A Buddhist Perspective on Artificial Intelligence

Kamran Karimi

Department of Computer Science
University of Regina
Regina, Saskatchewan
Canada S4S 0A2

karimi@cs.uregina.ca

Michelle D. Bryson

Department of Physics
University of Regina
Regina, Saskatchewan
Canada S4S 0A2

brysonmi@uregina.ca

Kiana Karimi

Department of Computer Science
Concordia University
Montréal, Québec
Canada H3G 1M8

k_karimi@cs.concordia.ca

Abstract

Artificial Intelligence (AI) has been a multi-disciplinary domain from the beginning. While technology has always played a big role in the amount of progress made by AI researchers, philosophy has had a big impact on the directions they take. In this paper we present a number of Buddhist views concerning enlightenment that we feel have serious implications for AI research. AI researchers have been using the human mind as a model for building their systems. This is natural because the human mind is the most intelligent thing we know. We show that the Buddhist beliefs are in stark contrast to the current approaches in AI because in Buddhism the ultimate goal is to stop the calculating and planning mind and thus achieve enlightenment. An enlightened person does not show the signs we traditionally attribute to intelligence. He does not consider the world to be made of interacting parts, does not build mental models to accommodate his observations, does not classify, does not plan, and does not care if his actions have any particular result (success or failure). An enlightened person thus looses intelligence as it is commonly understood, and still manages to function appropriately and survive in the world. This is contrary to the direction taken by many AI researchers, who equip their systems with either heuristics derived from their own experiences, or make the system learn from its own past observations. All artificial systems use a small amount of information to make their decisions, thus dividing the world into relevant and irrelevant parts. While these approaches have succeeded in developing useful programs and robots that help us in our lives, they seem to be at odds with what Buddhism claims to be the natural state of our minds. This paper is a brief examination of the theoretical consequences of Buddhist views on current AI approaches.

1. Introduction

One of the main teachings in Buddhism is the principle of change. Buddhism [6, 13] claims that everything is in constant flux, and nothing remains the same from one moment to the next. For this reason, Buddhism considers it impossible to come up with rules and regulations that can always be applied with success. No rule can capture the constant change, and no phrase can solidify the reality. Any concept is bound to become useless sooner or later, if not from the beginning, because reality can not be captured into a concept or be expressed by words in any language. The same applies to the mind and to the ego we perceive as the unchanging "me" that defines us. This is against common sense for many people, because it implies that there are feelings, but no feeler, and there are perceptions, but no perceiver. As we will see, Buddhism's sometimes counter-intuitive concepts have implications for the AI researchers too.

Buddhism does not consider the reality to be composed of parts. Everything is intimately connected to everything else. In a way, everything *is* everything else. Breaking up the reality into parts thus will inevitably lead to wrong results. This idea has been investigated under the tile of chaotic systems [2]. A chaotic system is a complex system that cannot be divided into parts with limited, controlled

interaction among them. In such a system, a small change in the input may lead to non-proportionally big and unexpected changes in the output. The weather is a chaotic system. It is often simulated by arbitrarily segmenting the area of interest into smaller sections and applying simplified rules to each section. The results are then combined to calculate the effects of each part of the other parts. The failure of even the strongest supercomputers to produce predictions that are always correct is a well-known fact. From a Buddhist perspective, which matches the common sense, no one can ever accurately predict the weather unless it is considered as a whole system. This includes anything that could ever have a direct or indirect effect on the weather, such the motion of a person, or the light of a distant star that increases the energy of a few air molecules at night.

In this paper we discuss some of the implications of the Buddhist way of thinking on AI. The human mind, in its normal form is always segmenting and modeling. AI heavily depends on mimicking the human mind to achieve many of its goals. Thus Buddhist views are bound to have serious philosophical implications for the AI practitioner, as explained in the rest of the paper. Section 2 gives an overview of the attempts to come closer to the creation of an artificial mind that shows the same signs of intelligence as a human being. Section 3 discusses how Buddhism's different approach to mind and intelligent behavior completely changes the problem we are to solve. It argues that according to Buddhism there is no way to simulate a mind using any kind of data structure or any amount of knowledge simply because this is against the nature of reality.

2. Knowledge and Representation of Knowledge

Historically, research into AI has been divided into two branches. Hard-AI concerns the emulation of human-level intelligence, or an artificial human mind, while Soft-AI concerns the development of techniques to solve limited and simplified everyday problems. These techniques include search algorithms and data representation methods, among others. Hard-AI's goals have not been achieved, but as a result of it many computer applications have been successfully developed to aid people in their jobs. From early on it was obvious to researchers that for a computer to work on a problem, it should have access to a representation of that problem. This entailed the coding of information about the real world and turning it into a form that is processable by the computer. This brought about the problem of knowledge representation and modeling.

