International Journal of Research
Available at https://edupediapublications.org/journals

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

Data Mining Framework using provision for Community healthcare decision-making

K.Rajesh Khanna[1], Dr.D.Suresh Babu [2], Dr. Vaibhav Bansal[3]

[1] Research Scholar, Department of Computer science, OPJS University, Rajasthan.
[2]Professor & Head, Computer Science Department, Kakatiya Government College, Warangal.
[3]Associate Professor, Department of Computer science, OPJS University, Rajasthan.

ABSTRACT: Comprehensive Assessment for Tracking Community Health (CATCH) provides systematic methods for communitylevel assessment that is invaluable for resource allocation and health care policy formulation. CATCH is based on health status indicators from multiple data sources, using an innovative comparative framework and weighted evaluation process to produce arank-ordered list of critical community health care challenges. The community-level focus is intended to empower local decision makers by providing a clear methodology for organizing and interpreting relevant health care data. Extensive field experiencewith the CATCH methods, in combination with expertise in data warehousing technology, has led to an innovative application of information technology in the health care arena. The data warehouse allows a core set of reports to be produced at a reasonablecost for community use.

KEYWORDS-Health care information systems; Data warehousing; Data staging.

I. INTRODUCTION

Data Mining is one of the maximum essential and motivating location of research with thegoal of locating significant data from massive data sets. In gift technology, DataMining is turning into popular in healthcare discipline due to the fact there is a want of efficient analytical methodology for detecting unknown and valuable facts in fitness information.In fitness enterprise, Data Mining affords several advantages which include detection of the fraudin medical health insurance, availability of scientific way to the sufferers at decrease price,detection of reasons of diseases and identification of scientific treatment strategies. It additionally allows the healthcare researchers for making green healthcare guidelines, building drug advice systems, developing fitness profiles of individuals and so on. [1]. The data generated by way of the health businesses is very vast and complicated due to which it's fardifficult to analyze the statistics to be able to make vital decision regarding affected personhealth. This facts includes info concerning hospitals, sufferers, scientific claims,treatment price and so on. So, there's a need to generate a powerful device for analyzing and extracting crucial information from this complex records. The analysis of health information improves the healthcare by way of improving the overall performance of patient control obligations.The final results of Data Mining technology are to offer advantages to healthcare company for grouping the sufferers having comparable type of sicknesses or health troubles sothat healthcare organisation offers them powerful treatments. It also can useful forpredicting the duration of live of sufferers in health facility, for scientific prognosis and making plan for effective data device control. Recent technologies are used inmedical subject to beautify the scientific offerings in fee effective way. Data Miningtechniques also are used to research the various factors which are responsible for sicknessesas an example type of meals, unique operating surroundings, education degree, residingconditions, availability of pure water, fitness care services, cultural ,environmental andagricultural elements as proven in Figure 1.

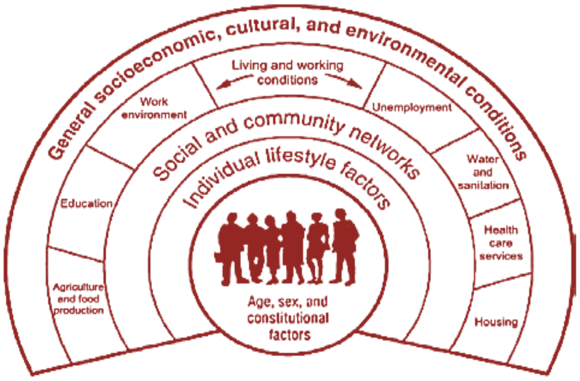
General socioeconomic, cultural, and environmental conditions
Living and working conditions
Work environment
Unemployment
Social and community networks
Individual lifestyle factors
Water and sanitation
Education
Health care services
Agriculture and food production
Housing
Age, sex, and constitutional factors

Figure 1. Factors Responsible for Diseases [2]

