

Submission Title: “**Julia for Knowledge Mining in Industry 4.0**”

Submission Type: Lightning Talk

Abstract:

Industry 4.0, simply I4.0 or I4, refers to the “Fourth Industrial Revolution” that's the new digital industrial technology for transforming industries into smart/intelligent industries (*iIndustry*) by connecting machines with intelligent robots and Industrial Internet of Things (IIoT) devices. In this talk, we have addressed and proposed several issues for knowledge mining from Industrial Big Data (*iBigData*) in Industry 4.0 using Julia programming language.

Description:

Industry 4.0 engenders and analysis data across the machines in *iIndustry* to produce high-quality products at low costs, and changes traditional production relationships among suppliers, producers, and customers. Industry 4.0 amalgamates nine technologies to transform industrial production, which includes: (1) Big Data Analytics, (2) Autonomous Robots/ Robotics, (3) Simulation, (4) Horizontal & Vertical System Integration, (5) Industrial Internet of Things (IIoT), (6) Cybersecurity, (7) Cloud Computing, (8) Additive Manufacturing (such as 3-D printing), and (9) Augmented Reality. I4.0 uses Decision Support Systems (DSS) incorporating with knowledge mining techniques to know what actions need to take in future that help manufacturers to optimise their operations quickly. The fourth revolution ameliorates the industries with intelligent computing fuelled by data with Machine Learning (ML) and Data Mining (DM) technologies. In this talk, we have addressed several issues for knowledge mining process in Industry 4.0 using Julia programming language. Knowledge mining is the process of extracting hidden information/patters from Industrial Big Data (*iBigData*) to lucid market trends, customer preferences and other information that's useful to businesses. Industrial Big Data is extremely large that we can't store all the data into a single computer/machine; so, we need more scalable and robust learning approach to deal with *iBigData*. We have collected the data set with 1067371 instances named “Online Retail II” from UCI Machine Learning Repository (<https://archive.ics.uci.edu/>) and implemented RainForest and BOAT (Bootstrapped Optimistic Algorithm for Tree construction) learning algorithms using Julia. RainForest and BOAT are basically decision tree (DT) based supervised learning algorithms for classifying Big Data. We have presented a new decision tree merging approach that addresses the repetition and replication problems in tree pruning. Industrial Big Data is multivariate, high-dimensional, noisy, and also the characteristics of data can be changed over the time (e.g. concept drifting in data streaming environment). In this talk, we also discussed the how we can handle the noisy and streaming data; find the most informative training instances, so that we can build a learning model with minimum number of instances. For selecting informative training instances, we have used simple partition-based clustering approach and implemented clustering algorithm in Julia.

Notes:

We have used Julia for implementing both RainForest and BOAT algorithms. We have used “DataFrames.jl” (Julia Data Frames package) and “MLJ.jl” (Julia open-source ML toolbox) for understanding data and building learning models.

Authors:

Dewan Md. Farid, Aicha Sekhari Seklouli, Ouzrout Yacine, Swakkhar Shatabda

Institutes:

1. Decision and Information Systems for Production systems (DISP) Laboratory, IUT Lumière - Université Lyon 2, 160 Boulevard de l'Université 69676 Bron CEDEX, France
2. Department of Computer Science and Engineering, United International University, Bangladesh

Acknowledgements:

We appreciate the support for this research received from the Building Skills 4.0 through University and Enterprise Collaboration (Erasmus+ Shyfte 4.0) project, which aims to develop Skills 4.0 in association with Higher Educational Institutes (HEIs) and industrial Small and Medium-sized Enterprises (SMEs) (<http://shyfte.eu>).