© 2017 IJSRSET | Volume 3 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Performance Enhancement of Information Retrieval via Artificial Intelligence

Lovee Sharma*[1], Vivek Srivastava[2]

*[1]Department of Computer Science and Engineering, AK Technical University, Lucknow, Uttar Pradesh, India
[2]Department of Mechanical Engineering, BBD University, Lucknow, Uttar Pradesh, India

ABSTRACT

Big data is an ocean of data, most of it unstructured, growing exponentially, requires proper strategies and techniques to deal with it. But it is not the major concern of data scientist whereas cost effectiveness, dealing with unstructured data, managing the data effectively, controlling its growth, making raw data useful, its storage are the major challenges. There has been an increasing interest in the application of AI tools to IR in the last few years. Concretely, the machine learning paradigm whose aim is the design of system able to automatically acquire knowledge by them. In the light of requirement of intelligent processing of the big data so as to retrieve the information as per the business requirement, authors have proposed a novel architecture. Proposed architecture may help in faster information retrieval with better accuracy and recall.

Keywords: Information Retrieval, Artificial Intelligence, Big Data, Map Reduce

I. INTRODUCTION

The intelligence exhibited by machines or software is known as **Artificial intelligence** (**AI**). As human brain perceives its environment and then takes decision similarly research in Artificial Intelligence also aims to make intelligent agent that can understand the environment and then able to take decisions, through artificial neural network. Artificial Intelligence research is deeply divided into subfields that sometimes fail to communicate with each other. Some division is due to social and cultural factors, also divided on the basis of technical issues, like some subfields concerned about the solution of specific problems and others concentrate on approaches or use of particular tool or towards the accomplishment of particular applications. Reasoning, knowledge, planning, learning, natural language processing (communication), perception , the ability to move and manipulate objects are the parts of **Artificial intelligence** .Tools through which research in Artificial Intelligence is implemented versions of search and mathematical optimization, logic, methods, based on probability and economics and many others.The Artificial Intelligence field is interdisciplinary involve number of sciences and professions converge like computer science, mathematics, psychology, linguistics, philosophy and neuroscience, artificial psychology.

Goal of Artificial Intelligence to make software and computers like human brain for this, deep analysis of natural brain is necessary, for this many contribution of all the above written fields become very important.

Today it has become an essential part of the technology industry, providing the heavy lifting for many of the most challenging problems in computer science. Basically Big data is a huge collection of digital information which is growing continuously through many channels like e-mail, social media, user generated content like photos and videos, customer and B2B transactions, GPS-equipped cell phones, mobile devices, social networking apps (whatsapp, hike etc), system logs of all kinds. These sources giving continuous stream of digital information without any definite limit hence it is called as Big data .Big data can act as a driving force that can affect organization's profit. Through Big data one can achieve a better understanding of their customers, employees, partners

and operations. It also provides a futuristic view and many opportunities. For example according to a study of University of Texas if a median Fortune 1000 business increased the usability of its data by only 10%, it would earn an extra $55,900 per employee- or $2+ billion a year in added revenue Mckinsey Global institute says that retailers can also be benefitted by their margins 60% by squeezing more and more value from their Big Data and the U.S healthcare system could save a shocking $200 billion a year. Inspite of being so valuable, all organizations are not able to afford it because of its unstructured data, infrastructure and cost expenses. For its implementation all the departments in an organization have to undergo many changes which is very tedious job and also users across the organization must have power to work with diverse data sets through self-service tools to acquire Big data productivity. After the adoption of Big data, the process of data analysis will become important part of business while it was considered as separate part earlier and not given so much importance. Big data analysis, not only facilitates the company to acquire better understanding or what's happening with the business and why, but also what else is possible.

It is like moving from writing letter through pen and paper to using e-mail or whatsapp. Main objective of data scientist is to convert data into valuable information which can yield rich, wide ranging more accurate and actionable insight that can help address customer needs, operational risks and performance opportunities within the enterprise and the extended supply chain.

II. LITERATURE REVIEW

Thought of artificial beings came from Greek myths, such as Talos of Crete, the bronze robot of Hephaestus and Pygmalion's Galatea. It is believed that research on intelligence were done in every major civilization, like animated cult images were worshipped in Egypt, Greece and humanoid automatons were built by Yan shi, Hero of Alexandria and Al-Jazari.

By the 19th and 20th centuries, artificial beings had become important part of fiction, their stories and fates discuss fears and ethical concerns that are presented by Artificial Intelligence. All these images and stories were the base of Artificial Intelligence but study of logic led directly to invention of the programmable digital electronic computer, by the great mathematician Alan Turing. According to turing's theory of computation by shuffling of 0 and 1 a machine could simulate any conceivable act of mathematical deduction. This research and concurrent discoveries in Neurology Information theory and Cybernetics inspired researchers to possibility of building an electronic brain.

