IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITP4867

Managing Large Healthcare Database for Decision Support System

Bambang Parmanto[1,2], Fei-Ran Guo.[1,2,3], Pamela J. Alan [1], Syarief A. Ahmad [1], and Christopher Lombard[1]
[1]Department of Health Information Management & [2]Center for Biomedical Informatics, University of Pittsburgh and [3]Department of Family Medicine, National Taiwan University, Taiwan, ROC.
Email: parmanto@pitt.edu

INTRODUCTION

Today, the healthcare industry is faced with many tasks including how to improve the quality while reducing costs. Healthcare organizations have accumulated enormous amounts of information related to patients and the medical care they receive. The success of relational database and its query language (SQL) has resulted in a huge amount of data being accumulated in the last thirty years. This type of database management system is latter known as the on-line transaction processing (OLTP). OLTP applications typically automate daily clerical data processing tasks that are the lifeblood of healthcare operations such as patient admission, billing, treatment, and outcome management. OLTP applications are traditionally managed by the relational database systems. While OLTP has been extremely successful in supporting daily operations, it is not designed for supporting fast and complex queries that are needed for data analysis and decision support system. Many of the databases reside in disparate information systems used to manage hospital affairs on a daily basis. These systems are not capable of storing large amounts of historical patient data nor are they built to support complex user queries. Without consolidating all of the information, the data is only of historical significance.

Data warehouse is the natural response to the needs for extracting information for decision support from the mountains of data. In contrast with OLTP, data warehouses are targeted for decision support. As such, aggregated and historical data are more important than detailed, individual transactions. To assist analysis and decision modeling, a data warehouse is designed to support intensive, complex, and ad-hoc queries. Query performance is more important than transactional performance. Since decision modeling requires complex analysis and visualization, the data in a data warehouse usually is modeled multidimensionally. In addition to multidimensional data model, OLAP operations also support rollup and drill-down, slice-and-dice, and pivot.

MANAGING LARGE HEALTHCARE DATABASE WITH OLAP

Data warehousing has been utilized and benefiting companies within the corporate world for the past ten years. Healthcare organizations can also benefit by developing and utilizing the information containing within a data warehouse. Much knowledge can be gained from the stored data such as identifying individuals at risk for targeted diseases, outcomes management, and decision support. A data warehouse integrates operational and historical data from many disparate systems across the organization and preserves it. By consolidating these data sources, users are able to utilize the information to make decisions about the organization and patients.

Center for Rehabilitation Service (CRS) of the University of Pittsburgh Medical Center is the leading provider of outpatient rehabilitation in the Greater Pittsburgh Metropolitan Area (Pennsylvania, USA). The center offers physical, occupational and speech therapies in over 40 clinics throughout Southwestern Pennsylvania.

Patients come to the clinics after being referred to by physicians based on diagnoses or disabilities. A therapist will be assigned to evaluate the patient's physical condition during the first visit and the therapist will be responsible for managing the course of the therapy (we will refer this course of therapy to as episode of care). A typical course of therapy consists of three to ten visits. To monitor the progress of the therapy, an outcome evaluation is conducted for every visit. Thirteen types of self-reported functional indicators are used to measure the physical condition of the patient, including ADLS, SF-36, Low back pain questionnaire, etc. When the goal of the treatment is achieved, the patient is discharged form the therapy.

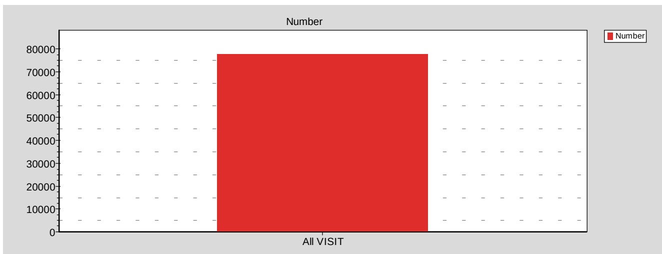
The center currently maintains two separate online transactional databases: claim and outcome. The claim database is primarily used for supporting insurance reimbursement and contains patient's identification, gender, age, diagnosis, treatment, the date of visit, the clinic, and therapist providing the services. The outcome database contains patient's identification, type of functional scores, the measurement of scores, body area with problem, and the date the measurement was reported. This database is used for quality control and for research. An information system that maintains the databases was implemented in 1997. When the decision support system was developed in 2001, a total of more than two million transactions of patient visits have been recorded from a total of more than 69,000 patients.

