Sparse Representation and Active Learning for Marine Ecology.

Samah Hijazi
† LISIC Laboratory, Université Littoral Côte d'Opale (ULCO),
E-mail: Samah.Hijazi@mel-etu.univ-littoral.fr

Domaine: Traitement de l'image

1. Abstract

In the context of marine ecology, phytoplankton plays an important role in assessing the quality of water. This quality assessment is increasingly carried out by high-frequency measurements coming from various sensors installed on sites or embedded high resolution cameras, and even qualitative or quantitative information acquired by experience. Managing and organizing these data to extract relevant, useful and robust information became a challenge especially when these data are abundant, noisy, missing, and even aberrant. The aim of my thesis is to use original approaches for the recognition and discrimination of phytoplankton species, the search for efflorescence states, and the detection of potentially harmful species. Particular attention will be paid to the quality of the data. It is clear that, in a context of abundant data, there are "no quality results without quality data".

Our purpose is to develop advanced techniques in data processing by sparse representation and active learning with the integration of contextual knowledge of the field of application. Since the performance of learning algorithms is often penalized by the very large dimension of the data space it is more wise to construct a data space of reduced size. Moreover, this performance depends not only on the quality of the data available but also on the integration of contextual knowledge such as full labeling, partial labelling, etc.

Indeed, labeling information is expensive, certain constraints generated by the comparison between pairs of objects are easy to be formalized by the analyst: two data are similar and must therefore be grouped together (must-link), or not similar, and must belong to distinct groups (cannot-link). However, these constraints are not always error free, and using them in attributes selection process can degrade the algorithm’s performance.

Thus as the first step of my thesis, we started studying the effect of erroneous side pairwise constraints on the sensitivity of available feature selection algorithms. The experimental results we obtained so far can be helpful in choosing which algorithm is robust enough against erroneous constraints.

2. Bibliography


