Abstract. This paper describes the MineFleet® distributed vehicle performance data stream mining system designed for commercial fleets. The MineFleet Onboard analyzes high throughput data streams onboard the vehicle, generates the analytics, and sends them to the remote server over the wireless networks. The paper describes the overall architecture of the system, business needs, and shares experience from successful large-scale commercial deployments. MineFleet is probably one of the first distributed data stream mining systems that is widely deployed at the commercial level. The paper discusses an important problem in the context of the MineFleet® application—computing and detecting changes in correlation matrices in a resource-constrained device that are typically used onboard the vehicle. The problem has immediate connection with many vehicle performance data stream analysis techniques such as principal component analysis, feature selection, and building predictive models for vehicle subsystems.

1 Introduction

The wireless and mobile computing/communication industry is producing a growing variety of devices that process different types of data using limited computing and storage resources with varying levels of connectivity to communication networks. The rich source of data from the ubiquitous components of businesses, mechanical devices, and our daily lives offers the exciting possibility of a new generation of data intensive applications for distributed and mobile environments. Mining distributed data streams in an ubiquitous environment is one such possibility. Several years of research on distributed data mining and data stream mining have produced a reasonably powerful collection of algorithms and system-architectures that can be used for developing several interesting classes of distributed applications for lightweight wireless applications. In fact an increasing number of such systems are being reported in the literature. Some commercial systems are also starting to appear.

This paper reports the development of MineFleet®, a novel mobile and distributed data mining application for monitoring vehicle data streams in real-time. MineFleet is designed for monitoring commercial vehicle fleets using
onboard embedded data stream mining systems and other remote modules connected through wireless networks in a distributed environment. Consider a nationwide grocery delivery system which operates a large fleet of trucks. Regular maintenance of the vehicles in such fleets is an important part of the supply chain management and normally commercial fleet management companies get the responsibility of maintaining the fleet. Fleet maintenance companies usually spend a good deal of time and labor in collecting vehicle performance data, studying the data offline, and estimating the condition of the vehicle primarily through manual efforts. Fleet management companies are also usually interested in studying the driving characteristics for a variety of reasons (e.g. policy enforcement, insurance). Monitoring fuel consumption, vehicle emissions, and identifying how vehicle parameters can be optimized to get better fuel economy are some additional reasons that support ample return of investment (ROI) for systems like MineFleet.

The MineFleet is widely adopted in the fleet management industry. Similar applications also arise in monitoring the health of airplanes and space vehicles. There is a strong need for real-time on-board monitoring and mining of data (e.g. flight systems performance data, weather data, radar data about other planes). The MineFleet system can also be applied to this aviation safety domain where it monitors planes and space vehicles instead of automobiles and trucks.

The main unique characteristics of the MineFleet system that distinguish it from traditional data mining systems are as follows:

1. Distributed mining of the multiple mobile data sources with little centralization of the data.
2. Onboard data stream management and mining using embedded computing devices.
3. Designed to pay careful attention to the following important resource constraints:
   (a) Minimize data communication over the wireless network.
   (b) Minimize onboard data storage and the footprint of the data stream mining software.
   (c) Process high throughput data streams using resource-constrained embedded computing environments.
4. Respect privacy constraints of the data, whenever necessary.

Section 2 presents an overview of the MineFleet system. Section 3 describes the need for monitoring the correlation matrices in MineFleet. Section 4 describes the computing problem. Sections 5 and 6 describe the algorithm. Section 7 reports some experimental results. Finally, Section 8 concludes this paper.

2 MineFleet®: An Overview

MineFleet® is a mobile and distributed data stream mining environment where the resource-constrained “small” computing devices need to perform various non-trivial data management and mining tasks on-board a vehicle in real-time.