E-Magazine 2020-21

Volume 12

Banarsidas Chandiwala Institute of Information Technology

Affiliated to GGSIP University,
Kalkaji, New Delhi.

Vision

"To be a World Class Institution and to nurture

Value based Professionals."

Mission

"To be student Centric through innovation in Pedagogy, Research,

Faculty Development,

Collaboration with Industry, Institutions and

High Quality Infrastructure."

From the Secretary's Desk

I welcome you all to this new edition of BCIIT Magazine. Banarsidas Chandiwala Sewa Smarak Trust Society is working with a mission "To provide yeomen service in the field of Health and Education". BCIIT was created by the society to fulfill its mission and the societal needs of higher technical education in the developing discipline of Computer Science in 1999. The society is aware of its responsibility to provide education to the youth of India.

The society has provided a serene environment of teaching/learning at BCIIT with state of the art infrastructure comprising:

- Wi-Fi Enabled Campus
- MOU with MNC for scholarship
- SPOKEN Tutorial Center by IIT Bombay
- Air Conditioned Classrooms.
- Computer Labs equipped with latest computers and all legal software.
- Air Conditioned Library with Digital Library with latest journals & magazines.
- On campus, well-furnished Hostel.
- Multi-Specialty Campus Hospital

The thrust areas of BCIIT are practical intensive teaching/learning and all round personality development of the students so that they may be ready to accept the challenges of contemporary professional life of IT industry. I am sure that students of BCIIT are well equipped with necessary skills to deliver what is expected from them by the industry.

Dr. Bhuwan Mohan

From the Director'sDesk

I welcome you all to this new edition of BCIIT Magazine. In the information age, science and technology are the corner stone's on which the structure of society rests. The rapid advances in Information and Communication Technologies (ICT) has made the world increasingly hyper- connected and competitive, offering new challenges and opportunities, thus bringing fundamental transformation in society.

The Banarsidas Chandiwala Institute of Information Technology (BCIIT) has taken this unique initiative to encourage the innovative thoughts of its faculty and students to be put in the form of articles in e-magazine. These articles are put on the Institute website so as to be available to more people for their references, use and comments. This e-magazine is a regular annual feature of the Institute since the first issue in 2011. Some of the faculties and students who ultimately wish to pursue the Ph. D program get lot of inspiration and initiate their research in the area of interest.

One of our dreams is to see that BCIIT stands tall among the other institutes of GGSIP University making an impact with value added contributions in the form of high standard and quality articles through its online endeavor. At our end we feel that we have highly experienced and inspired faculty and excellent and academically brilliant students who can contribute a lot in this manner.

I hope our humble effort will go a long way in putting the resourceful thoughts of our faculty and students in improving the quality of education through technology. It is the genuine and sincere attempt of our faculty and students who are constantly putting their heart and soul to achieve the results.

I pray and wish them good luck in their endeavor.

Dr. Ravish Saggar
Director, BCIIT

Table of Contents

Sr. No	Title and Author	Page Nos.
1	"Network Simulation using Cisco Packet Tracer " Dr. Vandana Sharma	6-10
2	"Artificial Intelligence Techniques for Cyber Security" Ms. Nidhi Grover	11-14
3	"A Survey on Load Balancing in Cloud" Ms. Pooja Arora	15-19
4	"Impact of Social Media on Youth" Ms. Rama Bansal	20-22
5	"Understanding Blockchain: Architecture, Applications and Challenges" Mr. Sandeep Jain	23-28
6	"A Review of Machine Learning Techniques" Sakshi Mittal & Dr. Sushma Bahuguna	29-35
7	"Sentiment Analysis: A Review" Dr. Anu Taneja	36-38

BCIIT e-Magazine 2021

www.bciit.ac.in

Network Simulation Using Cisco Packet Tracer

Dr. Vandana Sharma
Assistant Professor
Banarsidas Chandiwala Institute of Information Technology
Kalkaji, Delhi
vandana@bciit.ac.in

Abstract: *In today's world, Networking has become extremely necessary for providing e-mail, telnet, and other services. In these services, Networks play a very important role to communicate, administrate, and process the information from point to another point. Hence networks considered the basics of proper functionally in most of the companies and organizations. In this paper, cisco packet tracer is used as a network simulator tool to simulate and design a network project of a Multinational company. This paper discusses the interconnection between routers, switches and other components in the communication network. Further it demonstrates the extension of a LAN in a wide network setup for a Multi National Company.*

Keywords: *Cisco Packet Tracer, Networking, VLAN, Network Configuration*

I. Introduction

Network can be defined as several computer systems and other computing hardware devices that linked together for sharing information in form messages, files, and databases in an organization that may be in one building or spread over a large campus.

Network connections devices are connected by using any type of communication media such as the copper coaxial cable, twisted pair cable, optical fiber cable, and wireless. While the network devices include Network interface cards (NIC), Hubs, Switches, Routers, Gateways, and Modems. Each one of these devices has various properties than others from jobbing, security, and application. This paper highlights the VLANs (Virtual Local Area Network) and VTP (VLAN Trunking Protocol) switch as well as EIGRP (Enhanced Interior Gateway Routing Protocol). Switch ports can be assembled into different VLANs on a single switch and multiple interconnected switches. After creating multiple virtual LANs, switches are used to establish multiple domains. These domains are capable of broadcasting within a LAN. VTP allows rename and modify multiple VLANs on a single switch. Further this information is then propagated to all other switches in the same domain. EIGRP is a Cisco[1] proprietary classless routing protocol that is essentially an enhanced distance vector protocol.

Network simulation software like Cisco Packet Tracer is used to perform and analyze various network activities. It is used to implement various topologies, sub netting and analysis of various network configuration and selection of optimum path based on various routing algorithms, creation of appropriate servers, and troubleshooting commands.

II. Network Scenario

Authors have assumed three branch offices of a company in three different locations. Each branch consists of three buildings where each building has three floors and each floor contains three switches, where the connection between them is a mesh topology [2]. The various LANs are set up in the MNC comprising of Engineers, Accountants, Salesman, Workers, Chairman and 4 different Phone branch and servers. The Class C network

addresses are assigned to the staff systems creating virtual subnets. The range of addresses assigned from 192.168.1.0 to 192.168.10.0., the staff distribution in the whole company.

III. Network Configuration

A. Switching Configuration

In the main Switch configuration [3] set the name of the main switch and enable them with secret password of all lines (console and auxiliary and VTY lines) in the switch. The main switch is made as VTP server mode and set VTP domain to be able to create, modify, and delete VLANs because any changes made to a switch in server mode will be propagated throughout the VTP domain. Each of the VLANs is defined as a name. The commands of type show running configuration, VTP and VLAN are executed to configure the main switch of the network. The secondary switches are defined as VTP client mode. All are set in the same domain of VTP server. All Ethernet interfaces are divided among other VLANs. The mode of connections between switches is established. A set of tracking mode is configured and the connections between the switch and PC are in access mode.

B. Router Configuration

Router1 considers the controlled exchange of information between several buildings and other routers in different places. First of all, the router1 is also renamed and supported with secret password of all lines (console and auxiliary and VTY lines). The router1 is connected to the other switch using subnet-interfaces on port Ethernet interface 0/0 and all the IP networks of VLANs are added. EIGRP protocol routing is used to achieve the connection between router1 and router 2 by listing the IP network. The serial interfaces (s0/0/0 and s0/0/1) must be given IP addresses.

C. Phones Configuration

The phone configuration, the router1 created the DHCP pool with name (as voice1) and defined the IP network of the DHCP and IP default router1. The configuration of the telephone service in router1 must be defined as the maximum numbers of the directory numbers, maximum numbers of phoned IP address source, automaticity assign number to bottoms phones and gave a number to each phone. The mode of connections between the switch and phones is a voicing mode.

D. Servers Configuration

Four servers are used to design four different websites. All of the websites are defined in the (DNS and branch 3) server by clicking it and go to the DNS option in the service tab and then add the name and IP address of all websites. While the other servers will add only the name and IP address of one website. The DHCP option in the server tab is added to the IP address of the DNS servers and IP address of the default gateway. The router1 made the DHCP pool with name server to can be used to load the website by a device [4] – [6].

Interface	IP-Address	OK?	Method	Status	Protocol
FastEthernet0/0	unassigned	YES	unset	up	up
FastEthernet0/0.10	192.168.1.1	YES	manual	up	up
FastEthernet0/0.20	192.168.2.1	YES	manual	up	up
FastEthernet0/0.30	192.168.3.1	YES	manual	up	up
FastEthernet0/0.40	192.168.4.1	YES	manual	up	up
FastEthernet0/0.50	192.168.5.1	YES	manual	up	up
FastEthernet0/0.60	192.168.6.1	YES	manual	up	up
FastEthernet0/0.70	192.168.7.1	YES	manual	up	up
FastEthernet0/0.80	192.168.8.1	YES	manual	up	up
FastEthernet0/0.90	192.168.9.1	YES	manual	up	up
FastEthernet0/1	unassigned	YES	unset	administratively down	down
Serial0/1/0	unassigned	YES	unset	down	down
Serial0/3/0	11.11.11.1	YES	manual	up	up
Serial0/3/1	unassigned	YES	unset	administratively down	down
FastEthernet1/0	unassigned	YES	unset	up	up
FastEthernet1/1	unassigned	YES	unset	administratively down	down
Vlan1	unassigned	YES	unset	administratively down	down

Figure 1: Router1 Port configuration and IP addresses

```
hostname BCIIT
!
!
!
enable secret 5 $1$mERr$H7PDxl7VYMqaD3id4jJVK/
!
!
!
ip dhcp pool server
 network 192.168.10.0 255.255.255.0
 default-router 192.168.1.1
ip dhcp pool voice1
 network 192.168.7.0 255.255.255.0
 default-router 192.168.1.1
 option 150 ip 192.168.1.1
ip dhcp pool net1
 network 192.168.6.0 255.255.255.0
 default-router 192.168.1.1
 option 150 ip 192.168.1.1
ip dhcp pool voice2
 network 192.168.8.0 255.255.255.0
 default-router 192.168.1.1
 option 150 ip 192.168.1.1
ip dhcp pool voice3
 network 192.168.9.0 255.255.255.0
 default-router 192.168.1.1
 option 150 ip 192.168.1.1
!
```

