# Polytech network form for PhD Research Grants from the China Scholarship Council

This document describes the PhD subject and supervisor proposed by the French Polytech network of 14 university engineering schools. Please contact the PhD supervisor by email or Skype for further information regarding your application.

| Supervisor information | |
|---|---|
| **Family name** | RAGOT ; RAMEL |
| **First name** | Nicolas ; Jean-Yves |
| **Email** | nicolas.ragot@univ-tours.fr ; ramel@univ-tours.fr |
| **Web reference** | https://www.univ-tours.fr/site-de-l-universite/m-nicolas-ragot--199416.kjsp?RH=LI |
| **Lab name** | LIFAT, RFAI |
| **Lab web site** | www.li.univ-tous.fr ; www.rfai.li.univ-tours.fr |
| **Polytech name** | Polytech Tours |
| **University name** | Université de Tours |
| **Country** | France |

| PhD information | |
|---|---|
| **Title** | **Continuous learning of pattern recognition system for classification in data streams** |
| **Main topics regards to CSC list (3 topics at maximum)** | I7, I12, I1 |

| Required skills in science and engineering | Data analysis, computer science, programming langages (C++, Java, Python, Matlab), machine learning |
|---|---|

# Subject description (two pages maximum)

**Context & motivations**

Internet and digital technologies have led to the continuous generation of large amounts of data of all kinds (statistical indicators on different domains, multimedia data, document digitization, etc.). While the processing of this data has become a priority, particularly through data mining, many problems remain unsolved. In particular, the supervised classification tools used to identify patterns or elements in these data streams are based primarily on standard learning methods. These ones are limited in this context because they often require some *a priori* knowledge that makes these systems relatively rigid, static and therefore unable to evolve over time. It is only very recently that research has focused on the development of truly dynamic recognition systems with lifelong abilities. But, due to the complexity of the task, many scientific obstacles remain. This is particularly the case if we consider the evolution of concepts to identify and their interdependencies (concepts that are likely to change, disappear, merge or divide, etc.) but also and especially if we consider the description space of concepts that often remains fixed and common to all concepts.

**Goals of the PhD thesis**

In this thesis, we wish to continue work initiated in the laboratory on the learning of dynamic systems dedicated to supervised classification of data whose underlying concepts are evolving over the time. In particular, very little work has been done to change the description space of the data. While it is commonly accepted that concepts are likely to change over time for many problems, high recognition rates can only be preserved over the time if the description space evolves jointly with the concepts.

In this thesis, three points in particular will be studied. The first will focus on the nature of classifiers and algorithms that can absorb such changes in the representation space, without having to perform a complete retraining. Next, methods will be developed to automatically detect and change the representation space using feature selection and feature combination approaches. Finally, we will study under which conditions and in what ways the user can interact to make or guide these modifications, while ensuring a good consistency of the decision-making mechanism. From this point of view, we can also try to verify the usability and robustness of our approaches when several users are involved.

On the experimental level, the theoretical tools will be implemented and validated initially on classical benchmarks used by the community. They will also be tested on specific application areas such as classification or annotation in a digital document stream, or on data mining problems.

**References**

A. K. Ngo Ho, N. Ragot, J.-Y. Ramel, V. Eglin, N. Sidere, "Document Classification in a non-stationary environment: A One-Class SVM Approach", in Proc. of 12th International Conference On Document Analysis and Recognition (ICDAR 2013), August 25-28th, Washington DC, USA, 2013.

R.Elwell, R.Polikar: "Incremental Learning of Concept Drift in Nonstationary Environments". IEEE Trans on Neural Networks 22(10): 1517-1531 (2011).

M.Muhlbaier, A.Topalis, R.Polikar, "Learn++.NC: Combining Ensemble of Classifiers Combined with Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes," IEEE Trans on Neural Networks, vol. 20, no. 1, pp. 152 – 168, 2009.

M.Karasuyama and I.Takeuchi, "Multiple incremental decremental learning of support vector machines", IEEE Trans. on Neural Networks, Vol. 21 Issue 7, pp. 1048-1059, 2010.

Ghazal Jaber, "An approach for online learning in the presence of concept change", PhD Thesis from Université Paris Sud 11, 2013.