From: AAAI Technical Report SS-93-04. Compilation copyright © 1993, AAAI (www.aaai.org). All rights reserved.

Massively Parallel Artificial Intelligence and Grand Challenge AI Applications

Hiroaki Kitano

Center for Machine Translation
Carnegie Mellon University
5000 Forbes
Pittsburgh, PA 15213 U.S.A.
hiroaki@cs.cmu.edu

Software Engineering Laboratory
NEC Corporation
2-11-5 Shibaura, Minato
Tokyo 108 Japan
kitano@ccs.mt.nec.co.jp

Abstract

Proliferation of massively parallel machines have undergone the first stage where researchers learn to know what it is like. Now it come to the second stage in which researchers are asked to show visions for real applications. The author argues that **Grand Challenge AI Applications** should be proposed and pursued. These applications should have significant social, economic and scientific impact and serve as showcases of accomplishments of massively parallel AI. For the grand challenge to succeed, massive computing power, massive data resource, and sophisticated modeling would be critical. At the same time, it is thought that such efforts shall be promoted as international projects.

1. Introduction

This paper describes the role of massively parallel artificial intelligence[Waltz, 1990, Kitano et. al., 1991] for grand challenge AI applications. Grand challenge AI applications are applications with significant social, economic, engineering, and scientific impacts. The term "Grand Challenge" was defined as "a fundamental problem in science or engineering, with broad economic and scientific impact, whose solution will require the application of high-performance computing resources" in the High Performance Computing Act of 1991. To develop socially significant applications, use of high performance computing is not a prerequisite, but it is likely to be a necessary condition for success.

Before the discussion of the grand challenge AI applications, let us remind the research approach taken in the vast majority of AI research to date. Most research has been carried out in the way that researchers develop a highly intelligent system in a very restricted domain. It is a toy system, or a prototype system. Researchers hoped that such systems to be scaled up with larger funding. However, experiences in expert systems, machine translation systems, and other knowledge-based systems indicate that scaling up is extremely difficult for many of the prototype (Figure 1). One of the manifestation of this syndrome is the *knowledge acquisition bottleneck.* Not only it is difficult to acquire knowledge, but also maintenance of knowledge is prohibitively difficult due to complexity of interactions and the size of the whole system. It is the software engineering problem rather than the problems of the first principle of AI.

The alternative approach is to build a simple, but a useful system which can be operational in the real-world and pays-off. Then, incrementally improve the level of intelligence. The advantage of this approach is that the system can be deployed in the real-world at the early stage of the research so that the investment risk could be hedged. The problem is that the incremental improvement of the intelligence may not be that easy so that the system with higher level of intelligence may have to be developed from scratch. This may results in investment overhead. As this approach has not been considered seriously in the past, it is not clear whether there are unexpected difficulties. However, it seems that there are reasonable chance that we can go through this approach. At least it worth trying.

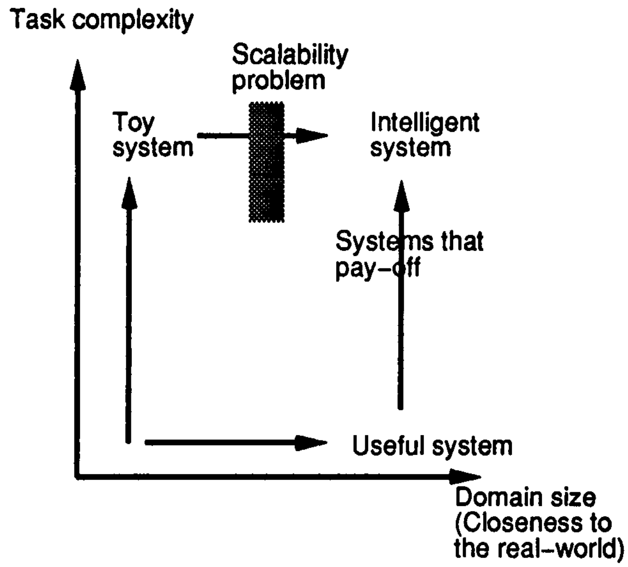

Figure 1: Approaches for Building Intelligent Systems

2. Grand Challenge AI Applications

It is not an easy task to make a list of significant applications. However, there are certain applications which could bring us obvious benefits. Also, there are a group of applications which can be a symbol of engineering and scientific accomplishments.