Figure 2: DHCP Protocol on Router1 (main router)

```
telephony-service
 max-ephones 9
 max-dn 9
 ip source-address 192.168.7.1 port 2000
 auto assign 1 to 9
!
ephone-dn 3
 number 1234
!
ephone-dn 4
 number 1235
!
ephone-dn 5
 number 1233
!
ephone-dn 6
 number 1244
!
ephone-dn 7
 number 123456
!
ephone-dn 1
 number 123458
!
ephone-dn 2
 number 1234567
!
line con 0
 password 123456
 login
!
line aux 0
 password 123456
 login
!
line vty 0 4
 password 123456
 login
line vty 5 15
 password 123456
 login
```

Figure 3: Telephony service on router1 (main router)

```
VLAN Name                              Status    Ports
---- --------------------------------- --------- -------------------------------
1    default                           active    Fa0/1, Fa0/6, Fa0/7, Fa0/8
                                                 Fa0/9, Fa0/10, Fa0/11, Fa0/12
                                                 Fa0/13, Fa0/14, Fa0/15, Fa0/16
                                                 Fa0/17, Fa0/18, Fa0/19, Fa0/20
                                                 Fa0/21, Fa0/22, Fa0/23, Fa0/24
10   engineers                         active    Fa0/5
20   accountants                       active
30   salesmen                          active
40   workers                           active
50   chairmen                          active
60   ass.charimen                      active
70   phone1                            active
80   phone2                            active
90   phone3                            active
100  VLAN0100                          active
1002 fddi-default                      act/unsup
1003 token-ring-default                act/unsup
1004 fddinet-default                   act/unsup
1005 trnet-default                     act/unsup
```

Figure 4: VLAN Services on main Switch

IV. Conclusion

The packet tracer is used to implement the network of the project (MNC) and clarity the conception of the VLANs, DHCP, phone, website server, and router configurations. Networking devices are expensive so the packet tracer is easy and best to implement the structure of the network before implementing it on the real ground. Also, in this paper, the VLANs provide the security, broadcast control and physical layer transparency while VTP reduce configuration and integrate VLAN management for any changing on VTP server then

it will be distributed to other switches in the same VTP domain, therefore the time of configuration the same VLAN is reduced.

References

[1] Todd Lammle, "CCNA" Routing and switching study guide exam 200-125.

[2] Nathaniel S. Tarkaa, Paul I. Iannah, Isaac T.Iber "design, and simulation of local area network using cisco packet tracer", the International Journal of Engineering and Science (IJES).

[3] Edward E. Ogheneovo, Ibiba S." Modelling Network Router, Switches, And Security Using Cisco and OPNET Simulation Software" International organization of Scientific Research (IOSR) of computer Engineering, volume 4, issue 7, July 2014.

[4] Behrouz A. Fourouzan, "Data Communications and Networking", fourth edition, 2007.

[5] Garima Jain, Nasreen Noorani, Nisha Kiran, Sourabh Sharma "Designing and Simulation of Topology Network Using Packet Tracer", International Research Journal of Engineering and Technology (IRJET).

[6] CiscoPacketTracer. http:// cisco.netacad.net /cnams/ content/templates/LibraryHome.jsp?#/ resource/lcms/cnams_site/english/generic _site_areas/library/course_ catalog/PTCacheInclude.html.

BCIIT e-Magazine 2021

www.bciit.ac.in

Artificial Intelligence Techniques for Cyber Security

Nidhi Grover
Assistant Professor
Banarsidas Chandiwala Institute of Information Technology
Kalkaji, Delhi
nidhigrover.bciit@gmail.com

***Abstract**: Data innovation and the world of web are expanding at an awfully quick pace and so are increasing the violations related with cyber world. Cyber frameworks are much inclined to different sorts of dangers, intrusions and powerfully advancing dangers related with them. Program based on ordinary security algorithms and simple and human inclusion is lacking for ensuring complete cyber security. In this way, there's an increasing prerequisite for more capable and cleverly cyber defense frameworks to serve the reason of providing cyber security. The imaginative hones of Fake Insights are getting better known in assisting clients to battle violations and related issues in cyber space. The reason of this paper is to review cyber security, its related dangers and the progressions made so distant by applying Artificial Intelligence strategies in cyber defense. The paper too points in illustrating the appropriateness and effectiveness of these AI procedures in show situation.*

***Keywords:** artificial intelligence, cyber space, cyber security, artificial expert system, artificial neuron, system, artificial intelligence system, artificial immune system.*

I.Introduction

In developing world of computers and web the problem relate with them are too expanding. The technological progression within the field of web and telecommunication has brought the world in front of new kind of issues known as cyber violations and cyber terrorism. The term like cyber security came into existence since of the rate like cyber wrongdoing or cyber war. Cyber foundation is much defenseless and poses risk to countries' in general improvement [1]. Cyberspace has ended up a modern stage for seething war or dread for numerous non-state actors. Now days, Cyber space may be a source from where a individual sitting in different continent can make fear in other landmasses by one click .Through the internet some person can devastate not only the respectful or government framework but it can destroy the atomic framework moreover. Nowadays, cyber weapons and apparatuses have gotten to be so capable that typical human administered security frameworks are unable to ensure against them.

Problems of cyber security

In final decade ,a modern word "cyber" came into existence and make a entirety modern string of words Illustrations of terms that surface in scholarly papers incorporate cyber society, cyber assaults, cyber security, hostile cyber capabilities and issues of cyber fear mongering [3]. The parent term of the internet is "cybernetics", this word is to begin with presented by Nobert wiener for his work in communication and control science [4]. Cyber world is an environment in which communication over computer systems happens [4]. In 21st century technology has moved to supply a stage where people can associated, trade thought, share information, provide social back, conduct commerce, coordinate action, play diversion lock in political talk and so on, using a virtual space or say worldwide arrange or cyber space. In an time of persistent development of cyber connectivity with ever expanding number of online applications from buying a needle to being a portion of a team conveyed to damages it comes

fundamental to be conscious of potential dangers approaching over the cyber space. It got to be exceptionally fundamental to secure this cyberspace for potential dangers. With expanding number of Computers and improvement of telecommunication networks number of cyber security issues are also increasing .Assaults like backdoor, zombies computer attack ,infections , worms, Trojan ,D Dos, intrusion attacks , phishing focusing on people ,trade world and indeed government moreover. What make the internet so valuable for a nation is since thousand's of GB data get handled and passes on computer and computing framework.

Artificial Intelligence

Fake insights are the computer science that is concern with making computer carry on like humans. This machine insights developed within the frame of summer inquires about venture of Dartmouth College in July 1956. But the thought of AI begun back in 15th or 16th century. AI can be depicted in two ways:

I. As the science of creating brilliantly machines.
II. As science of finding strategies of tackling the problem with more complexity that cannot be unraveled without applying a few intelligence.

The moment definition of AI that symbolizes a system as that has the capability of taking its possess decisions without impedances of others [5]. Some characteristics that a Manufactured intelligence should exhibiter [5, 6]: Conclusion, thinking, issue understanding (embodied agents, neural network) Information representation Arranging (multi operator planning) Learning (machine learning) Characteristic dialect handle (data retrieval) Movement and manipulation Discernment (discourse acknowledgment).This paper is going to toss the light on the taking after applications of AI [7]: Fake Neural Organize: The thought of framework is based on neural organize technology. Counterfeit Brilliantly Framework: It's a framework that has Qualities like pro-activeness, understanding of agent Communication language Fake Safe Framework: Thought of Artificial Immune Framework based on normal Resistant system. Master Framework: It could be a framework which is utilized to find answers to issues made by clients.

II.Artificial Neural Network

From thousands of year or say from the starting of civilization, human body is one of the biggest mysteries. But, from the centre of 19th century facts and mystery around human body begun getting revealed by a few researchers and like other revelations or inventions human begun utilizing its information for making their life's comfortable. Neural framework of brain is the thought utilized by two specialists in their field's mathematician Walter Pitts and Warren McCulloch in 1943 who composed a paper on how neurons might work. For appearing and portraying the work of neuron working in brain, they modelled a straightforward neural network utilizing electrical circuits [9, 10].The ANN (Artificial neural arrange) is parallel distributed process which tries to imitate like natural brain system [11]. The most reason for utilizing of ANN (Artificial neural organize) and its ubiquity in cyber defence is it's high-speed.

A.Intrusion etection and Prevention System

The title tell its story itself ,it's a program or hardware that secure a organize ,it works like a smoke detector ,it raise a hail of peril when it feel security is get bleached and a few pernicious movement is going on in the framework or arrange. The work of anticipation system is to allow reaction to this fade in organize and take necessary step to secure the information of arrange from this breakout .In this kind of circumstance framework like artificial neuron

framework can offer assistance the anticipation framework and it may too offer assistance the interruption discovery framework quick and reliable [14]. The properties of manufactured neuron system like to memorize, handle disseminated adjust information and self-organize are pertinent to tackling problems that require considering conditionality, imprecision and equivocalness at the same time[14,15],help in detecting as well taking counter degree in realtime world.

B. DDOS attack

DDOS is clarified as Dispersed denial-of-service attacks which has gotten to be one of the most internet security issues over the final decade [15]. DDOS is a sort of DOS assault where numerous compromised systems, these computers are frequently tainted with a Trojan; these are utilized to target a single framework causing a Dissent of Benefit (DOS) assault [17].SOM (Self Organizing Outline) is an manufactured neural network, which is based on the competitive learning.

Virus Detetion using Artificial Neural Networks

Computer infection may be a kind of risk that causing billions of dollar to numerous companies and government. These are really malevolent programs that are made to reproduce itself and causes harm to the host computer. The issue with today's antivirus programs is they distinguish infection on premise of known pattern of infection, so discovery of unused infection is troublesome. Again the models like SOM (Self-Organizing Outline) of ANN help in tackling the issue as SOM capability of gridding the closely resembling information and quick speed offer assistance in the location of unused infection.

III.Artificial Immune System & Expert Systems

Artificial Immune System is created to do the work like biological immune system i.e. to adapt itself in the changing environment and releasing the antibodies against the dangerous threat to a computer or network.. A safe Framework is made to do the work like natural resistant framework i.e. to adjust itself in the changing environment and discharging the antibodies against the unsafe danger to a computer or network. It is additionally required to see into science for a little inspiration. From thousands of year immune system of human being is battling from a few viruses and interlopers, and the reason behind it is the efficiency of body in keeping on learning and making strides itself along with the expanding ability of gate crashers. Another reason that produces safe framework so adaptable is its ability to separate between self and others and this makes it so effective [15]. Basically Manufactured resistant framework have following properties: Location: classification or location take put in immune framework when contaminated components get attached with tactile cell surface. Expert framework is one of the foremost broadly utilized A.I tools. A Master framework program planned to discover answers to various application space questions postured either by a software or a few user [12]. It is ordinarily utilized to support some decision-making errands. At show there are many expert framework that's being utilized at display in many organizations to illuminate the Complex and sophisticated problems .The master framework are the primary effective form of master framework utilizing AI method. Master frameworks are of two sorts specifically: Deduction Motor and Knowledge Base frameworks. A Information base framework speaks to facts and rules whereas Induction Motor is utilized to apply the rules to the know truths. Inference Motor may too include explanation and investigating capability. With the development of computer application and network technology number of cyber assaults are too increasing. Commercial or say open source interruption detection systems are autonomously not able to fathom this problem alone.