In the 1950s, when AI was a new discipline and in full motion, many people believed that we would succeed in creating an artificial human mind, characterized by human-level intelligence, only if we have a suitable knowledge representation format. An enormous amount of effort was spent in subsequent years to address this issue. Out of these efforts came Situation Calculus, Semantic Networks, Production Rules, Scripts, and Frames, among others [12]. Each of these schemes tries to model the world, and then use the model, which is much simpler than the real world, to mimic some intelligent behavior. Though this modeling process can (and has been) used to solve many practical problems, so far it has been unable to come even close to emulating human level intelligence. Most researchers now agree that there is no universal knowledge representation scheme that is appropriate for all cases, and that each application requires its own special scheme.

In the 1970s, more and more attention was diverted to the knowledge necessary to solve a problem. This lead to an explosive interest in Expert Systems [12], which used specific knowledge about a narrow domain to solve certain problems. Thus came systems that though could outperform a human being in a specific field such as diagnosing blood infections, were in no way able to do anything else remotely intelligent. Another wave in Hard-AI, which started in the 1980s, made the assumption that the key to human-level intelligence is in common sense and general knowledge. Proponents of this school argue that what makes humans superior is not just their brain structure (knowledge

representation) but what they know. Maybe if we just add more information to a system, it will come closer to behaving like an intelligent human being. Cyc is a prime example of this approach [8]. In this on-going project, common-sense knowledge is gathered and processed. Its developers feed it with the kind of knowledge that every human being would have gathered from the experience of living. They then make the relationships among these knowledge explicit, allowing Cyc to start drawing its own right and wrong conclusions. The wrong conclusions are then removed, so Cyc would not repeat them. Cyc has produced many byproducts, including an inference engine that can manage huge knowledge bases.

In an effort to build useful robots that could successfully interact with an environment, researchers suggested that there is no absolute need for a data representation and reasoning engine [1]. The reactive robots that were built using this approach were simple and effective, but could hardly adapt, because they were following a set of never-changing rules. Following this trend, since the early 1990s Embodied AI (EAI) [5] has become an active field of research. The basic assumption in EAI is that intelligence can only arise with interaction with the outside world, as in a body that can sense the real world and react to the stimuli provided to it.

Decision making is the most important outcome of any intelligent system. The common factor among all these approaches is that they use their prior knowledge, whether learned or hard-coded, to make a decision as to what to do next. The intelligent system may change the knowledge that is uses for decisions making as time progresses, but it still uses information gathered prior to the moment of decision making to choose a new action. Thus the decision is based on stale information. Another common trait among all AI approaches in that they simplify the world, and consider some information as useful, and some others as useless. Thus in an EAI system, for example, the "body" has a limited number of sensors. This breaks the reality into relevant and irrelevant parts.

3. Enlightenment and Intelligent Behavior

At least some schools of Buddhism have no problem with a system being conscious. His Holiness the Dalai Lama was asked if it is possible for an artificial system to ever become conscious, and the answer was positive. The Dalai Lama actually said it might be possible for the scientist working on the system to later reincarnate as the system [3], so it is advisable for all scientists to try and make their conscious systems as nice as possible!

But consciousness is not the same as intelligence. Intelligence has always been linked with the mind and with conception. The general assumption is that we cannot understand something without first conceptualizing it. We always break up a problem that we consider to be too complex into sub-problems, build a model in our minds, and then use that model to solve a problem. So if this is what we humans do, maybe we can program our machines to do the same. The problem, from a Buddhist point of view, is that we never really understand the world, exactly because we conceptualize it and turn it into a model. This stems from an inherent inability to freeze the reality by words and models. Buddhism maintains that reality can be perceived, but never conceived. In other words, we can see the reality of our surroundings and our selves, but we cannot describe it or represent it in anyway. This is not a purely spiritual observation. Kurt Gödel's ground breaking work on logic systems showed that any system that is based on axioms is always incomplete [4]. In other words, there are theorems that are true in that system, but can never be proved either right or wrong. This is an intrinsic property, and holds no matter how we change the axioms of the system. This is a very serious inadequacy in one of humanity's most prized achievements: Logic. This inability to capture the whole truth in a logical system has been used by Roger Penrose as the main argument against the possibility of creating an artificial human mind [10, 11].

This may still seem like a small problem. Even assuming that the reality can never be conceptualized and broken down into pieces, we seem to be getting by as intelligent people quite well. It is argued that we cannot make our machines do the same and expect them to show signs of achieving human capabilities such as creativity.. For example, the school to which Roger Penrose belongs believes that the fact that no conceptualization and modeling can capture the whole truth is enough to discredit claims about mimicking human beings. Our claim, by comparison, is smaller. We suggest that even though researchers may come very close to emulating a human mind, the result will probably never gain enlightenment. This is because enlightenment cannot be achieved with an intelligent mind that calculates and follows rules.

In Buddhism enlightenment is realized when we give up all attempts to conceptualize and build models. An enlightened person, also known as a Buddha, will thus no longer live in error and confusion. The outdated and limited models are gone, and he can see the whole truth moment after moment as it unfolds. At this stage no attempt is being made by a Buddha to understand things and the relationships among them. An enlightened person supposedly realizes that the Universe is a whole, and that nothing and no relationship lasts long enough to fit any model.

