

PhD Thesis Proposal

University of Nice Sophia Antipolis
Nice, FRANCE

Advisors:

Cédric RICHARD

cedric.richard@unice.fr

www.cedric-richard.fr

Rémi FLAMARY

remi.flamary@unice.fr

remi.flamary.com

Period to apply: May, 1 to June, 6

See: <http://ed-sfa-unice.fr/?q=content/thèse-sur-contrat-établissement-uns>

Please contact C. Richard and R. Flamary for assistance

Distributed estimation over multitask networks

Distributed adaptive estimation allows a collection of interconnected nodes to perform preassigned learning tasks from streaming measurements. For online parameter estimation, among various strategies such as consensus (Nedic'09, Kar'09), diffusion LMS (Lopes'08, Cattivelli'10) is an efficient algorithm that is particularly attractive due to its enhanced adaptation performance and wider stability ranges. Its variants and performance have been extensively studied in the literature, under various scenarios (Zhao'12, Chen'12'13).

The working hypothesis for earlier studies on distributed estimation algorithms, in particular diffusion LMS strategies, is that the nodes cooperate with each other to estimate a single parameter vector. We shall refer to problems of this type as single-task problems. However, many problems of interest happen to be multitask-oriented in the sense that there are multiple optimum parameter vectors to be inferred simultaneously and in a collaborative manner. The multitask learning problem is relevant in several machine learning formulations (Chen'09, Chapelle'11). In the distributed estimation context, which is the focus of this work, there exist many applications where either agents are subject to data measurements arising from different models or they are sensing data that varies over the spatial domain. This situation arises, e.g., in remote sensing for assessing, monitoring, and managing risks related to natural disasters, in radioastronomy conducted with large radio antennas arrays, to cite a few. It is then important for cooperation to occur only among agents with similar or related objectives to avoid biased solutions.

A handful of works have considered before particular problem formulations that deal with multitask scenarios (Chen'13, Bertrand'10). It is generally assumed in these studies that the nodes have some prior knowledge about clustering or about the parameter space. In the first part of this thesis, we shall derive unsupervised clustering strategies that allows each node to select, via adaptive adjustments of combination weights, the neighboring nodes with which it should collaborate to improve its estimation accuracy. In the related work (Chen'13'14'15), we formulated the multitask problem directly over networks with connected clusters of nodes. In that supervised context, the clusters were assumed to be known beforehand and no clustering strategy was proposed. In the second part of the thesis, we shall consider the problem of noisy, compressed or censored data corrupting local interactions between nodes and/or clusters of nodes. We shall examine how a connected network of agents, with each one of them subjected to a stream of data with partial regression information, can cooperate with each other to estimate the underlying model parameters.

Estimation distribuée sur réseaux multitâches

Les méthodes d'estimation adaptatives distribuées ont pour vocation de permettre à un ensemble de noeuds interconnectés constituant un graphe d'effectuer une tâche d'apprentissage à partir d'un flux de données. Pour l'estimation distribuée d'un vecteur de paramètres, parmi les multiples stratégies existantes, citons par exemple les méthodes de recherche de consensus, les méthodes de type diffusion LMS s'avèrent particulièrement attractives en raison de leur caractère adaptatif vis-à-vis de possibles évolutions des hypothèses et de leur grande stabilité. Plusieurs variantes en ont été proposées dans la littérature, et leurs performances largement étudiées sous différentes hypothèses.

Une hypothèse de travail largement retenue, pour ne pas dire toujours, dans la recherche et l'étude de méthodes d'estimation distribuées est que les noeuds du graphe coopèrent pour estimer un vecteur de paramètres unique et commun à tous. On peut qualifier le problème de simple tâche dans ce cas. Toutefois, de nombreux problèmes sont par nature multitâches au sens où il y a alors plusieurs vecteurs de paramètres à inférer simultanément de manière coopérative. Des problèmes d'apprentissage multitâches ont déjà été définis et traités dans le domaine du machine learning. Dans un contexte d'estimation distribuée, qui constitue le cadre de travail de la thèse, ce type de problème se pose lorsque des agents ou clusters d'agents du réseau observent des phénomènes distincts, potentiellement liés, compte tenu de l'étendue spatiale du réseau d'observation par exemple. Parmi les applications potentielles existantes, citons la surveillance par réseaux de capteurs d'environnements naturels pour la gestion des risques, la radioastronomie par de grands réseaux d'antennes, etc. Il est alors primordial que la coopération entre les noeuds s'organise en fonction des objectifs de chacun.

Peu de travaux jusqu'ici se sont intéressés à des scénarios multitâches. Ces quelques études considèrent généralement que les noeuds disposent d'informations a priori sur leur organisation en clusters dédiés chacun à une tâche. Dans (Chen'13'14'15), nous avons étudié ce problème multitâche dit supervisé.

Dans une première partie de la thèse, nous nous affranchirons de cette hypothèse et nous nous intéresserons à des stratégies de clustering non-supervisées des noeuds d'un réseau permettant à chacun d'eux de sélectionner, via des ajustements adaptatifs des poids des arêtes, les noeuds de leur voisinage avec qui coopérer afin d'améliorer la précision de l'estimation. Dans une seconde partie de la thèse, nous nous intéresserons à des problèmes supervisés et non-supervisés en présence d'informations corrompues (bruitées, compressées, censurées) transitant entre les noeuds ou clusters du réseau. Nous étudierons en quoi elles affectent les performances des méthodes existantes d'estimation distribuées et proposerons des stratégies alternatives.