SOUTENANCE DE THÈSE

M. Yonatan Carlos Carranza Alarcón

Soutiendra sa thèse de **doctorat** sur le sujet **:**

Distributionally robust, skeptical inferences in supervised classification using imprecise probabilities

Unité de recherche : ***Heudiasyc – UMR CNRS 7253***

Le mardi 8 décembre 2020 à 9h à l'université de technologie de Compiègne, amphi L101, centre Pierre Guillaumat et en suivant ce lien :
https://utc-fr.zoom.us/j/89683269171

Devant le jury composé de :

M. Frédéric Pichon, professeur, membre examinateur
Université d'Artois, Laboratoire de génie informatique et d'automatique de l'Artois, Béthune
M. Thierry Denoeux, professeur, président du jury
Université de technologie de Compiègne, laboratoire Heudiasyc, Compiègne
M. Benjamin Quost, maître de conférences, membre examinateur
Université de technologie de Compiègne, laboratoire Heudiasyc, Compiègne
M. Nicolas Verzelen, chargé de recherche, membre examinateur
INRA, Laboratoire de recherche MISTEA, Montpellier
Mme Ines Couso, professeur, membre rapporteur
Université d'Oviedo, département statistiques et recherche opérationnelle, Oviedo, Espagne
M. Willem Waegeman, professeur associé, membre rapporteur
Université de Ghent, department of Data Analysis and Mathematical Modelling, Ghent, Belgique
M. Jesse Read, professeur, membre rapporteur
Ecole polytechnique LIX, Maching Learning, Palaiseau
M. Sébastien Destercke, chercheur, directeur de thèse
Université de technologie de Compiègne, laboratoire Heudiasyc, Compiègne

Decision makers are often faced with making single hard decisions, without having any knowledge of the amount of uncertainties contained in them, and taking the risk of making damaging, if not dramatic, mistakes. In such situations, where the uncertainty is higher due to imperfect information, it may be useful to provide set-valued but more reliable decisions.

This works thus focuses on making distributionally robust, skeptical inferences (or decisions) in supervised classification problems using imprecise probabilities. By distributionally robust, we mean that we consider a set of possible probability distributions, i.e. imprecise probabilities, and by skeptical we understand that we consider as valid only those inferences that are true for every distribution within this set. Specifically, we focus on extending the Gaussian discriminant analysis and multi-label classification approaches to the imprecise probabilistic setting.

Regarding to Gaussian discriminant analysis, we extend it by proposing a new imprecise classifier, considering the imprecision as part of its basic axioms, based on robust Bayesian analysis and near-ignorance priors. By including an imprecise component in the model, our proposal highlights those hard instances on which the precise model makes mistakes in order to provide cautious decisions in the form of set-valued class, instead.

Regarding to multi-label classification, we first focus on reducing the time complexity of making a cautious decision over its output space of exponential size by providing theoretical justifications and efficient algorithms applied to the Hamming loss. Relaxing the assumption of independence on labels, we obtain partial decisions, i.e. not classifying at all over some labels, which generalize the binary relevance approach by using imprecise marginal distributions. Secondly, we extend the classifier-chains approach by proposing two different strategies to handle imprecise probability estimates, and a new dynamic, context-dependent label ordering which dynamically selects the labels with low uncertainty as the chain moves forwards.