II. BACKGROUND WORKS

In present era various public and private healthcare institutes are producingenormous amounts of data which are difficult to handle. So, there is a need of powerful automated Data Mining tools for analysis and interpreting the useful information from this data. This information is very valuable for healthcare specialist to understand the cause of diseases and for providing better and cost effective treatment to patients. DataMining offers novel information regarding healthcare which in turn helpful for making administrative as well as medical decision such as estimation of medical staff, decision regarding health insurance policy, selection of treatments, disease prediction etc., [8-11]. Several studies identified with primary focus on various challenges and issues ofdata mining in healthcare [12, 13]. Data Mining are also used for both analysis andprediction of various diseases [14-23]. Some research work proposed an enhancementin available Data Mining methodology in order to improve the result [24-26] and somestudies develop new methodology [27, 28] and framework for healthcare system [29-33]. It is also found that various Data Mining techniques such as classification,clustering and association are used by healthcare organization to increase their capability for making decision regarding patient health.

III. PROPOSEDWORK

Fig. 2 shows a 2-by-2 comparison matrix based on state averages and peeraverages. Community indicators that demonstrateunfavorable comparisons on all dimensions are highlighted as community health challenges. After thiss imple comparison, the health care challenges areprioritized using a set of five filters.

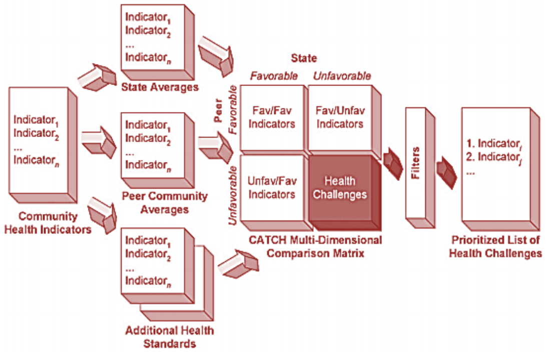

Fig. 2. The CATCH process.

Number Affected—number of persons in thecommunity affected by the indicator. Economic Impact—an estimate of the direct costper case for individuals affected by the indicator.Availability of Efficacious Intervention—an estimate of the relative degree to which treatment orprevention is likely to be effective. Magnitude of Difference—the degree to which the community indicator is worse than the dimensionalcomparisons.

Trend Analysis—for a 5-year period is the trendfavorable or unfavorable and what is the magnitude
of change in the trend direction?The community stakeholders are given an opportunity to weight the importance of each of the abovefactors. The final product of the CATCH methodologyis a comprehensive, prioritized listing of communityhealth care challenges.

The goals of the CATCH data warehouse includethe support and enhancement of the CATCH methods,the provision of cost-effective and thorough reports tocommunities, and the creation of a rich environmentfor more detailed research into critical health careissues. In addition, a focus on data quality makes thedata warehouse an especially valuable asset over timeas a rich and trustworthy historical repository is built.Lastly, the data warehouse lends itself to a variety ofdissemination strategies based on hardcopy reports,interactive access, and Web-enabled informationdelivery.

The different access technologies allow adiverse group of community planners and stakeholders to investigate important health care issues usingcomparable data. All of these characteristics make the CATCH data warehouse a unique application oftechnology in the field of public health. In fact, theimplementation of this type of data warehouse and its use in monitoring, as well as improving health status, will become a primary role of public health agenciesin the future.

The CATCH data warehouse includes a variety ofcomponents arranged in three broad categories: reporting tables for direct support of the CATCHmethods, aggregated dimensional structures, andfine-grained or transaction-oriented dimensional structures. In the sections that follow, examples of thesedata warehouse components are presented. All of thecomponents draw on the dimensional model or starschema, some components with more than a dozendimensions and some with a few simple dimensions.

A. The dimensional model

Important missions of a data warehouse include thesupport of decision-making activities and the creation of an infrastructure for ad-hoc exploration of very large collections of data. Decision-makers should beable to pursue many of their investigations using browsing tools, without relying on database programmers to construct queries. The emphasis on end-userdata access places a premium on an understandable database design that provides an intuitive basis for navigating through the data. The star schema ordimensional model has been recognized as an effective structure for organizing many data warehouse components [12,15,19]. The star schema is characterized by a center fact table, which usually contains numeric information that can be used in summaryreports. Radiating from the fact table are dimension tables that provide a rich query environment. This structure provides a logical data cube, with dimensions such as time and location identifying a set of numeric measurements within the cube. Fig. 3 contains a fragment from the hospital discharge transaction-oriented star schema discussed in this paper.

