AI Research in the People’s Republic of China: A Review

Jiang Xinsong
Song Guoning
Chen Yu

Shenyang Institute of Automation
Academia Sinica
People’s Republic of China

Editor's note: The AI Magazine is initiating a series of articles about AI research efforts from around the world. We are soliciting overview articles about research efforts in Western European countries, the U.S.S.R. and the Pacific Basin. If you would like to contribute, please contact Robert Engelmore (address in the Table of Contents) or myself. Thank you. —Claudia Mazzetti

Abstract

Since the 1970’s AI research has become very active in China and certain results have been achieved. This article is intended to review briefly what was and is going on in the AI field in China.

IN CHINA AI RESEARCH was begun early in the 1960’s, but it was confined to a small circle until the end of the decade. After that, many university departments awarding majors in computer science were organized. They tried to expand the applications of the computer and to develop theories. Meanwhile, the computer was also expected to improve process control. Considerable efforts were made for this purpose, but these brought few notable results, as a practical process is too complex to identify. However a veteran worker or a technician often manages somewhat better than a computer in process control. This fact suggested to scientists that AI might be a more successful approach. To tackle some obviously fascinating problems linguists and psychologists were also drawn into AI. Activity in AI has thus been increasing and certain results have been achieved. This article is intended to be a brief review of the following topics:

1. machine translation and natural language understanding;
2. theorem proving;
3. expert consulting systems;
4. computer aided instruction;
5. other areas.

Machine Translation

Machine translation was the earliest AI research performed in China and has experienced its thriving and declining times. The following three systems, developed in the Science-Technology Information Institute and Computer Technology Institute, Academia Sinica, are the typical ones. The ECTA English-Chinese Automatic Translation System (Li, 1982), directly outputting Chinese characters, translated a science-technology text of 1600 words (25 words/sentence as average, 60 words/sentence as the longest) in June, 1981, with no error. The accuracy was 90.8% when it translated a 3600-word article, and most errors can be corrected by revising the dictionary or changing the program slightly. The system uses the following procedures:
sentence segmentation—according to prepositions, conjunctions, articles and punctuation marks to segment a sentence;

template matching—to match each segment with its template and label it with feature data;

segment tree forming—to link segments with each other forming a tree structure;

tree scanning—to turn out the translated sentence.

English-Chinese Title Translation System (Jiang, 1982) which translated 5,000 metallurgic items with 80% accuracy and 100 items/60 min. speed. Among the errors 60% were lexical (easy to correct) and 40% were grammatical.

JF-111, A Universal Machine Translation System (Geng, 1982), which can theoretically translate any language into another if the grammars of the two languages are input as data.

Natural Language Understanding

Research on natural language understanding mainly focused on Chinese.

(Li, 1982) The system RJD-80 (Fan, 1981), developed in the Institute of Linguistics, Chinese Academy of Social Sciences, is a typical one. It is a Q-A system run on a small knowledge base of Chinese literature. The following are a few sentences as an example:

Q (man):
"Is the author of the drama "Teahouse" Laoshe?"
A (machine):
"Yes."

Q (man):
"Has Bajin ever written any drama?"
A (machine):
"No."

This system is a linguistic model understanding modern Chinese on the basis of linguistic theory. Its working principle is as follows:

1. Sentence parsing—250 words and 30 sentence patterns are stored. Parsing is based on Chomsky’s transformational generative grammar and Wood’s augmented transition network (ATN). For example, the sentence (Has Bajin ever written any drama?) would generate the structure:

\[(S((NP Bajin)(VP(V has)(NP Drama))) Q\),

here \(Q\) stands for alternative question;

\[(S((NP LAOSHE is a writer)\(\lor\)(CAOYU is a writer))) Q\pm),

or

\[(S((NP ?writer of TEAHOUSE)(VP(V is)(NP ?person))) Q)\).

The last NP "?Person" requires filling through searching the knowledge base.

3 Discourse analysis—to identify each pronoun appearing in different places. It is context-dependent analysis.

4 Simple reasoning and deduction—e.g. Caoyu is a dramatist \(\rightarrow\) a dramatist is a writer \(\rightarrow\) caoyu is a writer.

