IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

Data Mining and Analytics: A Proactive Model

Vishesh S[1], Manu Srinath[1], Vivek Adithya[2], Sumit Kumar Sourav[2], Ashwani Kumar Singh[3], Vyshnav L[4]

B.E, Department of Telecommunication Engineering, BNM Institute of Technology, Bangalore, India [1]

Student, Department of Computer Science Engineering, BNM Institute of Technology, Bangalore, India [2]

B.E, Department of Computer Science Engineering, VTU, Belgaum, India [3]

B.E, VTU, Belgaum, India [4]

Abstract: Data mining is the automated extraction of hidden predictive data from huge databases. It is a prospective approach rather than the traditional retrospective approach followed in many other applications. The main goal of data mining is simplification and automation of overall statistical process, from data source(s) to model application. In this paper, we present to you the importance, goals, applications, advantages and techniques of data mining. We also draw limelight on the improvement on the three major technologies which have enabled data mining to become more apt and productive in the present day scenario.

Keywords: Data mining, hidden predictive data, automation, techniques of data mining.

I. INTRODUCTION

There is a huge amount of data available in the information industry. This data is of no use until it is converted into useful information. Data mining is defined as extracting information or knowledge from data following a prospective approach rather than the traditional retrospective approach. [1]

Data mining also involves other processes such as data cleaning, data integration, data transformation, pattern evaluation and data presentation.

Data mining is very useful in the following domains: [2]

- Market analysis
- Fraud detection
- Customer relation
- Production control
- Science exploration

Data mining system can be classified into the following criteria: [3]

- Database technology
- Statistics
- Machine learning
- Information science
- Visualisation and other disciplines

Apart from these, a data mining system can also be classified based on the kind of

- Databases mined
- Knowledge mined
- Technique utilized
- Application adapted

Figure 1 shows classification of data mining.

II. DATA MINING – THE CONVERGENCE OF THREE TECHNOLOGIES

The improvements of the three major technologies which have enabled data mining to become more apt and productive in the present day scenario are:

- Increasing computing power
- Improved data collection and management
- Statistical and learning algorithms

In fact, data mining stands undefeated at present due to these three technologies and advancement in the same. Figure 2 shows the convergence of the three technologies to enable data mining.

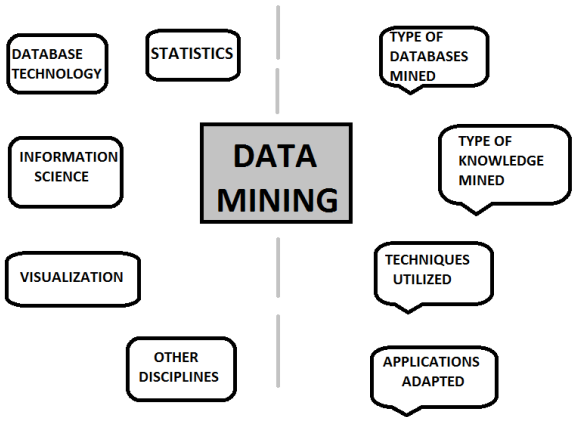

Figure 1 Classification of Data Mining

A. Increasing Computing Power

Data mining is the automated extraction and analysis of hidden predictive data and converting that into useful information. The amount of information to be processed in the current world is humongous and a very tedious task. This requires machines or systems or processors with very high computational capability.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

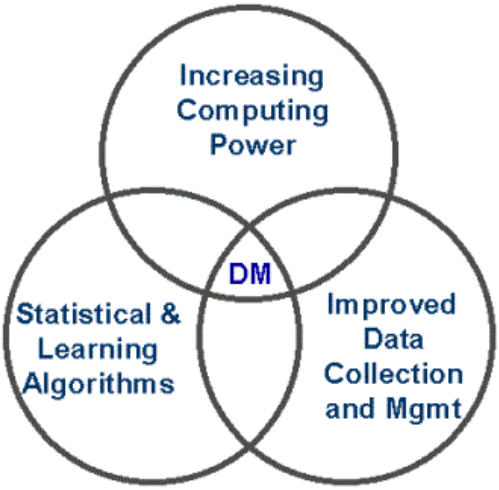

Figure 2 Convergence of three technologies to enable data mining

Hence data mining requires an extremely small, very large scale integration of components of an integrated circuit.
According to the Moore's Law, the number of transistors in a dense integrated circuit doubles approximately every two years. Although the Moore's prediction [4] has deviated from the predicted rate of integration, but with parallel processing, and multi core processing environment has kept Moore's law running [5] (not as in the previous decades).

B. Improved Data Collection and Management
Well chosen and well implemented methods for data collection and analysis are essential for data mining and analytics.
Before decisions are made about what data to collect and how to analyse them, the purpose of evaluation of must be decided. Good data management includes developing effective processes for consistently collecting and recording data, storing data security, cleaning data, transferring data, effectively presenting data and making data accessible for verification and use by others. [6]

C. Statistical and Machine Learning Algorithms
Techniques have often been waiting for computing technology to catch up. Machine Learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed. Within the field of data analytics, machine learning is a method used to design complex models and algorithms that land themselves to prediction. These analytical models allow researchers, data scientists, engineers and analysts to "produce reliable, repeatable decisions and results".

