

Ansaf Salleb-Aouissi, PhD

Associate Research Scientist

Center for Computational Learning Systems (CCLS)
Columbia University, New York, NY 10115
Phone: 212-870-1281 - Fax: 212-870-1285
Email: ansaf@ccls.columbia.edu
<http://www1.ccls.columbia.edu/~ansaf/>

RESEARCH INTERESTS

Machine Learning and Data Mining: Rule induction, Association Rules, Frequent Patterns, Ranking, Recommender Systems, Unsupervised Learning, Classification, Data Mining Applications, Multi-Relational Mining.

EMPLOYMENT

- | | |
|---------------------|--|
| 2006-Present | CENTER FOR COMPUTATIONAL LEARNING SYSTEMS(CCLS) COLUMBIA UNIVERSITY, NEW YORK.
ASSOCIATE RESEARCH SCIENTIST |
| 2004-2005 | FRENCH NATIONAL INSTITUTE OF COMPUTER SCIENCE AND CONTROL (INRIA) RENNES, FRANCE.
INRIA POSTDOCTORAL RESEARCH FELLOW |
| 2002-2004 | UNIVERSITY OF ORLÉANS, SCIENCES DEPARTMENT AND LIFO.
TEMPORARY ASSISTANT PROFESSOR AND RESEARCHER |
| 1998-2002 | UNIVERSITY OF ORLÉANS, SCIENCES DEPARTMENT AND LIFO.
TEACHING ASSISTANT |
| 1996-1998 | MICRORAMA CORPORATION, ALGERIA.
ENGINEER IN COMPUTER SCIENCE |

PROFESSIONAL ACTIVITIES

PC Membership:

- Knowledge Extraction and Management (Extraction et Gestion des Connaissances), , **EGC 2010**.
- “Advances in Knowledge Discovery and Management”, a refereed book of chapters, published by Springer, **AKDM 2009**.
- The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining **KDD 2009**,
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases **ECML/PKDD 2008**,
- Quality Measures for Data Mining Book, Springer-Verlag **QMDM2006**,
- Data and Knowledge Quality Workshops **DKQ2005, 2006**,
- Qualité des Données et des Connaissances workshops co-located with EGC **QDC2007-2009**,

Miscellaneous:

- Member of the research group GAfoDonnees at CNRS, France (2001-2005).
- Member of the working group on “Data Mining and Quality Measures: GaFoQualité”, sponsored by STIC-CNRS, Nantes, France. 2002.
- Member of the organization team for the French speaking conference on machine learning CAP 2002, Orléans France.
- Initiation of children to counting in the binary system at “Initiation to science”, University of Orleans, France (2003).
- Member of the working group on the "PKDD Discovery Challenge 2004" sponsored by STIC-CNRS. Lyon and Paris, France.
- Member of the Columbia University Research Officers committee 2006-2007.
- Organizer of the CCLS journal Club 2008.
- Organizer of the CCLS Technical Reports series 2007-present.
- Co-organizer of the CCLS retreat 2008. October 22, 2008 at the Wave Hill Center, New York.

EDUCATION

USTHB Algiers, Algeria	Computer Science	Engineer 1996
University of Orléans, France	Computer Science	Masters 1999
University of Orléans, France	Computer Science	Ph.D. 2003

REFEREED PUBLICATIONS

- [2009] B. Passoneau, **A. Salleb-Aouissi**, and N. Ide
Making Sense of Word Sense Variation
To appear in Semantic Evaluations: Recent Achievements and Future Directions **SEW-2009**
- [2009] B. Huang and **A. Salleb-Aouissi**
Maximum Entropy Density Estimation with Incomplete Presence-Only Data
In the 12th International Conference on AI and Statistics **AISTATS 2009**. JMLR Proceedings
Volume 5, pages 240-247.
- [2008] **A. Salleb-Aouissi**, Bert C. Huang and David L. Waltz
Vers des Machines a Vecteurs de Support "Actionables": Une Approche Fondée sur le Classement
Toward Actionable Support Vector Machines: A Ranking-based Approach
Best Paper Award in *Knowledge Extraction and Management (Extraction et Gestion des Connaissances)* **EGC 2008**, pp. 285-295 Sophia Antipolis, France.
- [2008] P. Gross, **A. Salleb-Aouissi**, H. Dutta and Albert Boulanger
Ranking Electrical Feeders of the New York Power Grid
3rd Annual Machine Learning Symposium at the New York Academy of Sciences **NYAS 2008**,
New York, October, 2008.
- [2008] H. Feng, H. Dutta and **A. Salleb-Aouissi**
On Improving Probability Estimate Trees
3rd Annual Workshop for Women in Machine Learning **WiML 2008** held in conjunction with
Neural Information Processing Systems (NIPS), Vancouver, B.C., 2008.