IV.Conclusion

The internet opens the unused entryways for businesses, governments and common individuals to realize new heights in their work. This paper gives a glimpse of the cyber assaults, interruption etc. cyber wrongdoings. These situation strengths to see up to more capable techniques such as Fake Insights based approaches to combat cyber violations. The ever expanding DDOS attacks, computer infections, worms, Trojans and logical bombs etc. donate rise to the advancement of devices such as manufactured neural systems, Cleverly Specialists, Expert Systems and Fake Safe Framework to settle or avoid these issues. The arrange based frameworks are still under threat and show security procedures are wasteful to protect against these destructive dangers. As thousands of GBs of information voyages through organize, unimportant human supervision and conventional security approaches are unable to coordinate the capacity and effectiveness of falsely intelligent systems.

References

[1] Mary Ellen O'Connell, "Cyber Security without Cyber War, To Defend The Web: Using ArtificialIntelligence As Online Security"

[2] Rain Ottis and Peeter Lorents," Cyberspace: Definition and Implications" , Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia.

[3] Selma Dilek1, Hüseyin Çakýr2 and Mustafa Aydýn3,"Applications Of Artificial Intelligence Techniques to Combating Cyber Crimes",International Journal of Artificial Intelligence & Applications (IJAIA),

[4] Jyothsna S Mohan1, Nilina ,"Prospects of Artificial Intelligence in Tackling Cyber Crimes", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14, Impact Factor (2013): 4.438, Department of Computer Science and Engineering, College of Engineering, Vadakara, Kozhikode, Kerala, India.

[5] Artificial Intelligence in CyberDefense,Enn TyuguR&D Branch,Cooperative Cyber Defense Center of Excellence (CCD COE)and Estonian Academy of Sciences2011 3rd International on Cyber Conflict.

[6] Raul Rojas, Neural Networks Systematic introduction, book. Prospects of Artificial Intelligence in tackling cyber crime.

[7] John A. Bullinaria , 2004. Introduction to Neural Networks : Lecture 16

[8] Himali Jani1 ,Sathvik Shetty1 ,Kiran Bhowmick ,"Virus Detection using Artificial Neural Networks Shivani Shah1".

[9] Enn Tyugu ,"Artificial Intelligence in Cyber Defense", R&D Branch ,Cooperative Cyber Defense Center of Excellence (CCD COE), Intelligent_agent.

[10] Jyothsna S Mohan1, Nilina ,"Prospects of Artificial Intelligence in Tackling Cyber Crimes", International Journal of Science and Research (IJSR).

[11] Raul Rojas, Neural Networks Systematic Introduction, book. Prospects of Artificial Intelligence in tackling cyber crime.

[12] P.Brangetto,M.Maybaum,J.Stinissen (Eds.),Caitríona H. Heinl,"Artificial (Intelligent) Agents and Active Cyber Defence"2014 6th International Conference on Cyber Conflict, Research Fellow, Centre of Excellence ForNational Security (CENS) S. Rajaratnam School of International Studies Singapore ,"International,Journal of Science and Research (IJSR)", ISSN (Online): 2319-7064, Index ,Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438.

[13] "Artificial Intelligence in Tackling cyber crime", Vijayalakshmi, Palanichamy Yogesh and Arputharaj Kannan Intelligent feature selection and classification techniques for intrusion detection in networks, Sannasi Ganapathy*, Kanagasabai Kulothungan, Sannasy,Muthurajkumar, Muthusamy

[14] Igor Kotenko , Alexander Ulanov ,"Agent-Based Simulation of Ddos Attacks And Defense Mechanisms St.-Petersburg Institute for Informatics and Automation of Russian Academy of Sciences, 39, 14th Liniya.

[15] By Shelly Fan," How Artificial Immune Systems May Be the Future of Cybersecurity", on Dec-27-2015, Artificial Intelligence, Computing, Featured, Tech.

[16] Hua Yang,1,2 Tao Li,1 Xinlei Hu,1 Feng Wang,1 and Yang Zou1 "A Survey of Artificial Immune System Based Intrusion Detection".

BCIIT e-Magazine 2021

www.bciit.ac.in

A Survey on Load Balancing in Cloud

Pooja Arora
Assistant Professor
Banarsidas Chandiwala Institute of Information Technology
Kalkaji, Delhi
pooja@bciit.ac.in

Abstract: *As a new prevalent commercial paradigm, cloud computing has attracted attention from both academic and industrial communities. Thanks to the advanced development of cloud computing, many enterprises and individuals are allowed to outsource the considerable amount of data to cloud instead of building and maintaining local data centers. This paper proposes a load balancing technique, named Elephant Herd Grey Wolf Optimization (EHGWO) for balancing the loads. The proposed EHGWO is designed by integrating Elephant Herding Optimization (EHO) in Grey Wolf Optimizer (GWO) for selecting the optimal VMs for reallocation based on newly devised fitness function. The proposed load balancing technique considers different parameters of VMs and PMs for selecting the tasks to initiate the reallocation for load balancing. Here, two pick factors, named Task Pick Factor (TPF) and VM Pick Factor (VPF), are considered for allocating the tasks to balance the loads.*

Keywords: *Cloud computing, load balancing, Elephant Herding Optimization, Grey Wolf Optimizer, reallocation, pitch factors*

I. Introduction

Cloud computing has emerged as a mainstream service-oriented architecture. The large-scale application of cloud computing also brings increasing number of tasks and surging amount of workloads [17][6]. However, due to different computing capacities of nodes and uneven task scale, some computing nodes within cloud may be underutilized while others may be overloaded, resulting in unbalanced load distribution [15][6].Therefore, it is imperative to spread the loads across computing nodes to take full advantage of cloud computing system and consequently improve user satisfaction [16][6]. The cloud users may also enjoy various types of computing services offered by public cloud [12].Cloud computing has been defined by NIST as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or cloud provider interaction. Cloud computing can be considered a new computing paradigm insofar as it allows the utilization of a computing infrastructure at one or more levels of abstraction, as an on-demand service made available over the Internet or other computer network. Because of the implications for greater flexibility and availability at lower cost, cloud computing is a subject that has been receiving a good deal of attention [11].

The aim of load balancing (LB) is to clearly understand the consumer requirements, the data and information can be sent and received without taking more time. LB in CC is one of the major problems without load balancing users could delays, and provide time-consuming system responses, the load can be network load, memory and CPU loads etc4. LB is the method of increasing the performance of Distributed System (DS). It is the method of transferring the load among different processors of DS to improve job response time and resource utilization while also ignoring a condition where few processors are overloaded while other processors are idle or doing under loaded work at any given instant of time in the system. LB calculates various terms like minimizing communication delays, minimizing execution time, maximizing

throughput and maximizing resource utilization [10]. A typical distributed system such as Expert Cloud includes the number of distributed HRs. This HRs can be connected to each other to achieve the high performance and execute the task that one HR is not able to do it by itself. To reduce the task execution time by the HR, the workload should be distributed based on the HRs strength. This makes the load balancing necessary. The main purpose of the load balancing is that it facilitates networks and resources by providing a maximum throughput with minimum response time. Dividing traffic among users, and as a result data can be sent and received without major delay [13][4].Load balancing redistributes workloads among cloud with carefully designed strategies that guarantee to maximize resource utilization and avoid overload [14][6].Load balancing is a methodology which provides methods to maximize throughput, utilization of resources and performance of system. As a part of its services, it gives easy and flexible process to keep data or files and make them available for large scale of users [9]. If too many tasks are crowded on certain nodes, tasks could be switched from heavy-burdened nodes to light-burdened ones to reduce the waiting time of tasks at nodes, which is called load balancing. It is commonly stated that task allocation and load balancing are crucial to the distributed systems [16].

To make the use of resources most efficiently in cloud system, there are several load balancing algorithms. Load balancing algorithms can be categorized into two types: static and dynamic, depending on whether load balancing decision is based on current load state or not. Static algorithms cannot adapt to dynamic changes of system once task execution starts. In contrast, dynamic algorithms make balancing decisions according to load status during runtime and adjust accordingly to redistribute workloads as necessary. This leads to a wide range of researches on dynamic algorithms for better performance. Diffusion methods and meta-heuristic-based methods are two main types of dynamic load balancing algorithms. Diffusion mechanism [18][6] refers to the collective strategy penetrations of some agents on others. The interaction between neighboring nodes within cloud can be abstracted as a diffusion process [19][6]. However, the diffusion method can only achieve locally optimal result. Metaheuristic- based approaches were raised by offering near-optimal solutions in acceptable time using particle swarm optimization (PSO), genetic algorithm (GA), ant colony optimization (ACO) and so on. A meta-heuristic called Grey Wolf Optimizer (GWO) is developed in [20] inspired by grey wolves (Canis lupus). The GWO algorithm mimics the leadership hierarchy and hunting mechanism of grey wolves in nature. Here, four types of grey wolves such as alpha, beta, delta, and omega are employed for simulating the leadership hierarchy. Meanwhile, a new kind of swarm-based metaheuristic search method, called Elephant Herding Optimization (EHO), is proposed in [21] for solving optimization tasks. The EHO method is inspired by the herding behavior of elephant group. In EHO, the elephants in each clan are updated by its current position and matriarch through clan updating operator. However, performance of the incorporation relies highly on the matching level between meta-heuristic mechanism and system model.

II. Challenges

- Load balancing is one amongst the most challenges in cloud computing. It's needed to distribute the dynamic work equally across all the nodes to attain a high user satisfaction and resource utilization [8].
- Response Time (RT) is described in distributed system how much time takes to response for particular load balancing. In the load balancing algorithms, dynamic algorithm take more response time as opposite static algorithm take shorter response time [10].

- Cloud computing has shown enormous advantages in service-oriented computation. However, load balancing for multi-class system resources is a challenge due to the difficulties in load data fusion [6].
- The challenge is to remove and avoid drawbacks of cloud computing system when the load is dynamically distributed by virtual machine [9].
- Performance (PR) is very important for any computing environment for overall efficiency of the system. In load balancing algorithm, all parameter are improved then can be improve the overall system performance [10].
- In [6], FWPFC-LB method is developed for load balancing. Even though the inclusion of feature weight preferences into clustering process enhances the precision of virtual machine classification during data fusion, memory, CPU, and system utilization are lower than the existing methods. This indicates the performance of FWPFC-LB needs further improvement.

