Artificial Intelligence and Cognitive Analytics Approaches towards Efficient Predictions for Business Intelligence

Salil Kanetkar
B.E. Computer Engineering
Computer Engineering Department,
VES Institute of Technology, Mumbai-74
India

Neha Chanchlani
B.E. Computer Engineering
Computer Engineering Department,
VES Institute of Technology, Mumbai-74
India

ABSTRACT

Business Intelligence techniques no longer just evolve around traditional databases and warehouses, but have leaped into the era of big data. The aim is now to extract as much hidden knowledge from this rich data. The data is highly varying and evolves from various sources. Business improvement strategies usually revolved just around board room discussions; however these discussions are now more often aided with mining strategies. Machine learning, Artificial Intelligence techniques are of great advantage in predicting efficient strategies and business patterns. Apart from this, cognitive analytics and natural language processing are such techniques which could drastically change how business intelligence has been perceived. They try to emulate the human mind and train the computer to think accordingly. In this paper, we highlight the above listed methods in detail.

General Terms

Artificial Intelligence, Machine Learning, Cognitive Computing, Natural Language Processing, Business Intelligence.

1. INTRODUCTION

The term Business Intelligence is an integral part of Management Information Systems (MIS) literature [2]. It is one the managerial weapons which contributes to the overall progress and development of the organization. No doubt that BI strategic and analytical importance for an organization, but over the period of time its methodology and techniques have varied largely.
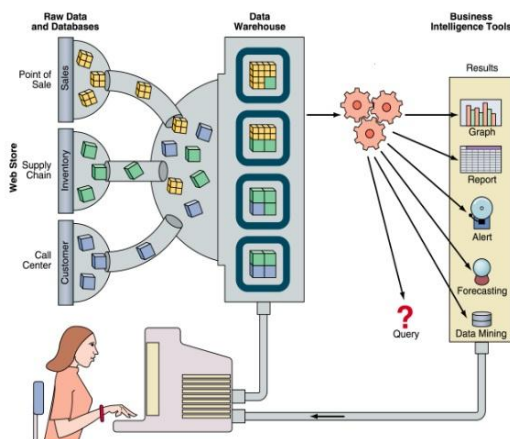

Fig 1: Business Intelligence Flow Overview

The Fig 1 shows the overall flow of business intelligence in an organization. Data collection is one of the most important aspects of an organization. This data is spread over multiple sources and myriad forms. It is the job of the data scientist to clean and present the appropriate data. This collected data is stored in various warehouses as history. The core part of the business intelligence has been the analysis of the data, on which we intend to focus on, in this paper. Apart from the traditional mining and forecasting techniques many newer ways are also being adopted to extract hidden and resourceful knowledge from the data [10].

Artificial Intelligence has been one such technique being used for the same. Something unique about this method is that it thinks according to humans. Machine Learning is one such branch of AI, which provides innumerable learning algorithms whereby a system can learn and adapt according to the changing situations [1]. This gives the dynamic aspect to the analytics part. Another important technique is that of artificial neural networks. They have the ability to transform the systems into appropriate mathematical functions. ANN is also off great importance in detecting and assimilating relationships between large numbers of variables [17]. Apart from these methods,   Learning is one such method, which involves itself with the nonlinear set of data. Natural Language Processing has been one such method which has been application of this field. We explain this in detail ahead in the paper.

Big Data has been one such term that has been gaining huge clamor over the past few years. Something that really differentiates big data from the rest of data are the three V's: Volume, Variety and Velocity. The size and quantity of data has tremendously increased over the years. Not only this, also the variety in which data is available has seen some drastic changes. Structures, semi structured and unstructured are the basis for classification of data. Multimedia data, streaming data are vastly expanding. Apart from volume and variety, velocity is also another vital aspect of big data. The rate at which data is being generated is also a parameter to look after.

Cognitive Analytics is a field gaining significance off late. It is in a way similar to the previously mentioned techniques, but tries to closely analyze every aspect as per the human brain [5].

In the further sections, we have explained the mentioned procedures in much detail and their respective mapping towards efficient mapping towards robust and effective business intelligence predictions.

2. COGNITIVE COMPUTING

Cognitive computing is the simulation of human thought processes in a computerized model. It involves self-learning systems that use data mining, pattern recognition and natural language processing to mimic the way the human brain works [6]. The goal of cognitive computing is to create automated IT systems that are capable of solving problems without requiring human assistance. Cognitive computing systems continually acquire knowledge from the data fed into them by mining data for information. The systems refine the way they look for patterns and as well as the way they process data so they become capable of anticipating new problems and modeling possible solutions [8]. While they'll have deep domain expertise, instead of replacing human experts, cognitive computers will act as a decision support system and help them make better decisions based on the best available data, whether in healthcare, finance or customer service.

