Data Science

PANEL DISCUSSION
moderated by
E. Le Pennec, M. Vazirgiannis

ORSAY – 30/06/2014

DATA SCIENCE

[If Facebook Can Profit from Your Data,
Why Can't You?*](#)

"Data Science" has become a popular (buzz)word...

- Data Science
- Data Science
- Education Training

Data Science…

- Types
  - Structured
  - Semi/unstructured (web, text)
  - Graphs (social nets, bio/med, …)
- Scale
  - Raw data - PB of satellite images downloaded each day
  - Users - Youtube viewers (2011) spent ~324.000 years per month viewing videos!
    - Each user produces yearly ~ 3GB of usage data
- Heterogeneity/Dimensionality
  - DNA data, Physics

Data Science...

- Data generate value
  - Scientific - Physics...
  - Economic – Marketing/advertising
  - Social – Health, smart cities

- Open Data...
  - Provide a leading vision for how governments approach open data.
  - Increasea awareness of open data and related issues.
  - Support the development of the global open data community.
  - Gather, expand, and amplify the evidence base for open data

http://globalopendatainitiative.org/

Data Science…

Data science life cycle

- Data pre-processing - normalization, feature selection & creation, descriptive data quality

- Data exploratory analysis (summary statistics, correlation …)

- Visualization/Navigation

- Machine learning (supervised & unsupervised),

- Bigdata Engineering issues

Data Science…

Methodological Challenges – in the context of bigdata..

- Advanced matrix factorization and dimensionality reduction

- Causality vs. Correlation + Interpretations

- Bigdata Learning theory

- Deep Learning

- Privacy/security

Engineering Challenges

- Advanced Bigdata technology (Spark, NoSQL databases, distributed file systems, object storage systems),

- Big distributed learning.

- Adapting to the domain…

Inter-disciplinarity challenge

- (Applied) Math, CS, Engineering, Physics

- Domain Sciences…

- Common approach/different technologies & terminologies ?

Data Science Ecosystems

- Many international conferences
  - IEEE Bigdata
  - ACM KDD 2014 - Data Science for Social Good
  - NETSCI conference

.....

- Lots of Bigdata*, Data Science* forums, blogs blocks
- Open data initiative…

Data Science Education/Training

- Numerous - Master programs (local as well..)
- MOOC courses (i.e. J. Hopkins @ Coursera)
- Even a doctoral school (Univ. of Edinburgh)

http://www.inf.ed.ac.uk/student-services/data-science-cdt

Further challenges ?

- Data Science for all…?
  - Initial
  - Continuous

  i.e. **D**ata **S**cience **S**tarter **P**rogram training programme in Polytechnique

- Saclay/France coordination …

Proposed Discussion Topics

Open Data

Interdisciplinarity: specificity vs. generality

- Common Terminology?

- Convergence of technology/algorithms

- New Algorithms frameworks (deep learning) ?

Education/Training

CDS

- International positioning

- Organization

- Tools

Further topics …?