

Christos Faloutsos

Professor

Member, CMU Database Group, Parallel Data Lab

Machine Learning Department, Electrical and Computer Engineering

2:30, August 13, 2007

Research Topic

Data Mining

Research Problem

How can we determine patterns in large datasets?

Problem Statement

Given an extremely large dataset, use power laws to construct a valid mathematical model for its behavior.

Operational Definition

Model: A formal abstraction of anything that uses mathematical and logical terminology to provide a definition.

Power Law: A mathematical function that does not change based on scale – generally x raised to a constant exponent summed with another a second function that shrinks asymptotically as x increases or decreases.

Problem Description

Much research has been done into small data sets and their behavior, but what of large datasets? How much does size matter? While datasets are often expected to display a normal distribution, extremely large datasets –such as links between websites- apparently obey power laws in their distributions, dropping off exponentially between graphs. Is this behavior universal, is there a scalable method for pattern detection, and how can we develop pattern detection for graphs that evolve over long periods of time?

Computer Science Perspective

This research involves the development of new algorithms for extrapolating patterns from data, something with which Computer Science is fundamentally concerned.

Disciplines Actively Involved

Computer Science

Operational Definition

Actively Involved Disciplines: A Discipline from which a member would be acknowledged in the research paper or any discipline involved for which there is a professional association that fosters more knowledge of it through the scientific method.

Description of Disciplines Involved

Computer Science: While many fields may be affected by this research, the actual nature of the work solely concerns the development of new algorithms for faster analysis and predictive analysis, both of which are the concern of Computer Science.

References

Presenter web page:

<http://www.cs.cmu.edu/~christos/>

By plandweh

Updated plandweh