Karim Sayadi

PhD

25 rue de la fontaine au roi.
75011, Paris.
☎ *0761440783*
✉ *karim.sayadi@ephe.sorbonne.fr*
CHArt Laboratory EA 4004 EPHE

Current Position

2016–2017 **Temporary Lecturer and Research Assistant**, *EPHE*, Ecole Pratique des Hautes Etudes, *Paris, France.*

Education

2013–2016 **PhD in computer science**, *UPMC*, University Pierre and Marie Curie, *Paris, France.*
Grants from the French Institue of Research and Development. International Doctoral Program.

2011–2013 **Master degree in Natural and Artificial Cognition (Computer Science)**, *Ecole Pratique des Hautes Ètudes*, *Paris, France..*
With highest honour.

2008–2011 **Bachelor Degree in Computer Science**, *Institute of Computer Science and Applied Mathematics*, *Monastir, Tunisie.*
With honors.

Computer Skills

Programming languages	Python, Java, R, html, C, MySQL, SQL Server
Tools	Matlab, Anylogic, Scilab, Scikit learn, Hadoop, Spark, Lucene, Tensorflow, Caffe, Keras, Weka
Operating systems	UNIX, Linux, Microsoft Windows
Software production	Pretopological Semantic Analyzer (PSA for health). Patented it under the french national agency of software 18/10/2012.

Languages

Fluent	Basic	Bilingual
English	**German**	**Arabic and French**

Community Life & Hobbies

2012–2014 Events and logistic manager for cultural activities in the International University Campus of Paris.

2010 Founder and honorary chairman of the free software club (ISIMUX)Institue of Computer Science and Applied Mathematics. Monastir, Tunisia.

Hobbies

Water Polo, Cycling, Cooking, Philosophy

PhD Thesis (defended on 28/03/2017)

Titre *Text-Based and Image-Based Classification : a Machine Learning Approach*

Description Different disciplines in the humanities, such as philology or palaeography, face complex and time-consuming tasks whenever it comes to examining the data sources. The introduction of computational approaches in humanities makes it possible to address issues such as semantic analysis and systematic archiving. The conceptual models developed are based on algorithms that are later hard coded in order to automate these tedious tasks. In the first part of the thesis we propose a novel method to build a semantic space based on topics modeling. In the second part and in order to classify historical documents according to their script. We propose a novel representation learning method based on stacking convolutional auto-encoder. The goal is to automatically learn plot representations of the script or the written language.

Keywords *Topic Modeling*, *Probabilistic Graphical Model*, *Machine Learning*, *Deep Learning*, *Digital Humanities*, *Natural Langage Processing*

Jury

Examiners :	DR. CNRS. Thierry Poibeau	-CNRS-ENS, Paris, France.
	DR. IE ILSP. Vassilis Katsouros	-Institute ILSP, Athena, Greece.
PhD Advisor :	Pr. Marc Bui	-University Paris 8-EPHE, France.
Experts :	Pr. Marcus Liwicki	-University of Fribourg, Switzerland
	Pr. Jean Daniel Zucker	-University of Paris 6, France
	Pr. Vu Duong	-Institute JVN, Ho Chi Minh city, Vietnam
	Associate Pr. Sofian Ben Amor	-University of Versailles, France

Research Activities

The research work that I conducted in my thesis was essentially about text document classification. The text was considered in its ASCII form factor or in its pixels form factor. For the former, my main approach was based on topic modeling and probabilistic graphical models (PGM). I implemented the Gibbs sampling algorithm to perform an approximate inference on the latent variables of the Latent Dirichlet Allocation model and therefore extracted the topics. I have also distributed the Gibbs sampling using Spark/Hadoop. When text came in the form factor of pixels embedded in an image, my main approach was based on features learning using Stacked Convolutional Auto-Encoders and classical method from Deep Learning like the Convolutional Neural Network. My ongoing research interest are the Variational Auto-Encoders(AE) where the AE learn a probability distribution from the input instead of a simple representation. My aim is to couple the Variational AE with tools from PGM and therefore learn the latent variables instead of using approximate inference.