III. Proposed Methodology

The primary intention of this research is to design and develop a technique for load balancing to improve the efficiency of the cloud computing system. Here, load balancing is performed to remove the tasks from over loaded VMs and assigning them to under loaded VMs without affecting the system performance. Accordingly, Elephant Herding-based Grey Wolf Optimizer (EHGWO) will be newly developed algorithm to perform the load balancing. Initially, the capacity and loads of the virtual machine will be found based on the executed tasks, then, the balance of the cloud system will be checked. When the load is found unbalanced, the capacity with load will be checked to take the decision whether load balancing can be done or not. In the other case, find the tasks to be removed by checking the two constraints like, load of the task and cost of the load balancing. Then, the removed tasks will be added in other VMs by optimally finding the VMs for the task execution. The optimal finding of VMS for executing of the removed task will be found out using the proposed EHGWO, which will be designed by combining Elephant Herding Optimization (EHO) [21] and Grey Wolf Optimizer (GWO) [20]. The implementation of the proposed approach will be in JAVA with Cloudsim tool. The performance of the proposed technique for load balancing will be evaluated with different cloud set up for makespan, and the results attained will be compared with that of existing works [1] [2] and [3].

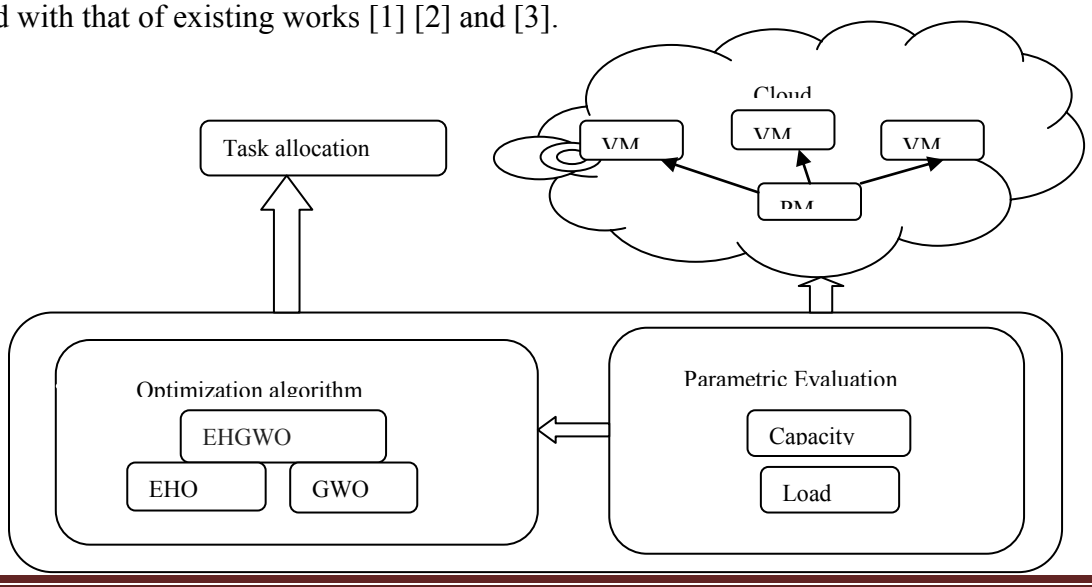

Figure 1. Block diagram of theproposed EHGWO based load balancing method in Cloud computing platform

IV. Expected Outcome

In [3], the performance of the LB-BCmethod is evaluated using makespan as one of the metrics, wherein the minimum makespan is approximately 550 sec. However, the proposed load balancing method will be implemented in such a way to reduce the makespan than that in [3].

References

[1]Shang-Liang Chen, Yun-Yao Chen ,Suang-Hong Kuo,"CLB: A novel load balancing architecture and algorithm for cloud services",Computers & Electrical Engineering,Vol:58, pp:154-160, February 2017.

[2] Qi Liu, Weidong Cai, Jian Shen, Xiaodong Liu, Nigel Linge, "An Adaptive Approach to Better Load Balancing in a Consumer-centric Cloud Environment," IEEE Transactions on Consumer Electronics, Vol: 62, no: 3, pp: 243 - 250, August 2016.

[3] Jia Zhao, Kun Yang, Xiaohui Wei, Yan Ding, Liang Hu, and Gaochao Xu, "A Heuristic Clustering-based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment", IEEE Transactions on Parallel and Distributed Systems, Vol: 27, no: 2, pp: 305 -316, February 2016.

[4] Shiva Razzaghzadeh, Ahmad HabibizadNavin, Amir Masoud Rahmani, and Mehdi Hosseinzadeh, "Probabilistic Modeling to Achieve Load balancing in Expert Clouds", Ad Hoc Networks, January 2017.

[5]Ranesh Kumar Naha and Mohamed Othman, "Cost aware service brokering and performance sentient load balancing algorithms in the cloud", Journal of Network and Computer Applications, Vol: 75, pp: 47–57, November 2016.

[6]Weihua Huang, Zhong Ma, Xinfa Dai, Mingdi Xu and Yi Gao, "Fuzzy Clustering with Feature Weight Preferences for Load Balancing in Cloud", International Journal of Software Engineering and Knowledge Engineering,Vol: 28, no:5,pp:593–617,2018.

[7]Narander Kumar and Diksha Shukla,"Load Balancing Mechanism Using Fuzzy Row Penalty Method in Cloud Computing Environment",Information and Communication Technology for Sustainable Development, pp: 365-373,2017.

[8]Santanu Dam, Gopa Mandal, Kousik Dasgupta and Parmartha Dutta,"An Ant-Colony-Based Meta-Heuristic Approach for Load Balancing in Cloud Computing", Applied Computational Intelligence and Soft Computing in Engineering, pp: 29,2018.

[9]Shalini Joshi and Uma Kumari, "Load Balancing in Cloud Computing:Challenges& Issues," In proceedings of 2nd International Conference on Contemporary Computing and Informatics (IC3I), December 2016.

[10]MahfoozAlam and Zaki Ahmad Khan," Issues and Challenges of Load Balancing Algorithm in Cloud Computing Environment", Indian Journal of Science and Technology, Vol: 10, no:25, July 2017.

[11]Wayne Jansen and Timothy Grance, "Guidelines on security and privacy in public cloud computing",pp: 800-144, 2011.

[12] Jianting Ning, Zhenfu Cao, Xiaolei Dong, Kaitai Liang, Hui Ma and Lifei Wei," Auditable -Time Outsourced Attribute-Based Encryption for Access Control in Cloud Computing",IEEE Transactions on Information Forensics And Security, 2017.

[13]Zenon Chaczko, Venkatesh Mahadevan, ShahrzadAslanzadeh and Christopher Mcdermid," Availability and Load Balancing in Cloud Computing", In proceedings of International Conference in Computer and Software Modeling, IPCSIT, Vol:14 , 2011.

[14]Daraghmi, E.Y. and Yuan, S.M.,"A small world based overlay network for improving dynamic load-balancing", Journal of Systems and Software, Vol: 107, pp:187-203,2015.

[15]A. S. Milani and N. J. Navimipour," Load balancing mechanisms and techniques in the cloud environments: Systematic literature review and future trends", J. Netw. Comput. Appl. Vol:71, no:1, pp: 86–98,2016.

[16]Y. C. Jiang, "A survey of task allocation and load balancing in distributed systems", IEEE Trans. Parallel Distrib. Syst.,Vol: 27no:2,pp: 585–599,2016.

[17]G. Mateusz, G. Alicja and B. Pascal," Cloud brokering: Current practices and upcoming challenges", IEEE Cloud Comput.,Vol: 2,no:2,pp: 40–47,2015.

[18]I. D. Falco, E. Laskowski, R. Olejnik, U. Scafuri, E. Tarantino and M. Tudruj, "Extremal optimization applied to load balancing in execution of distributed programs", Appl. Soft Comput.,Vol: 30,no:3,pp: 501–513,2015.

[19] Y. Jiang, "Concurrent collective strategy diffusion of multi agents: The spatial model and case study", J. Netw. Comput. Appl.,Vol: 39,no:4,pp: 448–458,2009.

[20]Mirjalili, S., Mirjalili, S.M. and Lewis, A.,"Grey wolf optimizer" Advances in engineering software,Vol: 69, pp.46-61,2014.

[21]G. Wang, S. Deb and L. d. S. Coelho, "Elephant Herding Optimization, "In proceedings of 3rd International Symposium on Computational and Business Intelligence (ISCBI),pp:1-5,2015.

BCIIT e-Magazine 2021

www.bciit.ac.in

Impact of Social Media on Youth

Rama Bansal
Assistant Professor
Banarsidas Chandiwala Institute of Information Technology
Kalkaji, Delhi
rama.bansal987@gmail.com

Abstract: *Students use social media and get encouraged by way of it. It relies upon on the usage that whether or not they are inspired in right or wrong route. Youngsters typically don't discuss about knowledge; they just share it or like it and the problem just stays focused on web only. Despite the superb benefit of rapid sharing of information, social media permits people to create fake identities and superficial connections, causes despair and is a number one recruiting device of criminals and terrorists.The overuse of those social media sites on each day has many negative consequences on the physical and intellectual health of college students making them lethargic and unmotivated to create contact with the human beings. This study is focused on the impact of social media on students and youth which can be negative or positive.*

Keywords: *Social Media, Impact, www, Internet, Youth*

I. Introduction

Students prefer spending a plentiful quantity of time on those social networking web sites which maintains them far from their very own reason of existence and interacting. Their social gatherings are hampered because browsing these social networking websites holds them extra concerned for which they're bound to ignore other social activities in their lives. However a few positives consequences are also there like meeting humans you may not have met outside the social media boards, sharing ideas past the geographical boundaries. Some of the studies which I reviewed imply the bad impact of social media in phrases of Time spend, existence style, fitness danger, privacy and so on.

Teens are becoming extra privy to the social troubles in particular from Facebook like sites. However as a substitute of getting the information the teens usually don't discuss them, they just share it or love it and the difficulty just stays centered on internet only. Most of the people suppose that teenagers can play a fantastic role in changing our society. Some research monitor that, Facebook and WhatsApp are the most famous among many of the youths; it offers people with a manner of retaining and strengthening social ties which can be beneficial to both social and academic sports.

