

Datamining and Telemedicine Challenges and Opportunities

Dr Jayanth G Paraki¹,

Founder – Omega Associates 123,
Malleshwaram, Bangalore 560055
jparaki@gmail.com

Abstract. Of all the many interesting things possible to do on the Internet there is none as exciting and rewarding as writing a 50,000 word novel titled 'I want to Buy Google' for NanoWrimo. It is even more significant and relevant when it is connected with datamining and research to discover the etiology of an elusive disorder known to the medical fraternity as Autism. Research into the etiology, pathogenesis and treatment of Autism requires collection of huge amounts of data from many research centers across the globe. Data mining is ready for application in the healthcare sector because it is supported by three technologies that are now sufficiently mature: Massive data collection, Powerful multiprocessor computers and Data mining algorithms.

Keywords: Google, Datamining, Information Retrieval, Telemedicine, KM

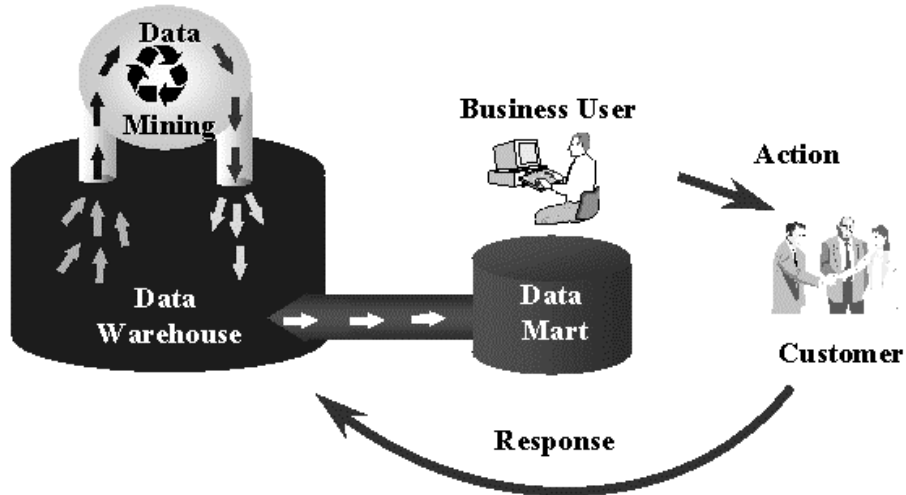
1 Introduction to Datamining

Datamining [5][6] describes a collection of techniques that aim to find useful but undiscovered patterns in collected data. The goal of data mining is to create models for decision making that predicts future behavior based on analyses of past activity. Data mining supports knowledge discovery, defined by William Frawley and Gregory Piatetsky-Shapiro [1][4] as "...the nontrivial extraction of implicit, previously unknown, and potentially useful information from data..." This definition holds the key to the substantial interest in data warehousing and data mining. The method used today in data mining, when it is well thought out and well executed, consists of just a few very important concepts. The first of these is the concept of finding a pattern in the data. In many cases, this just means any collection of data that occurs a surprising number of times. Usually "surprising" is better defined, but in general it means any sequence or pattern of data that occurs more often than one would expect it to if it were a random event. Model deployment is one of the main objectives of the data mining process. Datamining is integral to deployment of Telemedicine projects [3]. A suitable model for deployment of R&D in Autism and employing Telemedicine requires computational power to analyse billion or more data sets on a regular basis month after month from several geographical locations adding immense complexity

to the task and needing involvement of several computers to perform the computing and derive accurate results. This is both a challenge and an opportunity for scientists.

2 Data Mining: The Global Context

Data Mining Process



Data Mining tools [2] can enhance inference process and speed up design cycle but it cannot substitute for statistical and domain expertise. Data mining is given more importance in today's knowledge-driven world as the commercial organisations are increasingly looking to transform the data into business intelligence [5], thus gaining an edge over the competitors in the market. Data mining, the process of extracting patterns from large data sets by combining methods from statistics and artificial intelligence with database management, is primarily used by companies with a strong consumer focus such as those in the retail, financial, communication, and marketing areas. It is also used in scientific discovery, and fraud detection. Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Comprehensive data warehouses that integrate operational data with patient, clinician, and hospital information have resulted in an explosion of medical information. Correct diagnosis requires timely and sophisticated analysis on an integrated view of the patient's medical history and the hospital information system. However, there is a growing gap between more powerful storage and retrieval systems and the clinicians' ability to effectively analyze and act on the information they contain. Both relational and OLAP technologies have tremendous capabilities for navigating massive data warehouses, but brute force navigation of data is not enough. A new technological leap is needed to structure and prioritize information for specific patient problems. The data mining tools can make this leap. Quantifiable business