Two workshops were organized to discuss grand challenge AI applications. In Feburary 1992, NSF sponsored "The Workshop on High Performance Computing and Communications for Grand Challenge Applications: Computer Vision, Natural Language and Speech Processing, and Artificial Intelligence."

In October 1992, the author have organized "The Workshop on Grand Challenge AI Applications" at Tokyo, participated by a group of leading Japanese researchers[Kitano, 1992].

While there are interesting differences in type of applications and key technologies proposed in between American and Japanese workshops, the common thread was the need for massive computing power, massive memory storage, massive data resource, and ultra high band-width communication networks. Here is a part of the list:

- Speech-to-speech translation system
- Human genome analysis
- Global Architecture for Information Access (GAIA), a highly intelligent information access system
- Shogi and Go system which beats 'Meijin' (Grand Master)
- Intelligent robots and programs which undertake daily tasks
- Intelligent Vehicles

Some of these are obvious and some projects are already running (Human genome[NRC, 1988], speech-to-speech translation, and some robot works).

GAIA needs some explanations. The idea of GAIA was proposed by the author in 1992. It is an extended and more intelligent version of the electric library. GAIA access all possible information sources, such

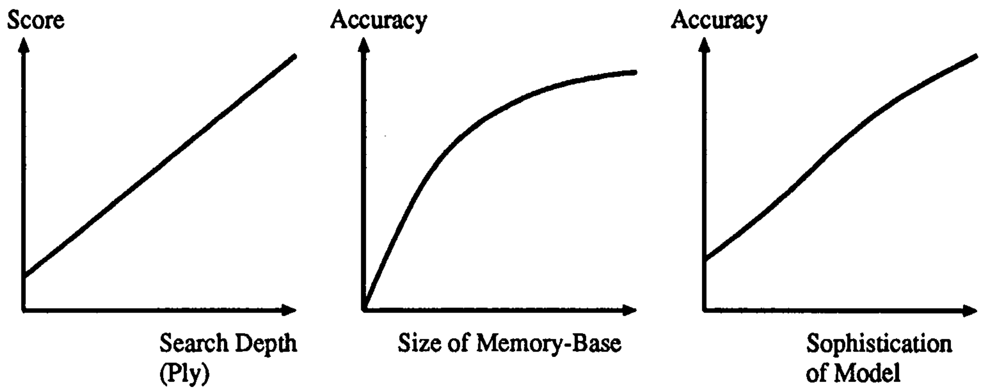

Figure 2: Improvement of the quality of Solution vs. Increase in Computing Power, Data, and Modeling Quality

as electric dictionary[EDR, 1988], image, newspaper, full-text data, databases, knowledge-bases[Lenat and Guha, 1989], audio-visual data, TV, and provides users with flexible and intelligent access to information. Flexible and intelligence access means that GAIA performs simple retrieval, similarity-based retrieval, scanning, summary, translation, trend analysis, and other functions which are necessary to access information. Thus, GAIA can also be read as *General Architecture for Intelligent Access*. The author believes that such system can be a basic infrastructure for various range of intelligent systems.

3. Computing, Memory, and Modeling

Grand challenge AI applications are hard tasks. The author argues, however, that there are ways to accomplish these challenges. In this section, we examine one possible approach. The approach emphasizes massive computing, massive data, and proper modeling. These are three major factors which leads us to successful intelligent systems developments.

The importance of computing power can be represented by Deep-Thought Chess machine[Hsu et. al., 1990, Hsu, 1990]. Deep-Thought demonstrates the power of computing using Chess. It was once believed that strong heuristic approach is the way to build the grand master level chess machine. However, the history of chess machine indicates that the computing power and the strength of the chess machine has almost direct co-relation (Figure 2-Left). Deep-Thought-II consists of 1,000 processors computing over a billion moves per seconds. It is expected to beat the grand master. Deep-Thought exemplify the significance of the massive computing.