It's very critical to overcome this problem. How can parents alleviate the poor components of social media whilst enhancing upon the fine consequences? Moderating their get admission to social media is one super technique. Maximum of the bad aspects may be conquering by using reducing the amount of time spent on social community sites. Being attentive to their educational progress and addressing any troubles will move an extended way in the direction of preserving the negative components of social media from influencing them. Offer enough time for face-to-face social interplay, like having some circle of relatives entertainment time in which you speak their studies in a comfy surroundings or inviting friends and own family over for get-together, providing a laugh, face-to-face social interaction with cherished ones. All this we help us to reduce the terrible impact of social media on the students with a view to in flip advantage our younger era.

It focused on understanding digital conduct, and perceptions of chance and protection amongst these active, young users of virtual and social media. As this turned into ordinarily a qualitative observe, the findings are not always representative of Kenyan young humans at large.

II. Literature Review

Khurana N (2015) located that "The goal organization prefers spending a plentiful quantity of time on those social networking web sites on a median of greater than 2 hours a day which keeps them faraway from their personal motive of existence and interacting with their own herbal environment. Their social gatherings are hampered because surfing those social networking sites maintain them extra concerned for which they are bound to disregard different enormous social activities in their lives."

Dr. A. Jesu Kulandairaj (2014) stated that Social Networking web sites influence the way of life of youth in order that the brands and corporations can make the most the space of Social Networking sites to create loyalty amongst kids. If the Social Networking web sites promote a healthful life fashion via its posts, videos and messages, so that it will help to increase a healthful young era.

Syed Muhammad et al. (2014) unearths that the immoderate users inside the instructional computer labs use the social media forms for comments, chatting, image and video sharing and texting and many others. This common touch the nearly half of the sampled populace. This shows that they ignore their number one cognizance on their take a look at and studies related activities while making use of the ability of internet in connecting with their buddies on the social media networking boards with their average utilized time between 30 to 60 mins.

Rita Njoroge (2011) examines the troubles of the relation between social media and its effect on behavior alternate of the teens.

Keli Wheeler (2015) observed that "negative impacts of social media are starting to surface making procrastination easier and dozing tougher. these issues can handiest grow and boom the manner that the net does. It's a incredible idea to display teens usage of media to understand what they are doing, in which they may be going, and how much time is simply being spent on networks."

Annapoorna Shetty et al. (2015) stated high quality use of social media can expand the teenagers's instructional profession, their capabilities, higher dwelling style, to adopt new tendencies, fashion, and anthropology so on.

Parvathy J., Suchithra R. (2015) founds that every era has its positives and negatives and those who are using has to extra cautious in using them and asked most effective use them for right purpose.

Ms. Shabnam et al. (2014) discovered that children are getting greater privy to the social issues specially from fb. but instead of having the expertise the youth typically don't speak them, they simply percentage it or find it irresistible and the issue simply remains targeted on web most effective. most people assume that teenagers can play a effective position in changing our society that is represent in most of the responses to distinctive queries.

Selasi Kwame Ocansey et al. (2016) concluded that the teens should be knowledgeable on better utilization of social media as a way to reduce time wastage on chatting and different irrelevant engagements that are not of predominant significance on their lives. also, stakeholders and network companies need to come up with way of filtering records that reaches the younger humans thru social media structures. this could curtail exposing them to pornographic and other unwarranted materials. To communicate efficiently to the young people, we recommend that colleges and authorities corporations need to embody social media as one among their manner of verbal exchange.

Whitney Sue Thoene(2012) concluded that average, college students who use social media extra frequently get hold of extra correspondence from corporations, and those college students then use the promotions. consequently, businesses have to use Facebook and Twitter to achieve the patronage of college students however be cautious of overloading them with too much data.

Dr. Pooja Deshmukh et al. (2014) found that Social Networking sites are very popular most of the youths with most people of them pointing out that they are energetic members of social networks.

Dr. Indrajit Roy chowdhury , Mr. BiswajeetSaha. (2015) said that thru this social networking site platform offers a chance among the young generation in Kolkata however also its useful blessings to customers are friendlier and they're frequently meet for some time whether they don't have enough time to meet to make a platform for eco-communication.

III. Research Methodology

The research is exploratory in nature. Research papers have been retrieved from the web journals of SAGE, technology Direct, Emerald and different online resources like Google, Google pupil, review and so on. since the examiner turned into focused at the social media effect amongst college students, so the key phrases like social media, youngsters, college students, verbal exchange with own family were used to discover the relevant papers for analysis.

IV. Conclusion

Social, Media, People, youth, online, information, sites etc. words are most frequently occurring words. Literature represents the social media impact on youth which can be positive and negative both. Study represents high relation between social & Media; People, use, online & sites; Youth, networking & information. This represents that we have researches on social media with relation to youth, networking and information. Women factor with respect to social media is rarely considered by the researchers. Women is very important factor which has to be considered in relation with social media.

References

[1] Boyd, D. (2008). Why Youth Heart Social Network Sites: The Role of Networked Publics in Teenage Social Life, 119-142. Doi: 10.1162/dmal.9780262524834.119

[2] Bryant, J.A., Sanders-Jackson, A., Smallwood, A. (2006). Iming, Text Messaging, and Adolescent Social Networks. Journal of Computer Mediated Communication, 577- 592.

[3]Clarke-Pearson, K.J., O'Keeffe, G. (2011). The Impact of Social Media on Children, Adolescents, and Families. American Academy of Pediatrics, 800-804. Doi: 10.1542/peds.2011-0054

[4] Duggan, J., Heath, N., Lewis, S., Michal, N. (2012). Non-Suicidal Self-Injury, Youth, and the Internet: What mental health professionals need to know. Child and Adolescent Psychiatry and Mental Health, 6(13), 1-9.

[5] Hedinger, K. (2014). Retrieved from: Karina Hedinger Kaiser, V.J. (2010). Generation M2: Media in the Lives of 8-18 Year Olds. The Kaiser Family Foundation.

[6] Larsen, M.C. (2007). Understanding Social Networks: On Young People's Construction and Co-Construction of Identity Online, 1-38.

[7] Lenhart, A. (2012).Teens, Smartphones,and Texting. Pew Internet and American Life Project, 1-33.

[8] Lenhart, A., Madden, M., Purcell, K., Rainie, L., Smith, A., Zickuhr, K. (2011).

[9] Lenhart, A., Madden, M., Purcell, K., Smith, A., Zickuhr, K. (2010).

[10] The Role of Social Networking Sites in Early Adolescents Social Lives. The Journal of Early Adolescence, 36(8), 348-371.

[11] Tanner, E. (2015). Girls, Instagram, and the Glmaorization of Self-Loathing, 4(1), 105-117. Toft, J. (2014).

BCIIT e-Magazine 2021

www.bciit.ac.in

Understanding Blockchain: Architecture, Applications and Challenges

Sandeep Jain
Assistant Professor
Banarsidas Chandiwala Institute of Information Technology
Chandiwala Estate, Kalkaji, New Delhi.
sandeep@bciit.ac.in

***Abstract**: Blockchain, the inspiration of Bitcoin, has received extensive attentions recently. Blockchain is an immutable ledger which allows transactions happen during a decentralized manner. Blockchain-based applications are arising, covering numerous fields including financial services, reputation system and Internet of Things (IoT), and so on. However, there are still many challenges of blockchain technology like scalability and security problems waiting to be overcome. This article shows the comprehensive overview on blockchain technology. We provide an overview of blockchain architecture, features, characteristics and challenges of blockchain.*

***Keywords:** Blockchain, Bitcoin, Cryptocurrency, Decentralized application*

I. Introduction

A blockchain is actually a distributed, decentralized database of public ledger of all transactions or digital events that are executed and shared among participating parties. In 2008 blockchain came into existence and in 2009 it was first implemented. In blockchain, information is chronologically stored in a continuously growing chain of data blocks, implemented in a decentralized network in a way that creates data integrity, trust, and security for the nodes, without the necessity for central authorities or intermediators [1].
The blockchain contains a particular and verifiable record of each single transaction ever made. The blockchain is immutable that is once entered, information can never be erased. Bitcoin, the decentralized peer to peer digital currency, is the hottest example that uses blockchain technology [2]. Businesses that need high reliability and honesty can use blockchain. Blockchain is a cryptographically secured record of transactions stored on decentralized network, where each block contains various transactions which are approved after complex consensus algorithm [1][2].

II. Types of Blockchain

A. Public Blockchain

A public blockchain may be a non-restrictive, permission-less distributed ledger system. Anyone who has access to the web can check-in on a blockchain platform to become a licensed node and be a neighbourhood of the blockchain network [3]. A node or user which may be a part of the general public blockchain is permitted to access current and past records, verify transactions or do proof-of-work for an incoming block, and do mining [2][3]. The basic use of public blockchains is for exchanging cryptocurrencies and mining. Thus, the foremost common public blockchains are Bitcoin and Litecoin blockchains [3]. Public blockchains are more secure. However, it's only risky when the participants don't follow the safety protocols sincerely.Example: Bitcoin, Ethereum, Litecoin.

B. Private Blockchain

A private blockchain may be a restrictive or permission blockchain operative only during a closed network. Private blockchains are usually used where only some members are participants of a blockchain network [3]. The level of security, authorizations, permissions, accessibility is in the hands of the controlling organization

[3]. Thus, private blockchains are similar in use as a public blockchain but have a little and restrictive network.
Examples: Multichain and Hyperledger projects (Fabric, Sawtooth), Corda etc.

C. Consortium Blockchain

A consortium blockchain may be a semi-decentralized type where quite one organization manages a blockchain network [3]. This is contrary to what we saw during a private blockchain, which is managed by only one organization. Consortium blockchains are typically employed by banks, government organizations, etc.Examples: Energy Web Foundation, R3, etc.
Table 1 shows the different blockchain platforms having different architecture and works with different programming languages and consensus mechanism.

Platform	Type	Architecture	PL	CM
Achain	Public	Parallel chain	Glua	PoW,PoS, BFT
Azure	Consortium	BaaS	JS, Solidity	PoA
Bitcoin	Public	Single Chain	C++	PoW
Ethereum	Public, Opensource	Single Chain	Solidity	PoW,PoS
Hydrachain	Private	Parallel Chain	Python	BFT, PoI
Hyperledger	Private, Consortium	Single Chain	C++, Solidity	BFT, PoET
IOTA	Public	DAG	Python	PoW
Chain Core	Private	Single Chain	Java	PoA
Eos.io	Opensource	Single Chain	C++	PoS,BFT
Stratis	Private, Consortium	Baas	Java	PoW,PoS
Corda	Opensource	Single Chain	Java	BFT

Table I Different types of Blockchain with their programming language and consensus mechanism.