- [2007] B. Huang, **A. Salleb-Aouissi**
 Maximum Entropy Density Estimation with Incomplete Data
Poster and Abstract at the 2nd Annual Machine Learning Symposium NYAS, New York Academy of science, Oct 2007.
- [2007] **A. Salleb-Aouissi**, C. Vrain and C. Nortet
 QuantMiner: A Genetic Algorithm for Mining Quantitative Association Rules
 In the Proceedings of the *20th International Conference on Artificial Intelligence IJCAI 2007*, pp. 1035-1040. Hyberadad, India.
- [2007] **A. Salleb-Aouissi**, C. Vrain
 A Contribution to the Use of Decision Diagrams for Loading and Mining Transaction Databases
Fundamenta Informaticae Journal 2007. Volume 78, Number 4, pp. 575-594. IOS Press.
- [2007] B. Duval, **A. Salleb**, C. Vrain
 On the Discovery of Exception Rules: A Survey
 Reviewed **Book chapter** appeared in “Quality Measures in Data Mining” book. In F. Guillet and Howard J. Hamilton editors Springer’s Lecture Notes in Artificial Intelligence, Volume 43/2007 pp. 77-98. Springer January 2007.
- [2006] J. Fortuna, M. Arias and **A. Salleb-Aouissi**
 Datamining towards Improving Performance in Predicting Electricity Distribution Feeder Failures
Best Student Poster Award at the **First annual Computer Science research** - Columbia University.
- [2006] M.-O. Cordier, V. Masson, **A. Salleb**, R. Trepos, C. Gascuel, F. Garcia
 SACADEAU project : recommending actions from simulation results
 Poster **BESAI 2006** - 5th Workshop on Binding Environmental Sciences and Artificial Intelligence, a workshop held in conjunction with the 17th European Conference on Artificial Intelligence (ECAI 2006) August 28th-September 1st, 2006. Riva del Garda, Italy
- [2006] R. Trepos, **A. Salleb**, M.-O. Cordier, V. Masson, C. Gascuel
 Une approche fondée sur une distance pour la recommandation d’actions.
 In *15e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (15th French Speaking Conference on Pattern Recognition and Artificial Intelligence) RFIA 2006* Tours, France.
- [2005] **A. Salleb** and C.Vrain
 Estimation of the Density of Datasets with Decision Diagrams
In Proceedings of 15th International Symposium on Methodologies for Intelligent Systems ISMIS 2005. In M.-S. Hacid, N. V. Murray, Z. W. Ras, editors, Springer’s Lecture Notes in Artificial Intelligence LNAI 3488, pp. 688-697, 2005.
- [2005] R. Trepos, **A. Salleb**, M.-O. Cordier, V. Masson, C. Gascuel
 A Distance-Based Approach for Action Recommendation.
 In Proceedings of the *9th European Conference on Machine Learning ECML 2005 Porto portugal* Springer’s Lecture Notes in Artificial Intelligence LNAI 3720, pp. 425-436, 2005.
- [2005] C. Nortet, **A. Salleb**, T. Turmeaux and C. Vrain
 Extraction de Règles d’Association Quantitatives Application à des Données Médicales.
 Proceedings of the *5th French speaking conference on Knowledge Discovery and Knowledge Management EGC 2005*. Revue des Nouvelles Technologies de l’Information RNTI, Ed. Cépaduès, Volume II, Pages 495-506. January 2005, Paris France.
- [2005] C. Nortet, **A. Salleb**, T. Turmeaux and C. Vrain
 Le rôle de l’utilisateur dans un processus d’extraction de règles d’association.
 Proceedings of the *French speaking workshop on Data and Knowledge Quality DKQ 2005 co-located with EGC 2005*. pages 35-42. January 2005, Paris France.