5 Generating answer sentence—to make up a sentence according to specified sentence pattern. For the purpose of flexibility it is possible to choose different sentence forms

Another successful project concerning natural language is the interface for the Rice Resource Retrival System (Kong, 1998), developed at Zhejiang University. It is possible to query the system in English, so a user need not remember a long boring string of commands. In order to combine several typical questions in retrieval procedure, such as: "Find out the value \(V_1\) of attribute \(A_1\) corresponding to the value \(V_2\) of attribute \(A_2\)," i.e. \(Q(A_1 = A_2)\). The system works on extracting key words and is supplemented by syntactic and semantic analysis. The ATN network method is employed in parsing (syntactic analysis) and each node represents a key word. For example, when you want to find the mean and the variance of the GRWT of all varieties of early rice in Zhejiang and Jiangsu provinces (GRWT refers to weight of 1,000 grains of rice). The key words are: mean, variance, GRWT, variety, rice, Zhejiang, Jiangsu. Through ATN analysis we get the following standard forms:

\[(mean of GRWT),

(Variance of GRWT),

when (TYPE EQ early rice),

and (OR1 EQ Zhejiang Jiangsu.)

and then retrieval from the data base gives the answer.

Zhejiang University has also developed another similar system, a language interface for a Warehouse Management System. With this interface one can inquire of the system in any sort of language if one trains the system beforehand.

To understand Chinese language by machine some institutions have set out to explore certain specific methods and fundamental corresponding theories. For instance, Qinghua University has proposed a 4-level hierarchy structure: sentence-component-structure-word (Guo, 1982).
Theorem Proving

Prof. Wu Wenjun, Institute of System Science, Academia Sinica, has established a novel approach on decision problems and the mechanization of theorem-proving in elementary geometry (Wu, 1982). Its basic idea, restricted to theorems with betweenness out of consideration and based on an entirely different principle, aims at giving a mechanical procedure which permits proof of quite non-trivial theorems in elementary geometry. Wu’s algorithm has solved certain practical problems that have been hard nuts to crack.

During the years of 1978-1982, Jilin University proposed a Generalized Resolution Principle (Wang, 1980), where it is more natural to retain the matrix forms in the statement of a theorem than to transform them into several disjunctive clauses. It is obvious that Robinson’s Resolution is a specific case of Generalized Resolution. In resolution, each time resolution is applied to two clauses, one must do it not only to the clauses themselves, but also to their factors. The proposed a strategy in resolution uses the reductant of a clause instead of the factor of a clause (Wang, 1982). In order to improve PI-Resolution and OL-Resolution, a Lock-Semantic Resolution Principle was proposed. Lock-Semantic Resolution in fuzzy logic and the deletion strategy for Generalized Resolution have been discussed elsewhere (Liu, 1978, 1980, 1982). Because OL Resolution is not complete, the modified MOL-Resolution was proposed (Huang, 1981).

In 1981, at the Mathematics Institute, Academia Sinica, proposed Ordered Unit Resolution and Ordered Input Resolution for a set of Horn clauses, and proved that they are complete (Liu 1981a, 1981b).

In 1981, Wuhan University experimented with the Unit Binary Resolution on a mini-computer PDP-11-03, and proposed some techniques for the implementation of resolution (Ceng 1981).

Expert Consulting System

By the end of the 1970’s there were only certain small prototype Expert Systems. After a few years of groping and accumulating experience, R & D work on some large practical systems was begun. Most of these systems, in the beginning, were medical systems. This resulted from the fact that a traditional Chinese medical doctor diagnoses a disease mainly by experience. To sum up the experience of an outstanding doctor and simulate his diagnostic procedures has aroused a great deal of interest among AI researchers. The first system of this kind, developed jointly by Beijing Academy of Traditional Chinese Medicine and the Institute of Automation, Academia Sinica, is used to diagnose hepatitis. Fuzzy set theory was applied in the system to make decisions in the process of diagnosis. This system has been successfully put into practice, diagnosing dozens of patients a day. The program is written in BASIC.

Another similar system, General Diagnosis System of Traditional Chinese Medicine, was developed by Shanghai Institute of Computing Technology. The main idea of the system is the concept of a “consciousness package.” The program is divided into several small pieces which are called consciousness packages. Once a certain condition has been fitted the package can also be activated by other packages. This activity is meant to simulate the process of associative thinking by human beings. A consciousness package is a basic unit of knowledge. For example, a dose of Chinese herbal medicine can be described as a package in which its property, what kind of disease to be cured and contraindication, etc., is involved.

