ScholarWorks

Application of Data Pipelining Technology in Cheminformatics and Bioinformatics

Login (/login)

Application of Data Pipelining Technology in Cheminformatics and Bioinformatics

Mao, Linyong

Name: Linyong Mao.pdf
Size: 889.0Kb
Format: PDF

View/Open

Abstract:

Data pipelining is the processing, analysis, and mining of large volumes of data through a branching network of computational steps. A data pipelining system consists of a collection of modular computational components and a network for streaming data between them. By defining a logical path for data through a network of computational components and configuring each component accordingly, a user can create a protocol to perform virtually any desired function with data and extract knowledge from them. A set of data pipelines were constructed to explore the relationship between the biodegradability and structural properties of halogenated aliphatic compounds in a data set in which each compound has one degradation rate and nine structure-derived properties. After training, the data pipeline was able to calculate the degradation rates of new compounds with a relatively accurate rate. A second set of data pipelines was generated to cluster new DNA sequences. The data pipelining technology was applied to identify a core sequence to represent a DNA cluster and construct the 95% confidence distance interval for the cluster. The result shows that 74% of the DNA sequences were correctly clustered and there was no false clustering.

Description:

Submitted to the faculty of the University Graduate School in partial fulfillment of the requirements for the degree Master of Sciences in the School of Informatics Indiana University December 2002

This item appears in the following Collection(s)

- Informatics Graduate Theses and PhD Dissertations (/handle/1805/303)
- Informatics School Theses and Dissertations (/handle/1805/954)

Show Statistical Information (#)

My Account

- Login
- Register

Statistics

- Most Popular Items
- Statistics by Country
- Most Popular Authors