Chapter 8.19
Seismological Data Warehousing and Mining:
A Survey

Gerasimos Marketos
University of Piraeus, Greece

Yannis Theodoridis
University of Piraeus, Greece

Ioannis S. Kalogeras
National Observatory of Athens, Greece

ABSTRACT

Earthquake data composes an ever increasing collection of earth science information for post-processing analysis. Earth scientists, local or national administration officers and so forth, are working with these data collections for scientific or planning purposes. In this article, we discuss the architecture of a so-called seismic data management and mining system (SDMMS) for quick and easy data collection, processing, and visualization. The SDMMS architecture includes, among others, a seismological database for efficient and effective querying and a seismological data warehouse for OLAP analysis and data mining. We provide template schemes for these two components as well as examples of their functionality towards the support of decision making. We also provide a comparative survey of existing operational or prototype SDMMS.

INTRODUCTION

For centuries, humans have been feeling, recording and studying earthquake phenomena. Taking into account that at least one earthquake of magnitude $M < 3$ ($M > 3$) occurs every one second (every ten minutes, respectively) worldwide, the seismic data collection is huge and rapidly

increasing. Scientists record this information in order to describe and study tectonic activity, which is described by recording attributes about geographic information (epicenter location and disaster areas), time of event, magnitude, depth, an so forth.

On the other hand, computer engineers specialized in the area of Information & Knowledge Management find an invaluable "data treasure", which they can process and analyze helping in the discovery of knowledge from this data. Recently, a number of applications for the management and analysis of seismological or, in general, geophysical data, have been proposed in the literature by Andrienko and Andrienko (1999), Kretschmer and Roccatagliata (2000), Theodoridis (2003), and Yu (2005). In general, the collaboration between the data mining community and physical scientists has been only recently launched (Behnke & Dobinson, 2000).

Desirable components of a so-called *seismic data management and mining system* (SDMMS) include tools for quick and easy data exploration and inspection, algorithms for generating historic profiles of specific geographic areas and time periods, techniques providing the association of seismic data with other geophysical parameters of interest, such as geological morphology, and top line visualization components using geographic and other thematic-oriented (e.g., topological and climatic) maps for the presentation of data to the user and supporting sophisticated user interaction.

In summary, we classify users that an SDMMS should support in three profiles:

- **Researchers of geophysical sciences**, interested in constructing and visualizing seismic profiles of certain regions during specific time periods or in discovering regions of similar seismic behavior.
- **Public administration officers**, requesting for information such as distances between epicenters and other demographical entities (schools, hospitals, heavy industries, etc.).
- **Citizens ("Web surfers")**, searching for seismic activity, thus querying the system for seismic properties of general interest, for example, for finding all epicenters of earthquakes in distance no more than 50Km from their favorite place.

The availability of systems following the proposed SDMMS architecture provides users a wealth of information about earthquakes assisting in awareness and understanding, two critical factors for decision making, either at individual or at administration level.

The rest of the article is organized as follows. Initially, we sketch a desired SDMMS architecture, including its database and data warehouse design. The section that follows, presents querying, online analytical processing (OLAP) and data mining functionality an SDMMS could offer, putting emphasis on the support of decision making. Furthermore, we survey and compare proposed systems and tools found in the literature for the management of seismological or, in general, earth science data. Conclusions are drawn in the last section.

THE ARCHITECTURE OF A SEISMIC DATA MANAGEMENT AND MINING SYSTEM

Earthquake phenomena are instantly recorded by a number of organizations (e.g., Institutes of Geodynamics and Schools of Physics) worldwide. The architecture of a SDMMS might allow for the integration of several remote sources. The aim is to collect and analyze the most accurate seismic data among different sources. Obviously, some sources provide data about the same earthquakes though with slight differences in their details (e.g., the magnitude or the exact timestamp of the recorded earthquake). SDMMS should be able to integrate the remote sources in a proper way by refining and homogenizing raw data.

Related Content

Data Mining in Diabetes Diagnosis and Detection
Indranil Bose (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 1817-1824).*
www.irma-international.org/chapter/data-mining-diabetes-diagnosis-detection/7734

Unsupervised Mining of Genes Classifying Leukemia
Diego Liberati, Sergio Bittantiand Simone Garatti (2005). *Encyclopedia of Data Warehousing and Mining (pp. 1155-1159).*
www.irma-international.org/chapter/unsupervised-mining-genes-classifying-leukemia/10771

Survival Analysis and Data Mining
Qiyang Chen, Alan Oppenheimand Dajin Wang (2005). *Encyclopedia of Data Warehousing and Mining (pp. 1077-1082).*
www.irma-international.org/chapter/survival-analysis-data-mining/10756

Material Acquisitions Using Discovery Informatics Approach
Chien-Hsing Wuand Tzai-Zang Lee (2005). *Encyclopedia of Data Warehousing and Mining (pp. 705-709).*
www.irma-international.org/chapter/material-acquisitions-using-discovery-informatics/10688

Classification Of 3G Mobile Phone Customers
Ankur Jain, Lalit Wangikar, Martin Ahrens, Ranjan Rao, Suddha Sattwa Kunduand Sutirtha Ghosh (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 2558-2565).*
www.irma-international.org/chapter/classification-mobile-phone-customers/7783