

228 Seminar 20-11-2014

Speaker: Enric Celaya ([celaya \(at\) iri.upc.edu](mailto:celaya@iri.upc.edu))

Title: *Online EM with Weight-Based Forgetting*

Abstract:

The expectation-maximization (EM) algorithm is an iterative procedure to estimate the parameters of a finite mixture so as to maximize the likelihood of a set of unlabelled data. Its primary use is classification of data, but it can also been used for other tasks such as probability density estimation, regression, or function approximation. The EM algorithm is a batch algorithm, so that all data must be processed at each iteration. When the dataset to be processed is large, or when data are supplied as a stream, the EM algorithm becomes impractical and incremental or online versions of the algorithm are more convenient. Sato & Ishii (2000) developed an online EM algorithm in which a time-dependent discount factor was introduced to forget the influence of old estimations obtained with earlier, inaccurate estimators. In their approach, forgetting is uniformly applied to the estimators of each mixture component at each iteration, irrespective of the weight attributed to each component for the current data sample. This causes an excessive forgetting in the less frequently sampled components that may cause instabilities in the estimations. To address this problem we propose a modification of the algorithm that involves a weight-dependent forgetting, specific for each mixture component, by which old estimations are only forgotten in accordance with the actual weight used to replace them with new data. A comparison of the weight-independent versus the weight-dependent approach shows that this last improves the accuracy of the estimation and exhibits a much greater stability.