Annual CDS Pitching Day

Rapport sur les contributions

ID de Contribution: **0**   Type: **Non spécifié**

Introduction

Orateur: KÉGL, Balázs (LAL)

ID de Contribution: **1**   Type: **Non spécifié**

RAMPs and Data Challenges

Orateur: KÉGL, Balázs (LAL)

ID de Contribution: **2** Type: **Non spécifié**

The linked data platform

Orateur: Mme RAFES, Karima (LRI)

ID de Contribution: **3**   Type: **Non spécifié**

The open software initiative

Orateur: GRAMFORT, Alexandre (Telecom ParisTech, CNRS)

ID de Contribution: **4**   Type: **Non spécifié**

Ultrawalls and interactive visualization

Orateur: PIETRIGA, Emmanuel (INRIA)

ID de Contribution: **5**   Type: **Non spécifié**

Apprentissage automatique pour une aide au codage PMSI

Orateur: TARIGHT, Namik (APHP)

ID de Contribution: **6**   Type: **Non spécifié**

Managing and analyzing analytical chemistry data sets

Orateur: TFAYLI, Ali (LIPSYS / UPSud)

ID de Contribution: **7**   Type: **Non spécifié**

Secure cytotoxic administration by non-invasive spectroscopy

Orateur: LE, Laetitia (LIPSYS / UPSud)

ID de Contribution: **8**   Type: **Non spécifié**

Efficiently Ranking big biological and biomedical data sets using rank aggregation techniques

Orateur: Dr COHEN-BOULAKIA, Sarah (Laboratoire de Recherche en Informatique)

ID de Contribution: **9** Type: **Non spécifié**

Causal discovery for CDS practitioners

Orateur: Mlle GUYON, Isabelle (UPSud / INRIA / ChaLearn)

ID de Contribution: **10**   Type: **Non spécifié**

Multi-level data fusion

Orateur: Dr BENAMEUR, Kaouthar (ONERA)

ID de Contribution: **11**   Type: **Non spécifié**

Global scientific positioning system for CDS

Orateur: CAILLOU, Philippe (LRI / UPSud)

ID de Contribution: **12**   Type: **Non spécifié**

Heterogeneous social data management and mining

Orateur: BENNACER, Nacéra (LRI)

ID de Contribution: **13**   Type: **Non spécifié**

Reconstructing the past: deep learning for population genetics

Orateur: CHARPIAT, Guillaume (TAO Team/ INRIA)

ID de Contribution: **14** Type: **Non spécifié**

A platform for decoding and recoding turbulence

Orateur: Dr PODVIN, Bérengère (CNRS)

ID de Contribution: **15**   Type: **Non spécifié**

A framework for sensor data management and analysis

Orateur: ZEITOUNI, Karine (DAVID / UVSQ)

ID de Contribution: **16**   Type: **Non spécifié**

Generic stereoscopic tools for planetary topography

Orateur: M. SCHMIDT, Frédéric (GEOPS / UPSud)

ID de Contribution: **17**   Type: **Non spécifié**

Tracking machine learning challenge

Orateur: Dr ROUSSEAU, David (ATLAS)

ID de Contribution: **19**   Type: **Non spécifié**

Innovative meta-model construction techniques for electro-nuclear scenarios simulation

mercredi 8 novembre 2017 10:00 (20 minutes)

Future of nuclear energy is very uncertain and actively discussed. To support these discussions and allows the choices to be informed, we need to simulate electro-nuclear scenarios with high level of precision. I will present the process used to simulate such scenarios and described the two key model needed : irradiation model and fuel creation model. For each of these two models, I will present what physics problem they have to solve and why the standard numerical meta-models are not sufficient and we would need in order to continue the development of our studies.

Summary

Orateur: ERNOULT, Marc (IPNO)

Classification de Session: Session 1

ID de Contribution: **20**   Type: **Non spécifié**

Flexible and Adaptive Data Acquisition

mercredi 8 novembre 2017 10:20 (20 minutes)

We investigate the possibility of flexible and adaptive settings of data acquisition for nuclear physics experiments(gamma-ray spectroscopy). The goal is to adapt rapidly the event selection settings for up to a few hundred detectors, knowing that the configuration changes with the experimental setup and that the behaviour of the electronics changes with the experimental conditions: we need to implement a feedback by learning from experimental data. This feedback should at least update the parameters of the online calculations, and possibly the algorithms.

Summary

Orateur: GENOLINI, Bernard (IPN Orsay)

Classification de Session: Session 1

ID de Contribution: **21**   Type: **Non spécifié**

Tracking Machine Learning Challenge

mercredi 8 novembre 2017 10:40 (20 minutes)

The Large Hadron Collider at CERN, where the Higgs boson has been discovered, is poised for a major upgrade for possible discovery of new particles, super-symmetric particles, dark matter or signs of extra-dimensions of space. The increase in yearly number of proton collisions recorded comes at a cost of a large increase of the recorded event complexity. Preliminary studies show that traditional algorithms suffer from a combinatorial explosion of the CPU time.

