

学术研究

面向服务的云数据挖掘引擎的研究

余永红, 向晓军, 高阳, 商琳, 杨育彬

1. 南京大学 计算机软件新技术国家重点实验室, 南京 210093
2. 南京邮电大学 通达学院, 南京 210003

收稿日期 修回日期 网络版发布日期 接受日期

摘要

数据挖掘算法处理海量数据时, 扩展性受到制约。在商业和科学研究的各个领域, 知识发现的过程和需求差异较大, 需要有效的机制来设计和运行各种类型的分布式数据挖掘应用。提出了一种面向服务的云数据挖掘引擎的框架CloudDM。不同于基于网格的分布式数据挖掘框架, CloudDM利用开源云计算平台Hadoop处理海量数据的能力, 以面向服务的形式支持分布式数据挖掘应用的设计和运行, 并描述面向服务的云数据挖掘引擎系统的关键部件和实现技术。依据面向服务的软件体系结构和基于云平台的数据挖掘引擎, 可以有效解决海量数据挖掘中的海量数据存储、数据处理和数据挖掘算法互操作性等问题。

关键词 [云计算](#) [Hadoop](#) [数据挖掘](#) [面向服务的体系结构\(SOA\)](#)

分类号

Research on Service-Oriented Data Mining Engine Based on Cloud Computing

YU Yonghong, XIANG Xiaojun, GAO Yang, SHANG Lin, YANG Yubin

1. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China
2. College of Tongda, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract

The scalability of data mining algorithms is restricted when dealing with large-scale data. There are significant differences in a wide range of application areas and requirements for knowledge discovery process. It is fundamental to provide effective formalisms to design distributed data mining application and support their efficient execution. This paper proposes a novel service-oriented data mining engine based on cloud computing framework, which is named as CloudDM. Differentiating from grid-based distributed data mining framework, CloudDM exploits the capacity of open source cloud computing platform—Hadoop for large-scale data analysis, supports the design and execution of distributed data mining applications according to SOA (service-oriented architecture). Moreover, it discusses and reports the key component functions and implementation technologies. According to the design principles of SOA and data mining engine based on cloud computing, the paper can solve the problems in massive data mining systems, such as big data storage, data processing and interactive operation of algorithms, etc.

Key words [cloud computing](#) [Hadoop](#) [data mining](#) [service-oriented architecture \(SOA\)](#)

DOI:

通讯作者

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(923KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [参考文献](#)

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ [本刊中 包含“云计算”的 相关文章](#)
- ▶ 本文作者相关文章

- [余永红](#)
- [向晓军](#)
- [高阳](#)
- [商琳](#)
- [杨育彬](#)