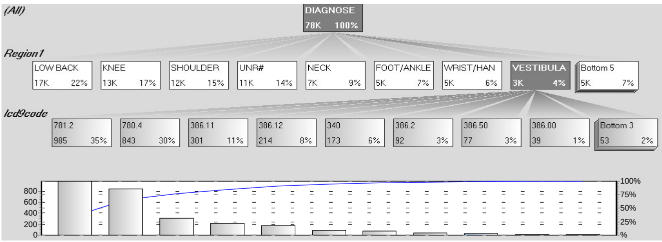
The goals of the decision support system includes predicting outcomes of patients with available variables, comparing the effectiveness of different strategies of treatment, and identifying clinics and therapists whose performances were different from the others. The decision makers at the center would like to compare the performance of therapists and centers, to see if there is underlining incompetence. The results could be used to identify the therapists who required enforced training and clinics whose quality of care should be improved. It is also crucial to know the optimal duration of a therapy and outcome of treatment, according to patients' demographic and clinic data. If the progress of patients could be estimated, therapists can discharge patients with optimal time and outcome to decrease the cost.

ANALYSIS AND CONCLUSION

Illustrations: Treatments and Outcomes Related to Vestibular

The barchart illustrates approximately 78,000 total number of episode of cares are stored in the database. These episodes of care are the results of visits from a total of 69,500 patients in the database. Approximately 8,500 episodes of care are repetitive cares from the same patient population.



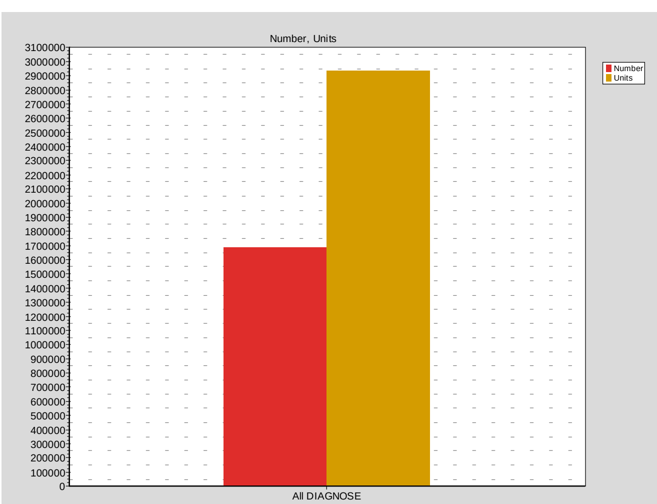


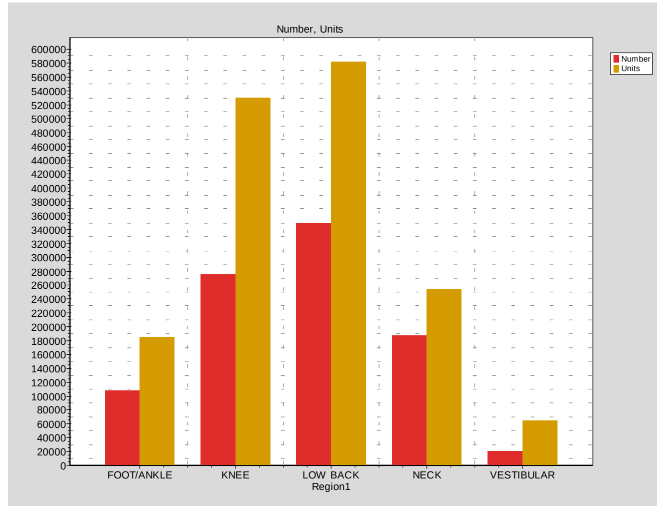
We will illustrate how this large database can be easily drilled down and sliced-and-diced to provide detailed analysis of a particular subject interest. For example, we would like to know the break down of all episode of care based on region of diagnoses. The above graph shows all the major regions of the diagnoses: low back (approximately 17,000 episode of care or 22% of the total database), followed by knee and shoulder. We would like to focus our attention on the vestibular area, which has more than 3,000 episodes of care or 4% of the cases.

The main function of the vestibular system is to keep track of the position and motion of a person's head in space. The vestibular system accomplishes three tasks. First, it contributes to a person's sense of equilibrium in relation to the force of gravity. This effects the person's sense of motion and spatial orientation. Second, inputs coming from the vestibular system transmit information to the body's muscles. Third, while a person's head and body are in motion, the vestibular system controls eye movements so that what the person is seeing remains steady and focused.

We can drill the database further down to get the detailed diagnoses of the vestibular area based on the ICD-9 code. About two-thirds of the diagnosis (65%) are from ICD9 code of 781.2 and 780.4. The fact that 35% of the vestibular diagnosis possibly resulting from head trauma came from ICD9 code 781.2 (abnormality of gait) makes sense. Whenever the vestibular system is affected, patients may have trouble walking because their equilibrium is off and the vestibular system is not conveying information correctly to the body's muscles. The next highest—30% of the vestibular diagnosis from ICD9 code 780.4 (dizziness and giddiness) also makes sense. When there is trauma to the vestibular system, some patients may describe a balance problem by saying they feel dizzy, light-headed, unsteady or giddy.