To reach out for Computer Science specialists, a Tracking Machine Learning challenge (trackML) is being set up for 2018.

Summary

Orateur: Dr ROUSSEAU, David (ATLAS)

Classification de Session: Session 1

ID de Contribution: **22**   Type: **Non spécifié**

Introduction

mercredi 8 novembre 2017 09:50 (10 minutes)

Orateur: KÉGL, Balázs (LAL)

Classification de Session: Session 1

ID de Contribution: **23**   Type: **Non spécifié**

RAMP on detecting Mars craters in satellite images

mercredi 8 novembre 2017 11:20 (20 minutes)

Impact craters in planetary science are used to date and characterize planetary surfaces and study the geological history of planets. It is therefore an important task which traditionally has been achieved by means of visual inspection of images. This talk will present a currently ongoing RAMP challenge with the goal to predict the location and size of craters on Mars based on a satellite image, and the pipeline that was set up to tackle this problem.

Summary

Orateur: VAN DEN BOSSCHE, Joris (INRIA)

Classification de Session: Session 2

ID de Contribution: **24**   Type: **Non spécifié**

Galaxy morphologies and deblending

mercredi 8 novembre 2017 11:40 (20 minutes)

Current and future generation of large scale astronomical surveys will have to deal with an increasing number of crowded fields due to their sensitivity. In such fields, a high number of objects (mostly galaxies) are "blended" together, which poses a challenge in terms of photometry (measure of individual fluxes) and morphology (shape measurements), both heavily related to the main science goals. This ramp would explore the use of deep learning techniques to tackle the deblending of galaxies (detection + segmentation) and ideally the measurement of their morphologies (regression).

Summary

Orateur: BOUCAUD, Alexandre (LAL)

Classification de Session: Session 2

ID de Contribution: **25**   Type: **Non spécifié**

Fake News Detection

mercredi 8 novembre 2017 12:00 (20 minutes)

Recently, there is a new phenomena of fake news and alternative facts. No one deliberately and consciously desires false information but paradoxically, people pervasively consume fake information. It is essential to restore people's confidence in reliable, fact-checking sources, and reduce media bias whether perceived or real. We believe that exploring artificial intelligence technologies could be leveraged to combat fake news and partly automate fact checking. We dive in American politics and propose a starting point for fake news detection.

Summary

Orateur: BOROS, Emanuela (CEA, LIST, Laboratoire Vision et Ingénierie des Contenus)

Classification de Session: Session 2

ID de Contribution: **26**
Type: **Non spécifié**

Decoding and recoding wall turbulence

mercredi 8 novembre 2017 12:40 (20 minutes)

Study of turbulence using numerical tools such as Direct Numerical Simulations, Large Eddy Simulations, etc.
leads to an analysis of large amount of data to discover the underlying physical mechanisms. Estimation of a
turbulent flow using few sensor measurements is yet another important problem in fluid mechanics, which also
involves big data analysis. In this project, we analyze turbulent channel flow using a sufficiently large numerical
dataset (around 9 TB) close to the wall region, with an aim of improving predictability and also understanding
the underlying physics.

Summary

Orateur: MURALIDHAR, Srikanth (limsi)

Classification de Session: Session 2

ID de Contribution: **27**
Type: **Non spécifié**

Switching from Matlab to Python

mercredi 8 novembre 2017 14:00 (20 minutes)

Atherosclerosis is an inflammatory disease of the arterial wall caused by the formation of an atheroma plaque in the vessels wall. Data analysis of single point spectra and images led to the identification of some spectral changes due to the murine macrophages J774 enrichment with fatty acids. Data processing was held using Matlab environment and we are looking to re-process this dataset in Python environment.

Summary

Orateur: TFAILI, Sana (Lip(Sys)², UFR Pharmacie, Université Paris Sud)

Classification de Session: Session 3

ID de Contribution: **28**   Type: **Non spécifié**

Ranking Big Data Sets using Rank Aggregation Techniques

mercredi 8 novembre 2017 14:20 (20 minutes)

The aim of (scientific) data ranking is to help users choose between alternative pieces of information especially when they are faced with huge amounts of data. However, ranking scientific data is a difficult task: various alternative quality criteria can be defined to order data items, depending on the data origin or even the way data have been obtained. As a consequence, it is very difficult to determine which ranking method (or which ranking criteria) to use. We present here a family of solutions named Rank Aggregation Techniques able to compute a consensus ranking from a set of input rankings. We will present the results obtained on our current applications.
We are open to any new collaboration with domain scientists having the need to make the most of alternative rankings.

