

Jinpeng Li

email: mr.li.jinpeng@gmail.com

Working experience

Research Engineer in CATI (Contract: INSERM)

LIB, Hôpital Pitié Salpêtrière, Paris, France, 03/2016 - present

- Software Engineering:
 - Main developer for CASA_Cloud : remote access to CASA tools via a graphical X session in a browser. https://github.com/cati-neuroimaging/casa_cloud
 - Implement the data processing webservice CARMIN (Common web API for Remote pipeline INvocation) for CATIDB.
 - Integrate the data processing Morphologist pipeline with CARMIN.
- Infrastructure Manager (chef de mission d'Infrastructure)
 - Infrastructure maintenance of CATI french national project (Neuroimage database based on alzheimer disease. Work on various techniques, such as Jenkins, OpenLDAP, Docker, etc.).
 - Infrastructure maintenance of Imagen European project (Neuro-genetic database).
 - Infrastructure maintenance of c-Veda india project (Neuroimage database based on externalizing disorder and addictions).

Research Engineer in the project France Life Imaging (Contract: Neurospin/CEA Saclay)

Saclay, Île-de-France, France — 09/2014-03/2016

- Develop OpenLDAP-based SSO (Single Sign On) system for all the webservices, sftp, and website.
- We have changed our web service technology from Cubicweb to Pyramid.
- Synchronize all the data roughly 13 Terabytes (cati_shared) between Neurospin and ICM every night.

Research Engineer in the project France Life Imaging (Contract: INSERM)

IM2A, Hôpital Pitié Salpêtrière, Paris, France, 03/2014 - 09/2014

- Develop web services allowing remote access to databases and processing tools hosted by the CATI platform.
- Connect these services to the IAM node infrastructure to allow interactions with other platforms.
- Create a demonstrator application combining several processing tools and databases.
- All our database is based on Cubicweb.

Machine Learning Research Engineer in project Brainomics in (Contract: Neurospin/CEA Saclay)

Saclay, Île-de-France, France — 03/2013-03/2014

Machine Learning and Distributed Computing Software Development on HPC (Torque/PBS and TGCC Curie Supercomputer). We compute the correlation between voxels of MRI images and DNA sequences in the project Brainomics. The system consumes 200 Nvidia GPUs (12.000 hGPU roughly) and we found two interesting snps in DNA sequence [2].

pylearn-epac: <https://github.com/neurospin/pylearn-epac>

Main developer 2013: pylearn-epac is a python-based and MapReduce-based machine learning workflow builder on local machine and on DRMS (Distributed Resource Management System) using soma-workflow. It uses cross-validation and permutation for model selection [3].

soma-workflow: <https://pypi.python.org/pypi/soma-workflow>

Contributor and maintainer 2013: A unified and simple interface to parallel computing resources based on DRMAA (Distributed Resource Management Application API). This package is implemented by python.

PhD, IRCCyN (CNRS UMR 6597)

Nantes, France — 10/2009-10/2012

Symbol and Spatial Relation Knowledge Extraction Applied to On-Line Handwritten Scripts.

Supervisors : C. Viard-Gaudin (christian.viard-gaudin@univ-nantes.fr) and H. Mouchère

(harold.mouchere@univ-nantes.fr)

Accomplishments

- Handwritten Symbol Shape Recognition
- Spatial Relation Analytics
- Stroke-order and stroke-direction invariant recognition
- Knowledge Extraction using the Minimum Description Length principle
- A MPI-based distributed and unsupervised system for handwritten symbol recognition (knowledge extraction)
- Recognition results are shown in <http://lijinpeng.org/PhDGallery>

Trainee, IRCCyN (CNRS UMR 6597)

Nantes, France — 02/2009-07/2009

I was working on the writer Identification using the grapheme distribution (tf-idf, term frequency-inverse document frequency).

Software Engineer, Guangzhou ABE Science & Technology Company (<http://www.iabe.cn/>)

Guangzhou, China — 03/2007-03/2008

Implementing a face detection and human face identification system based on ICA (Independent Component Analysis). This work was a small part in a remote video education system. The education system requires a specified people to finish lessons. (One or two days every week).

Other for fun

Chinese character recognition OCR based on Caffe

<https://github.com/JinpengLI> (stars 370, fork 156)

kdd2017 travel time competition rank 28/3574

<https://github.com/JinpengLI/kdd2017>

Education

PhD: see Working experience

Ecole de Polytechnique de l'Université de Nantes

09/2008-10/2009

Master 2 (Electronic and Communication System)

Guangdong University of Technology (GDUT)

09/2003-07/2008

Bachelor: Software Engineering

Master 1: Image Processing

Publications

Journal

[1] Jinpeng Li, Harold Mouchère and Christian Viard-Gaudin. An annotation assistance system using an unsupervised codebook composed of handwritten graphical multi-stroke symbols, Pattern Recognition

Conference

- [2] H el ene Lajous, Benoit Da Mota, Jinpeng Li, Edouard Duchesnay, Herv e Lema tre, Vincent Ducrot, Vincent Frouin, Voxelwise genome-wide association study on gray matter volume reveals a SLC39A8 variant associated with subcortical structures, SFGBM 2015
- [3] Jinpeng Li, Edouard Duchesnay, Mathieu Dubois, Laure Hugo, Benoit Da Mota, Embarrassingly Parallel Array Computing (EPAC), Workshop on Machine Learning Open Source Software, NIPS 2013, Nevada, USA
- [4] Jinpeng Li, Harold Mouch ere and Christian Viard-Gaudin. A Multi-Stroke Dynamic Time Warping Distance Based on A* Optimization, ICDAR 2013, Washington, DC
- [5] Jinpeng Li, Harold Mouch ere and Christian Viard-Gaudin. Reducing Annotation Workload Using a Codebook Mapping and its Evaluation in On-Line Handwriting, ICFHR2012, Italy
- [4] Jinpeng Li, Harold Mouch ere and Christian Viard-Gaudin. Quantifying spatial relations to discover handwritten graphical symbols, DRR, 22-26 January 2012, San Francisco, CA USA
- [5] Jinpeng Li, Harold Mouch ere and Christian Viard-Gaudin. Unsupervised Handwritten Graphical Symbol Learning Using Minimum Description Length Principle on Relational Graph, KDIR 2011, Paris, France
- [6] Jinpeng Li, Harold Mouch ere and Christian Viard-Gaudin. Symbol Knowledge Extraction from a Simple Graphical Language, ICDAR2011, Beijing, China

French Conference

- [7] Jinpeng Li, Harold Mouch ere and Christian Viard-Gaudin. Une distance entre deux ensembles de s equences avec la contrainte de continuit e (A distance between two sequence sets with the continuity constraint), CIFED 2012, Bordeaux, France
- [8] Jinpeng Li, Identification du scripteur utilisant les fr equences de graph emes-Application des donn ees en ligne (Writer identification using the frequencies of graphemes-Application in online data). CIFED-CORIA2010, Atelier Jeunes chercheurs (communication for young researchers), Sousse, Tunisie, pp.498-494, 2010