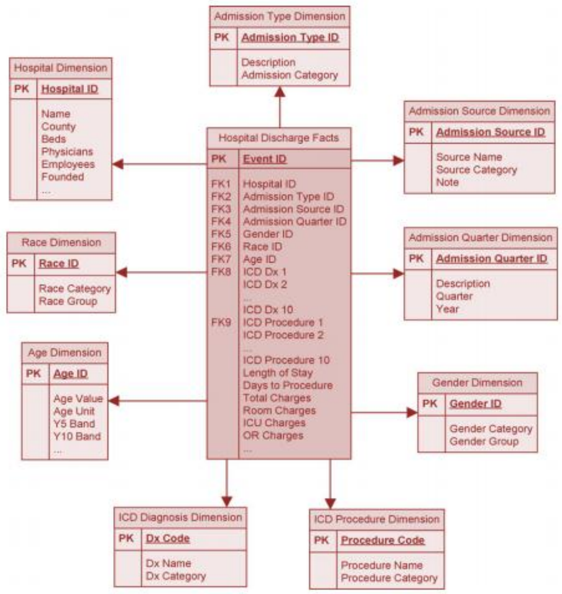

Fig. 3. Hospital discharge star schema

B. Fact tables

The most appropriate facts are additive numeric dataitems that can be summed, averaged, or combined inother ways across the dimensions to form summary statistics. The only way to compress the millions of data points and produce a reasonably sized answer set is to present some mathematical summarization. No humanwill want thousands, let alone millions, of items inanswer to their queries. As Kimball [19] points out, ‘‘the best and most useful facts are numeric, continuouslyvalued, and additive.’’ The CATCH data warehouse includes facts such as counts of hundreds of different health events, population-based rates, age-adjustedrates, and even fine-grained financial data in the caseof the hospital discharge data depicted in Fig. 3. Forexample, using the hospital discharge star it is possibleto focus on a single hospital (using the hospital dimension), select a single disease (using the ICD DIAGNOSISdimension), and investigate how the length of stay hasvaried over a specified time period. Using the hierarchical nature of the dimensions, it is also possible to ‘rollup’ to compare types of hospitals, disease categories,or even patient age bands. While the dimensionalstructure is simple and readily understandable, it supports a large and very useful universe of queries.

C. Dimension tables

The dimensions define the query environment, thericher the set of dimensions the more ways the data can be accessed via queries. Two of the important characteristics of dimensions are the richness of the attributes that describe the dimension and the hierarchical nature of the dimension. For example, the COUNTY dimension in the CATCH data warehouseincludes attributes that describe whether a county iscoastal, wealthy, urban, dense, large in area, or includes a military base. Therefore, the counties can be organized by any value in this attribute set. Some of the attributes lend themselves to hierarchical organization. In the case of COUNTY, there is natural geographic hierarchy that includes groups of counties that form regions within the state and the state itself.The county is also composed of finer geographic units. such as communities, ZIP codes, and census tracts. The dimension hierarchies enable roll-up and drill-down operations that control the level of detail in queries.These formally defined hierarchies also provide theframework for navigation or data browsing.In order to describe the dimension hierarchies succinctly to both end-users and developers, dimension hierarchy diagrams have been utilized in the CATCHdata warehouse design process. These diagrams show the hierarchical nature so that end-users have an uncluttered view of how they can navigate and designers can easily understand the dimensional structures.Fig. 4 illustrates an important health care dimension based on the International Classification of Disease(ICD) codes.

