

Introduction to the Minitrack: Databases, Data Warehousing, and Data Mining in Health Care

Donald J. Berndt and Paul J. Hu
{dberndt, paulhu}@coba.usf.edu
Information Systems and Decision Sciences
College of Business Administration
University of South Florida
Tampa, Florida, USA

Chih-Ping Wei
cwei@mis.nsysu.edu.tw
Department of Information Management
National Sun Yat-Sen University
Kaohsiung, Taiwan, R.O.C.

The nature of healthcare services is essentially information-based and can be greatly improved with effective information support, including data modeling, archive, retrieval and analysis. In this light, database and data warehousing/mining technologies are crucial to healthcare organizations' services as well as individual professionals' practices. The contemporary Information Age can be characterized by rapid information expansion/creation, which has initiated and propelled an increasing shift to a knowledge-based society where information organization/retrieval and knowledge generation/discovery become increasingly challenging. In response, we organized the first Databases, Data Warehousing and Data Mining in Health Care Minitrack, which intended to serve as a presentation and discussion vehicle for sharing interesting recent research work among researchers and practitioners from both information systems and health care communities.

This year, we have accepted five papers from among eight submissions. The first paper, "Hospital Discharge Transactions: A Data Warehouse Component," by D. Berndt, A. Hevner, and J. Studnicki reported on the Comprehensive Assessment for Tracking Community Health (CATCH) project and the ongoing construction of a data warehouse to support the assessment methodology. The methodology has been applied by hand over the past several years in over a dozen counties, but is an expensive and time-consuming process. The CATCH data warehouse is intended to support automated reporting, as well as customized investigations. This paper described a transaction-oriented data warehouse component based on a large volume of hospital discharge data with a rich set of dimensions for query formulation.

The second paper, "Inductive Learning Approach to Intelligent Patient Image Pre-fetching: Extension and Evaluation of CN2 Algorithm," by C. Wei, P. Hu, O. Liu Sheng, and Y. Lee, extended the CN2 algorithm to support radiologists' patient prior image reference needs

in primary reading. The learning performance of the extended algorithm was compared with that achieved by a back propagation neural network designed for the same application. The comparative results were satisfactory, suggesting that the extended CN2 algorithm may be appropriate for constructing the learning subsystem of an intelligent patient image pre-fetching system.

The third paper, "Database Description for A Pan-European Healthcare System," by Kovacs et al., discussed a description-driven distributed data collection and management system for capturing data in an engineering data warehouse. The proposed system provides desirable flexibility and interoperability support for managing complex and vast amounts of information that needs to be shared among geographically dispersed healthcare organizations. The system also can provide an environment where appropriate data mining techniques can be applied to generate/discover knowledge from the data repositories.

The fourth paper, "Mining Time Dependency Patterns in Clinical pathways," by F. Lin, S. Chou, and S. Pan, addressed case variations in patient clinical pathway management. Specifically, the authors proposed a data mining technique for discovering time dependency patterns in clinical pathway management for brain stroke patients. Provided with such patterns, care providers may become increasingly effective in clinical pathway management for newly admitted patients.

The fifth paper, "Metadata Design for Chinese Medicine Digital Library Using XML," by C. Yang and W. Chan, presented a XML metadata design for Chinese Medicine, which has received increasing attention around the globe. The authors detailed their modeling design and described a prototype system for supporting search and retrieval of XML documents stored in distributed Chinese medicine digital libraries.