DataSense: Data Intelligence

Céline Hudelot & Bertrand Thirion

MICS Laboratory & Inria Saclay

DigiCosme Research Days
Gif, 5–6/6/2018

DataSense: Data Intelligence

- Data are everywhere
  - ► The Web, information systems, sensors (smartphones, etc.)
  - ► Knowledge Graph, Wikipedia, WordNet...

- Big data makes new things possible
  - ► Statistical machine translation
  - ► e-science
  - ► Business intelligence
  - ► ...

Research Teams in DataSense

- DAVID — Université Versailles St Quentin
- Inria Saclay
- L2S — CentraleSupélec
- Li-PaRAD — Université Versailles St Quentin
- LIMSI — CNRS - Université Paris-Sud
- LIST — CEA
- LIX — École Polytechnique
- LMV — Université Versailles St Quentin
- LSV— ENS Paris-Saclay
- LRI — Université Paris-Sud
- LTCI — Télécom ParisTech
- MICS — CentraleSupélec
- SAMOVAR — Télécom Paris-Sud
- U2IS — ENSTA ParisTech

Plan

1. Structuring

2. Activities

3. Perspectives

DataSense 'Tasks'

1. Scalable, expressive and secure tools for large-scale data
2. Making sense of complex, heterogeneous data and knowledge
3. Machine learning: multi-task learning and meta-learning
4. Distributed decision making: reinforcement learning, partially observable processes and games
5. Interaction and Visualization

Global DataSense Activities

- Main global activities:
  - Research days (once a year)
  - Community building
  - Preparation of working groups and co-supervised theses
  - Missions doctorales
- Activities per 'Task'
  - Working groups
  - Co-supervised PhD theses
  - Induced collaboration, guest scientists
  - Post-doc and engineers
- Emerging projects:
  - SEDA: Breathing sense into data
  - Neural Meta Tracts
  - IoTA Internet of Things Analytics

DataSense Working Groups
Bottom-up creation of DigiCosme working groups

- **ERVEN** : Extraction, Représentation et Visualisation de connaissance pour l'Enseignement Numérique – 2018 / 2019
  - ▶ Anne-Laure Ligozat
- **E-santé** : Internet des Objets & E-santé – 2017 / 2019
  - ▶ Mehdi Amni
- **TAL & SEM** : TRAITEMENT SÉMANTIQUE DES DONNÉES TEXTUELLES – 2017 / 2018
  - ▶ Brigitte Grau, LIMSI
- **SDT** : Sécurité des Données Textuelles – 2016 / 2017
  - ▶ Cyril Grouin, LIMSI
- **D2K**: From Data to Knowledge – 2015/2017
  - ▶ Claire Nédellec, INRA; Chantal Reynaud, LRI
- **SSSL** : Sequential Structured Statistical Learning – 2015 / 2017
  - ▶ Oldaric Maillard, LRI

DataSense Working Groups
Bottom-up creation of DigiCosme working groups

- **SciCoSense:** Building a cartographic map of scientific communities – 2015/2016
  - ▶ Philippe Caillou, LRI
- **Human-Robot Interaction** – 2015/2016
  - ▶ Laurence Devillers & Jean-Claude Martin (LIMSI-CNRS)
- **Deep Learning and Distributed Representations** – 2014/2018
  - ▶ Alexandre Allauzen, LIMSI
  - ▶ Emmanuelle Frenoux, LIMSI
- **PASADENA**: Prédiction et Analyse de données structurées et hétérogènes – 2015/2018
  - ▶ Arthur Tenenhaus, CentraleSupélec
  - ▶ Maxime Sangnier, TPT
  - ▶ Flora Jay, LRI

Guest Scientists

- **Kevin Bretonnel Cohen** - Director, Biomedical Text Mining Group
  Computational Bioscience Program / University of Colorado School of
  Medicine - February / April 2016
  Contact : Pierre Zweigenbaum

- **Catherine Plaisant** - University of Maryland Institute for Advanced
  Computer Studies - June / July 2017
  Contact : Jean-Daniel Feketé (Inria)