benefits have been proven through the integration of data mining with current hospital information systems, and new products are on the horizon that will bring this integration to an even wider audience of users.

3 The Big Data Challenge

Data is necessary and invaluable to all major business enterprises- aerospace, banking, retail, real estate, healthcare and entertainment industries. Business enterprises vary in the manner they employ and use data for decision making. Datamining tools have various functions and the end-user employs them with care and diligence. Some of the important categories where datamining is prominent and profitable are:

1. Healthcare
2. Finance
3. Retail industry
4. Telecommunication
5. Text Mining & Web Mining
6. Higher Education

Data mining algorithms act on numerical and categorical data stored in relational databases or spreadsheets. What if you want to mine data items that are not numericals or categorical? There are many examples: web pages, document libraries, PowerPoint presentations, product specifications, emails, sound files, and digital images [7] to name a few. What if you want to mine the information stored in long character strings, such as product descriptions, comment fields in reports, or call center notes? Data that cannot be meaningfully interpreted as numerical or categorical is considered unstructured for purposes of data mining. It has been estimated that as much as 85% of enterprise data falls into this category. Extracting meaningful information from this unstructured data can be critical to the success of a business.

4 Example of Text Mining

The following is an example of text mining with Tanagra 1.4 .40 on a sample text containing glossary of terms in Telemedicine. Tanagra is a simple tool for academic and research purposes [2].

Dataset (Germany.txt)

Parameters

Database : D:\Documents and Settings\Jay\My Documents\Germany.txt

Results

Download information

Datasource processing

Computation time141 ms

Allocated memory 8 KB

Dataset description

1 attribute(s)

63 example(s)

AttributeCategoryInformations

GERMANY Discrete 52 values

Driven by aging population, increased medical requirements in remote locations and technology advancements, world market for telemedicine is projected to exceed \$18 billion by the year 2015. Application of telemedicine is expanding virtually across all the medical areas. Telemedicine currently finds application in majority of medical domains including radiology, cardiology, dermatology, psychiatry, dentistry, pediatrics and pathology, among others. Telemedicine has its early beginning's from NASA and is now existing virtually in every advanced nation in some form or other.

Acknowledgments. Ricco RAKOTOMALALA, "TANAGRA: a free software for research and academic purposes", in Proceedings of EGC'2005, RNTI-E-3, vol. 2, pp.697-702, 2005

References

1. Gregory Piatetsky-Shapiro, William J. Frawley (Eds.): Knowledge Discovery in Databases. AAAI/MIT Press 1991, ISBN 0-262-62080-4
2. Ricco RAKOTOMALALA, "TANAGRA: a free software for research and academic purposes", in Proceedings of EGC'2005, RNTI-E-3, vol. 2, pp.697-702, 2005
3. Paraki JG: Knowledge Management in Telemedicine. Handbook of Research in Informatics and Biomedicine. In: Athina Lazakidou. Idea Group Inc, 600 pages, ISBN=1-59140-982-9 (2006)
4. Jain, A.K. Duin, P.W. and Jianchang, Mao. Statistical Pattern Recognition: A Review. Pattern Analysis and Machine Intelligence, IEEE Transactions. January 2000
5. An Introduction to Data Mining. Pilot Software Whitepaper. Pilot Software. 1998.
6. Piatetsky-Shapiro, Gregory. The Data-Mining Industry Coming of Age. IEEE Intelligent Systems. 2000.
7. Manjunath, B.S. Huang, T. Teklap, A.M. and Zhang, H.J. Guest Editorial: Introduction to the Special Issue on Image and Video processing for Digital Libraries. Image Processing, IEEE Transactions. Volume: 9 Issue: 1. Jan. 2000