Memory aspects can be argued from a series of success in memory-based reasoning. Starting from the initial success of MBRtalk[Stanfill and Waltz, 1986], the memory-based reasoning has been applied to various domains such as protein structure prediction [Zhang et. al., 1992], machine translation [Sato and Nagao, 1990, Sumita and Iida, 1991, Furuse and Iida, 1992, Kitano and Higuchi, 1991a, Kitano, 1991], census classification[Creecy et. al., 1992], parsing[Kitano and Higuchi, 1991b, Kitano et. al., 1991b], weather forecast, etc. PACE, a census classification system, attained the classification accuracy of 57% for occupation code and 63% for industry code. AICOS expert system attained only 37% for occupation code and 57% for industry code. Example-Based Machine Translation (EBMT) developed by ATR Interpreting Telephony Laboratory translates "A no B" phrase in Japanese, which is considered to be very difficult, at the accuracy of 89%, whereas traditional machine translation systems only attained about 40-50% of accuracy. ATR also developed Transfer Driven MT (TDMT) which translates mixed-initiative dialogue in 88.9% of accuracy. These success can be attributed to the superiority of the approach which place memory as a basis of intelligence, rather than fragile hand-crafted rules. In the memory-based reasoning system, the quality of solution depends upon the amount of data collected. Figure 2-Center shows the general relationship between amount of data and quality of solution. Success of memory-based reasoning demonstrates the significance of the massive memory, or data-stream.

The importance of modeling can be discussed using SPHINX speech recognition system[Lee, 1988]. Use of massive computing and massive data-stream is not enough to build intelligent systems. The critical issue is how we model the domain of application. Figure 2-Right shows the improvement of recognition rate with sophistication of the modeling. Even if we use massively parallel machines and large data resources, if the modeling was not appropriate, only a poor result can be expected. SPHINX exemplifies the significance of the modeling.

4. Massively Parallel AI as a basis for Grand Challenge AI Applications

The approach described will play an important role in grand challenge AI applications. Take an example of spoken language translation system, it would have to be able to recognize speaker-independent, continuous speech in real time. Recently, some researchers extended it into a vocabulary-independent system. Speech recognition for this level of sophistication requires balk data processing and highly parallel processing to attain high recognition rate in real-time. Translation module will be developed using extended versions of memory-based reasoning. Success of EBMT, TDMT, and MBT-I demonstrate that the memory-based approach is the promising option to pursue. In fact, we have implemented EBMT on CM-2 and IXM2 as a part of the ATR-CMU joint research. This provides us with high quality and real-time translation.

One other example is the grand master Shogi and Go machine. Shogi and Go has significantly larger search space than Chess. Search space of Chess is about 10^{120}, but Shogi and Go has search space of order of 10^{200} and 10^{300}, respectively. Shogi is a game similar to Chess, but the men taken by the opponent can be redeployed as a men for the opponent. Thus, effective search space do not decrease even at near the end of the game. Go has completely different game style and vast search space.

The traditional question would be raised again whether computing alone could solve this problem. The author believes that use of massive computing and massive memory with a proper modeling and machine learning techniques will accomplish the grand master Shogi and Go machine. There are supporting evidences to back up this claim. Human grand master of Shogi and Go actually memorize scripts of all major matches in the past, and use these memory in the game. There is very high possibility that these grand masters are in fact performing memory-based reasoning. Human grand master, obviously, use other forms of inferencing and search, too. However, combination of memory-based approach and efficient search and learning technique will essentially simulates human grand master. Through this research, we can expect discovery of new search, memory access, and learning techniques as it was a case in Chess machine. It is also scientifically significant to test what level of Shogi and Go machine can we build just by using memory-based reasoning from all available scripts in the past. Such projects will inevitably shed light to nature of expert reasoning.

GAIA requires collective efforts to develop. While the preliminary system with no, or minimum, intelligence can be developed based on existing technologies such as massively parallel full-text retrieval [Stanfill et. al., 1989, Elexir, 1991], message understanding and information extraction as seen in TIPSTER and Message Understanding Conference (MUC), intelligent systems require break throughs in various aspects. Search on very large knowledge bases requires a set of massively parallel search technique [Evett et. al., 1990, Evett et. al., 1990, Geller, 1991]. Not only that intelligent processing scheme need to be developed, but also efficient ways to create and store very large knowledge, text, AV-base need to be developed. Some of these issues will be investigated in NOAH project [NOAH, 1992] and Real-World Computing Program by MITI. There are some research on image retrieval, text-retrieval, and message summary using memory-based approach. Again memory-based approach and combination with other approachs are expected to accomplish the challenge. Trend analysis and database mining requires massive computing power as complex operations to induce rules from database has to be performed on large data sets. Massive parallelism is the key technology for grand challenge.

5. High Performance Computing Support

The approach described so far requires massive computing power and memory. There is constant progress in hardware and device community to meet the demands[Gelsinger et. al., 1989, IEDM, 1991]. Emergence of 1,000 MIPS processor and spreading of RISC architecture are obvious examples. Memory is getting cheaper and larger as Giga-bits memory chips are in the mid-range target. High band-width interconnects, such as 200MB/Sec, are common in recent days. There are projects for Giga bits per second speed networks. In addition, there are major efforts going on in the area of mass secondary storage.