Based on similar idea, the Institute has also recently developed a control system for blast furnaces. Chemical reactions in blast furnaces are so complicated that one can not get the desired result with traditional identification methods, so the developers summed the experience of a chief of a working group and adopted his strategy to control some key parameters. The system has achieved the same operating level as a group chief.

There are still other examples of the medical diagnosis systems. One is for gynecological diseases. The system, developed at Jilin University (Fu 1981), diagnosed 204 patients 91.7% correctly. The system is programmed in BASIC, and occupies 17K main storage. Another system is for heart diseases. Developed at Zhongshan University, it achieved a 90% correct diagnosis rate (Hou 1981). A system for nephritis, developed by Nanjing Technology College (Chen 1982), provided 278 different prescriptions from 98 symptoms with a speed of 1 diagnosis/minute. There are also systems for lung disease, pediatrics diseases and other diseases. Most of these systems are still at their prototype stage, but some of them have been put into practice. As for the principal idea, most of them are simulating a doctor and supplementing by some math processing. Some of them are MYCIN-like systems.

Zhejiang University has attempted to solve certain problems in agriculture and sideline production with the use of Expert Systems approach. One such problem is the domain of crossbreeding in sericulture (Wang 1981). The conventional approach to crossbreeding is to select parents to generate a hybrid and then, with the new crossbred generation, hybridize again to bear a second generation of hybrid, and so on. After several generations of hybridization a desirable variety would be yielded. There are two major disadvantages to following such a method: long period and low rate of success. A computer aided crossbreeding system embodies certain experiences of veteran technicians, not only selects the parents based on knowledge rules but also describes the properties of the new generation beforehand. Thus one can make a comparison before choosing the best scheme. This system adopts MYCIN backward reasoning. To express various cross-breeding schemes, take the if:

\[ W = (W1 * W2) = ((A * R) * (C * D)) \]
It represents: A and B hybridize, generate W1 (first generation); C and D hybridize, generate W2 (first generation); W1 and W2 hybridize, generate W (second generation).

Experimental results with this system: among 25 varieties of silkworms the system selected 15 combinations to crossbreed and described their properties. In a conventional way 25 combinations were selected. Among them, 4 combinations were good, and 2 combinations coincided with what the computer did.

Zhejiang University also developed systems for wheat crossing and designing figure patterns printed on cloth (Gao 1981; Pan 1981). In the later system they adopted a formal language to describe pattern structure rules.

A Traditional Chinese Medicine Expert Consultation System—ZRK82 (Guan, Liu & Fu 1982)—has been designed at Jilin University. This research not only pays attention to how the system diagnosis/treatment result tallies with that of the expert, but also is concerned with the extent to which the system's conclusions of diagnoses and treatments are good, and 2 combinations coincided with what the system's entire inference course is consistent with that of the expert. The system, which can diagnose and treat the diseases of eczema and dermatitis, has been run on the WANG 2200 computer with BASIC-II. The rate of accord between the system’s conclusions of diagnoses and treatments and the expert’s conclusions is 96.82 %. The System has six main features:

- useful (the system is now being applied in the clinic);
- educational when appropriate;
- able to explain its behaviors and results in Chinese;
- able to understand and respond to simple questions stated in natural language-Chinese;
- able to acquire new knowledge by being told; and
- able to modify itself.

A researcher at Jilin University recently visited Rutgers University, and built an expert system called ESMES (Y. & C. no date) for Experimental Strategy Mini-Expert System. The system’s characteristics are:

- Very small size;
- Ease of use;
- Ease of implementation;
- It can be implemented on a microcomputer;
- Multiple strategies of Reasoning. Its reasoning mechanism is similar to EXPERT, MYCIN, INTERNIST and PROSPECTOR. It allows the use of alternative strategies to see if any one of the schemes gives better results when tested against real problem cases; and
- Domain-independent.