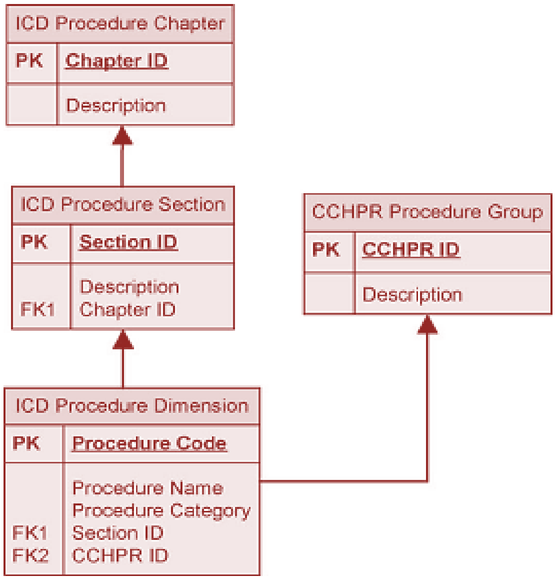

Fig. 4. ICD PROCEDURE dimension hierarchy

Inorder to meet these goals, the data warehouse designincludes several levels of data granularity, from the coarse-grained data used in generic report production to actual event-level data, such as hospital discharges.The data warehouse design includes major components at all three levels of granularity as illustrated inthe data access pyramid found in Fig. 5.Report indicators—Reporting tables with derivedor highly aggregated data are used to support the coreCATCH reports, including comparisons between at arget county and peer counties. These tables also provide fast response for interactive access via data browsing tools and can provide the foundation for simple community-wide Internet access. In addition,the metadata play an important role at the reporting level, providing indicator definitions, state or federalgoals, and expert domain knowledge for priorityfilters (e.g., economic impact and treatment availability). This report level of the data warehouse may notbe needed in all data warehouse applications but provides important support for rapid generation ofcommunity CATCH reports.

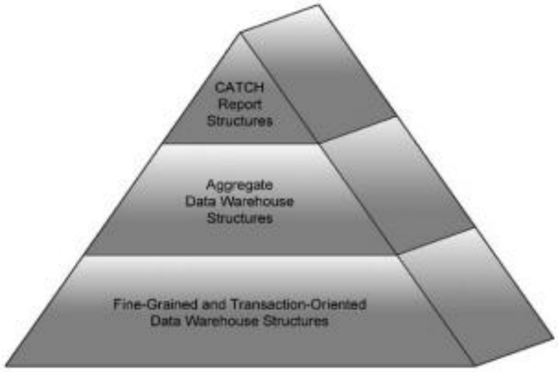
CATCH Report Structures
Aggregate Data Warehouse Structures
Fine-Grained and Transaction-Oriented Data Warehouse Structures

Fig. 5. Data access pyramid.

Aggregate data—There are families of star schemas that provide true dimensional data warehousecapabilities, such as interactive roll-up and drill-down operations. These components have carefully designed dimensions that can be utilized by more sophisticated data browsing tools. The star schemas are populate dusing thorough data staging and quality procedures that usually involve processing detailed data sets extracted by various health care agencies and organizations.Typically, the data are aggregated and transformed for loading into a family of related star schemas—a constellation—that share important dimensions and support interactive online analytic processing (OLAP)techniques.

Transaction data—For certain types of information, the design calls for retaining very fine-grained oreven event level data. An example is the hospitaldischarge data that includes each hospital dischargeevent for the more than 200 hospitals that are mandated to report such information in Florida. These dataare retained at the transaction level because of the richset of facts and dimensions available for analysis andthe density of potential aggregations that result innegligible space savings.

IV. CONCLUSION

In this paper, we have described some of thetechnical challenges faced in designing and implementing a data warehouse for health care information. Wehave presented innovative research contributions in theareas of data warehouse design, data staging for ETLprocessing, data quality assurance, and health care datawarehouse applications.The CATCH data warehouse is now fully functional. For example, it has been recently used toproduce a comprehensive CATCH report for Miami–Dade COUNTY, Florida's largest county.

REFERENCES

[1] H. C. Koh and G. Tan, "Data Mining Application in Healthcare", Journal of Healthcare InformationManagement, vol. 19, no. 2, (2005).