Feb-June 2016 Invited researcher in the Document Image Voice Analysis (DIVA) Group. University of Fribourg, Switzerland. (Pr. Rolf Ingold, Pr. Marcus Liwicki) (5 months)

Nov-Dec 2015 Invited researcher in the Big Data and High Performance Computation Group (HPC JVN). Institute John Von Neumann. National University of Ho Chi Minh City, Vietnam. (Pr. Vu Duong) (1 month)

Sept 2015 Workshop on Historical Document. German Research Centre for Artificial Intelligence (DFKI). Kaiserslautern, Germany. (Pr. Marcus Liwicki) (1 week)

Dec 2014 Workshop on social ecosystem modeling for the Moorea Island. University of Berkeley, California, United States of America. (Dr. Joachim Claudet, Dr. Neil Davies) (1 week)

Nov-Dec 2014 Invited research in the Information Technology and Communication Group. Polytechnic Institute of Hanoi, Vietnam. (Associate Pr. Ha Quoc Trung) (1 month)

Publications

Apr 2017 Karim Sayadi, Mansour Hamidi, Marc Bui, Marcus Liwicki and Andreas Fischer. *Characer-Level Dialect Identification in Arabic Using Long Short-Term Memory.*, To appear in CICLing proceedings Avril 2017 intl conference. Budapest, Hungary, 17-23 April 2017.

Apr 2017 Quang Vu Bui, Karim Sayadi and Marc Bui,. *Combining Latent Dirichlet Alllocation and K-means for Documents Clustering : Effect of Probabilistic Based Distance Measures*, Intelligent Information and Database Systems, Springer. 9th Asian Conference, ACIIDS 2017, Kanazawa, Japan, April 3-5, 2017.

Dec 2016 Karim Sayadi, Quang Vu Bui and Marc Bui. *Distributed Implementation of the Latent Dirichlet Allocation on Spark*, Proceedings of the Sixth International Symposium on Information and Communication Technology, Ho Chi Minh City, Vietnam, 08-09 December 2016.

Apr 2016 Karim Sayadi, Marcus Liwicki, Rolf Ingold, Marc Bui,. *Tunisian Dialect and Modern Standard Arabic Dataset for Sentiment Analysis : Tunisian Election Context*, IEEE-CICLing (Computational Linguistics and Intelligent Text Processing) Intl. conference, Konya, Turkey, 7-8 April 2016.

Dec 2015 Quang Vu Bui, Karim Sayadi, Marc Bui. *A multi-criteria document clustering method based on topic modeling and pseudoclosure function*, ACM-SOICT (Symposium on Information and Communication Technology) Intl. conference, Hue, Vietnam, 3-4 December 2015. (Extended version accepted in Informatica journal revised on June 2016)

July 2015 Karim Sayadi, Quang Vu Bui, Marc Bui. *Multilayer classification of web pages using Random Forest and semi-supervised version of the Latent Dirichlet Allocation*, IEEE-I4CS (International Conference on Innovations for Community Services) Intl. conference, Nuremberg, Germany. 8-10 July 2015.

July 2013 Karim Sayadi, Marc Bui, Michel Lamure. *Predictive topic modeling : Complex Networks approach using dynamics of author's communities*, EURO INFORMS (Operational Research), Rome, Italy, 1-4 July 2013.

May 2012 Karim Sayadi, Marc Bui, Vigile Hoareau, Sofian Ben Amor, *Une approche prétopologique pour la catégorisation des données de microblogging*, Conférence nationale VSST'12 (Veille Scientifique et Technologique), Ajjaccio, Corse, France, 24-25 Mai 2012.

Research Report

Oct 2013 Marc Bui, Karim Sayadi. *Modèle de recommandation de contacts basé sur l'analyse thématique des échanges*. Rapport du projet n 45 Nexboo. Convention Techno Pole de la Réunion. 20 Octobre 2013

Research Projects

Feb 2016 HisDoc DIVA Group http ://bit.ly/hisdoc2. In this project I worked with Mathias Seuret under the supervision of Pr. Marcus Liwicki to propose a solution to have an unsupervised workflow for script identification on historical documents.

Dec 2014 Moorea IDEA. http ://mooreaidea.org. This project intend to propose a simulation based on the big data collected around the Moorea island ecosystem.

Teaching Activities

Temporary Lecturer (2016-2017) at École Pratique des Hautes Études

The courses presented below were carried out within the training program provided by the École Pratique des Hautes Études (EPHE). The courses and tutorials that I assured were aimed at intorducing an algorithmic approach to students from the humanities field.

Materials	Hours
Machine Learning with Python	30h
Introduction to programming with Python	35h
Data Analysis with Python	35h
Database Management with MySQL	35h
Document Layout with LaTeX	30h
Image Processing with OpenCV	27h
Total	192h

Other Interventions

2015-2016 Tutorial on LaTeX. École Pratique des Hautes Études (8h).

2014-2015 Tutorial on the use of Processing (software for learning how to code with sketchbooks). École Pratique des Hautes Études (12h).

2014-2015 Tutorial on the use of Unified Modeling Language. École Pratique des Hautes Études (8h). Master 1 CNA-PC.

2014-2015 Tutorial on spatial information processing. University of Paris 8 (16h). Master 1 geomatic.

2013-2014 Course on machine learning applied on text data. École Pratique des Hautes Études (8h). Master 2 CNA-PC.