Tutorial on Machine Learning and Data Mining

DR DAVID LUBINSKY
Department of Computer Science
University of the Witwatersrand, WITS
E-Mail: david@cs.wits.ac.za

Abstract: The ability to learn from observations and to modify our understanding of the world based on experience is an essential aspect of intelligent behavior. Machine learning has thus become an important sub-area of artificial intelligence research and has yielded a number of interesting results about how it is possible to build computer systems that learn as well as insights into the process of learning. In addition, many new technologies have been developed which have been applied in the area of automatic learning from large databases.

Searching large databases for hidden relationships is the focus of the new area known as data mining, which is being applied increasingly to search large databases for hidden nuggets of information.

In this tutorial we will quickly introduce the related fields of machine learning and data mining. The tutorial will run for two hours. The first hour will be an introduction to machine learning, and the second will be a more in depth look at induction and data mining and how these fields extend from machine learning.

Outline

Part 1 Introduction to Machine Learning

Why is the problem important.
Supervised and unsupervised learning.
Machine learning and statistics.
What kind of thing do we want to learn.
The languages that can be used to represent learned knowledge.

> Logical languages
> Horn Clauses
> Trees
> Numeric functions
> Neural nets

Part 2 Data Mining

What kind of data is available.

> Corporate databases
> Medical databases
> Financial databases
> Automatically collected data

What are the nuggets we are trying to find in data.

Popular types of data mining systems

> Tree based methods
> High search methods
> Statistical methods

Evaluation of learned information

> Cross validation
> Prediction

Open areas and problems with the current methodologies.