ALRA: Active Learning in Real-World Applications

http://www.nomao.com/labs/alra

Chairs

Laurent Candillier (Nomao)
Max Chevalier (IRIT)
Vincent Lemaire (Orange Labs)

Aims of the workshop

This workshop aims to offer a meeting opportunity for academics and industry-related researchers, belonging to the various communities of Computational Intelligence, Machine Learning, Experimental Design and Data Mining to discuss new areas of active learning, and to bridge the gap between data acquisition or experimentation and model building. How **active** sampling, **incremental** learning and data acquisition, can contribute towards the design and modeling of highly intelligent machine learning systems?

Machine learning indicates methods and algorithms which allow a model to learn a behavior thanks to examples. Active learning gathers methods which select examples used to build a training dataset for the predictive model. All the strategies aim to use a set of examples as small as possible and to select the most informative examples.

When designing active learning algorithms for real-world data, some specific issues are raised. The main ones are scalability and practicability. Methods must be able to handle high volumes of data, and the process for labeling new examples by an expert must be optimized.

We encourage papers that describe applications of active learning in real-world. The industrial context, the main difficulties met and the original solution developed, shall be described. Contributions on the following challenge, that proposes such a practical application of active learning, will also be welcome.

Associated challenge to the workshop

As a search engine of places, Nomao collects data coming from multiple sources on the web and aggregates them. The **deduplication** process consists in detecting what data refer to the same place. To automate this process, using Machine Learning is well suited, and to optimize the creation of the training dataset, using Active Learning is appropriate.

However, in such a real-world case, labeling data is costly but large amounts of unlabeled data are available. So this raises specific issues: the main ones are scalability of the proposed method, representativity of the training dataset (e.g. learning when test and train inputs can have different distributions), and practicability of the labeling process (e.g. purchase of data labels by batches).

Today, 29,104 examples have already been labeled, each example being characterized by 120 features. This training dataset is available on the Nomao Challenge page, along with a test set of size 1,985. A huge dataset of 100,000 unlabeled examples will also be provided. Then two active campaigns will be organized, each participant being allowed to ask for the labeling of a given number (e.g. 100) of the unlabeled examples by an expert. And a test campaign will be carried out to evaluate the different approaches proposed, each participant being asked to label a given set of examples, and their predictions being compared to the known true labels. The challenge is completely described on the web page: http://www.nomao.com/labs/challenge. And the data of the challenge will also be disseminated.

Papers that address this issue will be welcome. Authors will thus contribute to the confrontation of proposed solutions and to discussions during the workshop.

Challenge Prize

Author of the best results will receive a free registration for the conference and workshop.

Topics of interest to the workshop include (but are not limited to)

•Active Learning
•Experimental Design
•Incremental Learning
•On-line learning
•Case Studies of Active Learning

Key dates

•First active campaign: Friday, June 1, 2012
•Second active campaign: Friday, June 8, 2012
•Final test campaign: Friday, June 15, 2012

•Paper submission deadline: Friday, June 29, 2012
•Paper acceptance notification: Friday, July 20, 2012
•Paper camera-ready deadline: Friday, August 3, 2012
•Workshop: Friday, September 28, 2012

Biographies of the organizers

Laurent Candillier
Nomao – Ebuzzing Group
1 avenue Jean Rieux
31500 Toulouse
Tel.: +33 6 64 35 08 72
Email: laurent@nomao.com
Bibliography: http://lcandillier.free.fr

Laurent Candillier is currently director of Research & Development at Nomao – OverBlog – Ebuzzing. He studied Computer Science at the University of Lille 1, in France, and at Aston University, in Birmingham, England. He got his PhD from the University of Lille 3 in 2006, on the domain of Machine Learning and Data Mining. He then joined the Research & Development team of Orange Labs, becoming expert on the domain of Recommender Systems. Since he joined the Ebuzzing Group, he also tackled the important domains of Information Retrieval and Natural Language Processing. He participated to many international conferences and workshops, got teaching responsibilities, and is now supervising two PhD thesis.

Max Chevalier
IRIT - SIG
118, route de Narbonne
31062 Toulouse cedex 09
Tel.: +33 5 61 55 74 43
Email: Max.Chevalier@irit.fr
Bibliography: http://www.irit.fr/~Max.Chevalier

Max Chevalier is associate professor in Computer Science at the Toulouse 3 University (France) since 2003 and member of the Generalized Information System research group at IRIT. He obtained his PhD in computer science from the Toulouse 3 University in 2002. Then, he defended his Research Accreditation (HDR) in Computer Science titled “users and information retrieval” from the University Toulouse 3 in 2011.
Max Chevalier research activities are related to adaptive information access & retrieval. His research is particularly based on analysis of user activities related to information (annotations, social information-based activities...). The angular stone of his work concerns user and context modeling and learning. Results are principally integrated in enhanced information retrieval systems and recommender systems.

