — SES Telecom and FdI
- Social Networks — Augure
- Diet vs Socio-demographic vs Health — IRS Nutriperso

Algorithm Selection and Configuration

- Continuous optimisation — Thalès TRT, winner 1-obj track, BBComp2017
- Combinatorial optimisation — IRT SystemX, winner OASC 2017