Three Open Problems in AI

RAJ REDDY

Carnegie Mellon University

We present three open problems in AI which, if solved, would bring us closer to achieving the goal of Human Level AI. Common to all these systems is the ability of a system to learn from examples, observations and errors. Indeed, one might say that these represent necessary conditions before we can aspire to reach Human Level AI.

1. *Read a Chapter in a Book and Answer the Questions at the End of the Chapter*

Reading and understanding books is a quintessentially human activity. It is the process by which much knowledge transfer occurs from generation to generation. For example, we test students' understanding of a given subject by asking them to answer questions at the end of the chapter in a textbook. In general, we are satisfied if a student correctly answers 80 to 90% of the questions. When a computer can do the same task, we will have arrived at a significant milestone.

A human being who starts reading at age 4, lives to be 100 years old, and reads a book a day every day could complete 35,000 books in a lifetime. By many estimates, the total number books ever written in all languages is under 100 million. Harvard library has around 12 million volumes. The Library of Congress has fewer than 30 million volumes. All the unique titles in the OCLC member libraries is under 42 million. However, US libraries do not have most of the books published in other languages internationally, thus leading to an approximate estimate of 100 million books ever published.

Once a computer can read one book and prove that it understands it by answering questions about it correctly, then, in principle, it can read all the books that have ever been written. This has led to the speculation that once computers can read, understand, and share knowledge with each other, without the limitations that biology imposes, they will begin to exhibit super human intelligence.

For a machine to read a book, understand it and answer questions about it, it needs mechanisms

- —for converting (analog) paper-based information into machine processable form;
- —to read and understand text with all the implied ambiguity and imprecision of natural language and interpret the intention of the author;
- —to convert the understanding into executable representation of the knowledge;
- —to interpret and represent questions into initial conditions and the desired end goal; and
- —for problem solving to apply the knowledge extracted from the chapter and previously known (acquired) knowledge including qualitative common sense knowledge to solve the problem at hand.

What does it mean to "read a book"? First, the book needs to be in a machine-readable format. Many books are "born digital", that is, their content is available in machine processable form. Books published before the advent of digital publishing would have to be scanned resulting in a digital image (essentially a digital photograph) of the page in the computer. Optical character recognition systems exist today that process text, tables, graphs, and grey scale and color images in a page and produce processable text and table with less than 1% error. Graphs and images are usually left uninterpreted. But these OCR systems are never perfect. Thus, any system that attempts to read and "understand" must also cope with errors and typos.

Once we have machine-readable text, the real problem begins. At present, we do have systems that can translate from one language to another with mixed results. This and other forms of understanding (such as information retrieval or document classification) can be done with only a superficial understanding of the meaning of words, phrases and sentences. The implication of the proposed task is that to answer question at the end of the chapter, the knowledge within has to be distilled into an executable representation, to be used in conjunction with reasoning and problem solving methods to solve the problems at the end.

Herbert Simon (with Dorothea Simon) attempted to solve this problem in the early 1970s. Their problem solving tool was a Production System. They reasoned that if all the knowledge in a chapter of a physics textbook could be represented as "productions" (also known as "production rules") that then were used to solve many of the problems at the end of the chapter, that would represent true understanding. Lacking the tools for "Natural Language Understanding", they read the chapter themselves and represented their understanding as "productions", which they were able to successfully use to solve the problems.

Having partially solved the big problem, Simon had a succession of students work on the rest of the problem of getting a machine to read and understand—without success. The main difficulty was "to understand the meaning of a sentence" in the absence of systems with common sense knowledge.

Since the 1970s, we have made great strides in "Natural Language Understanding". These advances combined with a systematic approach to transforming "understanding" into executable productions should result in significant progress. The "systematic approach " we propose is to extend the Simon paradigm of getting humans to do "what machines cannot do" and creating a research agenda of all such human interventions, to be explored by generations of graduate students ultimately approaching "nirvana".

Once we can read one book and demonstrate deep understanding by solving problems at the end, we are not at the end. There is the other unsolved problem of assimilation and integration of knowledge from multiple sources, that is, from all the books. That is not the end either; we also have to integrate knowledge from other life experiences such as auditory knowledge and visual knowledge. Once we have mastered the concepts of visual knowledge, we will be in a position to interpret diagrams and pictures in books along with other tasks such as knowing how to repair a robot from observation, which is the next open problem.

2. *Remote Repair*

The second open problem we propose is the task of getting a machine to learn how to repair a Mobile Robot and successfully demonstrate the capability by repairing one on Mars (or with appropriate simulated time delay on earth) after observing a human being repairing a similar Robot in the lab.