He is co-heading the francophone scientific Association on Information Retrieval and Applications (ARIA). He is participating as program committee member to many international conferences (e.g. ECIR, DEXA) and member of the editorial board of ISI journal (Ingénierie des Systèmes d'Information).

Vincent Lemaire
Orange Labs
2 avenue Pierre Marzin
22300 Lannion
Tel.: +33 2 96 05 31 07
Email: vincent.lemaire@orange.com
Bibliography: http://perso.rd.francetelecom.fr/lemaire/

Vincent Lemaire was born in 1968 and he obtained his undergraduate degree from the University of Paris 12 in signal processing and was in the same period an Electronic Teacher. He obtained a PhD in Computer Science from the University of Paris 6 in 1999. He thereafter joined the R&D Division of France Télécom where he became a senior expert in data-mining. His research interests are the application of machine learning in various areas for telecommunication companies with an actual main application in data mining for business intelligence. He developed exploratory data analysis and classification interpretation tools. Active learning, incremental learning and data-space exploration are now his main research interests. He obtained his Research Accreditation (HDR) in Computer Science from the University of Paris-Sud 11 (Orsay) in 2008. He is also a member of the AFIA steering committee.

Previous workshop and special session organization:
•Active and Incremental a Learning (AIL) - ECAI 2012 [...]
•Incremental classification and novelty detection - CIDN 2012 [...]
•Active, Incremental and Autonomous Learning: Algorithms and Applications (AIAL) - IJCNN 2012 [...]
•Workshop on Unsupervised and Transfer Learning - ICML 2011 [...]
•Autonomous and Incremental Learning (AIL) - IJCNN 2011 [...]
•Active and Autonomous Learning (AAL) - IJCNN 2010 [...]
•Fast scoring on a Large Database - KDD 2009 [...]

Program Committee

1.Mahmoud Abou-Nasr (Ford Motor Company, USA)
2.Cesare Alippi (Politecnico di Milano, Italia)
3.Albert Bifet (University of Waikato, Hamilton, New Zealand)
4.Zalan Bodo (Babeg Bolyai University, Cluj-Napoca, Romania)
5.Laurent Candillier (Nomao – Ebuzzing Group, Toulouse, France)
6.Max Chevalier (Institut de Recherche en Informatique de Toulouse, France)
7.Lehel Csato (Babeg Bolyai University, Cluj-Napoca, Romania)
8.Gideon Dror (Academic college of Tel-Aviv Yaffo, Israel)
9.Hugo Jair Escalante (National Institute of Astrophysics, Optics and Electronics, Mexico)
10.Matthieu Geist (IMS Research Group, Supélec, Metz, France)
11.Liang Lan (Temple University, Philadelphia, USA)
12.Vincent Lemaire (Orange Labs, Lannion, France)
13.Chris Lovell (University of Southampton, UK)
14.George Runger (Arizona State University, Tempe, AZ, USA)
15.Burr Settles (Carnegie Mellon University, USA)
16.Fabien Torre (INRIA, Lille 1 University, France)
17.Ming-Hen Tsai (National Taiwan University)
18.Ioannis Tsamardinos (University of Crete, Greece)
19.Slobodan Vucetic (Temple University, Philadelphia, USA)

Invited speaker

•A budget is dedicated to the invitation of an expert on the subject

Program (tentative)

- 9:00 - 9:30 Opening
- 9:30 - 10:30 Invited Talk
- 10:30 - 11:00 Coffee Break
- 11:00 - 12:30 Papers presentations
- 12:30 - 14:00 Lunch Break
- 14:00 - 15:30 Challenge presentation and results
- 15:30 - 16:00 Presentation of the winner of the challenge
- 16:00 - 17:00 Papers presentations
- 17:00 - 17:45 Participants open discussion

Submissions

Submitted papers must be written in English and formatted according to the Springer LNAI guidelines. Instructions for authors and paper stylesheet files can be downloaded at:
http://www.springer.de/comp/lncs/authors.html
The maximum length of papers should not exceed 16 pages.

The papers will have to be submitted via Easy Chair:
http://www.easychair.org/conferences/?conf=alraecml2012

Papers will normally be reviewed by three referees. The review process is single-blind (reviewer identities unknown to authors) and there will be no opportunity for author rebuttal. This decision was made to minimize reviewer workload and to concentrate it in time, which may ultimately result in better review quality and decisions. If necessary, a discussion will take place among the reviewers of a paper until a decision is reached.

Follow-up

If the workshop and/or challenge is a success in terms of participation and quality of submissions, a special issue of an international journal or a book will be proposed.