

Distributed Data Classification in Sensor Networks

Ittay Eyal

Idit Keidar

Raphael Rom

Department of Electrical Engineering, Technion — Israel Institute of Technology

Abstract

Low overhead analysis of large distributed data sets is necessary for current data centers and for future sensor networks. In such systems, each node holds some data value, e.g., a local sensor read, and a concise picture of the global system state needs to be obtained. To this end, we define the *distributed classification problem*, in which numerous interconnected nodes compute a *classification* of their data, i.e., partition these values into multiple collections, and describe each collection concisely.

We present a *generic algorithm* that solves the distributed classification problem and may be implemented in various topologies, using different classification types. For example, the generic algorithm can be instantiated to classify values according to distance, like the famous k-means classification algorithm.

However, the distance criterion is often not sufficient to provide good classification results. We present an instantiation of the generic algorithm that describes the values as a *Gaussian Mixture* (a set of weighted normal distributions), and uses machine learning tools for classification decisions.

We prove that any implementation of the generic algorithm converges over any connected topology, classification criterion and collection representation, in fully asynchronous settings.