This is in stark contrast to the current major directions of research in AI. The researchers have realized that if a system wants to work in the real world, it should be able to deal with change. They keep coming up with new ways to represent things and concepts, to find relationships among things and concepts, and to update the things and the relationships as they change. This adaptability (being able to learn how to learn new things, concepts and relationships that were not predicted by the original designer) is very attractive. Ironically from a Buddhist point of view, new knowledge representation schemes and algorithms are bound to be useless, because they are simply another way to freeze the reality. The only way to deal with the nature of reality (change) is not to use any structure or algorithm, and just perceive. No rules should be extracted from past experiences because they will create conditioning and color the future experiences. There should also be no distinctions made among the parts of a system, because that rests of the wrong assumption that those individual parts exist independent of each other.

An enlightened person is supposed to act with no will. There are no planning systems at work inside him. He does not consider the consequences of his actions, and he would not check to see if all the preconditions for his action are met before he performs it. These all go against the most basic teachings in any AI book. Actually, many people would consider these as signs of a mental problem. We dissect, learn, classify, draw conclusions, plan, and execute our plans. We also suffer when our plans fail, because we realize that the world does not match the model we have in our mind. This is exactly how we expect our AI systems to behave (except that we assume they do not suffer when their plans fail!), and this is how we program them to behave. The problem is not just with complex AI systems that learn, but even with very simple reactionary programs that respond to the outside stimulation by a set of pre-programmed and unchanging rules written down by the system designer. Not thinking may seem to us as a sign of inferior intelligence, but somehow an enlightened person is always supposed to do the "right" action, which is not the same as the action that brings the most rewards or maximizes a utility function. A Buddha may do something that leads to his/her own destruction, and this is not considered wrong. If there is a utility function that is being maximized by the actions of an enlightened one, we are not aware of it.

Following any rule that is derived from past experiences (either learnt automatically, or put in there by the developer) is doomed to fail to match the reality of the present moment. But maybe random action holds the key, because a truly random sequence of actions is never influenced by the past. True

randomness in choosing what to do next may thus lead to the "right" action. In [7] an artificial agent living in an Artificial Life environment [9] uses random choice to select among a set of possible actions because there the creature does not know of any heuristics for navigation in the domain. One could continue to use the same random method, even after the artificial domain has been explored, instead of extracting rules from past observations. This approach is quite contrary to most rational minds, because randomness is usually seen as neutral and without purpose, and thus undesirable. However, neutrality and purposelessness are among the cornerstones of Buddhist thought.

4. Concluding Remarks

Creating an artificial human mind seems to be the holy grail of AI. The common approach has been for the researcher to come up with data structures and algorithms that he predicts will solve a series of problems. This approach has been hugely successful in solving many everyday problems, and is sure to bring more fruits irrespective of our motivations in creating them.

Buddhism states that reality can not be captured by any given set of rules. This view is supported by Kurt Gödel's proven results, which have been used by many people to predict that the efforts of people who want to build an artificial human mind is doomed to fail. For a Buddhist, we can really recreate a human mind when we can give it the ability to not only think, but also become enlightened. Buddhism suggests that this is impossible with any form of conceptualization. It is not obvious how our current approaches to AI systems development should be modified to accommodate enlightenment if we cannot represent the outside world, cannot derive information from observations, and cannot plan for reaching a goal, and check for the success of our actions.

Is there a lesson here for the computer scientist? It all depends on what kind of mind we want to emulate. While we do not expect people to abandon their endeavors in traditional computer science because of the Buddhist views, we hope they will consider them as food for thought.

References

[1] Brooks, R.A., Intelligence without Representation, *Artificial Intelligence Journal (47)*, 1991.
[2] Casti, J.L., *Complexification*, Harper Perennial, 1994.
[3] The Dalai Lama and Cutler, H.C., *The Art of Happiness: A Handbook For Living*, Riverhead Books, 1998.
[4] Dawson, J.W., *Logical Dilemmas: The Life and Work of Kurt Gödel*, A K Peters Ltd, 1997.
[5] Franklin, S., Autonomous Agents as Embodied AI, *Cybernetics and Systems 28(6*), 1997.
[6] Hagen, S., *Buddhism Plain & Simple*, Broadway Books, 1999.
[7] Karimi, K. and Hamilton, H. J., Finding Temporal Relations: Causal Bayesian Networks vs. C4.5, *The 12th International Symposium on Methodologies for Intelligent Systems (ISMIS'2000)*, Charlotte, North Carolina, October 2000.
[8] Lenat, D. B., *Cyc: A Large-Scale Investment in Knowledge Infrastructure, Communications of the ACM 38, no. 11*, November 1995.
[9] Levy, S., *Artificial Life: A Quest for a New Creation*, Pantheon Books, 1992.
[10] Penrose, R., *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford University Press, 1996.
[11] Penrose, R., Shimony, A., Cartwright, N., Hawking, S., *The Small, The Large, and the Human Mind*, Cambridge University Press, 1999.
[12] Russel, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.
[13] Watts, A., *Buddhism: The Religion of No-Religion*, Writers Club Press, 1999.