



It has become very easy to create, publish, and collect data in digital form. The volume of structured and unstructured data is increasing at a tremendous pace. This has led to a whole new set of applications that can be build to solve the problem of turning raw data into valuable information. Possible applications include everything from discovering new trends out of a stream of weblog entries, to automatic learning approaches that supplementing market research processes for new products. Machine learning provides tools for building these applications. A large community of researchers has been working on the topic of learning from data. Although solutions to common problems are publicly available, scaling these solutions into the range of terabytes and petabytes is an

Web: <http://lucene.apache.org/mahout/>



It has become very easy to create, publish, and collect data in digital form. The volume of structured and unstructured data is increasing at a tremendous pace. This has led to a whole new set of applications that can be build to solve the problem of turning raw data into valuable information. Possible applications include everything from discovering new trends out of a stream of weblog entries, to automatic learning approaches that supplementing market research processes for new products. Machine learning provides tools for building these applications. A large community of researchers has been working on the topic of learning from data. Although solutions to common problems are publicly available, scaling these solutions into the range of terabytes and petabytes is an

Web: <http://lucene.apache.org/mahout/>

open issue. To scale algorithms to such dimensions it is vital to distribute data as well as computation. The mission of the Apache Mahout project is to build a suite of scalable machine-learning algorithms that can cope with today's quantities of data. Mahout is built on top of Apache Hadoop.

Algorithms within Mahout:

Classification

Naive Bayesian, Random Forests

Clustering

Canopy Clustering, k-Means, Fuzzy K-Means, Mean Shift, Dirichlet Process Clustering, atent Dirichlet Allocation

Evolutionary Algorithms

Recommendation Learning

User-based Recommender, tem-based Recommender , Slope-One Recommender

Download:

<http://www.apache.org/dyn/closer.cgi/lucene/mahout/>

User list: mahout-user@lucene.apache.org

Dev list: mahout-dev@lucene.apache.org

open issue. To scale algorithms to such dimensions it is vital to distribute data as well as computation. The mission of the Apache Mahout project is to build a suite of scalable machine-learning algorithms that can cope with today's quantities of data. Mahout is built on top of Apache Hadoop.

Algorithms within Mahout:

Classification

Naive Bayesian, Random Forests

Clustering

Canopy Clustering, k-Means, Fuzzy K-Means, Mean Shift, Dirichlet Process Clustering, atent Dirichlet Allocation

Evolutionary Algorithms

Recommendation Learning

User-based Recommender, tem-based Recommender , Slope-One Recommender

Download:

<http://www.apache.org/dyn/closer.cgi/lucene/mahout/>

User list: mahout-user@lucene.apache.org

Dev list: mahout-dev@lucene.apache.org