3. BIG DATA ANALYTICS

With data growing incrementally and the recent technology advances that make it possible to manage and process this data, Big Data has now moved into the executive's agenda for every organization [7]. Leveraging all the available data to provide actionable business insights is no more a luxury, but a necessity for survival. With Big Data, organizations have the opportunity to touch and enhance a wide range of business processes while gaining a comprehensive 360 degree understanding of their organization in relation to their customers, products, services, suppliers and employees by leveraging every data touch point [12]. Enterprises today are seeking business transformation capabilities from IT providers by adopting big data for:

3.1 Risk Management

Inaccurate risk assessment can lead to poor decision making, high costs and scrutiny from regulators. With so much data to analyze, companies need a systemic approach to effectively identify and assess all of the risk exposures, known and unknown, that their business faces. Organizations must identify every risk, from threats to the company's reputation to data breaches and risk of non-compliance with regulations, and weigh them against business opportunities. By analyzing with big data, they can strike a balance between risk and opportunity [13].

3.2 Driving R&D

Big Data for R&D is less about velocity and more about variety, viability and sometimes volume. The key analytics capability for data is the ability to visualize relationships and patterns. By combining real-world outcomes data with clinical data and through the mining of genetic data and a broader understanding of regional and population data, analytically savvy organizations can begin to recognize research failures faster, design more efficient trials and speed the discovery and approval of new innovation while lowering costs along the way [14].

3.3 Analyzing Consumer Behavior

The emergence of internet, e-commerce and social media has radically altered the landscape of available consumer behavior data. Cash registers and Point-of-Sale (POS) systems are being replaced by e-commerce sites that record every move consumers make, even when they don't buy something. Casual telephone conversations about recent purchases are being replaced by tweets that can be scanned and analyzed by anyone who follows those Twitter feeds. All of this data on actual consumer behavior and experiences is there to be measured and analyzed [15].

4. ARTIFICIAL INTELLIGENCE

Artificial Intelligence is that domain of computer science that focuses on development of systems that think like humans, act like humans, think rationally and act rationally. For a long time now, there has been a huge focus on development of automated agents or systems that can do the job which a human mind can do [9]. AI tries to imitate a human brain and takes decisions just like a human brain would have in that particular situation. AI alone is too vast a field. It has various sub-domains which can help us explore business intelligence techniques. Machine Learning and Data Mining algorithms are techniques which help us analyze huge amounts of data. Artificial Neural Networks is also one such area using which we try developing computational models useful in machine learning and pattern recognition in data.

4.1 Machine Learning

As the name suggests, machine learning is engineering systems that can learn based on experience and recognize various trends in the data. The strength of any business lies in effectively exploiting its data to the fullest. An appropriate algorithm is necessary which will maps the available data to the needs of the organization. Machine learning algorithms are mainly of two types: Supervised learning and Unsupervised learning algorithms. Supervised Learning algorithms are those which operate on labeled data, while unsupervised learning algorithms are those which run on unlabeled data.

Machine Learning or Statistical Machine Learning has experienced many changes in the past few years. It is no longer being looked after just from an academic point of view, but is widely being applied to help companies and organizations make profits [16].

The importance of Machine Learning algorithms can be attributed to the fact that they can predict various uncertain situations. Classifiers are an integral part of BI and are nothing but machine learning algorithms. Support Vector Machine, Fuzzy Decision Tree, Naive Bayes classifier and Neural Network are different types of classifiers used widely in the industry.

Tactical BI (green filed) in the Fig 2 below signifies the importance of Machine Learning in BI. With the complexity of systems and data, analysis becomes difficult. The need is not just to build systems that provide current status, but also learn to make accurate predictions.
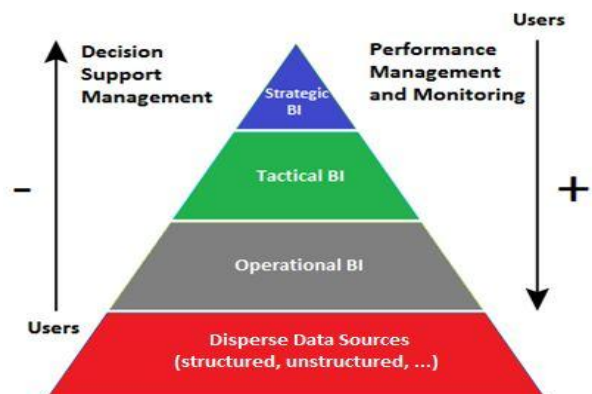