Computer-Aided Instruction

Jilin University has developed some computer instruction systems. A factoring system was developed which can factorize polynomials with positive integer coefficients and exponential (Wang 1981). It provides 16 standard formulas which can be solved directly, such as:

\[(a^2 - b^2), (a^3 \pm b^3), (a^4 + 4b^4), (a^4 \pm 2ab + b^2), \ldots\]

If a polynomial to be solved is not a standard formula, mentioned above, the system factorizes it, step by step, until a standard formula is found. This system is capable of showing the whole process of calculus in detail, so that students understand the whole course of solving a problem. 90 % of the exercises in a junior middle school can be solved with this system and the method is similar to the way humans work. The SSH System (Sun 1981) simulates a student solving various applied arithmetic problems. The HDES (Higher Degree Equation Solver) (Liu 1981) system finds the accurate solutions for higher degree equations in some algebra books (e.g. college Algebra by H.B. Fine, etc.) and even some more difficult problems which are not in these books and which cannot be solved by means of existing methods. A Proving System for Trigonometry Equality (Wang 1982) can prove the trigonometry equalities in a high school course. A Factorization Teaching System (Sun 1982) can give an explanation and a justification for a polynomial factorization, and therefore can be used for teaching and study.

Others

1. **Program Synthesis.**
   Beiging Aeronautical Engineering Institute has developed a system for synthesizing recursive programs using a structure inducing approach (Sun 1982). Some simple functions have been synthesized with this system, such as SORT (J) and REVERSE (J).

2. **Robotics.**
   Shenyang Institute of Automation, Academia Sinica, has developed a playback-type robot (Song et. al. 1981; Ma 1982). It has 5 freedoms: arm stretch out, draw back, pitching movement and turn round; wrist turn round and swing; it has been implemented by computer control in two ways: point to point and linear track and speed. In point-to-point control the playback accuracy is less than 1.5 mm, in linear-track control retracking accuracy is not more than 10 mm. The maximum weight grasped by the hand is about 35 kg.

3. **Model Logic about “Knowing”**
   Prof Ma Xiwen, Beiging University, has established a predicate calculus of modal logic about “knowing”, (Ma 1982) including its formal system W and its semantic interpretation JS. He has discussed certain principal features of the system W-JS, with some typical examples, such as the well-known puzzle “Mr. S and Mr. P”.

References


Huang Bingchao (1981) OL-guijie de fanlie ji gaijin de MOL-guijie ("OL-Resolution and Improved MOL Resolution"). Jilin University Journal (Jilin Daxue Xuebao), No. 4.


Liu Ruling (1981a) HORN-ji de xiaoji dingli ("A Resolution Theorem for HORN Clauses"). Science Sinica (Zhongguo Kexue), No. 7.


Liu Xuhua (1978) Yizhong xin de yuyi guijie yuanli ("A New Type of Semantic Resolution Principle"). Jilin University Journal (Jilin Daxue Xuebao), No. 2.


Pitman’s Advanced Publishing Program announces a new venture

Research Notes in Artificial Intelligence

Topics to be covered:

Automatic programming, expert systems (ICAI/medicine/VLSI), learning, logic programming, natural language, planning, representation, vision.

Nature of Publication:

- Rapid publication – usually within 3-4 months of receipt of final copy.
- Publication of monographs, collections of related papers, polished versions of theses and lecture notes on advanced topics.
- International distribution, yet reasonably priced at under $20 per volume

Editors:

David Barstow
Sclumberger-Doll Labs

Hans Berliner
CMU

Wolfgang Bibel
Univ of Munchen
Germany

Jaime Carbonell
CMU

Randall Davis
MIT

Takeo Kanade
CMU

Doug Lea
Stanford University

Drew McDermott
Yale University

Tom Mitchell
Rutgers University

Makoto Nagao
Kyoto University

Japan

Erik Sandewall
Linkoping University

Sweden

Ted Shortliffe
Stanford University

Yorick Wilks
Essex University

United Kingdom

Richard Young
APU, Cambridge
United Kingdom

Main Editors:

Derek Sleeman
Dept. of Computer Science
Stanford University
Stanford, CA 94305
Phone: (415)497-3257
ARPA address: Sleeman@SUMEX

N S Sridharan
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903
(201)932-2894
ARPA address: Sridharan@Rutgers

Contact the editors or the publisher for a copy of the prospectus or additional details.

Pitman Publishing, Ltd. 128 Long Acre London WC2E 9AN England