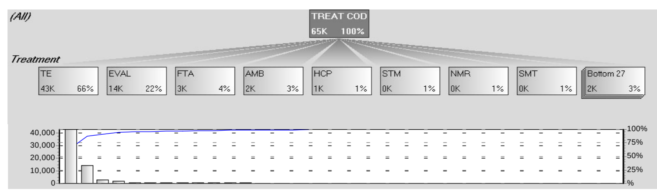
The database also maintains information related to the treatments given to patients during the episodes of care. The total number of treatments is 1,687,083, or approximately 21 treatments for each episode of care. One treatment can be applied more than one unit, bringing the total to 2,934,968 units of treatments.





We can drill down the treatment database based on the regions of diagnoses. The above barchart shows the makeup of regions for all the treatments.

We can drill down further by focusing on one of the body areas, for example the Vestibular area. The result is the decomposition tree that illustrates the types of treatments given for diagnoses related to Vestibular area.



The treatment category TE (therapeutic exercise) would include exercises that may focus on helping a patient develop their sense of balance, steadiness, etc. EVAL (evaluation with written report) is just an assessment of the patient. This would probably appear in the upper two or three treatments of every diagnosis.

We have demonstrated how a new database technologies, namely OLAP and data warehouse, can be used to manage large healthcare database. The data warehouse provides an easy graphical interface for information browsing that allows clinicians, researchers, and administrators to drill down, rollup, slice and dice, and to look at information from different dimensions. These types of information are currently obtained using manual system with the help of statistical packages. In much more sophisticated environments, statistical database has also been used. Therefore, comparisons between the two systems with OLAP will be briefly discussed.

Comparison of OLAP and Statistical Databases

A lot of relevant work has been done on both statistical databases and OLAP**.** In a comparison of OLAP and statistical databases, Shoshani (1997) showed the many differences and similarities between both concepts. Comparisons were made in respect to conceptual similarity, application areas, conceptual model structures, conceptual operators, physical organization, access methods, and privacy issues.

The conceptual structure of both OLAP and statistical databases have the same components which include summary measures, summary functions, one or more dimensions, and zero or more classification hierarchies. Statistical databases are typically used in socio-economic areas such as census and economic areas and databases that monitor natural resources. OLAP, on the other hand, is typically used in sales and stock market analysis and healthcare maintenance organizations. While both are used for different application areas, the problems they tackle, while similar, have a different emphasis. Statistical databases emphasizes complex multi-level classification structures; regional, spatial and geographic dimensions; privacy issues; and conceptual modeling. OLAP emphasizes simple structures; temporal dimensions and a greater efficiency of large datasets.

Another distinction is that while both statistical databases and OLAP contain base data (microdata) and summary data (macrodata), statistical databases usually present users the macrodata only, while OLAP presents users with both microdata and macrodata.

The operations that can be performed in both OLAP and statistical databases are very similar only differing in the terminology used to describe the operation. For example, a drill-down in OLAP is the same operation as an S-aggregation operation in statistical databases.

OLAP also places more emphasis on the physical organization of the database and the efficiency of accessing information from a large data set. As mentioned previously, statistical databases are more concerned with the summary data. OLAP, on the other hand, contains both the base data and the summarized data.

Why use OLAP over statistical software packages?

One benefit of utilizing OLAP over statistical software packages, is that OLAP provides user-friendly interfaces and involves no programming experience. Statistical software packages such as SPSS and SAS require training, statistical and programming experience to operate. Another benefit is OLAP runs on the server, thus providing immediate results to user queries and calculations. Statistical software packages run as client processes. This means that when the client requests data from the database on the server, the server downloads the data to the client to be processed or analyzed. This results in slow processing time because of inefficient computer resources. Not to mention the number of hours or even days it takes to get the information using current statistical packages that can be displayed in seconds in this OLAP systems.

REFERENCES

Han, J. and Kamber, M, (2001). *Data Mining: Concepts and Techniques*. Academic; 1993:237-257.

Gardner, S. R. (1998). Building the Data Warehouse, *Communications of the ACM*, 41(9), Sep. 1998, 52-60.

Shoshani, A. (1997) OLAP and Statistical Databases: Similarities and Differences. In *Proceedings of the Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 185—196, Tuscon, Arizona, May 1997. http://citeseer.nj.nec.com/shoshani97olap.html