Chapter 13
Sensor Field Resource Management for Sensor Network Data Mining

David J. Yates
Bentley University, USA

Jennifer Xu
Bentley University, USA

ABSTRACT

This research is motivated by data mining for wireless sensor network applications. The authors consider applications where data is acquired in real-time, and thus data mining is performed on live streams of data rather than on stored databases. One challenge in supporting such applications is that sensor node power is a precious resource that needs to be managed as such. To conserve energy in the sensor field, the authors propose and evaluate several approaches to acquiring, and then caching data in a sensor field data server. The authors show that for true real-time applications, for which response time dictates data quality, policies that emulate cache hits by computing and returning approximate values for sensor data yield a simultaneous quality improvement and cost saving. This “win-win” is because when data acquisition response time is sufficiently important, the decrease in resource consumption and increase in data quality achieved by using approximate values outweighs the negative impact on data accuracy due to the approximation. In contrast, when data accuracy drives quality, a linear trade-off between resource consumption and data accuracy emerges. The authors then identify caching and lookup policies for which the sensor field query rate is bounded when servicing an arbitrary workload of user queries. This upper bound is achieved by having multiple user queries share the cost of a sensor field query. Finally, the authors discuss the challenges facing sensor network data mining applications in terms of data collection, warehousing, and mining techniques.

DOI: 10.4018/978-1-60566-328-9.ch013
INTRODUCTION

Applications for mining sensor network data vary in scale from monitoring and controlling microscopic manufacturing equipment, to implementing an earthquake early warning system for a country like Japan. Sensor network data mining requires bridging the gap between low-level data that is acquired in sensor fields and high-level knowledge that is useful to real-world applications. Our research describes a new approach to sensor field resource management, which is necessary but not sufficient to bridge this gap.

There are many performance metrics of interest in sensor networks for data mining. We focus on two that are common to the vast majority of applications that mine sensor network data:

1. The accuracy of the data acquired by the mining application from the sensor networks; and
2. the total system end-to-end delay incurred in the sequence of operations needed for an application to obtain sensor data.

Although almost all sensor network applications have performance requirements that include accuracy and system delay, their relative importance may differ between applications. We therefore define the quality of the data provided to data mining applications to be a combination of accuracy and delay. Measuring the quality of sensor network data is important for data mining applications since the data quality determines the level of confidence in the knowledge extracted from the mining process. As in most systems, improved quality usually comes at some cost. For current wireless sensor networks, the most important component of cost typically is the energy consumed in providing the requested data. In turn this is dominated by the energy required to transport messages through the sensor field. This cost versus quality trade-off has recently been an active area of research (Boulis et al., 2003; Hu et al., 2006; Sharaf et al., 2004; Son et al., 2005; Tilak et al., 2002; Yu et al., 2004).

To perform our research, we construct a model for a system that acquires and mines sensor network data. We then develop novel policies for caching sensor network data values in sensor field gateway servers, and then retrieving these values via cache lookups. We also propose a new objective function for data quality that combines accuracy and delay. So that we can compare data quality across different data mining applications and systems, this objective function normalizes data quality for a given system to values in the range [0,1]. Finally, we use our system model to assess the impact of several factors on data quality and query cost performance:

- Our caching and lookup policies;
- the relative importance of data accuracy and system end-to-end delay; and
- the manner in which the sensed data values in the environment change.

This assessment evaluates seven different caching and lookup policies by implementing them in a simulator based on CSIM 19 (Schwetman, 1990, 2001).

The remainder of this chapter is organized as follows. In the next section we provide the background for sensor network data mining, outlining the three data management phases common to such applications and the resource requirements for each phase. This chapter’s main focus is on the first phase (sensor network data acquisition). In the subsequent section we describe our approach to acquiring and caching sensor network data. This approach addresses the problem of balancing data quality and the cost of acquiring sensor network data. We evaluate the performance of our approach and discuss our results, including their implications for sensor field resource management. We also identify the challenges of sensor network data mining, including showing how processing sensor data streams is fundamentally...