[2] R. Kandwal, P. K. Garg and R. D. Garg, "Health GIS and HIV/AIDS studies: Perspective and retrospective",Journal of Biomedical Informatics, vol. 42, (2009), pp. 748-755.

[3] D. Berndt, R. Satterfield, Customer and household matching:resolving entity identity in data warehouses, Proceedings ofAeroSense 2000, Conference on Data Mining and KnowledgeDiscovery, Orlando (April 2000).

[4] Center for Disease Control and Prevention, Principles of Community Engagement, 1997, Atlanta.

[5] D. Chrislip, C. Larson, Collaborative Leadership: How Citizens and Civic Leaders Can Make a Difference, Jossey-Bass,San Francisco, 1994.

[6] M. Corey, M. Abbey, Oracle Data Warehousing, Oracle Pressand Osborne McGraw-Hill, New York, 1997.

[7] S. Cropper, Collaborative working and the issue of sustainability, in: C. Huxham (Ed.), Creating Collaborative Advantage,SAGE Publishers, London, 1996.

[8] A. Dennis, Information exchange and use in group decisionmaking: you can lead a group to information but you can'tmake it think, MIS Quarterly 20 (4) (1996) 433– 458.

[9] A. Dennis, K. Hilmer, N. Taylor, Information exchange anduse in GSS and verbal group decision making, Journal of MIS14 (3) (1998) 61– 88.

[10] D. Dey, S. Sarkar, P. De, A probabilistic decision model forentity matching in heterogeneous databases, Management Science 44 (10) (October 1998) 1379–1396.

[11] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy(Eds.), Advances in Knowledge Discovery and Data Mining,The AAAI Press, Menlo Park, CA, 1996.

[12] P. Gray, H. Watson, Decision Support in the Data Warehouse,Prentice-Hall, Englewood Cliffs, NJ, 1998.

[13] J. Han, M. Kamber, Data Mining: Concepts and Techniques,Morgan Kaufmann Publishers, San Francisco, 2001.

[14] V. Harinarayan, A. Rajaraman, J. Ullman, Implementing datacubes efficiently, Proceedings of the 1996 ACM SIGMOD,Montreal (June 1996).

[15] W. Inmon, Building the Data Warehouse, Wiley, New York,1992.

[16] Institute of Medicine, Summary of recommendations, in: W.Waterfall (Ed.), The Future of Public Health, National Academy Press, Washington, DC, 1988.

[17] Institute of Medicine, Healthy Communities: New Partnerships for the Future of Public Health, National Academy Press,Washington, DC, 1996.

[18] Institute of Medicine, Measurement tools for a communityhealth improvement process, in: J. Durch, L. Bailey, M. Stoto(Eds.), Improving Health in the Community, a Role for Performance Monitoring, National Academy Press, Washington,DC, 1997.

[19] R. Kimball, The Data Warehouse Toolkit, Wiley, New York,1996.

[20] H. Mintzberg, The Nature of Managerial Work, Harper andRow, New York, 1973.

[21] H. Nakajima, Editorial: new players for a new era, WorldHealth 50 (3) (1997) 3.

[22] J. Srivastava, P. Chen, Warehouse creation—a potential roadblock to data warehousing, IEEE Transactions on Knowledgeand Data Engineering 11 (1) (1999) 118–126

[23] S. Gupta, D. Kumar and A. Sharma, “Data Mining Classification Techniques Applied For Breast CancerDiagnosis And Prognosis”, (2011).

[24] K. S. Kavitha, K. V. Ramakrishnan and M. K. Singh, “Modeling and design of evolutionary neural networkfor heart disease detection”, IJCSI International Journal of Computer Science Issues, ISSN (Online): 1694-0814, vol. 7, no. 5, (2010) September, pp. 272-283.

[25] S. H. Ha and S. H. Joo, “A Hybrid Data Mining Method for the Medical Classification of Chest Pain”,International Journal of Computer and Information Engineering, vol. 4, no. 1, (2010), pp. 33-38.