

Data Mining: From Standalone Algorithms to Integrated Components and Ready-made Internet Services

Sunita Sarawagi

Indian Institute of Technology, Bombay.

e-mail: sunita@it.iitb.ernet.in

Abstract

Founded on the traditional fields of statistics and machine learning, data mining, as is known today, has come a long way. The past seven years have seen considerable research effort in the development of scalable algorithms that have since transitioned into standalone commercial tools. However, these tools have failed to achieve the anticipated level of market acceptance. Several reasons have been cited for this, including, non-triviality of model selection in the discovery process, unavailability of high-quality data, incognizance of prior knowledge, and, impedance mismatch with applications.

In this talk we will discuss recent technical developments that are seeking to avoid these shortcomings. Some examples are: closer integration of mining operations with conventional querying systems; mining services to let data rich mining experts share their mining models on the internet; customized mining operations embedded in vertical applications like customer relationship management; and, user-cognizant mining. The talk will address one of these issues at length.