III.Architecture

Blockchains are secure intentionally and are an example of a distributed computer system with high byzantine fault tolerance. It solves the double spending problem without the necessity of a trusted authority or central server.The block time is that the average time it takes for the network to get one extra block within the blockchain [1]. Some blockchains create a replacement block as frequently as every five seconds [1][6]. The included data becomes verifiable before the time of block complete.

A. *Timestamping*

Cryptocurrencies use various timestamping schemes to avoid the necessity for a trusted third party to timestamp transactions added to the blockchain ledger [3][6].

B. Proof-of-work schemes

The first timestamping scheme invented was the proof-of-work scheme [2]. The most widely used proof-of-work schemes are supported SHA-256 and scrypt. The latter now dominates over the planet of cryptocurrencies, with a minimum of 480 confirmed implementations [3][6].
Some other hashing algorithms that are used for proof-of-work include CryptoNight, Blake, SHA-3, and X11.

C. Proof-of-stake and combined schemes

Some cryptocurrencies use a combined proof-of-work/proof-of-stake scheme. The proof-of-stake may be a method of securing a cryptocurrency network and achieving distributed consensus through requesting users to point out ownership of a particular amount of currency [3]. It is different from proof-of-work systems that run difficult hashing algorithms to validate electronic transactions. The scheme is essentially hooked in to the coin, and there is currently no standard sort of it [2][3].

D. Mining

In cryptocurrency networks, mining may be a validation of transactions. For this effort, successful miners obtain new cryptocurrency as a gift [1][6]. The reward decreases transaction fees by creating a complementary incentive to contribute to the processing power of the network. However, with more people venturing into the planet of virtual currency, generating hashes for this validation has become much more complex over the years, with miners having to take a position large sums of cash on employing multiple high-performance ASICs [1][3]. Thus, the worth of the currency obtained for locating a hash often doesn't justify the quantity of cash spent on fixing the machines, the cooling facilities to beat the big amount of warmth they produce, and therefore the electricity required to run them.

IV.Characteristics of Blockchain

A. *Decentralization*

In centralized transaction systems, each transaction must be validated through the central trusted agency like Banking, inevitably resulting to the value and therefore the performance bottlenecks at the central servers [6]. Contrast to the centralized mode, third party is not any longer needed in blockchain. Consensus algorithms in blockchain are wont to maintain data consistency in distributed network [6].

B. Persistency

Transactions are often validated quickly and invalid transactions wouldn't be admitted by honest miners. it's nearly impossible to delete or rollback transactions once they're included within the blockchain. Blocks that contain invalid transactions might be discovered immediately [6].

C. Anonymity

Each user can interact with the blockchain with a generated address, which doesn't reveal the important identity of the user [1][6].

D. Auditability

Bitcoin blockchain stores data about user balances supported the Unspent Transaction Output (UTXO) model [2]: Any transaction has got to ask some previous unspent transactions. Once the present transaction is recorded into the blockchain, the state of these referred unspent transactions switch from unspent to spent. So, transactions might be easily verified and tracked.

V.Applications of Blockchain

The first application of the blockchain technology was the digital currency Bitcoin, but the blockchain might be a way bigger opportunity than bitcoin. Blockchain technology is one among the primary identifiable large implementations of decentralization models that have the potential to “reorganize all manner of human activity” thanks to their ability to supply frictionless and trust less interaction between people and technology. Below more information is provided for more possible uses of the blockchain technology.

A. Cruptocurrency

Bitcoin was first described during a 2008 paper as "A peer to see Electronic Cash System". Within the beginning of 2009, the primary cryptocurrency became a reality with the mining of the genesis block and therefore the confirmation of the first transactions[4]. The depression within the Eurozone area and particularly in Cyprus forced the Bitcoin price to travel up to $1216.73 [2]. Since then, its price has bogged down steadily, thus the cryptocurrency was characterized by many as a financial bubble [2][4]. During 2013 and 2014, fraud incidents and hacking attacks made both investors and users question their trust to the currency. At that point, the worth was a touch below $300 with the daily volatility being around 3.5% when gold's volatility was 1.2% which of other major currencies between 0.5 and 1% [5]. The reactions globally varied from country to country. Today most developed countries allow the utilization of Bitcoin, but consider it "private money" (USA) or "property" (Germany). However, there are some countries, most of which are in Asia and South America, along with India and Russia where Bitcoin is taken into account illegal.

B. Smart Contracts

An emerging use of blockchain during the last years is that the creation of "smart contracts". The term "smart contracts" appeared in 1994 when Nick Szabo described a computer virus with if-then structure interacting with the important world [4]. Different developers using Bitcoin overlay protocols so as to integrate their activities further expanded this concept. More specifically, they were created platforms where the users can purchase derivatives with Bitcoin (derivatives), issue their own currency (Colored Coins) and use the Bitcoin network as credit for his or her exchanges (Blockstream) [4]. The important revolution on this sector came from the Ethereum Project. The most purpose of this project is to make an independent platform where, employing a programming language, the users can create a virtual contract between them for any purpose they need. A same project named Codius is additionally under developing from the Ripple Labs [4]. The contract is executed when the loop of the commands if then is coming into a result from real-world data [4].

C. Voting

Voting procedures remain in many countries a controversial topic, as incident of electoral fraud (invalid or inaccurate vote, multiple registration) and therefore the big percentage of abstention often shape the ultimate result. The use of blockchain technology to run a voting campaign seems as an efficient solution. The members could hook up with a PC-based system through their computer, laptop or smartphone, using open-source code that's hospitable editing employing a quite authentication (biometric, written) prove their identity to the program. Then, they enter their private key to access their right to vote and using their public key to pick their preference and ensure it [4][5]. So far, some projects are founded that promote voting through blockchain systems. The primary is BitCongress that uses the Ethereum platform to develop its idea supported the scenario that each voter has access to at least one "votecoin" that enhances him to vote just one time and his vote are going to be recorded on the blockchain after the system verifies it [4][5]. Other similar projects are Remotengrity, which provides every physical vote with a cryptographic code so as to verify the authenticity of each vote [2]. Generally, the transformation of the electoral system from paper-based to digital will increase its reliability and therefore the convenience that gives to the voters.

D. Intellectual property rights

Blockchain might be wont to enforce to prove property rights. An honest example of this is often "Proof of Existence". The location allows you to upload a file to certify that you simply had custody of it at a given time. Neither its contents nor your own personal information is ever revealed — rather, all the info within the document gets digested into an encrypted number. Proof of Existence is made on top of the Bitcoin

blockchain.Therefore, the thousands of computers thereon network have now collectively verified your file [4][5].

E. Smart Property

A less known field that the blockchain technology could have a serious impact is that of "smart property". The web of Things (IoT), a technology that connects every home device within the global network, is consistently growing [4]. Smart property may be a combination of IoT and therefore the Blockchain infrastructure that's defined as a physical asset whose ownership is controlled via blockchain with the utilization of contracts [4].

F. Finance

The foremost widely accepted application for the blockchain technology is within the field of finance, because it ensures the much-valued transparency between the trading parties [4]. Every transaction publicly or private equities, stocks, bonds or derivatives might be transcriptwithin the blocks and afterwards be confirmed by the agency for its legitimacy. From now, it's easier to detect fraud cases or concealment through stock market moves [4]. Aside from finance within the traditional form, the blockchain could also improve the contemporary sorts of financing. Crowdfunding also can be improved through blockchain adaptation [2][4].

VI.Challenges in Blockchain

A. Security and Privacy

For any blockchain, a key evaluation parameter is how well the safety and privacy conditions meet the need of the blockchain. Analysing the safety and privacy problems with blockchain may be a broad research area, and a few studies are conducted during this area. Which shows that the term Information utilized in the above context can have multiple meanings like data within the database, smart contract data or transactions [5]. Privacy is often defined as data privacy and user privacy (anonymity). Some cryptographic mechanisms for achieving security and privacy of data subjected to different blockchain layers [5].

B. Scalability Issues

The dimensions of blockchain are continuously growing, and scalability is becoming an enormous problem within the blockchain domain [5]. Scalability depends on the underlying consensus, network synchronization and architecture. To scale the blockchain, the computational power and therefore the bandwidth capabilities should be high for every node within the blockchain, which is practically infeasible [5]. Most of the present blockchains grant limited scalability.

C. Forking

A blockchain fork is actually caused when two miners find a block at almost an equivalent time thanks to a software update or versioning. During a blockchain network, each device or computer is taken into account as "a full node" which runs software to stay the blockchain secure by verifying the ledger [5]. The software is updated to regulate some parameters and to put in new features within the blockchain [5]. This updated software might not be compatible with the old software. Consequently, the old nodes which haven't updated their software and therefore the new nodes which have performed a software update can cause a fork within the blockchain once they create new blocks [5]. There are two sorts of forks: one which isn't compatible with previous software version, called a tough fork, and another one which is compatible with the previous version (backward-compatible), called a soft fork [4][5]. A tough fork happens when there's a big change within the software like change of block parameters or change of consensus mechanism, within the case of Ethereum, a tough fork will occur when it'll migrate from Proof of labor to Proof of Stake.

D. ***Throughput***

It's a measure of the number of blocks appended in blockchain per second which effectively means the number of transactions processed per second. Throughput depends on many factors like underlying consensus algorithm, number of nodes participating in consensus, network structure and node behaviour. Considering these primary factors, attaining high throughput may be a bit hard in blockchain [4][5]. However, to realize high throughput, the dimensions of the transaction are often reduced by excluding some shared blockchain blocks from the nodes which follow old rules. The throughput is often increased by increasing the block size and the bandwidth of the network till a particular level [4][5].

E. Energy Consumption

The mining process of blockchain consumes tons of energy [4][5]. Most of the PoW puzzle based consensus protocols waste an enormous amount of energy. Various consensus algorithms are introduced which use less energy than Bitcoin's PoW like Proof of Stake and PBFT [5]. Energy is additionally consumed during communication over the network. Some cryptographic mechanisms also consume high energy therefore the selection of a correct cryptographic mechanism should be based not only on the memory requirement and the computational load but also on the quantity of energy consumption.

VII.Conclusion

From this article it's clear that the blockchain technology is yet unexplored and has much to supply in many fields. Blockchain can change our life with very imported application used for banking business, voting and trading. Further research suggestions would come with the economic implications and impact of such non-Bitcon blockchain applications.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf.