The field of AI research was founded at a conference on the campus of Dartmouth College in the summer of 1956. The attendees, including John McCarthy, Marvin Minsky, Allen Newell, Arthur Samuel, and Herbert Simon, became the leaders of AI research for many decades. By their continuous efforts we saw many astonishing results like computers were solving word problems in algebra, proving logical theorems, winning at checkers and speaking English.

By the middle of the 1960s, research in the U.S. was heavily funded by the Department of Defense and laboratories had been established around the world. AI's founders were profoundly optimistic about the future of the new field: Herbert Simon predicted that "machines will be capable, within twenty years, of doing any work a man can do" and Marvin Minsky agreed, writing that "within a generation ... the problem of creating 'artificial intelligence' will substantially be solved".

In 1974, research in Artificial Intelligence faced lots of setbacks, both the U.S. and British governments cut off all undirected exploratory research in AI, they called that period "AI winter", [30] a period when funding for AI projects was hard to find. In the early 1980s, AI research was revived by the commercial success of expert systems, a form of AI program that simulated the knowledge and analytical skills of one or more human experts. By 1985 the market for AI had reached over a billion dollars. At the same time, Japan's fifth generation computer project inspired the U.S and British governments to restore funding for academic research in the field. However, beginning with the collapse of the Lisp Machine market in 1987, AI once again fell into disrepute, and a second, longer-lasting AI winter began.

In the 1990s and early 21st century, AI achieved its greatest successes, albeit somewhat behind the scenes. Artificial intelligence is used in many areas like logistics, data mining, medical diagnosis and many other areas throughout the technology industry. The success was due to several factors: the increasing computational

power of computers (see Moore's law), a greater emphasis on solving specific subproblems, the creation of new ties between AI and other fields working on similar problems, and a new commitment by researchers to solid mathematical methods and rigorous scientific standards.[34]

On 11 May 1997, Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov. In February 2011, in a Jeopardy! quiz show exhibition match, IBM's question answering system, Watson, defeated the two greatest Jeopardy champions, Brad Rutter and Ken Jennings, by a significant margin.

Earlier people use tally sticks to keep record of data, these tally sticks are used in trading activity and record inventory. They used to keep same number of tally sticks as much items they want to count. But this strategy was not feasible for large calculations and data keeping. Invention of numbers and then computers played a vital role in data storage and calculations. Further scientists worked then invented abacus for calculations and first mechanical computer was developed in Greece.

According to research of Fremont Rider, Wesleyan University Librarian, the size of American University libraries becoming double in every years. With this growth rate Rider estimates that the Yale Library in 2040 will have approximately 200,000,000 volumes, which will occupy more than 6,000 miles of shelves. In 1958 Hans Peter Luhn drives the attention of business executives towards the Business Intelligence and he defines BI as "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal". Various Business Intelligence tools like spreadsheets, Reporting and querying software, OLAP etc providing facility for analyzing data and presenting actionable information to help corporate executives, business managers and other end users make more informed business decisions. The tools generally read data that have been previously stored, often, though not necessarily, in a data warehouse or data mart. Except for spreadsheets, these tools are provided as standalone tools, suites of tools, components of ERP systems, or as components of software targeted to a specific industry. The tools are sometimes packaged into data warehouse appliances. For securing data the US Government plans the world's first data center to store 742 million tax returns and 175 million sets of fingerprints on magnetic tape. First Data Corporation was incorporated in 1971. In 1980, American Express Information Services Corporation (ISC) bought 80% of First Data. The remaining 20% was purchased in 5% increments each subsequent year until June 1983.

First Data Corporation spun off from American Express and went public in 1992. In 1995, the company merged with First Financial Management Corp. (FFMC) and was then organized into three major business units serving card issuers, merchants and consumers. Western Union became part of First Data as a result of the merger with FFMC. After establishment of data center the question of management of data raised, for this Relational Database model was introduced by IBM mathematician Edgar F Codd. The Hierarichal file system allows records to be accessed using a simple index system. This means anyone can use databases, not just computer scientists. Now it was easy to retrieve previous data and store new information through database management system. Most MRP systems used in business units are software-based, while it is possible to conduct MRP by hand as well, therefore computer and data storage is used for everyday routine tasks. The birth of the internet in 1991 brings revolution in digital universe anyone can now go online and upload their own data, or analyze data uploaded by other people, further research in storage area reduced the price of digital storage to the point where it is more cost effective than paper. Google also launched their search engine in 1997 which quickly became the most popular in the world. According to the report Big Data: The Next frontier for Innovation, Competition and Productivity by McKinsey Global Institute the situation is, the average US company with over 1,000 employees is storing more than 200 terabytes of data. In 2010 Eric Schmidt, executive chairman of Google, tells a conference that as much data is now being created every two days, as was created from the beginning of human civilization to the year 2003 way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. RESULTS AND DISCUSSION

The proposed system is consists of several phases. These steps consist of:

1) Input the user query.

