Tom Mitchell
Fredkin Professor of AI and Machine Learning
Machine Learning Department
School of Computer Science

August 29, 2007 (11:00pm)

Research Topic
Machine Learning

Research Problem
How can we make learning systems take into account and aggregate different sources of data available on the web?

Problem Statement
Given a complete knowledge of code used on the web, develop a learning system that can parse web pages given as input and aggregate them into a single, unified knowledge base.

Operational Definitions
Knowledge Base: A collection of formal logical or mathematical statements that define the known facts about a given object, where "object" can be construed as broadly as possible (i.e. the world, a specific scenario, a data set, a chair, etc.).

Problem Description
Currently, learning systems are almost always run on a single dataset for each experiment. When a new experiment is performed, the old information is generally removed. Instead, Dr. Mitchell plans to create a system that continuously reads the Internet and builds a database of "facts" cross-referenced by a given ontology, which will operate with minimal – "10 minutes per day" – interaction with a human user.

Computer Science Perspective
The primary effort is a computer science problem. Moreover, it is attempting to explore a different area of machine learning, namely a continuous effort with accumulation of data and on a dataset which is not well formed.

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 the success of the research may have ramifications for many fields, the actual reparsing of web page information into a knowledge base concerns only subsets of Computer Science, like data mining and artificial intelligence.

References
Presenter's homepage:
http://www.cs.cmu.edu/~tom/
Additional links:
http://www.cs.cmu.edu/~tom/pubs/AImagazine-7-2005.html
http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-21/www/index.html

By Sduggan with edits from others
Updated plandweh