



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 8, August 2018

Artificial Intelligence and Its Related Application in the Classification of Big Data

Saurabh Burange¹, Priyen Dang², Anurag Singh³, Rudra Dave⁴, Bhavin Dave⁵, Soham Bhat⁶

Student, Department of Information Technology Engineering, St. Francis Institute of Technology, Mumbai,
Maharashtra, India^{1,2}

Student, Department of Information Technology Engineering, Thakur College of Engineering and Technology,
Mumbai, Maharashtra, India³

Student, Department of Electronics Engineering, K.J Somaiya College of Engineering, Mumbai, Maharashtra, India^{4,5,6}

ABSTRACT: This paper aims on explaining the importance of Artificial Intelligence and focus on its application in classification of Big Data. Artificial Intelligence has a vast concept and has several applications like in gaming, medical applications and many others but specifically in our paper we focus on its application in classification of Big Data. Through technology innovations, there has been a large increase within the utilization of Bigdata knowledge. To sustain Associate in Nursing current ascension of knowledge Bigdata there is a Associate in Nursing rising demand for a complicated content-based knowledge classification system. In recent times efforts has been made and dedicated in order to develop classification primarily based on the knowledge ways which may be helpful in successfully retrieving the knowledge of interest. Considering the restricted man-power it is expected to develop retrieval ways that use options mechanically extracted from massive knowledge and this can be done with the help of Architecture and Algorithm co-design for Bigdata processing Applications. Manycore processor consisting of classification of heterogeneous cores with stream process capabilities and zero-overhead inter-process communication through computer science with a hardware-software mechanism has been designed and the major motivation behind it is often for achieving superior and low-power consumption, particularly thus on cut back access needed for Bigdata processing Applications.

KEYWORDS: Artificial Intelligence, Application of Artificial Intelligence, Big Data, Pollination based optimization, Biogeography based optimization.

I. INTRODUCTION

Artificial Intelligence plays a vital and increasing role in research of management science and operational research areas. Basically, intelligence is considered as the ability to collect knowledge and information and aims to solve complex problems with proper reasoning. In the near future intelligent machines will replace human efforts and capabilities in many areas. Artificial intelligence is the study and developments of intelligent machines and software that can reason, learn, gather knowledge, communicate, manipulate and perceive the objects. It is the study of the computation that makes it possible to perceive reason and act accordingly. Artificial intelligence is different from psychology and computer science because it emphasis on computation and on perception, reasoning and its corresponding actions respectively. Its aims and helps to make machines more powerful, smarter and useful. It works with the help of artificial neurons and scientific theorems. Major areas of Artificial Intelligence are Classification of Bigdata, Expert Systems, Natural Language Processing, Speech Understanding, Robotics and Sensory Systems, Computer Vision and Scene Recognition, Intelligent ComputerAided Instruction, Neural Computing. The various techniques applied in artificial intelligence are Neural Networks, Fuzzy Logic, Evolutionary Computing, and also Hybrid Artificial Intelligence. Artificial intelligence is more consistent, permanent and less expensive. Specifically in this paper we are discussing application of Artificial Intelligence in classification of Bigdata. Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 8, August 2018

operational efficiencies, cost reductions, and reduced risks for the business. To harness the power of big data, you would require an infrastructure that can manage and process huge volumes of structured and unstructured data in real time and can protect data privacy and security. There are various technologies in the market in order to handle big data. While looking into the technologies that handle big data, we examine the following two classes of technology and take them in to consideration.

1.1 OPERATIONAL BIGDATA

Operational Bigdata includes systems like Mongo DB that provide operational capabilities for real-time, interactive workloads where data is primarily captured and stored. NoSQL Big Data systems are designed to take advantage of new cloud computing architectures that have emerged over the past decade to allow massive computations to be run inexpensively and efficiently. This makes operational big data workloads much easier to manage, cheaper, and faster to implement. Some NoSQL systems are capable of providing insights into patterns and trends based on realtime data with minimal coding and without the need for data scientists and additional infrastructure.

1.2 ANALYTICAL BIGDATA

Analytical Bigdata includes systems like Massively Parallel Processing (MPP) database systems and MapReduce that provide analytical capabilities for retrospective and complex analysis that may touch most or all of the data. MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL, and a system based on MapReduce that can be scaled up from single servers to thousands of high and low end machines. These two classes of technology are complementary and frequently deployed together.

II. LITERATURE SURVEY

Behrouz et. al. A combination of multiple classifiers leads to a significant improvement in classification performance. Furthermore, by learning an appropriate weighting of the features used via a genetic algorithm (GA), we further improve prediction accuracy. The GA is demonstrated to successfully improve the accuracy of combined classifier performance, about 10 To 12% when comparing to non-GA classifier. This method may be of considerable usefulness in identifying students at risk early, especially in very large classes, and allow the instructor to provide appropriate advising in a timely manner. Riccardo et al. [1] proposed cognitive, and behavioural aspects of distance students. Course Vis is presented in the paper, and several examples of pictorial representations generated by the tool. Luo et. al. [2] Efficient meaning for sampling of data, reduction of data also needed to develop. Newly develop mining technique and searching algorithms that are suitable for extracting more different or complex relationship between fields.