2) Apply the Artificial Intelligence based algorithm
3) Returning the data file.
4) Input the user query to the expert system.
5) Analyzer: Searching the user query if it was existed in the historical file or not
6) Return the results.

These steps can be explained in the following phases:
Phase 1: This phase includes selecting the database file to be inputted, to AI algorithm. The inputted database can be contained any type of data.
Phase 2: This phase comprises the selection of the best artificial intelligence based algorithm. The AI based algorithm depends on the best informed search through problem solving agent.

In *informed* search, a heuristic is used as a guide that leads to better overall performance in getting to the goal state. Instead of exploring the search tree blindly, one node at a time, the nodes that we could go to are ordered according to some evaluation function h(n) that determines which node is probably the "best" to go to next. This node is then expanded and the process is repeated (i.e. Best First Search). A* Search is a form of BestFS. In order to direct the search towards the goal, the evaluation function must include some estimate of the cost to reach the closest goal state from a given state. This can be based on knowledge about the problem domain, the description of the current node, the search cost up to the current node BestFS optimizes DFS by expanding the most promising node first. Efficient selection of the current best candidate is realized by apriority queue.

Phase 3: The results of the AI based algorithm are displayed in this phase.
Phase 4: At this phase the users entering the query data to expert system.
Phase 5: This phase includes analyzing the user query by searching in historical file if the user query exists or not.
Phase 6: This is the final phase consists of displaying the query answers through expert system and saving the user query in historical file then return the query answer to the user.
Our methodology is based on intelligent agents which observes through sensors and acts upon an environment using actuators (i.e. it is an agent) and directs its activity towards achieving goals but in this model we are using two types of agents i.e, model based and goal based agents. A model-based agent can handle a partially observable environment. Its current state is stored inside the agent maintaining some kind of structure which describes the part of the world which cannot be seen. This knowledge about "how the world works" is called a model of the world, hence the name "model-based agent". Similarly goal-based agents further expand on the capabilities of the model-based agents, by using "goal" information. Goal information describes situations that are desirable. This allows the agent a way to choose among multiple possibilities, selecting the one which reaches a goal state. Search and planning are the subfields of artificial intelligence devoted to finding action sequences that achieve the agent's goals. Basically the main objective of using agents is to keep track of the world state as well as set of goals it is trying to achieve and chooses an action that will lead to the achievement of its goal.

Function

SIMPLE-PROBLEM-SOLVING-AGENT(PERCEPT)
returns an action
Inputs: percept, a percept
Static: seq, an action sequence, initially empty
State, some description of the current world state
Goal, a goal, initially null
Problem, a problem formulation
state ← UPDATE_STATE(state, percept)
if seq is empty **then do**
goal ← FORMULATE –GOAL(state)
problem ←FORMULATE PROBLEM(state, goal)
seq ← SEARCH(problem)
action ← FIRST(seq)
seq ← REST(seq)
return action

Problem:- A problem can be defined formally by five components:
Initial state:- The initial state that the agent starts in.
Actions:- A description of the possible actions available to the agent. Given a particular state s, ACTIONS (s) returns the set of actions that can be executed in s.
Transition model:- A description of what each action does, the formal name for this is the transition model, specified by a function RESULT (s,a) that returns the state that results from doing action a in state s. We will

also use the term successor to refer to any state reachable from a given state by a single action.

State space:- Together, the initial state, actions, and transitions model implicitly define the state space of the problem- the set of all states reachable from the initial state by any sequence of actions.

The state space forms a directed network or **graph** in which the nodes are states and the links between nodes are actions.