- [2004] **A. Salleb**, T. Turmeaux, C. Vrain and C. Nortet
Mining Quantitative Association Rules in a Atherosclerosis Dataset.
PKDD Discovery Challenge 2004 (co-located with the *6th European Conference on Principles and Practice of Knowledge Discovery in Database*) Pisa, Italy, pp. 98-103, Sept. 2004.
- [2004] Z. Maazouzi **A. Salleb** and C. Vrain
BoolLoader : un chargeur efficace dédié aux bases denses.
In the *4th French speaking conference on Knowledge Discovery and Knowledge Management EGC 2004*. Revue des Nouvelles Technologies de l'Information RNTI, pp. 71-81, Ed. Cépaduès, January 2004, Clermont-Ferrand, France.
- [2003] T. Turmeaux, **A. Salleb**, C. Vrain and D. Cassard
Learning Characteristic Rules Relying on Quantified Paths
In the *7th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD 2003*, Lavrač et al. (Eds.) LNAI 2838, pp. 471-482, 2003. Springer-Verlag
- [2003] **A. Salleb**
Recherche de Motifs Fréquents pour l'Extraction de Règles d'Association et de Caractérisation. (Mining Frequent Patterns for the Discovery of Association and Characterization Rules).
Ph.D. thesis, *Laboratoire d'Informatique Fondamentale d'Orléans LIFO, University of Orléans. France. Advisor: Professor Christel Vrain.* December 2003
- [2003] T. Turmeaux, D. Cassard, **A. Salleb** and C. Vrain
Apprentissage de Règles Caractéristiques
In the *3rd French speaking conference on Knowledge Discovery and Knowledge Management EGC 2003* Hermès science publications, RSTI - RIA revue, volume 17, pages 437-448, January 2003, Lyon, France.
- [2002] **A. Salleb**, Z. Maazouzi and C. Vrain.
Mining Maximal Frequent Itemsets by a Boolean Approach.
In the *15th European Conference on Artificial Intelligence, ECAI 2002*, F. van Harmelen (Ed.) IOS Press Amsterdam, 2002, pages 385-389, Lyon, France, July 2002.
- [2002] **A. Salleb**, Z. Maazouzi
Approche Booléenne pour l'Extraction des Itemsets Fréquents Maximaux
In the *French speaking conference on machine learning CAP 2002*, PUG Press, pages 111-122, June 2002, Orléans, France.
- [2002] A. Lips, D. Cassard, V. Bouchot, M. Billa, **A. Salleb**, M. Gonzalez, L. Bougrain, C. Vrain, B. Tourlière, G. Stein and F. Alexandre.
Quantitative Assessments of a Continent-Scale Metallogenic GIS by Data-Driven and Knowledge-Driven Approaches to Construct Decision-Aid Documents.
Extended abstract **GIS in Geology 2002**, Moscou, November, 2002.
- [2002] **A. Salleb**, Z. Maazouzi and C. Vrain
Search Space Splitting for Mining Maximal Frequent Itemsets
In the International Conference on Information and Knowledge Engineering **IKE'02**, Las Vegas, USA, CSREA Press ed., June, 2002, pages 482-488
- [2001] D. Cassard, M. Billa, V. Bouchot, **A. Salleb**, G. Stein and B. Tourlière
Predictive Mapping with GIS Andes Datasets.
Abstract and poster session in the *European Union of Geosciences, Symposium FMF5 EUG XI*, Cambridge publications ISSN 1362-0886, Strasbourg France, April 2001.
- [2000] **A. Salleb** and C. Vrain
An Application of Association Rule Discovery to Geographic Information Systems
In the *4th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD 2000*, D.A. Zighed et al. (Eds.) LNAI 1910, pp. 613-618. Springer-Verlag ed.

- [1999] **A. Salleb**
Knowledge Discovery in Geographic Databases
Masters thesis (DEA), *Laboratoire d'Informatique Fondamentale d'Orléans LIFO, University of Orléans* September 1999

