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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

Approaches to Measuring the Intelligence of Machines by Quantifying them

Prerna Kapoor[1]

B.E. Student, Department of Information Technology, Thadomal Shahani Engineering College, Mumbai, India[1]

Abstract: Artificial intelligence is a very broad term which covers many fields and thus for measuring how far researchers have reached we need an accurate method of characterizing the intelligence of machines. The reason there isn't one common test for all systems is because intelligence depends on various factors which may vary depending on the environment.The most famous test for measuring intelligence, the Turing Test, has flaws attached to it.Thus, this paper highlights a few of the limitations of the Turing Test and also, describes a few of the other tests used by researchers to evaluate intelligent machines.

Keywords: Artificial Intelligence, Turing Test, intelligent systems, measure of intelligence.

I. INTRODUCTION

The terms Artificial Intelligence (AI), Machine Learning and Intelligent systems are all inter-related. According to Arthur Samuel (an American pioneer in the field of machine learning, computer gaming and artificial intelligence), the term'Machine Learning' is defined as a "Field of study that gives computers the ability to learn without being explicitly programmed". Human Intelligence is measured by IQs but no such method or test has been invented for machines. There are several reasons for this such as – Most IQ tests have a limited scope and thus systems can be programmed to ace the IQ test but it may not be truly intelligent. Secondly, few researchers state that an intelligent system is one which can perform a particular task well by learning new strategies on its own such as the Deep Blue [for chess playing strategies] but other researchers believe that an intelligent system should have a more general approach i.e. if it has learnt a complicated procedure it should easily be able to learn faster methods to solve easier puzzles.A machine can truly be called Intelligent if and only if it satisfies the latter condition.

II. PROBLEMS WITH MEASURING INTELLIGENCE

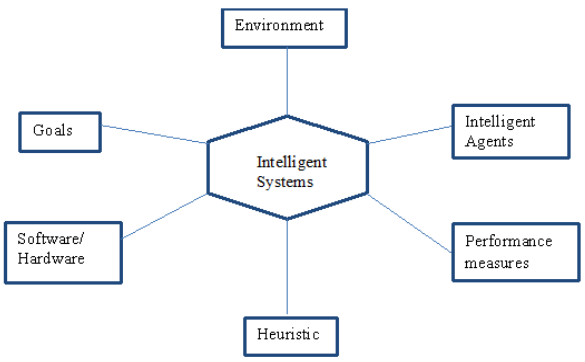

Figure 1. Factors affecting intelligence.

A fundamental difficulty in artificial intelligence is that nobody really knows what intelligence is, especially for artificial systems which may have senses, environments, motivations and cognitive capacities which are very different to our own. [1] Some researchers believe in the term "Universal Intelligence". Universal Intelligence is difficult to define because it doesn't depend on only one factor. An informal definition of intelligence is - "Intelligence measures an agent's general ability to achieve goals in a wide range of environments."[1]

As the diagram [figure 1] shows, intelligence of a system depends on the following factors:

i. Environment – Environment plays a crucial role in the level of intelligence of a system. The system usually learns in a known environment but, the true test of intelligence occurs when the machine uses the learnt concepts, in an unknown environment and pursues towards the goal state.
ii. Agents –An intelligent agent is an entity which observes through sensors, interacts with the environment and the knowledge base and directs the machine's activity to the goal state using actuators.The agent receives the feedback from the environment and thus learns if it is moving towards the goal or away from it.
iii. Performance Measures –The performance measures are relative to the AI's environment and goal, they keep changing in every situation. Performance measures are also a way of checking the level of intelligence of a machine.Only if an AI goes above and beyond the performance measures, true learning has occurred.
iv. Heuristics - A Heuristic technique is any approach to problem solving, learning, or discovery which uses a practical methodology which is not guaranteed to be optimal or perfect, but it is sufficient for the immediate goals.If an AI is provided with many complicated solutions it should select the simplest and the fastest.The intelligence of a system also heavily depends on heuristic search techniques used.
v. Goals – The goal is one of the primary measurements of a system's intelligence. A goal is what defines if the machine is taking one step towards successful learning

DOI 10.17148/IJARCCE.2015.41017

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

or not. Another reason why Universal Intelligence cannot be practiced is because different systems have different goals. In an ideal case, if a system is able to achieve a goal of a particular problem or scenario the system should also be able to solve similar or related problems.

vi. Software/Hardware – This is not the most important factor when it comes to AI but the goal cannot be achieved without them. For example, if a programmer has written an AI algorithm for the fastest word search but the processor is very slow it will definitely affect the performance of the system. Also, new programming languages only developed for AI such as – AIML, PROLOG, LISP etc. are recommended.

