

Apache Mahout

Alix Bernard

April 30, 2021

Overview

- 1 Introduction
- 2 ML techniques
- 3 Features
- 4 Examples
- 5 Relevant links

Introduction

Mahout: one who rides an elephant as its master.

Apache Mahout is an open-source project started in 2008; it is a **distributed linear algebra framework** used for **creating scalable machine learning algorithms**.

Introduction

The algorithms of Mahout are built on top of Hadoop and written in Java.

Apache Hadoop is an open-source framework allowing to **store and process big data in a distributed environment** across clusters of computers.

Introduction

Companies using Mahout include:

- Adobe
- Facebook
- LinkedIn
- Twitter
- Yahoo

ML techniques

Mahout is a project open to implementations of all kinds of machine learning techniques, in practice it focuses in three key areas of ML for now:

- Recommendation
- Classification
- Clustering

Features

Mahout features:

- Works well in a distributed environment and uses Hadoop library to scale effectively.
- Offers ready-to-use framework for data mining tasks on large volume of data.
- Has applications to analyze large datasets effectively.
- Includes several MapReduce enabled clustering implementations.
- Supports Naive Bayes classification implementations.

Examples of Mahout uses

Mahout can be used for:

- **Collaborative clustering**: mines user behavior and makes product recommendations.
- **Clustering**: takes items in a class and organizes them into naturally occurring groups, such that items belonging to the same group are similar to each other.
- **Classification**: learns from existing categorizations and then assigns unclassified items to the best category.
- **Frequent itemset mining**: analyzes items in a group and then identifies which items typically appears together.

For more details and examples:



<https://mahout.apache.org/>



<https://www.tutorialspoint.com/mahout/>



<https://www.acte.in/apache-mahout-tutorial/>

Thank you for your attention!