



Laure Berti-Équille

Title:

ML-Based Data Cleaning : Current Solutions and Challenges

Abstract:

With the success of machine learning (ML) techniques in database research, ML has already proved a tremendous potential to dramatically impact the foundations, algorithms, and models of several data management tasks, such as error detection, data cleaning, data integration, and query inference. Part of the data preparation, standardization, and cleaning processes, such as data matching and deduplication for instance, could be automated by making a ML model “learn” and predict the matches routinely. This talk will survey the recent trends of applying machine learning solutions to improve data cleaning as one of the most crucial and difficult task in data management and it will present the next research challenges in the convergence of machine learning and data management.

Bio:

Laure Berti-Équille is a Research Director at IRD, the French research institute for sustainable development, currently leading the research group DIAMS (Data Integration, Analysis, and Management as Services, <http://diams.lis-lab.fr>) at Aix-Marseille University. Before, she was a full professor at Aix-Marseille University (AMU), senior scientist at Qatar Computing Research Institute (Hamad Bin Khalifa University, Qatar), Associate Professor (with tenure) at University of Rennes 1 (France), and visiting researcher at AT&T Labs Research (USA) as a recipient of the prestigious European Marie Curie Outgoing Fellowship. Her interests are at the intersection of large-scale data analytics, and statistical machine learning with a focus on data quality, anomaly detection, and truth discovery, with more than 80 publications and three monographs. She initiated the very first workshop editions on information and data quality in information systems (IQIS 2005) and quality in databases (QDB 2009 and 2016) in conjunction with SIGMOD and VLDB respectively, and co-organized the first French workshops on Data and Knowledge Quality in conjunction with EGC (Extraction et Gestion de Connaissances) in 2005, 2006, 2010, and 2011. She has received various grants from the French Agency for National Research (ANR), the French National Research Council (CNRS), and the European Union.

<http://pageperso.lif.univ-mrs.fr/~laure.berti/>