Chapter 6

Grid Data Mining Strategies for Outcome Prediction in Distributed Intensive Care Units

Manuel Filipe Santos  
Universidade do Minho, Portugal

José Machado  
Universidade do Minho, Portugal

Filipe Portela  
Universidade do Minho, Portugal

António Abelha  
Universidade do Minho, Portugal

Miguel Miranda  
Universidade do Minho, Portugal

Álvaro Silva  
Centro Hospitalar do Porto, Portugal

Fernando Rua  
Centro Hospitalar do Porto, Portugal

ABSTRACT

Previous work developed to predict the outcome of patients in the context of intensive care units brought to the light some requirements like the need to deal with distributed data sources. Those data sources can be used to induce local prediction models, and those models can in turn be used to induce global models more accurate and more general than the local models. This chapter introduces a distributed data mining approach suited to grid computing environments based on a supervised learning classifier system. Five different tactics are explored for constructing the global model in a Distributed Data Mining (DDM) approach: Generalized Classifier Method (GCM), Specific Classifier Method (SCM), Weighed Classifier Method (WCM), Majority Voting Method (MVM), and Model Sampling Method (MSM). Experimental tests were conducted with a real world data set from intensive care medicine. The results demonstrate that the performance of DDM methods is very competitive when compared with the centralized methods.

DOI: 10.4018/978-1-4666-3667-5.ch006
INTRODUCTION

Recently, there is a significant progress in the research related to distribute data mining. Digital data stored in the distributed environments is doubling within a few years. More advanced and feasible distributed data mining algorithms and strategies are required in the current fast growing environment.

Learning Classifier System (LCS) is a concept formally introduced by John Holland as a genetic based machine learning algorithm (Santos, Mathew, Kovacs, & Santos, 2009). Manuel Santos (Santos, 1999) developed the DICE system, a parallel and distributed architecture for LCS. In his work he attempted to parallelize the genetic algorithm and LCS message operations to increase system’s performance. A. Giani, Dorigo and Bersini also did significant research attained in the experimental work research in the area of parallel LCS (Giani, Starita, & Vanneschi, 1999). Their implementation also tried to increase the performance of the system. All implementations of parallel LCS consider a single data and generate a single model.

This work is part of two major projects – the Gridclass project – whose main goal is to implement the UCS in a grid environment and – the INTCare project – whose main goal is to implement an intelligent decision support system for Intensive Care Units where the data distribution among different sites is an important issue. Gridclass system does not paralyze any part of the UCS. Various instances of the UCS are executed in different distributed sites with different set of data. All the experimental work was done using the Grid gain platform; a java based distributed computing middleware (Gain, 2006).

The key objective of this work is to construct a global data mining model from different local models of the grid and compare DDM and CDM methods. Grid computing architecture is considered the best distributed framework for solving the distributed data mining task (Luo, Wang, Hu, & Shi, 2007; Cannataro, 2004). Each node of the grid environment executes different UCS and those nodes send local data mining models to the central site for developing a global model. This work considers five different methods for merging local models from each distributed sites (Santos, et al., 2009; Santos, Mathew, & Santos, 2010; Santos, Mathew, & Santos, 2011). The different strategies are: Specific Classifier Method (SCM), Weighted Classifier Method (WCM), Generalized Classifier Method (GCM), Majority Voting Method (MVM) and Model Sampling Method (MSM).

The Intensive Medicine is a specific environment where the patients normally are in weak conditions. The decisions are normally made by some stress or by a necessity of quickly response. The doctors are very difficult make decision in this conditions especially when they don’t have the required clinical data about the patients. In order to help them some projects were created and INTCare (Gago et al., 2006; Manuel Filipe. Santos et al., 2011) is one of them. One of the main goals of INTCare is the outcome prediction in Intensive Care Units. In order to meet this objective, a new platform was developed that allows the clinical data collect in real-time and in electronic format. This data will used in a distributed data mining approach suited to grid computing environments based on a supervised learning classifier system.

Remaining sections of this paper are organized as follows: Section 2 gives the background details of the intensive care unit data and INTCare, section 3 describes the way of data acquisition from ICU and section 4 explains the global model construction methods. Section 5 shows the experimental set up and results of DDM and CDM. Section 6 discusses the performance of DDM vs. CDM. Further section 6 shows some related works and final section presents main conclusions.
Related Content

Patient-Doctor Interconnectivity: Improving Health Care Management and Patient Compliance with Web Technology

Rural E-Health Infrastructure Development

Web-Based Learning Environment for Medical Education: E-Fer, A Practical Tool for Diagnosis and Treatment of Chronic Wounds

A Risk-Based Classification of Mobile Applications in Healthcare