[2] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang , " An Overview of Blockchain Technology : Architecture, Consensus and Future Trends", 2017 IEEE 6th International Congress on Big Data.

[3] M. Raikwar, D. Gligoroski, K. Kralevska, " SoK of used Cryptography in Blockchain, Sept. 2019.

[4] G. Foroglou, A. Tsilidou, "Further Applications of the Blockchain", at https://www.researchgate.net/publication/276304843 , May 2015.

[5] P.Tasatanattakool, C. Techapanupreeda, " Blockchain: Challenges and Applications", at https://www.researchgate.net/publication/324725048, January 2018.

[6] S. S. Sarmah , " Understanding Blockchain Technology", Computer Science and Engineering 2018, 8(2): 23-29 DOI: 10.5923/j.computer.20180802.02, Aug. 2018.

BCIIT e-Magazine 2021

www.bciit.ac.in

A Review of Machine Learning Techniques

Sakshi Mittal
Banarsidas Chandiwala Institute of Information Technology
Affiliated to GGSIPU, New Delhi, India
sakshim969@gmail.com

Sushma Bahuguna
Banarsidas Chandiwala Institute of Information Technology
Affiliated to GGSIPU, New Delhi, India
sushma@bciit.ac.in

Abstract: *Machine Learning is an area of study and research. As the amount of data and information is going abundant, the machine learning topic is floating on the surface for a few years. Machine Learning is a technology where a machine gets trained and the main motive is to mimic humans. The way human can calculate by analyzing data, the same way machine learning tries to do so but the difference is human is not able to handle such abundant data, but when it comes to machine learning it can handle and analyse ample amount of data within few clicks.*

Keywords:*Machine Learning, Algorithms, Supervised Learning, Unsupervised Learning.*

I. INTRODUCTION

Machine Learning is the hottest catchword moving around. It earns to, as it is one of the greatest thought-provoking subfields of Computer Science. Machine Learning is a way in which a machine learns with its experience. Some amount of input is given as a task, the performance increases after more and more experience given to the machine. For example, when we do some task like playing some sport so at first, the player wouldn't be perfect he may break some rules. On the second try, the player will correct the previous mistakes but can make some new, so it's a continuous experience and learning. In the same way machine learning works. This technology works along with various algorithms that are suitable for different situations. There are various ways in which we can classify different learning approaches i.e. Supervised, Unsupervised, and Reinforcement Learning. All these learning approaches have various applications discussed further in this paper. Data plays a major role in Machine learning. The data is collected, refined andis divided into two parts, that is train data and test data. The training data is used to train the machine and test data is later used to check whether our model is working as expected or not. Machine Learning is not limited to a single domain. It may be applied to multiple domains.

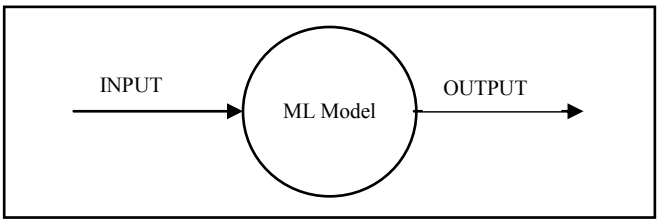

Figure 1: ML Model

The above diagram shows Machine learning model is based on sample input data also known as training data set, in order to make predictions. We give the training data set as input to a machine learning model (algorithm chosen) and then after analysing, we get the output as the result i.e. predictions.

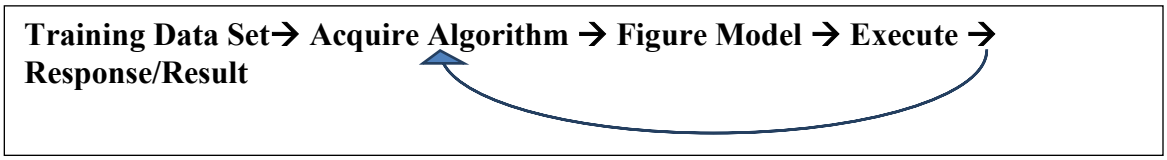

Figure 2: Machine Learning Flow

Figure 2 shows the flow of machine learning. As the last step is feedback, in case we don't get the desired feedback/result, we tend to go back to the second step and try another algorithm and see whether we get the required results or not.

II. A DEEPER DIVE

A. What is data? How it is handled?

Today, data is present in abundance and hence needs various ways to be collected, refined, and finally being analysed. This data abundance needs to be sorted, it cannot be analysed in its raw format. The data is first cleaned and processed so that there is no ambiguity remaining. The second most important step comes out to be dividing the data into training and testing data. This division is important because only this division trains the model and finally we check the accuracy of the model by test data. If the test data values as compared to train data comes out as near to each other, we can state that the model is working correctly.

a. Training Data

Training data is a big chunk of data from the whole dataset that is used to train the machine. It works as instructions to the machine which a machine follows and gets trained. The training data comprise of 75-80% of the total dataset that is for example if we have 1000 data entries, then training data will be around 750 to 800 entries. This division is done so that the machine has a different dataset at the time of testing. If we feed the same training data, the machine tends to detect a pattern and hence it would start giving over fitted results which are of no use.

b. Testing Data

Testing the remaining data, that is kept isolated from the training dataset to evaluate the actual performance of the model. The test data consists of 20-25% of the total dataset. For example, 1000 data entries are there, so the test dataset set will have 200-250 entries. The testing dataset is also known as Validation dataset.

B. Labelled and Unlabelled data

a. Labelled Data

This kind of data is already in a classified format. Classified data means that the data is already a part of some class and is with a tag, i.e. a name or a type etc. For example, if we see any Real Estate dataset, we see that the "number of rooms" is classified in a certain column, "date of construction" in a separate column and so on. The crux is no data that is classless data. Therefore we say that this data happens to be labelled data. For doing analysis on labelled data, we use regression analysis.

Regression analysis is a method of predictive modelling technique that explores the connection amid a **reliant on** (target) and **non-reliant variable (s)** (forecaster). This method is used for forecasting, time sequence modelling, and discovering the fundamental effect relationship among the variables. The purpose of regression analysis to foresee the value of the reliant on the variable for individuals for whom the data concerning the descriptive variables is existing, or in command to estimation the effect of some explanatory variable on the reliant variable.

b. Unlabelled Data

Just the opposite of the labelled data, unlabelled data is the data which does not belong to any class. The data exists in as a chunk, which is not known. What value belongs to what class or what kind of classes are there? So the data needs to be classified first and then used. For this kind of data **Classification** or **Clustering** is done.

Classification is a technique done to categorize unnamed data. It is a technique to put the unnamed chunk of data into a named class where it can be further analyzed. A classic example of classification can be: We may have many documents of type text / images and separating these documents into text document and image document is called as classification. Various algorithms are used for this process; the most popular one is Naïve Bayes Classifier.

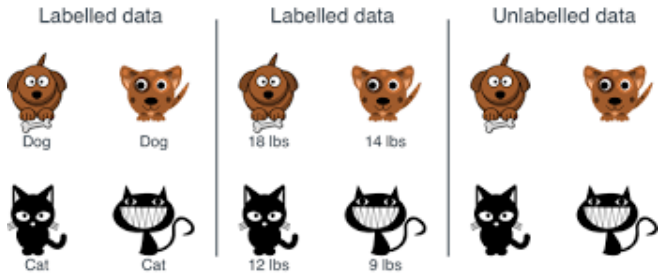

Figure 3: Difference between labelled data and unlabelled data.

Regression and classification, both are the part of Supervised Learning. The model is chosen according to the data.

III. Types of Learning

A. Supervised Learning

Supervised Learning is the type of learning which is applied to labelled data. It is the one, where the knowledge is directed by an educator. We have a dataset which acts as a tutor and its part is to train the model or the machine. Once the model gets trained it can jump start making a prediction or decision when new data is given to it. It is anywhere we take an input variable(X) and resultant variable(Y) and we use an algorithm to study the plotting function from the input to the resultant. Suppose we have a sample of images and we know that what sample of image it is like dog images, cat images or human images. For labelled data supervised learning is used.

X → Y

There are various algorithms under supervised learning [1], [4], [6]:

a. Linear Regression
b. Logistic Regression
c. Decision Tree
d. Random Forest
e. KNN
f. Naïve Bayes Classifier

a. Linear Regression

It is used to estimate real values like cost of house, salary prediction, number of sales in company etc. That means where we have a linear relationship of data and we have to predict the real values like what will be my salary after 10 years or what will be the cost of house after 15 years or what will be the sales figure if company increase the number of advertisements etc. In that case we use linear regression to predict the outcome.

b. Logistic Regression

We use logistic regression when we have to classify the data into two categories, for example we have only the binary categorization of data that is only 0 and 1 values are possible. Logistic regression is used to estimate discrete values such as binary values 0 and 1, yes/no, true/false. This type of regression algorithm can be used in spam detection, whether a person can buy a certain product or not etc.

c. Decision Tree

It is basically used for the classification problem where we have to classify the data into different categories. Decision Tree can be used for more than two categories.

d. Random Forest

Random Forest algorithm is very similar to decision tree algorithm but the only difference is, we use more than one decision tree in the random forest algorithm and hence it produces higher accuracy as compared to the Decision Tree algorithm. Random forest is nothing but the collection of numerous decision trees.

e. Naïve Bayes Classifier

This algorithm is based on Bayes Theorem. In this type of algorithm there is no relation between the different predictors. **Example** If anyone asks what is name of that fruit which is red in color, smaller in size and circular in shape? Some of us would say apple or strawberry or cherry or plum etc. But the output is not clear. So the probability is checked in this algorithm. In any of the output where we get the highest probability, that will be our output. In that case, color has no relation with the shape, shape has no relation size. Hence there is no relationship between the predictors. This is known as Naïve Bayes Classifier algorithm.

f. KNN (K-Nearest Neighbours)

An item is classified by a mainstream poll of its neighbors, or event built on a likeness measure. K is the number of neighbors. For example, there is a football team, and the team is just lacking a single player. There are 5 persons among which one has to be selected. There the captain will take the help of voting from the players already in team, the one who gets the maximum votes will be selected (output).

B. Unsupervised Learning

It is the training of a model using information that is neither classified nor labelled in which there is no explanation of data. Basically this type of learning is applied to that data which is not labelled where we don't know anything about the data. It is like we have a X but we don't have a Y. For example, when we download any kind of audio from internet and the name of the speaker is not mentioned anywhere, this kind of data is called as unlabelled data. Whenever we train unlabelled data, it is known as unsupervised learning [2], [3], [5]. In this type of learning we use the features of the data to find the label/output. This type of learning is also known as **Clustering Analysis**.

