Research Grants for PhD students from the China Scholarship Council

Information Form (please read the guidelines carefully on the website www-csc.utt.fr)

Supervisor's name : Beauseroy | Given names : Pierre

Status (prof., assistant prof., …) : Professor

Laboratory : LIST3N/M2S | Website address : http://lm2s.utt.fr/fr/index.html

Institution : UTT | Website address : https://www.utt.fr/

Scientific competence of the supervisor:

Artificial intelligence, Machine learning, detection, classification, feature selection, pattern recognition

Two major publications in the field proposed for the PhD :

1. Sofia Marino, Pierre Beauseroy, and Andre Smolarz. Weakly-supervised deep learning approach for potato defects segmentation. Engineering Applications of Artificial Intelligence, 85 :337-346, 2019
2. Martin Palazzo, Pierre Beauseroy, and Patricio Yankilevich. A pan-cancer somatic mutation embedding using autoencoders. BMC Bioinformatics, 20 :655-665, 2019

Website address of the personal page : http://lm2s.utt.fr/fr/_plugins/mypage/mypage/content/beausero.html

Supervisor's email : pierre.beauseroy@utt.fr

Description of the research work proposed for a PhD | ***Topic # (see list) :*** I.12

Title : Unsupervised multi-task learning from incomplete multisource data

Subject :

In recent years, thanks to deep learning, machine learning has made substantial advances in various domains such as object detection, classification and image/signal segmentation. Multi-tasks learning is based on the idea that solving multiple learning problems simultaneously helps to benefits of commonalities and differences across the considered problems. It can result in improved training efficiency and prediction accuracy for each problem model, when compared to solving the problems separately. Up to now little attention and efforts have been put to apply this type of approach in the context of unsupervised learning (when no labels are available for training). The goal of this research project is to study the cases of incomplete multiple sources of structured data. It corresponds to situations where all sources are not necessary always available to describe each observation. For example, biological data are typical of such situations. Each patient is examined by a series of tests. Each test gives a specific set of measurements. But the tests performed on patients are different from each other. This situation can be modeled as a case of missing data where the missing data differ for each patient and depend on the performed tests. The goal is to take advantage of all the patients data together in order to improve each test data model on one hand and to be able to design a global model that enables to reconstruct missing data on the other hand.

Keywords :

Artificial Intelligence, Machine learning, Deep learning, multi-tasks learning, unsupervised learning

Expected collaborations :

Quality control of agricultural products.
Data-mining for biologocal data.

Background required from the applicant :

Statistics, Machine learning
Matlab and python would be appriciated

Existence of a PDF file detailing the proposal ("yes" or "no") : YES

(see guidelines on the website www-csc.utt.fr)