Summary

Orateur: Dr COHEN-BOULAKIA, Sarah (Laboratoire de Recherche en Informatique)

Classification de Session: Session 3

ID de Contribution: **29**　　　　　　　　　　　　　　　　　　Type: **Non spécifié**

Calcium rhythms, Algorithms and markers of the culture media impact during fertilization

mercredi 8 novembre 2017 14:40 (20 minutes)

During fertilization in mammals, the egg emits a series of calcium oscillations that are specific to each individual and whose frequency and amplitude are modulated by the culture medium in which it is cultured. We have developed advanced microfluidic techniques to record, stimulate and analyze the calcium response during the first hours of in-vitro fertilization (IVF) and have hundreds of individual data depending on the composition of the culture medium. The construction, as part of a collaborative project with CDS, of a prediction tool based on algorithms and a mathematical formalism of the functioning of the egg would open new perspectives of knowledge for developmental biology.

Summary

Orateur: COTINOT, Corinne (INRA)

Classification de Session: Session 3

ID de Contribution: **30**
Type: **Non spécifié**

Learning and Image Processing

mercredi 8 novembre 2017 15:00 (20 minutes)

INRA/MaIAGE and INRA/BDR join their efforts to create efficient software for the early evaluation of embryo viability from time-lapse observation of embryos. Specifically, we are investigating the early development of bovine embryo which can be observed in 2D+Time light microscopy at different stages of the development. We have a database of hundreds of expert annotated embryos. More than one hundred qualitative and quantitative measures as well as original movies are available. A preliminary exploratory statistical analysis has been conducted and classification regression trees revealed discriminant features with respect to viability. Considering a restricted number of these features, we aim at their automatic evaluation from the movies.

Summary

Orateur: TRUBUIL, Alain (INRA)

Classification de Session: Session 3

ID de Contribution: **31** Type: **Non spécifié**

RAMP on predicting autism from resting-state functional MRI and anatomical MRI

mercredi 8 novembre 2017 15:20 (20 minutes)

This talk will present the ongoing preparation of a RAMP aiming at distinguishing subjects with Autism Spectrum Disorder (ASD) from typical control subjects. This analysis will use the Autism Brain Imaging Data Exchange (ABIDE I & II) database and data from Robert Debre Hospital based on R-fMRI and anatomical MRI. We will particularly focus on presenting the problematic, the typical pipeline answering this problem, and the current status of this RAMP.
This work is in collaboration with the Pasteur Institute (Neuroanatomy group of the Unit of Human Genetics and Cognitive Functions).

Summary

Orateur: LEMAITRE, Guillaume (INRIA)

Classification de Session: Session 3

ID de Contribution: **32**   Type: **Non spécifié**

End-to-end Deep Learning Approach for Demographic History Inference

mercredi 8 novembre 2017 16:00 (20 minutes)

Recent methods for demographic history inference have achieved good results, circumventing the complexity of raw genomic data by summarizing them into handcrafted features called summary statistics. We developed a new approach based on deep learning that takes as input the variant sites found within a sample of individuals from the same population, and infers demographic descriptor values without relying on these predefined summary statistics. By letting our model choose how to handle raw data and learn its own way to embed them, we were able to outperform a method frequently used in population genetics for the inference of three out of seven demographic descriptor values of a scenario with a bottleneck and two expansions. This is still a preliminary work and we are hopeful that future developments would allow us to tackle a broader range of demographic scenarios and outperform previous methods by developing more flexible artificial neural network architectures.

Summary

Orateur: SANCHEZ, Theophile (LRI)

Classification de Session: Session 4

ID de Contribution: **33**   Type: **Non spécifié**

A generic autocompletion of SPARQL in a multi-service context at Paris-Saclay

mercredi 8 novembre 2017 16:20 (20 minutes)

Dans cet article, nous démontrons, au travers d'une expérimentation, une approche permettant de proposer des complétions d'une requête en cours de rédaction en exploitant de nombreux types d'autocomplétion et ce dans un contexte multi-services. Cette expérimentation s'appuie sur un éditeur SPARQL auquel nous avons ajouté des mécanismes d'autocomplétion qui supportent une ontologie en perpétuelle évolution, ici avec la base de connaissances collaborative de Wikidata.

Summary

Orateur: Mme RAFES, Karima (LRI)

Classification de Session: Session 4

ID de Contribution: **34**   Type: **Non spécifié**

Discussion

mercredi 8 novembre 2017 17:00 (30 minutes)

Classification de Session: Session 4

ID de Contribution: **35**   Type: **Non spécifié**

Scientific Knowledge Graph Enrichment and Expansion

mercredi 8 novembre 2017 16:40 (20 minutes)

In the setting of the CDS2 of University Paris Saclay, the Data IT platform is a prominent initiative for moving forward a creation of a linked open data cloud dedicated for data science. In order to take part of this challenging goal, we propose in this project to be involved in the enrichment of the datasets, existing in Data IT platform, and in proposing tools contributing to the improvement of the data and knowledge quality that are already available and those which will be accessible through Data IT platform.

Summary

Orateur: SAIS, Fatiha (LRI)

Classification de Session: Session 4