III. REAL TIME SCENARIO

Consider a corporate company like Konigtronics Private Limited running business in the 21st century market. Konigtronics is involved in data analytics and is very keen in learning "which clients are most likely to respond to the next promotional email, and why?" This can be done by extraction of hidden predictive information from the database of the company. This will allow Konigtronics to predict the future trends and behaviours, allowing businesses to make proactive, knowledge driven decisions.

Figure 3 shows the company's efforts in data analytics and its proactive responses to the mined and analysed data.
The company can also employ statisticians or data scientists to do the same. This would be a better idea in case of the data exchanged being very small. Another solution to this is to make the system semi-automatic or outsourcing. But outsourcing has its set of disadvantages when it comes to security and confidentiality. A proactive response is derived which is prospective rather than retrospective.

For example, let there be 26 clients A, B, C,..., Z. Clients B, C and D have already adapted their technology to the latest release by Konigtronics. But, A and F have responded to all the latest technology updates by Konigtronics till date and not yet responded to the technology update that B, C and D have updated.
The probability that A and F will respond to the technology update notification sent to them is very high when compared to E, G,...., Z. Sending an update notification to B, C and D is not prescribed because they have responded to the auto-update. A proactive task must be framed to attract the clients E, G, H,..., Z to respond to the update.

Most likely to respond: A & F
No need to send notification: B, C & D
Lesser probability of updating the technology: E, G, H,...,Z.

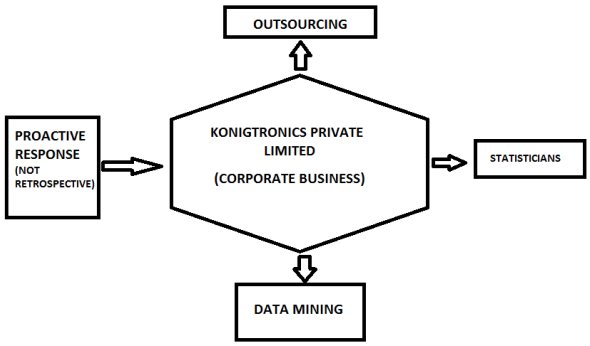

Figure 3Company's efforts in data analytics and its proactive responses to the mined and analysed data

IV. DATA MINING IN GENETICS

Modern biomedical research is uncovering the pathology of diseases once considered to be hopelessly complex and incurable. A great deal of this can be attributed to gene mapping i.e. localisation of disease susceptible genes to certain areas of the human genome by a combination of state of the art laboratories and computation methods.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

From a data mining or machine learning point of view, gene mapping can be seen as a classification problem.

V. CONCLUSION

Data mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands on analysis can now be answered directly from the data. Data mining tools sweep through databases and identify previously hidden patterns in one step. Data mining techniques can yield the benefits of automation on existing hardware and software platforms and can be implemented on new systems as existing platforms are updated and new products are developed.

REFERENCES

[1] Linking innovative product development with customer knowledge: a data-mining approach - CT Su, YH Chen, DY Sha - Technovation, 2006 – Elsevier
[2] 14 useful applications of data mining - Big Data Made Simple - bigdata-madesimple.com/14-useful-applications-of-data-mining/
[3] Data Mining: Practical machine learning tools and techniques - IH Witten, E Frank, MA Hall, CJ Pal - 2016 - books.google.com
[4] 50 Years of Moore's Law - Intel - www.intel.com/content/www/us/en/silicon-innovations/moores-law-technology.html
[5] Moore's Law and Multicore - College of Engineering | Oregon State ..web.engr.oregonstate.edu/~mjb/cs475/Handouts/moores.law.and.multicore.2pp.pdf
[6] Guidance on good data and record management practices - www.who.int/.../Guidance-on-good-data-management-practices_QAS15-624_160920...

BIOGRAPHIES

Vishesh S born on 13th June 1992, hails from Bangalore (Karnataka) and has completed B.E in Telecommunication Engineering from VTU, Belgaum, Karnataka in 2015. He also worked as an intern under Dr Shivananju BN, Department of Instrumentation, IISc, Bangalore. His research interests include Embedded Systems, Wireless Communication and Medical Electronics. He is also the Founder and Managing Director of the company Konigtronics Private Limited. He has guided over a hundred students/lecturers/interns/professionals in their research works and projects. He is also the co-author of many International Research Papers. Presently Konigtronics Private Limited has extended its services in the field of Real Estate, Webpage Designing and Entrepreneurship.

Manu Srinath hails from Bangalore (Karnataka); he has completed B.E in Telecommunication Engineering from VTU, Belgaum, Karnataka. His research interests include networking, image processing and cryptography. He is the Executive Officer at the company Konigtronics (OPC) Pvt. Ltd.

Vivek Adithya hails from Bangalore (Karnataka). He is currently pursuing B.E in Computer Science Engineering at BNM Institute of Technology. His research interests include Cyber Security, Machine Learning, Artificial Intelligence and Cloud Computing.

Sumit Kumar Sourav hails from Bangalore (Karnataka). He is currently pursuing B.E in Computer Science and Engineering at BNM Institute of Technology. His research interests include Data mining, Cognitive science, Computer networking.

Ashwani Kumar Singh hails from Bangalore (Karnataka); he has completed B.E in computer Science and Engineering from VTU, Belgaum. His research interests include Data mining, Data analytics, Data science, Data compression.

Vyshnav L hails from Bangalore, Karnataka. He has completed B.E from VTU, Belgaum. His research interests include Software Engineering and Data Analytics.