Exploring the Capabilities of Mobile Agents in Distributed Data Mining

U.P.Kulkarni \(^{(1)}\), K.K.Tangod \(^{(2)}\), S. R. Mangalwede \(^{(3)}\), A.R.Yardi \(^{(4)}\)

\(^{(1)}\) Faculty, Department of Computer Science and Engineering, SDM College of Engineering and Technology, Dharwad, India.
\(^{(2,3)}\) Faculty, Department of Computer Science and Engineering, Gogte Institute of Technology, Belgaum, India.
\(^{(4)}\) Principal, Walchand College of Engineering and Technology, Sangli, India.

Abstract

Over the past decade, data mining has gained an important role in analysis of large datasets and there by understanding the complex systems in almost all areas. Such datasets are often collected in a geographically distributed way, and cannot, in practice, be gathered in to single repository. Existing data mining methods for distributed data are of communication intensive. Many algorithms for data mining have been proposed for a data at a single location and some at multiple locations with improvement in terms of efficiency of algorithms as a part of quality but effectiveness of these algorithms in real time distributed environment are not addressed, as data on the web/network are distributed by very of its nature. As a consequence, both new architectures and new algorithms are needed. In this paper we introduce the software agent technology that supports building of distributed data mining architecture and explore the capabilities of mobile agents’ paradigm and will show by experiment that, it is suited for distributed data Mining compared to traditional approaches like client server computing.

1. Introduction

Progress in bar-code technology has made it possible for retail organizations to collect and store massive amounts of sales data, referred to as the basket data. The emergence of network-based computing environments has introduced a new and important dimension i.e distributed sources of data and computing. The advent of laptops, palmops, handhelds, embedded systems, and wearable computers are also making ubiquitous access to a large quantity of distributed data a reality. Advanced analysis of distributed data for extracting useful knowledge is the next natural step in the increasingly connected world of ubiquitous and distributed computing. Most of the popular data mining algorithms are designed to work for centralized data and they often do not pay attention to resource constraints of distributed and mobile environments. Recent research in this area has demonstrated that handling these resource constraints in an optimal fashion requires a new breed of data mining algorithms and systems that are very different from their centralized counterparts.

A mobile agent is a running program that can move from host to host in a network and created a new paradigm for data exchange and resource sharing in rapidly growing and continually changing computer network \([4,8,9]\). Mobile code promises to increase system flexibility, scalability, and reliability. To date, however, this promise has been only partially fulfilled. Among the reasons for the technology’s unmet potential are security concerns and incomplete knowledge of the possible consequences of mobile code use \([7]\). This paper explores the capabilities of Mobile Agents’ Technology for Distributed Data Mining activities.

2. Motivations and Related Work

The process of data mining is becoming harder for large datasets because of the following reasons:

- Data stored online doubles every year.
- Datasets are distributed geographically
- Datasets are immovable.

In traditional approach, data mining algorithms are applied to the data at a single location. When data is collected in distributed way; this means that data must
be transferred to single central place to enable conventional algorithms to be applied. This approach is costly in terms of communication, storage at central site. The various new approaches proposed [17,19,21,22,26,27,28] in the recent past aim at integrating the knowledge, which is discovered out of data at different geographically distributed sites, but it is necessary to have minimum amount of network communication, and maximum of local computation.

3. Proposed Work

Goal of this work is to develop a system using Mobile Agents and show quantitatively how Mobile Agents are better to mine the huge data available in the distributed environment where the component data are distributed among several sites. The experiments are conducted using scatter-gather style. In the first case agents are sent to remote data repository and large datasets are transferred to central site and then Data mining algorithm is applied on this collected large dataset at central site. Around 6MB data size was used for the faster data transfer because as data chunk size is increased the time taken by Mobile Agent to migrate reduces, as indicated in figure-4. In the second case agents are sent to each remote data repository. Each such agent computes local models, and are brought to central site for combining them to generate global model at central site. The comparative performances of these two cases are shown in figure-1. From figure-1, it is clear that Mobile Agents perform better in terms of time and network bandwidth usage compared with traditional system built over client server technology. The percentage improvement of performance increases with decrease in the number of patterns obtained in each site because, more the number of patterns, implies, more data to be carried by mobile agents during migration to central site. Mobile agents are not good for data transfer which is clear from figure-3, where performance of Mobile agents are compared with File Transfer Protocol (FTP) to transfer huge data over internet. The poor performance is due to serialization and deserialization process of objects involved in migration. However there is limitation on the capacity (Up to 6.004 MB + 514 Bytes) of the data/knowledge, which the agents carry with them, as indicated in Figure-2.
4. Conclusion

Compared to the existing applications of similar kind, Mobile Agents’ application proves to be one of the best and robust methods to handle distributed data and hence distributed data mining, in faster way. This also addresses the issue of handling of dynamically generated data, because the data is maintained at remote sites, which can be updated continuously, or as and when it is required. Since this approach does not require the huge amount of data transfer from remote to central site, the network resources are used optimally. Hence existing client server based distributed data mining systems needs to be rewritten using Mobile Agents paradigm, to take advantage they offer. The development of methodology for consolidating the collected knowledge from different sites is in progress.

5. Bibliography

[11] Jian Pei, Jiawei Han, et al, “Mining Frequent Item sets with Convertible Constraints”, Simon Fraser University, Canada.


[22] Shonali Krishnaswamy, Seng Wai Loke, Arkady Zaslavsky, “Supporting The Optimization of Distributed Data Mining by Prediction Application Run Times”.

[23] Naren Ramakrishnan, Ananth Y. Grama, “Mining Scientific Data”, Department of Computer Science, Virginia Tech, VA 24061 & Purdue University, IN.


[31] Frank Wang et al, “A Distributed and Mobile Data Mining System”, 0-7803-7840-7/03, 2003 IEEE.
