Call for Papers of a Special Issue on
"Online Real-Time Learning Strategies for Data Streams"
for *Neurocomputing* (impact: 2.3)

Aims and Scope:

Learning from on-line data streams is a research area of growing interest, because large volumes of data are continuously generated from multi-scale sensor networks, production and manufacturing lines, social media, the Internet, wireless communications etc., often with a high incoming rate. This desires the usage of real time learning and modelling algorithms. An important aspect in data stream mining is that the data analysis system, the learner, has no control over the order of samples that arrive over time --- they simply arrive in the same order they are acquired and recorded. Also, the learning algorithms usually have to be fast enough in order to cope with (near) real-time and on-line demands. This usually requires a single-pass learning procedure, restricting the algorithm to update models and statistical information in a sample-wise manner, without using any prior data.

Researchers in the area of computational intelligence have pioneered the notion of evolving (intelligent) systems (EIS) in order to address the challenges in data stream mining. EIS offer a plausible solution for learning from data streams, because they offer a flexible working framework, which adapts to any variation in data trends or expansion of system states by incremental, recursive parameter adaptation and evolution of structural components likewise. EIS are also scalable to be deployed in the online real-time scenario, because they work in single-pass mode. Apart from incremental and evolving learning methodologies for neural networks and machine learning models, EIS include the transparent trait of (neuro-) fuzzy systems via the human-like linguistic rule and emulates approximate reasoning merit of human being, which copes with uncertain, imprecise, and inaccurate natures of real-world problems. However, four learning issues in learning from large data streams have not been sufficiently explored in current EIS research:

1) Existing EISs still learn on all incoming data without being able to rule out superfluous data samples for model updates, thereby being computationally intractable for processing large data streams. In addition, labelling cost in the large-scale classification problem is huge, because it requires intensive operator intervention.

2) The issue of concept drift reflects a situation in which the input and/or output concepts do not follow a fixed and predictable data distribution. In the classification problem, the uncertainties in data distribution can be also perceived in the increase of class overlapping or a confusion of class labels in the feature space, which leads to a deterioration of classifiers in classifying data points. Existing EIS has matured to cope with abrupt drifts, because it adopts an open structure principle which captures rapidly changing environments. However, it weakly takes into account other types of concept drifts: gradual, incremental, cyclic. The gradual, incremental drifts are more

difficult to be handled, because they cannot be detected by global drift detection approaches in the structural learning scenario and cannot be addressed by the parameter learning scenario either. These drifts gradually interfere current data distribution, which undermines the predictive accuracy more severely. On the other hand, the cyclic concept drift often causes the catastrophic forgetting of previously valid knowledge, because current EISs lack of capability to perform the component recall mechanism. In other words, a component (e.g. a neuron, a rule, a branch in a decision tree etc.), pruned in the earlier training episode, cannot be reactivated again in the future.

3) Real-world data streams often portrays inexact, inaccurate and uncertain characteristics, which induces uncertainty in the data representation. This phenomenon occurs as a result of disagreements in expert knowledge, noisy measurement, and noisy data.
Existing EIS are mostly crafted in the context of crisp learning techniques such as neural networks and classical machine learning models. These are not sufficiently robust to overcome uncertainty in the data representations. Because the data do not truly represent the system dynamics, the identification of proper components and structures (such as type-1 or even type-2 fuzzy rules) representing such uncertainties becomes an extremely important task.

4) The curse of dimensionality is another underlying issue in mining large data streams. In addition to the increase of computational complexity and memory requirement, the curse of dimensionality makes a learning problem hard to be solved, because the input dimension is inherent to the sample size. It is evident that a model can achieve higher accuracy after performing the dimensionality reduction or the feature selection process due to a reduction of curse of dimensionality and over-fitting.
In realm of the EIS, the feature selection process is assumed to be carried out offline in the pre-training process using pre-recorded samples. Conversely, existing feature selection scenarios adopt the batched learning mode or necessitates the multi-pass learning procedure. Because an input feature cannot be recalled in the future once discarded, this problem causes the discontinuity of the training process, which at least warrants a retraining phase from scratch. To the best of our knowledge, this issue is an uncharted territory of any existing work.

Topics:
The main topics of this special session include, but are not limited to, the following:
[Basic Methodologies]
- Online real-time unsupervised learning and clustering for large data streams
- Online real-time supervised classification and regression for large data streams
- Online real-time time-series modelling for large data streams
- Recent advances in incremental learning of any types of neural networks (MLP, RBF, etc) from data streams
- Recent advances in evolving neuro-fuzzy and fuzzy systems
- Online real-time intelligent controller for large data streams

[Advanced Concepts]
- Appropriate handling of data uncertainty in various forms in learning from large data streams
- Tools and techniques for data stream mining in uncertain environments
- Computational intelligence methods for big data analytics and huge data bases

- Techniques to address drifts and shifts in data streams
- On-line dynamic dimension reduction in high-dimensional streams
- Feature selection and extraction techniques for large data streams
- Sample selection and active learning for large data streams
- Reliability in model predictions and parameters
- Domain adaptation, importance weighting and sampling
- Parameter-low and –insensitive learning methods
- On-line complexity reduction to emphasize transparent, more compact models

- *Real world applications of on-line data stream mining and modeling techniques such as: [Applications]*
    - Processing of huge data bases and Big Data
    - Data stream modelling and identification (supervised and unsupervised)
    - Online fault detection and decision support systems
    - Online media stream classification
    - Process control and condition monitoring
    - Modelling in high throughput production systems
    - Web applications
    - Adaptive chemometric models in dynamic chemical processes
    - Online time series analysis and stock market forecasting
    - Robotics, Intelligent Transport and Advanced Manufacturing
    - Adaptive Evolving Controller Design
    - User Activities Recognition
    - Cloud Computing
    - Multiple Sensor Networks

Timeline and Submission

Submission Deadline (extended): August 1st, 2016
Acceptance Deadline: December 1st, 2016
Expected Publication Date: March 1st, 2017

Papers will be evaluated based on their originality, presentation as well as relevance and contribution to the field of data stream mining techniques, suitability to the special issue, and overall quality. All papers will be rigorously refereed by 3 peer reviewers. Submission of a manuscript to this special issue implies that no similar paper is already accepted or will be submitted to any other journal.

Authors should consult the "Guide for Authors", which is available online at
https://www.elsevier.com/journals/neurocomputing/0925-2312/guide-for-authors,
for information about preparation of their manuscripts. Manuscripts should be submitted via the Elsevier Editorial System http://www.journals.elsevier.com/neurocomputing/.
IMPORTANT: Please choose "**SI: Data Stream Mining**" when specifying the Article Type.

Guest Editors (alphabetically):

1. Dr. Mahardhika Pratama, La Trobe University, Melbourne, Australia
2. Dr. Edwin Lughofer, Johannes Kepler University Linz, Austria, edwin.lughofer@jku.at
3. A/Professor Dianhui Wang, La Trobe University, Melbourne, Australia