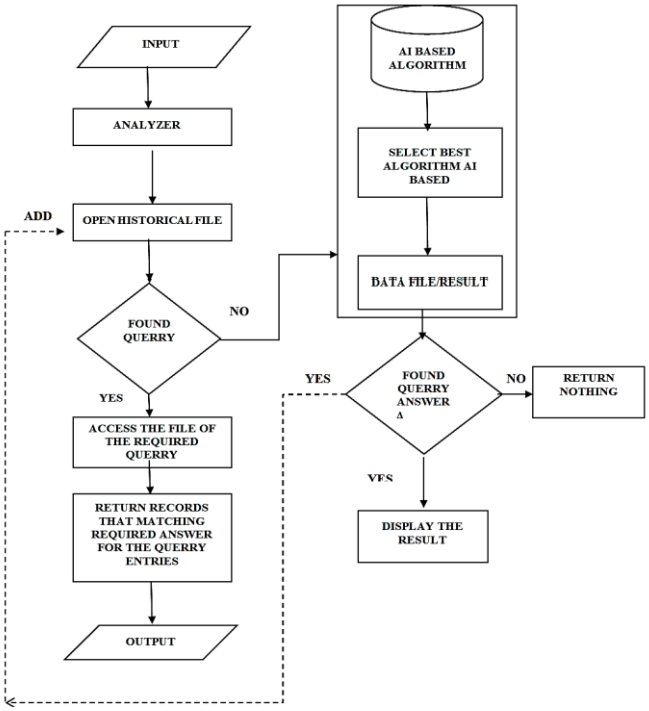

Figure 1: Proposed Architecture

Function

RECURSIVE-BEST-FIRST-SEARCH(problem)
returns a solution, or failure
RBFS(problem, MAKE-NODE(INITIAL-

STATE[PROBLEM]),∞)

Function RBFS (problem, node, f_limit) **returns** a solution, or failure and a new f-cost limit

If GOAL –TEST [problem] (state) **then return** node
Successors ← EXPAND (node, problem)

If successors is empty **then return** failure, ∞

For each s **in** successors **do**
f[s]←max(g(s) + h(s), f[node])
repeat
best← the lowest f-value node in successors
if f[best] > f_limit **then return** failure , f[best]
alternative ← the second –lowest f-value among successors

result, f [best] ← RBFS (problem, best, min (f_limit, alternative))

if result ≠ failure **then return** result

It works by maintaining on the recursion stack the complete path to the current node being expanded as well as all immediate siblings of nodes on that path, along with the cost of the best node in the sub-tree explored below each sibling. Whenever the cost of the current node exceeds that of some other node in the previously expanded portion of the tree, the algorithm backs up to their deepest common ancestor, and continues the search down the new path. In effect, the algorithm maintains a separate threshold for each sub-tree diverging from the current search path.

RBFS is robust and optimal (if the heuristic is admissible), but it still suffers from excessive node regeneration due to its low memory profile, which entails a long processing time. Given enough time, though, it can solve problems that A-star cannot solve because it runs out of memory.

IV. CONCLUSION

Future use of Big Data with the help of Artificial Intelligence is very bright. The use of artificial intelligence will lead to production of machines and computers, which are much more advanced than what we have today. Researchers are continuously working to handle growth of data as well as to convert it into valuable assets. One-day computer will be able to do any work. Speech recognition systems will reach much higher levels of performance and will be able to communicate with humans, using both text and voice, in unstructured English. There will be a great future some day for expert system applications in all aspects of health care, in both clinical and administrative areas, in improving patient care and in allocation of financial, social, and other resources. However, when it comes to the question of Artificial Intelligence creating machines, which are more intelligent than human beings are, no one seems to have the answer. In addition, even if it is possible, the amount of time it will take cannot be

predicted. What can be said with certainty is that machines having common sense will be developed, although it will pertain to specific niche areas only. It is also expected to have human brain features like learning from experience, cognition and perception.

V. REFERENCES

[1] Magoulas, Roger; Lorica, Ben (February 2009). "Introduction to Big Data". Release 2.0 (Sebastopol CA: O'Reilly Media)

[2] Jump up ^ "Manufacturing: Big Data Benefits and Challenges". TCS Big Data Study. Mumbai, India: Tata Consultancy Services Limited. Retrieved 2014-06-03.

[3] Top ten Big Data Security and Privacy Challenges 2012

[4] Best practices for cloud security ; A White Paper by Bloor Research; Fran Howarth ; January 2012

[5] The Future of Human Evolution; The FHE Team

[6] TURNING BIG DATA INTO BIG DOLLARS With Cloud Computing And Open Source Software; Anand Bhadouria Sr. IT Strategist; Enterprise Cloud Solutions

[7] Spatial Big-Data Challenges Intersecting Mobility and Cloud Computing; Shashi Shekhar, Michael R. Evans, Viswanath Gunturi, Kwang Soo Yang Computer Science & Eng. Faculty, University of Minnesota 200 Union Street S.E. #4192, Minneapolis, MN 55455

[8] Applications of big Data: Current Status and Future Scope ; 1Sabia, 2Sheetal Kalra Department of Computer Science & Engineering Guru Nanak Dev University Regional Campus, Jalandhar.

[9] Bart Selman,(moderator) AT & T, Rodney A. Brooks MIT, Thomas Dean (Brown University), Eric Horvitz (Microsoft), Tom M. Mitchell (CMU), Nils J. Nilsson (Stanford University); "Challenge problems for Artificial Intelligence" ; To appear in Proc. AAAI-96.

[10] Avneet Pannu; "Artificial Intelligence and its Application in Different Areas", ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 10, April 2015