OTHER PUBLICATIONS

- [2009] **A. Salleb-Aouissi**, C. Vrain and T. Turmeaux
Characterization in Relational Databases
Under preparation.
- [2009] R. Trepos, **A. Salleb-Aouissi**, M-O. Cordier, V. Masson and c. Gascuel
Building actions from classification rules
Under submission to Knowledge and Information Systems Journal.
- [2008] P. Gross, **A. Salleb-Aouissi**, H. Dutta, A. Boulanger
Susceptibility Ranking of Electrical Feeders: A case Study
Technical report number CCLS-08-04
- [2008] **A. Salleb-Aouissi**, R. Trepos, M-O. Cordier, V. Masson
From Classification rules to action recommendations
Technical report number CCLS-08-01
- [2008] **A. Salleb-Aouissi**, B. Huang, Dave Waltz,
Toward Actionable Support Vector Machines: A Ranking-based Approach
Technical report number CCLS-08-02
- [2004] Z. Maazouzi, **A. Salleb** and C. Vrain
Can We Load Efficiently Dense datasets?
Technical Report *RR-2004-02, Laboratoire d'Informatique Fondamentale d'Orléans* 2004
- [2000] V. Bouchot, B. Tourlière, Y. Itard, **A. Salleb**, D. Cassard, M. Billa, Y. Deschamps, N. Nicol and J.P. Milesi
Méthodologie pour la valorisation des SIGs par la production de cartes de prédictivité
Technical Report, *BRGM/RP-50629-FR, 2000*

PROJECTS

- 2009- 2010** Collaboration between Ansaf Salieb-Aouissi (Columbia University) and Frank Nielsen (École Polytechnique, Paris), coordinators of the project “Supervised Ranking: A Distance-based Approach”. This collaboration is partially funded by a “Columbia-Paris Alliance Joint Projects Grant”.
- 2008-2010** Member of the *Epilepsy project* involving CCLS and the Medical school (Columbia University). This collaboration aims at developing an “Early Warning” Device to Allow Epilepsy Patients to Live a More Normal Life. This project is funded by Research Initiatives in Science and Engineering (RISE) at Columbia University.
- 2006-2010** Member of the *CALM project* involving CONSOLIDATED EDISON OF NEW YORK and CCLS (Columbia University). This collaboration aims at developing and applying advanced Artificial Intelligence and Machine Learning methods toward a modern grid technology to improve public safety and system reliability.
- 2004-2005** Member of the *SACADEAU project* involving IRISA-INRIA and INRA (French Institute for Agronomy Research) in France. This project aims at building a Knowledge Acquisition System for Decision-Aid to Improve Stream-water Quality.
- 2003-2004** Member of the project “French-Russian Research Hub for the Search and Discovery of Super-Large Metallic Deposits”. It is a collaboration between the French Geological Survey BRGM and the Russian Academy of Sciences. This collaboration was supported by a NATO Science Program.
- 2000-2004** During my Masters and my Ph.D. thesis I was a member of a collaboration between the University of Orleans in France and the *French Geological Survey (BRGM) Service of Mineral Resources*.

TALKS & LECTURES

- 2008** **NEW YORK.** Two Lectures at the Columbia University Advanced Machine Learning Class **COMS 6772**. “*Mining Frequent Patterns and Association Rules*”.
- 2008** **NICE, FRANCE.** Talk at the 8th French speaking conference on Knowledge Discovery and Knowledge Management **EGC 2008**. “*Vers des Machines a Vecteurs de Support "Actionables": Une Approche Fondée sur le Classement*”.
- 2006-2007** **NEW YORK.** Lectures at ConEdison, New York “*Ranking of Electrical Components according to their susceptibility to failure*” during meetings with collaborators.
- 2005** **NEW YORK.** Talk at Columbia University “*Towards the discovery of interesting and actionable rules in real-life databases*”
- 2005** **SARATOGA SPRINGS NEW YORK.** Talk at ISMIS'05 Symposium on Methodologies of Intelligent Systems. “*Estimation of the Density of Databases with Decision Diagrams*”.
- 2005** **PARIS-FRANCE.** Talk at DKQ 2005 workshop. “*Le rôle de l'utilisateur dans un processus d'extraction de règles d'association*”.
- 2005** **PARIS-FRANCE.** Talk at EGC 2005 conference. “*Extraction de Règles d'Association Quantitatives Application à des Données Médicales*”.
- 2004** **RENNES-FRANCE.** Talk at INRA (French Institute for Agronomy Research). “*Classification and Clustering: application au projet Sacadeau*”.
- 2004** **LYON-FRANCE.** Talk at GDRI³ (a working group on Information, Interaction and Intelligence supported by the French Ministry of Research and the CNRS, Centre National de la Recherche Scientifique)
- 2004** **PISA-ITALY.** Talk at the PKDD Discovery Challenge 2004 (co-located with the 6th European Conference on Principles and Practice of Knowledge Discovery in Database, september 2004). “*Mining Quantitative Association Rules in a Atherosclerosis Dataset*”.
- 2004** **RENNES-FRANCE.** Seminar at INRIA/IRISA France. “*Recherche de motifs fréquents applications aux données géographiques*”.
- 2004** **VERSAILLES-FRANCE.** Seminar at PRiSM, Université de Versailles Saint-Quentin en Yvelines France. “*Recherche de motifs fréquents applications aux données géographiques*”.
- 2004** **ORLÉANS-FRANCE.** Talk at the Data Mining Workshop Journée Fouille de Données LIFO-MAPMO(Mathematics and Applications, Mathematical Physics of Orléans). “*Extraction de Règles dans les Bases de Données Géographiques : Applications et Perspectives*”.
- 2004** **CLERMONT-FERRAND-FRANCE.** Talk at the 4th French speaking conference on Knowledge Discovery and Knowledge Management EGC 2004. “*BoolLoader : un chargeur efficace dédié aux bases denses*”.

