

Chapter 3.21

A Framework for Data Warehousing and Mining in Sensor Stream Application Domains

Nan Jiang
Cedarville University, USA

ABSTRACT

The recent advances in sensor technologies have made these small, tiny devices much cheaper and convenient to use in many different applications, for example, the weather and environmental monitoring applications, the hospital and factory operation sites, sensor devices on the traffic road and moving vehicles and so on. The data collected from sensors forms a sensor stream and is transferred to the server to perform data warehousing and mining tasks for the end user to perform data analysis. Several data preprocessing steps are necessary to enrich the data with domain information for the data warehousing and mining tasks in the sensor stream applications. This chapter presents a general framework for domain-driven mining of sensor stream applications. The proposed framework is able to enrich sensor

streams with additional domain information that meets the application requirements. Experimental studies of the proposed framework are performed on real data for two applications: a traffic management and an environmental monitoring site.

INTRODUCTION

Networks of thousands of sensors present a feasible and economic solution to some of our most challenging problems, such as real-time traffic modeling, weather and environmental monitoring, and military sensing and tracking. Recent advances in sensor technology have made possible the development of relatively low cost and low-energy-consumption micro sensors, which can be integrated in a wireless sensor network. These devices - Wireless Integrated Network Sensors (WINS) - will enable fundamental changes in applications spanning the home, office,

DOI: 10.4018/978-1-60566-816-1.ch006

clinic, factory, vehicle, metropolitan area, and the global environment.

Concerning the needs of the user for knowledge discovery from sensor streams in these application domains, new data warehousing, data mining techniques have to be developed to extract meaningful, useful and understandable patterns for the end users to perform data analysis. Many research projects have been conducted by different organizations regarding wireless sensor networks; however, few of them discuss the sensor stream processing infrastructure, and the data warehousing and data mining issues need to be addressed in the sensor network application domains. There is a need for new methodologies in order to extract interesting patterns in a sensor stream application domain. Since the semantics of sensor stream data is application dependent, the extraction of interesting, novel, and useful patterns from stream data applications becomes domain dependent.

Some data warehousing and data mining methods have been recently proposed to mine stream data, for example in (Manku 2002, Chang 2003, Li 2004, Yang 2004, Yu 2004, Dang 2007), the authors proposed algorithms to find frequent patterns over the entire history of data streams. In (Giannella 2003, Chang 2004, Lin 2005, Koh 2006, Mozafari 2008), the authors use different sliding window models to find recently frequent patterns in data streams. These algorithms focus on mining frequent patterns with one scan over the entire data stream.

In (Chi, 2004), Chi et al considers the problem of mining closed frequent itemsets over a data stream sliding window in the Moment algorithm, and in (Li, 2006), the authors proposed the New-Moment algorithm which uses a bit-sequence representation of items to reduce the time and memory needed. The CFI-Stream algorithm in (Jiang, 2006) directly computes the closed itemsets online and incrementally without the help of any support information. In (Li, 2008), Li et al proposed to improve the CFI-stream algorithm with

bitmap coding named CLIMB (Closed Itemset Mining with Bitmap) over data stream's sliding window to reduce the memory cost.

Besides pattern mining in data stream applications, as the number of data streaming applications grows, there is also an increasing need to perform association mining in data streams. One example application is to estimate missing data in sensor networks (Halatchev, 2005). Another example application is to predict frequency of Internet packet streams (Demaine, 2002). In the MAIDS project (Cai, 2004), an association mining technique is used to find alarming incidents from data streams. Association mining can also be applied to monitor manufacturing flows (Kargupta, 2004) to predict failures or generate reports based on accumulated web log streams. In (Yang, 2004), (Halatchev, 2005), and (Shin, 2007), the authors proposed using two, three, and multiple frequent pattern based methods to perform association rule mining.

In general, these approaches have focused on mining patterns and associations in data streams, without considering an application domain. As a consequence, these methods tend to discover general patterns, which for specific applications can be useless and uninteresting. Stream patterns are usually extracted based on the concept of pattern frequency. With no semantic or domain information, the discovered patterns cannot be applied directly to a specific domain.

In this book chapter, we present a data warehousing and mining framework where the users give to the data the semantics that is relevant for the application, and therefore the discovered patterns will refer to a specific domain. We will also discuss the issues needed to be considered in the data warehousing and mining components of this framework for sensor stream applications.

The remaining of the chapter is organized as follows: in the background section, we present some basic concepts about sensor data applications and the data warehousing and mining issues needed to be considered in stream data applications. Following with the background information,

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/framework-data-warehousing-mining-sensor/44111

Related Content

Citizens' Voice and Adoption of Pakistani E-Government Services

Muhammad Ovais Ahmad, Jouni Markkula and Markku Oivo (2015). *Business Technologies in Contemporary Organizations: Adoption, Assimilation, and Institutionalization* (pp. 246-262).
www.irma-international.org/chapter/citizens-voice-and-adoption-of-pakistani-e-government-services/120762

Methodological Issues in the Evaluation of System Analysis and Design Techniques

Andrew Gemino (2005). *Business Systems Analysis with Ontologies* (pp. 305-321).
www.irma-international.org/chapter/methodological-issues-evaluation-system-analysis/6127

The Management Conversation- It Still Needs to Make Sense

Stephen J. Andriole (2005). *The 2nd Digital Revolution* (pp. 143-216).
www.irma-international.org/chapter/management-conversation-still-needs-make/30280

Ontological Analysis of Business Systems Analysis Techniques: Experiences and Proposals for an Enhanced Methodology

Peter Green and Michael Rosemann (2005). *Business Systems Analysis with Ontologies* (pp. 1-27).
www.irma-international.org/chapter/ontological-analysis-business-systems-analysis/6117

A Tutorial on RDF with Jena

Wan-Yeung Wong, Tak-Pang Lau, Irwin King and Michael R. Lyu (2009). *Services and Business Computing Solutions with XML: Applications for Quality Management and Best Processes* (pp. 197-216).
www.irma-international.org/chapter/tutorial-rdf-jena/28976