Thus, intelligence varies from system to system and is a very complex term.

III. TURING TEST: STILL ACCURATE?

Turing test was devised by the founder of modern computers Alan Turing during 1950, to check if a machine was intelligent or not. It involved a computer and a human giving text based responses to any question asked. The judge was a human, who had to decide if the replies were from a human or a computer. If the computer convinced the judge that it is human then it was considered truly intelligent. Many chatbots (such as cleverbot, ELIZA, Alice etc.) are created but none of them have still passed the Turing Test. Alan Turing believed that computers were a universal discrete machine because they could be used to implement any system or algorithm and thus were chosen to be the "Intelligent Systems". Though simple and clever, the test has attracted much criticism.

Firstly, it is only text based which makes it very easy for a computer to imitate a human as the tone, expression, pitch, emotions etc. parameters are eliminated. American philosopher John Searle (invented the Chinese argument) and N. Block argue that passing the test is not sufficient to establish intelligence. Essentially, they both reason by saying that a machine could appear to be intelligent without having any "real intelligence", perhaps by using a very large table of answers to questions. In this case, an unintelligent machine could, at least in theory, consistently pass the Turing test.[2] Also, even if a machine passes the Turing Test, the machine may not be able to perform other tasks expected by a general all-purpose AI such as learning to play chess or even tic tac toe, or find the shortest path between two nodes. Thus, even if machines pass the Turing test most researchers will agree that the intelligence of the machine is debatable.

Secondly, a more common criticism is that passing the Turing test is not necessary to establish intelligence. Usually this argument is based on the fact that the test requires the machine to have a highly detailed model of human knowledge and patterns of thought, making it a test of humanness rather than intelligence.[2] The chatbots have not increased intelligence rather done the opposite. The bots replied in an evasive manner or replied correctly only to topics within a limited scope. For example, if asked "Do you like Microsoft or Apple?" the chatbot would reply, "I have no opinion on that subject". The chatbot does reply but gives a vague response which cannot be considered intelligent in any way. The Loebner competition restricts the domain of the conversation to make the Turing Test easier. Even most successful Loebner entrants are even more focused on faking human fallibility, rather than anything resembling intelligence.[2]

Lastly, the decision whether a machine passed the Turing test or not is based on a human judge's capability and the questions asked. There have been instances where unintelligent machines have been classified as human. Thus, some consider this to bring the validity of the test into question. Though it is the most famous test for intelligence there have been few flaws attached to it.

Thus, the Turing Test is a binary test whether a machine is intelligent or not but is not an optimumtest to measure the level of intelligence that a machine has acquired.

IV. OTHER TESTS USED

Morgenstern

For Morgenstern, a machine will demonstrate intelligence only when it can show that once it knows one intellectually challenging task, it can easily learn another related task. She gives the example of AI chess players, which are able to play the game at a level few human players can match but are unable to switch to simpler games, such as checkers or Monopoly. "This is true of many intellectually challenging tasks," says Morgenstern. "You can develop a system that is great at performing a single task, but it is likely that it won't be able to do seemingly related tasks without a whole lot of programming and tinkering."[3] Morgenstern's test is logical because we are trying to get machines to imitate the human brain and this is how the human brain works. If a young child is taught all the rules of a sport say, football and he/she plays it, then a game of hockey or rugby will not be as difficult since they are similar environments.

Riedl

The origins come from the Lovelace test (named in honor of Ada Lovelace), where she felt that only when computers originate things – be creative – could they be believed to have minds.[4]This test concentrates more on creative intelligence and asks an AI machine to create a poem or drawing and the AI only passed the test if its creator was not able to explain how the machine came up with the output. It was very difficult to know if the test actually worked. A new version of this test called "Lovelace Test 2.0" was created by Mark O. Riedl, associate professor at the School of Interactive Computing, Georgia Tech. In this, he stated that the basic Lovelace Test is too vague and that more measurable outputs are required in order to compare the relative intelligence of different machines.[4] This test has 2 components - First, they ask for a creative artifact such as a story, poem, or picture and secondly, they provide a criterion. For example: "Draw me a picture of a man holding a penguin."[5] The human judge decides if the computer actually passes the test. If it did, then it is given a more difficult and challenging problem till the AI is not

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

able to solve the problem. Thus, this test is a good measure of comparison of different intelligent systems, rather than ascribing a single, quantifiable value for the intelligence of a system.