- **Timothy Miller** - Boston Children's Hospital & Harvard Medical
  School - Juin / Juillet 2017
  Contact : Aurélie Névéol (LIMSI)

- **Ramon Pino Pérez** - Department of Mathematics, University of the
  Andes (Venezuela) - April/ June 2018
  Contact : Isabelle Bloch (Télécom ParisTech)

PhD students

- **HiDimStat** : Statistical control of sparse models in high dimension - 2017
  Bertrand Thirion (Inria) & Joseph Salmon (LTCI)
- **Idiab** : Internet des Objets pour le suivi et la modélisation de la glycémie de patients diabétiques - 2017
  Mehdi Ammi (LIMSI) & XXX
- **AlCoMol** : Algorithmique de graphes pour l'aide á la décision dans la construction moléculaire - 2016
  Domnique Barth (DAVID)
- **OPALE** : Opérateurs monotones aléatoires et applications á l'optimisation stochastique - 2015
  W. HACHEM, P. BIANCHI, J. JAKUBOWICZ, LTCI/ SAMOVAR
- **COT** : Coréférence événementielle cross-document dans les dossiers électroniques patient - 2015
  A. NÉVÉOL, X. TANNIER, O. FERRET, LIMSI / CEA-LIST
- **SEDA** : BREATHING SENSE INTO DATA - 2015. Fabian Suchanek, LTCI
- **SensoMotor-CVE** : Murs d'images en contexte interactif complexe – 2015
  Patrick Bourdot, LIMSI
- **BIPIMA** : BIPolarité de l'Information Multimédia pour l'Annotation sémantique d'images dans un contexte de médias sociaux - 2014

Post-doc and engineers

- **MAEL** : MultimediA Entity Linking - 2018
  Contact : Hervé Le Borgne, CEA LIST

- **VASTE** :Veracity Assesment in Spatio-TEmporal heterogeneous data.
  Contact : Fatiha Sais, LRI

- **MetaTracts** : Parsimonious multi-resolution representations for statistically analyzing brain tractograms - 2018
  Contact : Pietro Gori, LTCI - LIX

- **PASADENA** – 2017
  Contact : Remy Boyer & Franck Nielsen, L2S

- **AMPHI** : Approximate Message Passing for HIgh-dimensional data - 2017
  Contact : Bertrand Thirion, Inria & Jospeh Salmon, LTCI

- **TAL & SEM** : Traitement Sémantique des Données Textuelles - 2017
  Contact : Brigitte Grau & Olivier Ferret, LIMSI

- **D2K** : De la Donnée à la Connaissance - 2017
  Contact : Jean-Daniel Feketé, Inria

- **SEDA** - 2015. Contact : Fabian Suchanek, LTCI

Doctoral missions

- **Plateforme Camomile d'annotation collaborative de documents multimédia** - 2016
  Direction : Hervé BREDIN, LIMSI
  `https://github.com/camomile-project/camomile-polymer-client`
- **Software for easier access to open M/EEG data repositories** - 2016
  Direction : Alexandre GRAMFORT; LTCI
  `https://github.com/jasmainak/bids-validator`

Plan

DataSense Perspectives

Within the STIC Domain

- Continuation of the support of collaborative research in the DataSense tasks: both fundamentals and applicative works.
- Stronger interactions with ComEx and SciLex
  - ▶ e.g. develop information-theoretic analysis of deep learning
  - ▶ e.g. formal methods meets machine learning, towards safe AI-systems, machine learning for optimizing ontology reasoning...
- Stronger coordination with DataIA (convergence institute) + Center for Data Science + Optimal Control & signal communities
  - ▶ joint calls
  - ▶ (junior) summer schools, software development
  - ▶ Future AI institute (?)

DataSense Perspectives

Beyond STIC: given the maturity of the domain, time to open to new communities

- Open Digicosme to new labs (e.g. INRA maIAGE; INRA mia)
- E-science: have more common projects with other departments
  - Life science
  - physics
  - (?) SHS
  - Maths, theoretical physics

THANK YOU!