



Hamilton Institute

On Continuous Counting and Learning
in a Distributed System

Dr Bozidar Radunovic

Microsoft Research, Cambridge, UK

Friday, 3rd August, 2012

Abstract:

Consider a distributed system that consists of a coordinator node connected to multiple sites. Items from a data stream arrive to the system one by one, and are arbitrarily distributed to different sites. The goal of the system is to continuously track a function of the items received so far within a prescribed relative accuracy and at the lowest possible communication cost. This class of problems is called a continual distributed stream monitoring.

In this talk we will focus on two problems from this class. We will first discuss the count tracking problem (counter), which is an important building block for other more complex algorithms. The goal of the counter is to keep a track of the sum of all the items from the stream at all times. We show that for a class of input loads a randomized algorithm guarantees to track the count accurately with high probability and has the expected communication cost that is sublinear in both data size and the number of sites. We also establish matching lower bounds. We then illustrate how our non-monotonic counter can be applied to solve more complex problems, such as to track the second frequency moment and the Bayesian linear regression of the input stream.

We will next discuss the online non-stochastic experts problem in the continual distributed setting. Here, at each time-step, one of the sites has to pick one expert from the set of experts, and then the same site receives information about payoffs of all experts for that round. The goal of the distributed system is to minimize regret with respect to the optimal choice in hindsight, while simultaneously keeping communication to the minimum. This problem is well understood in the centralized setting, but the communication trade-off in the distributed setting is unknown. The two extreme solutions to this problem are to communicate with everyone after each payoff, and not to communicate at all. We will discuss how to achieve the trade-off between these two approaches. We will present an algorithm that achieves a non-trivial trade-off and show the difficulties of further improving its performance.

Venue: Seminar Room, Hamilton Institute, Rye Hall, NUI Maynooth

Time: 2.00pm - 3.00pm

Travel directions are available at www.hamilton.ie