Special purpose hardwares will be regarded as practical option for very high performance computing. WSI-MBR, a wafer-scale integration for memory-based reasoning[Kitano and Yasunaga, 1992], can attain over 70 teraflops when produced on 8-inch wafer using 0.3 micro design rule. It is technologically feasible in 5 years. WSI-MBR is a hybrid analog/digital system optimized for MBR. It can store up to 2 million records on one wafer. Using the wafer stack technology, may be with the micro-pin technology, peta-flops and few hundreds million records systems are technically feasible.

CMOS associative memory may find its market nitch in the light of massively parallel AI and genetic algorithms[Higuchi et. al., 1991, Kitano et. al., 1991a, Stormon et. al., 1992]. Associative memory is a well established technology which can be commercially manufactured in the short run.

Neural network devices are one other option. Again, use of WSI for neural network device was alrady proven to be effective[Yasunaga et. al., 1991].

Interconnection speeds up constantly as GIGAswitch and other technologies mature. In the long run, optical connections with TeraHeltz calibration may come into play.

Therefore, although substantial efforts has to be invested, the author is optimistic about the hardware support. One thing to add, however, that the computing power should be regarded as national and international resource. Index such as *Gross National Computing Power (GNCP)* would be able to measure the industrial strength.

6. International Cooperation

For the grand challenge to succeed, some form of international cooperation need to be carried out. Already, Human Genome project is an international projects. ATR's interpreting telephony project involves U.S. and European groups to build a Japanese-English-German speech-to-speech translation system. The scale of grand challenge applications inevitably draw in international aspects as no single institution or country can afford development of truly significant systems with real-world scale. It is already happening now, and this will be enhanced, and it need to be enhanced. The author has been organizing the International Consortium for Massively Parallel Advanced Computing Technologies (IMPACT) to facilitate international exchange of information and coordination of joint research efforts.

7. Conclusion

This paper examined the relation between massively parallel AI and grand challenge AI applications. Although this paper merely focused on one of many aspects of grand challenge and massively parallel AI, ample evidences consistently indicate needs for massively paralle AI for grand challenge. The evidences also suggest that massive computing, massive data, and appropriate modeling are key for the success. International cooperation would be inevitable in many of grand challenge projects.

References

[Creecy et. al., 1992] Creecy, R., Masand, B., Smith, S., and Waltz, D., "Trading MIPS and Memory for Knowledge Engineering: Automatic Classification of Census Returns on a Massively Parallel Supercomputer," *Communications of the ACM,* 1992.

[EDR, 1988] Japan Electric Dictionary Research Institute, "EDR Electric Dictionaries," Technical Report, Japan Electric Dictionary Research Institute, 1988.

[NOAH, 1992] Systems Research & Development Institute of Japan, *A Plan for the Knowledge Archives Project,* 1992.

[Evett et. al., 1990] Evett, M., Hendler, J., and Spector, L., *PARKA: Parallel knowledge representation on the Connection Machine,* Tech Report CS-TR-2409, University of Maryland, 1990.

[Evett et. al., 1990] Evett, M., Hendler, J., Mahanti, A., and Nau, D., "PRA* A memory-limited heuristic search procedure for the connection machine," *Proceedings – Frontiers of Massively Parallel Computing,* 1990.

[Furuse and Iida, 1992] Furuse, O. and Iida, H., "Cooperation between Transfer and Analysis in Example-Based Framework," *Proc. of COLING-92,* Nantes, 1992.

[Geller, 1991] Geller, J., *Advanced Update Operations in Massively Parallel Knowledge Representation,* New Jersey Institute of Technology, CIS-91-28, 1991.

[Gelsinger et. al., 1989] Gelsinger, P., Gargini, P., Parker, G., and Yu, A., "Microprocessors circa 2000," *IEEE Spectrum,* Oct. 1989.

[Higuchi et. al., 1991] Higuchi, T., Kitano, H., Furuya, T., Kusumoto, H., Handa, K., and Kokubu, A., "IXM2: A Parallel Associate Processor for Knowledge Processing," *Proceedings of the National Conference on Artificial Intelligence,* 1991.