- 2003** **ORLÉANS-FRANCE.** Popularization of sciences at "Fete de la science", CNRS.
- 2003** **ORLÉANS-FRANCE.** Popularization of Science at "Sciences en Sologne" 2003. *"Extraction de Connaissances dans les Systèmes d'Information Géographiques"*.
- 2002** **LYON-FRANCE.** Talk at the 15th European Conference on Artificial Intelligence, ECAI 2002. *"Mining Maximal Frequent Itemsets by a Boolean Approach"*.
- 2002** **ORLÉANS-FRANCE.** Talk at the French speaking conference on machine learning, CAP 2002 *"Approche booléenne pour l'extraction des itemsets fréquents maximaux"*.
- 2000** **ORLÉANS-FRANCE.** Popularization of Science at "Sciences en Sologne" 2000 and Seminar at LIFO, University of Orléans. *"Extraction des connaissances dans les bases de données"*.
- 2000** **LYON-FRANCE.** Talk and poster presentation at the 4th European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD 2000 *"An Application of Association Rule Discovery to Geographic Information Systems"*.

TEACHING AND STUDENT SUPERVISION

Teaching:

- Columbia University: Advanced Machine Learning COMS 6772 (4h - Fall 2008),
- University of Orléans (France):
 - Data Mining I and II (Masters, 67h - Falls 2000, 2001, 2002, 2003),
 - Data Mining II (Masters, 24h - Springs 2003, 2004),
 - Operating systems (Graduate, 57h - Spring 2003),
 - Combinatorial mathematics (Undergraduate, 36h - Spring 2004),
 - Databases (Undergraduate, 90h - Springs 1999, 2000),
 - Algorithmic and C (Masters, 56h - Fall 2003),
 - Algorithmic and Pascal (Undergraduate, 272h - Falls 2000, 2001, 2002, 2003),
 - Visual Basic (Undergraduate, 66h - Springs 1999, 2000).

Students supervision:

- Summer 2009: Karthik Marudhachalam Ramasamy (M.S.), Xiangrong Kong (M.S.) Columbia University,
- Spring 2009: Tristan Naumann (GRA), Amar Patel (Undergraduate), Karthik Marudhachalam Ramasamy (M.S.), Xiangrong Kong (M.S.) Columbia University,
- Fall 2008: Tristan Naumann (GRA), Ruiyang Wu (GRA), Haoyun Feng (GRA), Karthik Marudhachalam Ramasamy (M.S.), Columbia University,
- Spring 2008: Eugene Kogan (Undergraduate), Columbia University.
- Fall 2007: Christos Angelopoulos, M.S., Columbia University,

- Summer 2007: Charles Naut, Summer internship Brooklyn Technical High school, ConEdison and Columbia University. (co-supervision with Albert Boulanger),
- Fall 2006: Joey Fortuna, M.S., co-supervision with Marta Arias at Columbia University,
- 2004: David Cossard, Engineer student, BRGM Orléans, France,
- 2003: Cyril Nortet, M.S. co-supervision with Daniel Cassard at BRGM and University of Orléans and Bagus Emir Ikhwanto M.S. University of Orléans, France,
- 2001: Andrea Legtchenko, M.S. University of Orléans, France.

Graduate advisees:

1. Bert Huang, Ph.D. candidate, Columbia University, Fall 2006-present.
2. Ronan Trépos, co-supervision between 2004-2005 with Prof. M.-O. Cordier Univ. of Rennes I, France. Defended in 2008.

Last update May 15, 2009.