Fig 2: Tactical BI

4.2 Artificial Neural Networks

Artificial Neural Networks are nothing but simulation of the human brain. They learn to recognize the pattern in the data when they are trained on a sample set of data. ANN's have indeed risen as a vital tool for business prediction and forecasting. Something really unique about ANN's is their inherent nonlinear structure which becomes useful in handling complex data. The features that make ANN really very robust and efficient are those of approximating complex functional relationships and data-driven non parametric methods [18].
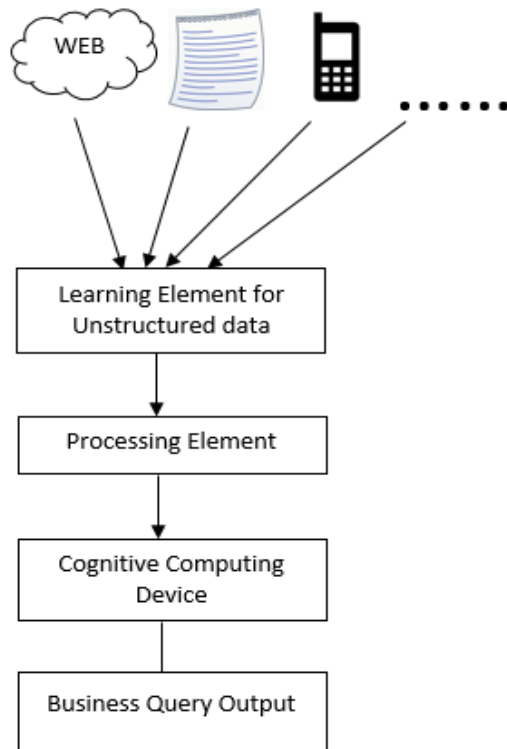
5. NATURAL LANGUAGE PROCESSING

Fig 3: Flow of NLP Technique for Business Intelligence

Natural Language Processing is one of the most widely applied human computer interaction and artificial intelligence method over the past few years. It deals with development of complex elements that have the potential to deconstruct, analyze and comprehend unstructured data and natural language conversations [4].

As shown in the flow chart in Fig 3, NLP can make it simpler for a person to interact with the system [11]. One of the biggest advantages is that the end user is not expected to write complex queries involving complex syntax. Even a layman can use NLP and get the expected output. For example a user can simply ask the system, "What is the expected profit for the next week?" as a substitute to writing cumbersome queries. The cognitive system later processes the speech of the user and displays the expected output. A lot of research is going on in this direction, however it is sure to metamorphose the way processing happens [3].

6. TOOLS

Many tools are currently available in the market for business intelligence. A gist of few of them is mentioned below.

6.1 Apache Hadoop

Apache Hadoop is an open source distributed software platform for storing and processing data. Written in Java, it runs on a cluster of industry-standard servers configured with direct-attached storage. Using Hadoop, one can store petabytes of data reliably on tens of thousands of servers while scaling performance cost-effectively by merely adding inexpensive nodes to the cluster. It is optimized for analytical workloads. The MapReduce programming model is designed for analytics and the Hadoop file system is optimized for sequential data access. On the other hand, traditional RDMS databases are purpose built and optimized for record storage and retrieval with random read and write access. Thus, Hadoop is magnitudes faster for analytics workloads that need to scan through all the data like joins and aggregations. The Fig 4 below gives an overview of the Hadoop Ecosystem. The various ecosystem components give unique functionality.
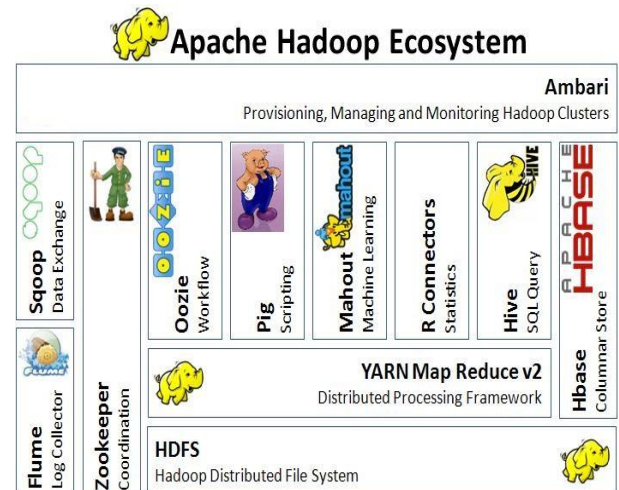

