Chapter 11
Query Optimisation for Data Mining in Peer-to-Peer Sensor Networks

Mark Roantree
Dublin City University, Ireland

Alan F. Smeaton
Dublin City University, Ireland

Noel E. O’Connor
Dublin City University, Ireland

Vincent Andrieu
Dublin City University, Ireland

Nicolas Legeay
Dublin City University, Ireland

Fabrice Camous
Dublin City University, Ireland

ABSTRACT
One of the more recent sources of large volumes of generated data is sensor devices, where dedicated sensing equipment is used to monitor events and happenings in a wide range of domains, including monitoring human biometrics and behaviour. This chapter proposes an approach and an implementation of semi-automated enrichment of raw sensor data, where the sensor data can come from a wide variety of sources. The authors extract semantics from the sensor data using their XSENSE processing architecture in a multi-stage analysis. The net result is that sensor data values are transformed into XML data so that well-established XML querying via XPath and similar techniques can be followed. The authors then propose to distribute the XML data on a peer-to-peer configuration and show, through simulations, what the computational costs of executing queries on this P2P network, will be. This approach is validated through the use of an array of sensor data readings taken from a range of biometric sensor devices, fitted to movie-watchers as they watched Hollywood movies. These readings were synchronised.

with video and audio analysis of the actual movies themselves, where we automatically detect movie highlights, which the authors try to correlate with observed human reactions. The XSENSE architecture is used to semantically enrich both the biometric sensor readings and the outputs of video analysis, into one large sensor database. This chapter thus presents and validates a scalable means of semi-automating the semantic enrichment of sensor data, thereby providing a means of large-scale sensor data management which is a necessary step in supporting data mining from sensor networks.

INTRODUCTION

We are currently witnessing a groundswell of interest in pervasive computing and ubiquitous sensing which strives to develop and deploy sensing technology all around us. We are also seeing the emergence of applications from environmental monitoring to ambient assisted living which leverage the data gathered and present us with useful applications. However, most of the developments in this area have been concerned with either developing the sensing technologies, or the infrastructure (middleware) to gather this data and the issues which have been addressed include power consumption on the devices, security of data transmission, networking challenges in gathering and storing the data, and fault tolerance in the event of network and/or device failure. If we assume these issues can be solved, or can at least be addressed successfully, we are then left to develop applications which are robust, scalable and flexible, and at such time the issues of efficient high-level querying of the gathered data becomes a major issue.

The problem we address in this chapter is how to manage, in an efficient and scalable way, and most importantly in a way that is flexible from an application developer or end user’s point of view, large volumes of sensed and gathered data. In this, we have a broad definition of sensor data and we include raw data values taken directly from sensor devices such as a heart rate monitor worn by a human, as well as derived data values such as the frame or time offsets of action sequences which appear in a movie. In the case of the former there would be little doubt that heart rate monitor readings are sensor values, whereas the latter still corresponds to data values, taken from a data stream, albeit with some intermediate processing (audio-visual analysis in this case). We now describe the motivation for our work.

Motivation and Contribution

To design a scalable system to manage sensor data, it is first necessary to enrich the data by adding structure and semantics in order to facilitate manipulation by query languages. Secondly, in order to improve efficiency, the architecture should be suitably generic to make it applicable to other domains. Specifically, it should not be necessary to redesign the system or write new program code when new sensor devices are added. Finally, when the number of sensor devices increases to very large numbers, the system should be capable of scaling accordingly.

The contribution of the research work reported here is the development of an architecture that is both generic, and has the capability to scale to very large numbers. In this respect, our XSENSE architecture facilitates the addition of new sensor devices by requiring that the knowledge worker or user provides only a short script with structural information regarding the sensor output. Scalability is provided in the form of a Peer-to-Peer (P2P) architecture that classifies sensors into clusters, but otherwise contains no upper limit on the numbers of sensors in the network.

The chapter is structured as follows: in §2, a description of sensor networks is provided and in particular the sensor network we use in our experiments, together with the issues involved
20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage: www.igi-global.com/chapter/query-optimisation-data-mining-peer/39548?camid=4v1


Related Content

Conceptual Modeling Solutions for the Data Warehouse

Using Data Mining for Forecasting Data Management Needs

Process-Based Data Mining

Diabetic Data Warehouses