III. PROPOSED SYSTEM AND METHODOLOGY

The first step is acquiring the data set of our choice. Data set can be in form of numeric or text or others. Then Association rule mining is done using Apriori algorithm. Apriori algorithm basically consists of two steps. First is CANDIDATE GENERATION phase and the second is count of SUPPORT factor with THRESHOLD. If Support factor is greater than Threshold, Positive rules are generated and are accepted, otherwise Negative rules are generated and are rejected. The rules which are Positive are valid rules and rules which are declared negative by Support and Threshold are invalid. Valid i.e. Positive rules are presented as a input for Pollination based optimization. Further, rules can be changed on the basis of reproduction factor which is R based on the formulation mentioned below:

$$R = [(A \times D) / (\alpha + A \times D)] + [(\alpha / (\alpha + A \times D) \times NP) / (AP + NP)] - C(N + D)$$

Optimization is a natural process embedded in the living beings. Pollination is a process of transfer of pollen from male parts of flower called anther to the female part called stigma of a flower. Self pollination results in the development of some flower seeds, when flower of one plant gives the pollen and pistil both then it is said to be a self pollination and



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 8, August 2018

when the pollen and pistil are from different flowers of different plants then it is said to be a cross pollination. Pollinators are responsible for movement of the pollens which are further responsible for the reproduction by setting of seeds. But pollinators don't have any information about the benefit of the plant. Pollination in plants is done for the energy requirement and to produce new plants. The floral display, fragrance and nectar lure pollinators and leads to pollination. Some species of plants optimize their nectar, display and fragrance producing resources. If pollination process is proceeding smoothly the plants spend average resources. If pollination process is above normal the plants reduce expenditure on resources for producing nectar, floral display and fragrance in the flowers. If the pollination success goes below normal, plants increase the resource expenditure such that more floral display fragrance and nectar to attract pollinator.

3.1 ALGORITHM

Step 1: Initialization of PBO parameter.

Step 2: Randomly generate vectors.

Step 3: Evaluate Reproduction vectors.

$R = [(A \times D) / (\alpha + A \times D)] + [(\alpha / (\alpha + A \times D) \times NP) / (AP + NP)] - C(N + D)$

Step 4: Update the number of season based on reproduction vector.

Step 5: Update N, D and A based on errors.

Step 6: Exit.

IV. FUTURE SCOPE

Work on security can be performed in order to enhance it. Hill cipher method and secure algorithm like RSA can be performed. Furthermore, also password secure method can be used to enhance the security.

V. CONCLUSION

The field of artificial intelligence gives the ability to the machines to think analytically, using concepts. Tremendous contribution to the various areas has been made by the Artificial Intelligence techniques from the past many years. Artificial Intelligence will continue to play an increasingly important role in the various fields. Artificial Intelligence has several applications like it can be used in Natural Language Processing, gaming, medical application and several others. It can also be used to build smarter products which will add to the benefit of people around. For example the construction of locomotion system for ALS patients requires Artificial Intelligence and it helps ALS patients to attain mobility just by using their thinking capability. Furthermore Artificial Intelligence can replace motion sensor based games as well in future. Specifically in this paper we have focused on application of Artificial Intelligence in classification of Big data. After evaluation of parameters like precision, recall and f-measure, we analyze artificial intelligence are better optimizer and can reduce calculation with efficient results for big data classification in comparison to traditional systems and methodology. Results of Apriori algorithm, Apriori with PBO, Apriori with BBO has been compared and the conclusion indicates that Apriori with PBO gives the best results. So it can be concluded that it is a unique blend of a classification algorithm, parallelism and artificial intelligence.

REFERENCES

- [1] Asmaa Benghabrit, Brahim Ouhbi, Hicham Behja "Text Clustering Using Statistical and Semantic Data" -IEEE 2013.
- [2] Marisa S. Viveros, John P. Nearhos, Michael J. Rothman - "Applying Data Mining Techniques to a Health Insurance Information System". 2012
- [3] Linna Li, Bingru Yang, Faguo Zhou "A Framework for Object-Oriented Data Mining" Fifth International Conference on Fuzzy Systems and Knowledge discovery 2010.
- [4] Dezheng Feng, Zaimei Zhang, Fang Zhou, Jianheng Ji [4] "Application Study of Data Mining on Customer Relationship Management in E-commerce".
- [5] Bhavani Thuraisingham "Data Mining for Malicious Code Detection and Security Applications" -IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops 2012.
- [6] He YueShun, Ding QiuLin "Application research of data mining architecture for intelligent decision" -Asia-Pacific Conference on Information Processing 2009



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 8, August 2018

- [7] Luo Fang, QiuQizhi "The Study on the Application of Data Mining Based on Association Rules" -International Conference on Communication Systems and Network Technologies (2012).
- [8] Xindong Wu "Data Mining: Artificial Intelligence in Data Analysis" -Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology 2004
- [9] Ricardo Mazza, Vania Dimitrova "Course Vis: Externalizing Student Information to Facilitate Instructors in Distance Learning" Proceedings of the International conference in Artificial Intelligence in Education.2004
- [10] Behrouz Minaei-Bidgoli, Deborah A. Kashy, GerdKortemeyer, William F. Punch "Predicting Student Performance: An Application Of Data Mining Methods With The Educational Web-Based System Lon-Capa" 33rd ASEE/IEEE Frontiers in Education Conference 2003