A solution to this class of problems will have significant societal impact. Remote repair technology will spawn new service industries such as remote mechanics, remote monitoring and diagnosis, telemedicine and telesurgery, teleagriculture, and operations in hazardous environments.

Systems that can successfully perform tasks in a real world environment must understand concepts of space and time, and approximate algorithms where re-execution of a program does not always give the same result.
To repair a mobile robot on Mars, we need

—a mobile repair platform with all the relevant tools and fixtures;
—a semi-autonomous system in which a human supervisor can provide guidance but not intimate teleoperation (Note that the time delay of 10 to 15 minutes, depending on the relative position of Earth and Mars, implies that most of the navigation and obstacle avoidance must be locally controlled);
—a system that can repair implies precision manipulation capability for disassembly and assembly of the disabled platform;
—a system that can learn from observation by looking over the shoulder of a human operator (requires a system with 3D vision, modeling of space, discovering and programming equivalent manipulation operators to the human actions); and
—a system that can engage in clarification dialog with humans to verify and validate the understanding of the observations of human operations.

Each of the tasks stated above seems likely to yield progressive advancement given sustained effort. Perhaps the hardest is likely to be the learning task, which is central to progress given that we cannot afford to program all such tasks.

At present we are not aware of any system that can perform these tasks at the implied generality of repairing a mobile electro-mechanical system. Professor Katsu Ikeuchi of CMU (currently at Tokyo University) demonstrated in 1988, a robotic system capable of learning from observation, a simple stacking operation in a blocks-world. Even the simple task of observing "picking up a block and place it on top of another block" and deriving (learning) the equivalent motion action from a sequence images proved to be difficult. Professor Ikeuchi, instead, chose to derive the robot actions required, by inferring (planning) the sequence of actions given the beginning and end states of the scene.

3. *Encyclopedia on Demand*

We are living in the age of "Wealth of Information and Scarcity of Human Attention" [Simon, 1995].

As of last year the information available on the web exceeds 100 terabytes. Information that is publicly available in libraries and other copyrighted forms exceeds 100 times the information on the web. The deep web consisting of all data on all the disks in all corporations and households is a million times larger. We have been facing the information glut for many years and it promises to get worse.

The third open problem we have chosen arises from what Professor Jaime Carbonell of CMU calls the "Bill of Rights for the Information Age", namely how to get the "Right Information" to the "Right People" in the "Right Language" in the "Right Timeframe" in the "Right Level of Granularity". Specifically (and narrowly) the right level of granularity task we propose is to "produce a 5000 word or less, encyclopedia style article, on a given subject, by summarizing from the relevant information available on the web in less than 24 hours"

Given the web has 100 terabytes, just reading all the data at 100 megabytes per second (current best bandwidth of a single disk) would take over 11 days!

Finding all related information using inverted index techniques will help to retrieve most of the discoverable data. Much of data retrieved will be disjointed, containing duplicate entries, and obsoleted by later web postings.

The task of creating an encyclopedia style article requires several new technologies such as

—document clustering to identify a group related articles;
—synthesis of information from all the related articles into a single merged document;
—summarization of the merged information into a convenient size; and
—language generation of a natural and intuitive sentences of the finished summary.

Conventional "google-like" keyword based retrieval systems often return thousands of hits. Finding entries that are duplicate or similar (as in reporting of the same story by different newspapers) and grouping them together is called document clustering. This is currently achieved defining a document-similarity metric based on a vector of the content words in each document.

Synthesis of information from multiple sources requires merging relevant information from different sources while removing duplicate or equivalent sentences, rationalizing conflicting information (as in the case of an evolving story on number deaths after an earthquake), and inserting background information that is often assumed to be known in a news story (as in information about Second World War in a story about the Berlin Wall). This is an area of current research interest but progress appears to be slow.

Summarization and abstraction is routinely done by trained human experts and on a less professional level by most of us. Summarization by machine involves selection of information-rich phrases and sentences from an article and producing a shorter article that hopefully captures the essence of the article. Information-rich content words and phrases are currently assembled from the document content and structure, and by giving extra weight to title, abstract and conclusions, and to chapter and section headings.

Language generation or sentence synthesis based on the intended meaning has proven to be a challenging task. It has usually been studied in the context of language translation and is equally important to our task. Human language provides many different ways of saying the same thing. Human beings seem to somehow choose the most succinct form and this fact has been called the "principle of least effort". Computers that can do the same are still in their infancy.

We appear to be making progress on all these technologies and yet we seem to be far away from being able to create an acceptable quality encyclopedia style article on a subject. We may be able to make substantial progress by accepting human assisted generation of encyclopedia-on-demand as an intermediate goal. Removing these human interventions can then become the research agenda for the next generation system.