Fig 4: Hadoop Ecosystem Overview

6.2 Jaspersoft BI Suite

Jaspersoft Business Intelligence software comprises of a broad BI suite that covers data visualization and data integration tools, online analytical processing (OLAP), reporting and data integration. Interactive web reports, dashboards and analysis tools are also among the features that Jaspersoft comes with. The JasperReports Server now offers software to suck up data from many of the major storage platforms, including MongoDB, Cassandra, Redis, Riak, CouchDB and Neo4j. Hadoop is also well-represented with JasperReports providing a Hive connector to reach the inside of HBase. Jaspersoft is expanding by making it easier to use these sophisticated reports with newer sources of data. It isn't offering particularly new ways to look at the data, just more sophisticated ways to access data stored in new locations.

6.3 Skytree Server

Skytree is a machine learning program intended to serve as a replacement for SQL databases, and has the ability to take even unstructured sets of data and crunch it faster than traditional methods. It can improve system performance to process big data 10,000 times faster. It is designed to connect quickly and easily with the existing IT infrastructure, making data setup a breeze. It can be configured to accept data

streams from multiple sources and compute near-instantaneous results (including real-time queries) on each one. The resulting analytics are returned through the same channels.

6.4 Pentaho Business Analytics

Pentaho is another software platform that began as a report generating engine; it is, like JasperSoft, branching into big data by making it easier to absorb information from the new sources. We can hook up Pentaho's tool to many of the most popular NoSQL databases such as MongoDB and Cassandra. Once the databases are connected, we can drag and drop the columns into views and reports as if the information came from SQL databases.

6.5 IBM Watson

Watson is an Artificial Intelligence computing system that IBM built to apply advanced natural language processing, information retrieval, knowledge representation, automated reasoning, and machine learning technologies to the field of open domain question answering. The goal of such a tool is to have computers start to interact in natural human terms across a range of applications and processes, understanding the questions that humans ask and providing answers that humans can understand and justify.

7. CASE STUDY: HEALTH CARE DOMAIN

A healthcare organization attempts to describe its many processes, inputs, outputs, and products using a language. One of the fundamental challenges of healthcare is that this language is often imprecise and ambiguous. The medical record, paper or electronic, largely remains a collection of free-form narrative serving as the communication language of healthcare delivery. Insights, attributes of processes, prognoses, evaluations, and clinical observations are locked up in narrative. Healthcare organizations therefore need an additional enhancement approach and technology – natural language processing (NLP). Also, Data of healthcare organizations is growing and moving faster; 80% of medical data is unstructured and is clinically relevant. This data resides in multiple places like individual EMRs, lab and imaging systems, physician notes, medical correspondence, claims, CRM systems and finance. Getting access to this valuable data and factoring it into clinical and advanced analytics is critical to improving care and outcomes, incentivizing the right behavior and driving efficiencies. Healthcare organizations are therefore leveraging big data technology to capture all of the information about a patient to get a more complete view for insight into care coordination and outcomes-based reimbursement models, population health management, and patient engagement and outreach.

7.1 Natural Language Processing (NLP) and Cognitive Analytics

NLP is data enrichment technology of particularly high-value to healthcare data management challenges. NLP technology at its core consumes free-form clinical narrative as raw data passes through the data management factory, and outputs suggested diagnostic and procedural coding of a controlled medical vocabulary system of choice. These codes in turn are managed, mapped, validated using the terminology services. Many of the very same vendors that offer terminology management engines also offer adjunct NLP modules and solutions or partner with best-of-breed NLP engines.

7.2 Big Data Analytics and Machine Learning

Successfully harnessing big data unleashes the potential to achieve the three critical objectives for healthcare transformation viz. Building sustainable healthcare systems, collaborating to improve care and outcomes and increasing access to healthcare. Analysis of this data using Machine Learning Algorithms can help in Inferring knowledge from complex heterogeneous patient sources, efficiently handling large volumes of medical imaging data and extracting potentially useful information and biomarkers, analyzing genomic data and capturing the patient's behavioral data through several sensors; their various social interactions and communications. It can provide many insights about making the right decisions at the right time for the patients. Efficiently utilizing the colossal healthcare data repositories can yield some immediate returns in terms of patient outcomes and lowering care costs. Data with more complexities keep evolving in healthcare thus leading to more opportunities for big data analytics.