[Hsu et. al., 1990] Hsu, F. H., Anantharaman, T., Campbell, M. and Nowatzyk, A., "A Grandmaster Chess Machine," *Scientific American,* Vol. 263, No. 4, Oct., 1990.

[Hsu, 1990] Hsu, F. H., *Large Scale Parallelization of Alpha-Beta Search: An Algorithmic and Architectural Study with Computer Chess,* CMU-CS-90-108, Ph.D. Thesis, 1990.

[IEDM, 1991] IEEE, *Proc. of the International Electron Device Meeting (IEDM-91),* 1991.

[Kitano and Yasunaga, 1992] Kitano, H. and Yasunaga, M., "Wafer Scale Integration for Massively Parallel Memory-Based Reasoning," *Proceedings of the National Conference on Artificial Intelligence (AAAI-92),* 1992.

[Kitano et. al., 1991a] Kitano, H., Smith, S. F., and Higuchi, T., "GA-1: A Parallel Associative Memory Processor for Rule Learning with Genetic Algorithms," *Proceedings of the International Conference on Genetic Algorithms (ICGA-91),* 1991.

[Kitano et. al., 1991b] Kitano, H., Moldovan, D., Cha, S., "High Performance Natural Language Processing on Semantic Network Array Processor," *Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI-91),* 1991.

[Kitano et. al., 1991] Kitano, H., Hendler, J., Higuchi, T., Moldovan, D. and Waltz, D., "Massively Parallel Artificial Intelligence," *Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI-91),* 1991.

[Kitano and Higuchi, 1991a] Kitano, H. and Higuchi, T., "High Performance Memory-Based Translation on IXM2 Massively Parallel Associative Memory Processor," *Proceeding of the National Conference on Artificial Intelligence (AAAI-91),* 1991.

[Kitano and Higuchi, 1991b] Kitano, H. and Higuchi, T., "Massively Parallel Memory-Based Parsing," *Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI-91),* 1991.

[Kitano, 1991] Kitano, H., "ΦDmDialog: An Experimental Speech-to-Speech Dialogue Translation System," *IEEE Computer,* June, 1991.

[Kitano, 1992] Kitano, H. (Ed.), *Proceedings of the Workshop on Grand Challenge AI Applications,* Tokyo, 1992. (in Japanese)

[Lee, 1988] Lee, K. F., *Large-Vocabulary Speaker-Independent Continuous Speech Recognition: The SPHINX System,* CMU-CS-88-148, Ph.D. Thesis, 1988.

[Lenat and Guha, 1989] Lenat, D. B. and Guha, R. V., *Building Large Knowledge-Based Systems,* Addison-Wesley, 1989.

[NRC, 1988] National Research Council, *Mapping and Sequencing the Human Genome,* National Academy Press, Washington, D.C., 1988.

[Sato and Nagao, 1990] Sato, S. and Nagao, M., "Toward Memory-based Translation," *Proceedings of COLING-90,* 1990.

[Stanfill et. al., 1989] Stanfill, C., Thau, R., and Waltz, D., "A Parallel Indexed Algorithm for Information Retrieval," *Proceedings of the SIG-IR-89,* 1989.

[Stanfill and Waltz, 1986] Stanfill, C. and Waltz, D., "Toward Memory-Based Reasoning," *Communications of the ACM,* 1986.

[Stormon et. al., 1992] Stormon, C., Troullinos, N., Saleh, E., Chavan, A., Brule, M., and Oldfield, J., "A General-Purpose CMOS Associative Processor IC and System," *IEEE Micro,* December, 1992.

[Sumita and Iida, 1991] Sumita, E., and Iida, H., "Experiments and Prospects of Example-Based Machine Translation," *Proceedings of the Annual Meeting of the ACL (ACL-91),* 1991.

[Elexir, 1991] Third Eye Software, *Elexir,* 1991.

[Waltz, 1990] Waltz, D., "Massively Parallel AI," *Proceedings of the National Conference on Artificial Intelligence (AAAI-90),* 1990.

[Yasunaga et. al., 1991] Yasunaga, et. al., "A Self-Learning Neural Network Composed of 1152 Digital Neurons in Wafer-Scale LSIs," *Proc. of the International Joint Conference on Neural Networks at Singapore (IJCNN-91),* 1991.

[Zhang et. al., 1992] Zhang, X., Mesirov, J., and Waltz, D., "Hybrid System for Protein Secondary Structure Prediction," *J. Mol. Biol.,* 225, 1049-1063, 1992.