We have abundant data which is not labelled at all. So by training that data using the data features, we try to tell the machine to cluster that data which has similar features. For example, we have a collection of documents which include image documents and text documents. So after analyzing their features, the unlabelled data will be divided into two clusters as the output. Cluster 1 can be of image documents and Cluster 2 can be of text documents or vice a versa.

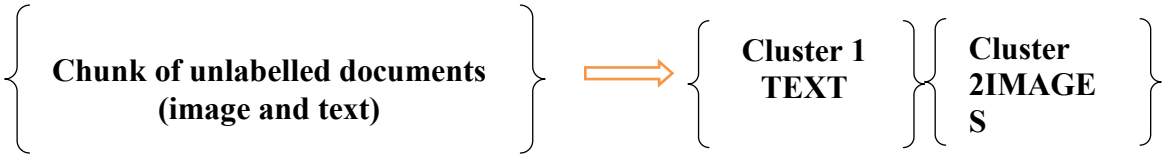

Figure 4: Clustering Analysis

C. Reinforcement Learning

It is an area of machine learning where a Remote Learning agent learns from the consequences of its actions, rather than being taught explicitly. It selects the actions on the basis of its past experiences (exploitation) and also by new choices (exploration). So reinforcement learning consist of exploitation and exploration, by doing these two, we can reach to a good output. Here, the machine takes the decision on its own. This kind of learning can also be called as reward and punishment learning where the output is decided by a past experience

For example, a child who wants to watch cartoon on TV but he doesn't know how to turn on the TV. So the child presses various buttons of the remote. Suddenly TV turns on, so the child remembered the button and learnt from the exploration that this particular button turns on the TV and he will not waste any while whenever the next time he would want to watch TV. Then, he wanted to watch Hungama channel but he didn't know how to change the channel. He again started pressing random buttons, the TV got turned off this time (exploitation), and the child got to know this particular button turns the TV off so from next time I will not press that button. So here we see the small child learnt from his experience and next time he will act (output) according to the experiences gained. The same way reinforcement learning works.

IV. APPLICATION AREAS OF MACHINE LEARNING

A. Virtual Assistant

- Chat Bots
- Siri
- Alexa

B. Image Recognition

- Face recognition

C. Social Media

- Sentiment Analysis
- Spam Filtration

D. Financial Services

- Algorithmic trading
- Fraud Detection

E. Healthcare

- Drug Discovery
- Disease Diagnosis

F. Ecommerce

- Customer Product Support
- Recommendation Systems
- Advertising

V. CONCLUSION

After seeing the meaning of machine learning and learning about the algorithms, we come to a conclusion that, Machine learning is not limited to just one domain but is a dynamic packet in itself. Machine learning is being used in health care, ecommerce, finance, social media, building applications like chat bots. Machine learning wide spread domain in which a lot of studies are still going on. Daily some astonishing features are coming out. In the near future we wish to explore it even more and list more of its commercial as well as exploratory applications.

REFERENCES

[1] Peter Harrington, “Machine Learning in Action”, Manning Publications Co.Development, ISBN 9781617290183, 2012

[2] Shai Shalev-Shwartz and Shai Ben-David, “Understanding Machine Learning: From Theory to Algorithms”, Cambridge University Press, ISBN 978-1-107-05713-5,2014

[3] UmangKejriwal, (May 20, 2020), "Machine Learning Training".YouTube https://www.youtube.com/watch?v=LA4OR6Y1KBM

[4] P. Harrington, "Machine Learning in Action",Manning Publications Co., Shelter Island, New York, ISBN 9781617290183, 2012.

[5] R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Machine Learning, Vol 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992.

[6] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., Indianapolis, Indiana Published simultaneously in Canada, ISBN: 978-1-118- 96174-2, 2015.

BCIIT e-Magazine 2021

www.bciit.ac.in

Sentiment Analysis: A Review

Anu Taneja
Assistant Professor
Banarsidas Chandiwala Institute of Information Technology
Chandiwala Estate, Kalkaji, New Delhi.
anutaneja@bciit.ac.in

***Abstract**: Sentiment Analysis has gained wide popularity due to increased use of social media. It is one of the challenging research areas as it is difficult to keep track of all the activities due to information overload problem. It is one of the important techniques used by businesses to keep track of the sentiments of the customers in the text. It helps in growth of the business and understands the customers better. This paper focuses on the importance, types, applications and challenges faced in the area. This paper would help the researchers as it acts as base foundations who are working in this domain.*

***Keywords:** sentiment, mining, social media, emotions, machine learning*

I. Introduction

Sentiment Analysis is another name of opinion mining that is a sub-branch of machine learning. The large amount of information is stored on various social media platforms like Twitter, Facebook, YouTube and various other e-commerce websites in the form of reviews, tweets, comments etc [3]. The user information is analyzed by such platforms to understand the customers better that help in the growth of business. This domain is gaining wide popularity as otherwise it is very difficult to manually process it and also people are influenced by other thoughts like a user can decide to watch a movie based on other decisions. Now it is used widely in other domains also like prediction of stock markets, election results, and reactions to terror attacks etc. The business platforms analyze such reviews that help to know the satisfaction level of users. For instance, users' opinions can be expressed in different forms like emotions in text that can be further categorized as positive, neutral or negative. Thus, it has become a powerful tool [1] to analyze the emotions hidden in conversations and understand the users in a better form.

II. Types of Sentiment Analysis

The sentiment analysis is broadly categorized into the following [2]:

A. **Fine Grained Sentiment Analysis:** It focuses on polarity of the sentiments, so as per requirements of the business it is being utilized. The polarity category includes very positive, positive, neutral, negative, and very negative. For example, the people express their opinions about the products in the form of ratings on 1 to 5 scale is a good example of fine grained sentiment analysis.
B. **Emotion Detection:** It focuses on detection of emotions like happiness, sadness, anger and other feelings expressed in the conversations. For the detection of emotions, various researchers have utilized lexicons, but the major limitation of lexicons is confronted as emotions can be expressed in different forms.
C. **Aspect-based Sentiment Analysis:** Instead of just analyzing the overall ratings it has become important to analyze the user comments based on features also like whenever a user submits his/her review about a certain product then users might comment about a particular feature of a product like

camera quality of this mobile phone is not much good. This comment clearly indicates that user in his comment is talking camera feature of a mobile phone.

D. **Multi-lingual Sentiment Analysis:** As people belong to diverse areas, thus comments can be submitted in different languages. Nowadays, the work is carried out in this domain as it involves large amount of pre-processing and require translation algorithms too.

III. Working of Sentiment Analysis

The sentiment analysis algorithms [4] fall broadly into three categories:

A. **Rule-Based:** The rule-based systems work based on the manually drafted rules to identify the polarity, opinion of the comment. The natural processing techniques are used to identify the polarity like if more positive words are encountered in a sentence then it is considered in the positive sentiment category. The major drawback is requiring large maintenance to fine-tune the systems.

B. **Automatic-based Algorithms:** It uses machine learning algorithms; sentiment analysis problem can be classified as a classification problem like text will be given as an input to the model and the classifier can categorize into positive, negative or neutral. It basically consists into two parts:

a) ***Training****:* In this phase, the algorithm learns based on the inputs fed to the system, initially, the features are extracted from the text using various approaches like bag-of-words, word embedding and many more.

b) ***Prediction:*** Then, classifier algorithms are applied like Naïve Bayes, Support Vector Machines, Linear Regression, Neural Networks and many more to generate the model.

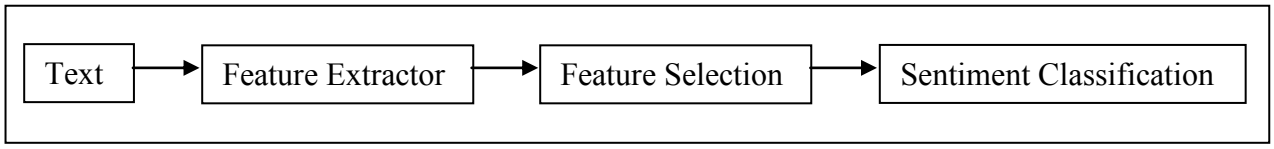

Figure 1: Working of Sentiment Analysis

C. Hybrid Approach: It combines the rule-based and automatic based algorithms, the major advantage of these systems is the results are more accurate.

IV. Applications of Sentiment Analysis

Nowadays, sentiment analysis is used in various areas [2], the few are listed below:

a) Social Media: The people post huge amount of reviews about different products that helps the sellers to analyze the sentiments of users and grow their business in the right direction.

b) **Market Research:** Sentiment Analysis help to know the future trends of the market as it provides deeper insights of the opinion about users like how they feel about a certain product.

c) **Predictions:** The sentiment analysis basically helps in the analysis of data posted in either of the forms like reviews, emojis, ratings applicable to various domains like products, hotels, food, election results

V. Challenges of Sentiment Analysis

It is itself a challenge to accurately analyze the sentiments; the few of the challenges [5], [6] are listed below:

a) **Subjectivity:** The sentences are either subjective or objective but subjective sentences have more explicitly stated sentiments, so it becomes difficult to analyze the sentiments of sentences that are objective in nature.
b) **Recognition of context:** It is difficult to analyze the sentiments without context and machine learning algorithms are not able to detect the context. If the context in which the sentiment is expressed is not identified properly, then the polarity might change [7].
c) **Negative Sentiments:** The users sometimes express the negative feelings in positive words, and if context is not well understood, it would ultimately change the polarity.
d) **Performing comparisons:** It is difficult to compare the sentiments in certain situations.

VI. Conclusion

This research paper describes the sentiment analysis and its importance. The workflow, its applications areas along with the challenges faced is highlighted. The broad categories of sentiment analysis are also briefed up. This domain majorly helps the business to grow as sentiment analysis provides deeper insights about customers' opinions, thoughts, and feelings expressed about a certain product or service utilized on the web. The researchers can work in this direction to overcome the challenges faced and new algorithms can be developed that takes into consideration the context too. This would help to achieve better results and hence the satisfaction level of users would be enhanced to a certain extent.

References

[1] Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415-463). Springer, Boston, MA.
[2] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, *89*, 14-46.
[3] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).
[4] Singh, P. K., & Husain, M. S. (2014). Methodological study of opinion mining and sentiment analysis techniques. *International Journal on Soft Computing*, *5*(1), 11.
[5] Osimo, D., & Mureddu, F. (2012). Research challenge on opinion mining and sentiment analysis. *Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment*, *508*.
[6] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, *28*(2), 15-21.
[7] Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing* (pp. 347-354).