Winograd Schema

Machine learning researcher Hector Levesque of the University of Toronto proposed a test called the Winograd schema which involved resolving ambiguous sentences and he believed that this is a behavior worthy of the name intelligence .A Winograd schema is a pair of sentences differing in only one or two words and containing an ambiguity that is resolved in opposite ways in the two sentences and that requires the use of world knowledge and reasoning for its resolution.[4]

Example:
The lawyer asked the witness a question, but he was reluctant to [repeat/answer] it. Who was reluctant?
Answer 0: the lawyer
Answer 1: the witness [6]

The example sentence given above, shows that humans can easily decipher that if the word is repeat the lawyer was reluctant and if the word is answer the witness was reluctant. Thus, the computer needs to make word associations and understand what the verb is referring to.

Matt Mahoney

Another interesting test called the "Compression Tests" was created by Matt Mahoney, a programmer and a data compression researcher stated a variation that replaces the Turing test with a text compression test.[2] By using more complex models that capture higher level features such as aspects of grammar, the best compressors are able to compress text toabout 1.5 bits per character for English. However, humans, who make use ofgeneral world knowledge, the logical structure of the argument etc., are able to reducethis down to about 1 bit per character. Thus, the compression statistic provides an easilycomputed measure of how complete a machine's models of language, reasoning and domainknowledge are, relative to a human.[2]One criticism the compression tests get is that just because an AI is a powerful compressor,it does not mean it will translate to a general all-purpose intelligent system.

V. DISCUSSION

Few computers are already extremely faster and efficient than normal humans can ever be. For example, to crack a code of 2^{16}characters it would take humans to a few hours to finish but a high speed computer with a good program can crack this in few minutes or less. Also, there are some computations which humans will not be able to finish in their whole lifetime whereas computers can finish these tasks in a few days. Thus, in a few ways computers are already way ahead of humans in terms of computation, speed and intelligence. However, this is limited only to specific tasks whereas humans even at an early age can be proficient at a number of tasks and machines are only being made intelligentby imitating humans. For example, there are few AI systems which havepassed the Turing test but are not able to solve questions which are solved by children in their 4^{th} grade.

VI. CONCLUSION

The pace at which computer intelligence is increasing is phenomenal due to arduous efforts by researchers aided by ever progressing technology.Measure of Artificial Intelligence will also keep changing and will never be static. Moreover, there is not a perfect universal method to quantify intelligence yet, but there are several tests to compare the intelligence of several machines.

REFERENCES

1. Shane Legg and Marcus Hutter, Universal Measure of Intelligence for Artificial Agents, IDSIA Switzerland. Link - ttp://www.hutter1.net/ai/siors.pdf
2. Shane Legg and Marcus Hutter, Universal Intelligence: A Definition of Machine Intelligence, arXiv: 0712.3329v1 [cs.AI] 20 Dec 2007.
3. Rewriting the Rules of Turing's Imitation Game. Link - http://www.technologyreview.com/news/535391/rewriting-the-rules-of-turings-imitation-game/
4. 5 Tests for Artificial Intelligence. Link - http://electronicdesign.com/test-measurement/5-tests-artificial-intelligence#slide-2-field_images-846491
5. How creative is your computer? Link - http://www.slate.com/articles/health_and_science/new_scientist/2014/12/lovelace_test_of_artificial_intelligence_creativity_better_than_the_turing.html
6. Hector J. Levesque, The Winograd Schema Challenge, Link- http://www.cs.toronto.edu/~hector/Papers/winograd.pdf

BIOGRAPHY

Ms. Prerna Kapoor is a Bachelor's studentpursuing her B.E in Information Technology from Thadomal Shahani Engineering College, Bandra (West), Mumbai, Maharashtra, India. Her fields of interest are artificial intelligence, machine learning, algorithm & analysis and data mining.