Design of an Integrated Digital Library System Based on Peer-to-Peer Data Mining

Mohammed Ammari, Mohammadia School of Engineers, University Mohammed V - Agdal, Rabat, Morocco

Dalila Chiadmi, Mohammadia School of Engineers, University Mohammed V - Agdal, Rabat, Morocco

ABSTRACT

Traditional libraries will evolve to digital libraries which are clearly superior at: Dissemination, sharing, linking, storing, and information variety. Therefore, one can say that electronic libraries have specific needs in terms of content, services and long-term preservation. In contrast, digital libraries suffer from several inherent constraints: storage limitation, performance, relevancy, decentralization, lack of semantic, fault tolerance, scalability. The main intention of this paper is to present a design of an integrated digital library system based on peer-to-peer data mining. This article aims also to prove that peer-to-peer mining, an emerging branch of distributed data mining, is a hot research area well suited to overcome intrinsic problems of digital libraries.

Keywords: Design Patterns, Factor Analysis, Integrated Library System, Ontology, Peer-to-Peer Data Mining

1. INTRODUCTION

Data mining (Han & Kamber, 2006) has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection, and customer retention, to production control and science exploration.

Distributed data mining (DDM) consists of mining data using distributed resources. DDM has emerged as a hot research area. Careful attention in the usage of distributed resources of data, computing, communication, storage and human factors in a near optimal fashion are paid by DDM.

Distributed Data Mining can be applied in several domains (Kargupta, 2007):

- Mining Large Databases from Distributed Sites: Grid data mining in Earth Science, Astronomy, Counter-terrorism, Bioinformatics
- Monitoring Multiple Time Critical Data Streams: Monitoring vehicle data streams

DOI: 10.4018/ijcee.2012070101
in real-time, Monitoring physiological data streams

- Analyzing Data in Lightweight Sensor Networks and Mobile Devices: Limited network bandwidth, Limited power supply
- Preserving Privacy: Security and safety related applications
- Peer-to-Peer Data Mining: Large decentralized asynchronous environments

DDM is gaining attention in peer-to-peer (P2P) systems which are emerging as a new distributed computing paradigm for many novel applications that involve connecting nodes via largely ad hoc connections without central coordinator. Grid-computing, massive data storage, instant messaging, videoconferencing, voice over IP, multi-player, P2P search engines, social networks, digital libraries are some examples. We can classify peer to peer networks as structured or unstructured, depending on the way they are connected and how the data they contain is arranged. In a structured network the connections between nodes are of some regular structure, which allows deterministic and optimal lookup hops (typically O (log N)). In contrast to structured networks, nodes in unstructured networks do not share a regular structure and a unified identifier space. Lookups are thus normally achieved by flooding and using replication in the network.

The main intention of this paper is to prove that P2P data mining applications may play a key role in the next generation of digital libraries. P2P data mining is, in fact, well suited to overcome intrinsic problems of digital libraries: storage limitation, performance, search pertinence, decentralization, lack of semantic, scalability. In the following section we discuss some related works. In the third section we present the motivation of the application of P2P data mining in the field of digital libraries, before closing on illustrative example.

2. RELATED WORKS

Our work can be classified in the field of Peer-to-Peer data mining:

In a work done by Datta et al. (2006), they offer an overview of DDM applications and algorithms for P2P environments, focusing particularly on local algorithms that perform data analysis by using computing primitives with limited communication overhead. The authors describe both exact and approximate local P2P data mining algorithms that work in a decentralized and communication-efficient manner.

Sunny T et al. (2012) provide an overview of DDM and P2P Data Mining. Their paper discusses the need for DDM, taxonomy of DDM architectures, various DDM approaches, DDM related works in P2P systems and issues and challenges in P2P data mining.

In a work done by Ammari et al. (2013), they propose a system that takes the form of a semantic layer which can be superimposed on top of any P2P infrastructure. This layer is subdivided as semantic categories through a Hilbert curve which has the merit of good preservation of locality. HyperCup structure was chosen to support semantic category in order to increase fault tolerance and overcome bottleneck problem. HyperCup was also selected, because of its efficient broadcast algorithm. Classification within a semantic category will be conducted through multidimensional data analysis. The classification process will be recursively repeated with ontology refinement. The main intention of the current paper is to instantiate the proposed system in the field of digital libraries in order to prove that P2P data mining applications may play a key role in the next generation of digital libraries. Indeed, the current paper presents a design of an integrated digital library system based on peer-to-peer data mining.
The Role of Metaphors in the Interpersonal Discourse of Online Forums
www.igi-global.com/chapter/role-metaphors-interpersonal-discourse-online/42859?camid=4v1a

From Town Hall to the Virtual Community: Engaging the Public with Web 2.0 and Social Media Applications
www.igi-global.com/chapter/from-town-hall-to-the-virtual-community/107790?camid=4v1a