8. CONCLUSION

Artificial Intelligence combined with Cognitive Analytics can indeed revolutionize the way businesses operate. The amount and variety of data is just going to keep on increasing over the years to come, and these techniques indeed provide the right way to tackle the problem. Computer Vision and Human Computer Interaction along with Natural Language processing will take the burden off the administrators to write complex queries. The ultimate goal is going to remain the growth of business with smart and efficient methods for correct predictions. And indeed the techniques mentioned in the paper will make it happen in a robust manner.

9. REFERENCES

[1] Tyson Condie, Paul Mineiro, Neoklis Polyzotis, Markus Weimer, "Machine Learning on Big Data", ICDE Conference 2013.

[2] Jayanthi Ranjan, "Business Intelligence: Concepts, Components, Techniques and Benefits", Journal of Theoretical and Applied Information Technology, Vol 9.No 1. (pp060 - 070).

[3] Tech Trends 2014: Inspiring Disruption, Deloitte University Press.

[4] M.A. Migut, J.C. van Gemert, M. Worring, "Interactive Decision making using Dissimilarity to visually represented Prototypes", IEEE Symposium on Visual Analytics Science and Technology, October 23 - 28, 2011, Providence, RI, USA.

[5] Qihui Wu, Guoru Ding, Yuhua Xu, Shuo Feng, Zhiyong Du, Jinlong Wang, Keping Long, "Cognitive Internet of Things: A New Paradigm Beyond Connection", IEEE Internet of Things Journal, Vol. 1, No. 2, April 2014.

[6] Stephen Rudolph, Anya Savikhin, David S. Ebert, "FinVis: Applied Visual Analytics for Personal Financial Planning", IEEE Symposium on Visual Analytics Science and Technology October 12 - 13, 2009, Atlantic City, New Jersey, USA.

[7] Rob van den Dam, "Big Data a sure thing for Telecommunications", 2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery.

[8] Tera Marie Green, William Ribarsky, Brian Fisher, "Visual Analytics for Complex Concepts Using a Human Cognition Model", IEEE Symposium on Visual Analytics Sciences and Technology, October 21 – 23, 2008, Columbus, Ohio, USA.

[9] Kristina Dervojeda, Diederik Verzijl, Fabian Nagtegaal, Mark Lengton & Elco Rouwmaat, "Big Data", Artificial Intelligence, Business Innovation Observatory, European Union, September 2013.

[10] Gerrit Lahrmann Frederik Marx Robert Winter Felix Wortmann, "Business Intelligence Maturity: Development and Evaluation of a Theoretical Model", Proceedings of the 44th Hawaii International Conference on System Sciences – 2011.

[11] Alexander Mikroyannidis, Babis Theodoulidis, "Ontology management and evolution for business intelligence", International Journal of Information Management, Volume 30, Issue 6, December 2010, Pages 559–566.

[12] Wixom, Barbara; Ariyachandra, Thilini; Douglas, David; Goul, Michael; Gupta, Babita; Iyer, Lakshmi; Kulkarni, Uday; Mooney, John G.; Phillips-Wren, Gloria; and Turetken, Ozgur (2014) "The Current State of Business Intelligence in Academia: The Arrival of Big Data," Communications of the Association for Information Systems: Vol. 34, Article 1.

[13] Yaochu Jin, Barbara Hammer, "Computational Intelligence in Big Data", IEEE Computational Intelligence Magazine, August 2014.

[14] Pavan Sridhar, Neha Dharmaji, "A Comparative Study on How Big Data is Scaling Business Intelligence and Analytics", International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463 Vol. 2 Issue 8, August-2013.

[15] Edmon Begoli, James Horey, "Design Principles for Effective Knowledg Discovery from Big Data", 2012 Joint Working Conference on Software Architecture & 6th European Conference on Software Architecture.

[16] Hardeep Singh, Bikram Pal Singh, "Business Intelligence: Effective machine learning for business administration", International Journal of IT, Engineering and Applied Sciences Research (IJIEASR), ISSN: 2319-4413, Volume 2, No. 1, January 2013.

[17] A. Maithili, Dr. R. Vasantha Kumari, Mr. S. Rajamanickam, "Neural Network Towards Business Forecasting", IOSR Journal of Engineering, Apr. 2012, Vol. 2(4) pp: 831-836.

[18] Vinod Sharma, "Artificial Neural Network Applicability in Business Forecasting", International Journal of Emerging Research in Management &Technology, ISSN: 2278-9359, December 2012.