

## 特別論説



## 情報処理最前線

## コンピュータは人間に勝てるか?†

小谷善行†† 松原 仁††† 大沢英一††††

## 1. はじめに

コンピュータが人間社会に現れて以来、それはさまざまな分野で、非常に便利な道具として使われてきている。そこでは人間とコンピュータが、おのおのの得意な仕事に合わせて、うまく役割分担されている。この場合、コンピュータの仕事を人間にやらせるのは非常に困難であったり、また非常に能率が悪かったりする。逆に、人間の仕事をコンピュータにやらせるのは、まったく不可能であったりする。

では、コンピュータと人間を同じ土俵にあげることができたら、コンピュータはどこまで人間に迫ることが可能だろうか？ コンピュータを人間のように振舞わせるのは、人工知能 (AI) の研究テーマであるが、まず直感的にもうまくいきそうなのは、チェス、将棋などのゲームにおいてコンピュータを人間のレベルにすることである。これらのゲームは問題が厳密に定義されているので形式的に表現することができるし、目標の達成度も人間と対戦することによってはっきり分かるからである。

本稿では、いくつかのゲームにおいて現在、コンピュータが人間に勝てるかどうか、そして勝てないとしてもどこまで人間に近づいたかについて検証してみることにする。

まずは、チェスについてみてみよう。コンピュータチェスの歴史は、ほとんどコンピュータ自身の歴史に重なる。数十年の歴史を経て、いまついにコンピュータが最強の人間 (国際グランドマス

ター) とほぼ同じ強さをもつようになったのである。このことは人間が月に到着したのと同程度以上に科学技術の歴史的意義があると思う。月に行くという物質的なことだけではなく、情報处理的なこと、内的世界のことにも意味があるはずである。両者は、直接的にはその事実自体が人間のために有益にみえないという点、そして長期間のさまざまな技術の蓄積と総合のもとに達成したという点で似ている。逆に、月に行くというのは国家的プロジェクトで行われたのに対し、コンピュータチェスは非常に多くの個人的な意思のもとに発展してきた。これは、特徴的な違いである。

次の話題はコンピュータ将棋である。上で述べたように、チェスにおいては、コンピュータが最強の人間に勝つのは時間の問題である。その時点で長い間 AI の研究対象であったチェスの役割もほぼ終わる。AI の研究対象としてチェスの次に来るのは将棋で、その次に来るのは囲碁であろう。チェスを強くする試みの中でゲーム木探索のさまざまな手法が提案され、AI の進歩に貢献してきた。チェスに比べると将棋はまだまだ人間のほうが強いが、コンピュータ将棋を強くするためには新たに多くの技術的ブレイクスルーが必要であり、その試みを通じてこれからの AI の進歩に貢献すると思われる。

そして、最後の話題のタイトルは「コンピュータは人間をだませるか？」という一見奇抜な質問である。このタイトルだけでは、何を述べようとしているのか、読者の方々にはさっぱり予想がつかないかも知れない。ここでの目的は、Alan M. Turing によってなされた、20 世紀末ごろのコンピュータの知能に関する予測について、いまの時点で検証してみることにする。これがこのタイトルとどのような関係になっているのかも徐々にご理解いただけることと思う。ここではまず、Turing

† Can Computers Defeat Humans? by Yoshiyuku KOTANI (Dept. of Electronic and Information Science, Tokyo University of Agric. and Tech.), Hitoshi MATSUBARA (Electrotechnical Laboratory) and Ei-Ichi OSAWA (Sony Computer Science Laboratory).

†† 東京農工大学電子情報工学科

††† 電子技術総合研究所

†††† ソニーコンピュータサイエンス研究所

が提案した計算機の知能テストの内容とその予測について述べることから始め、そしてこのテストを題材にしたコンテストの話題について触れる\*。

## 2. 人間対コンピュータ (その1)

### —チェス— (小谷善行)

#### 2.1 コンピュータチェスの歴史

コンピュータチェスの始まりは、コンピュータの始まりと同じく Charles Babbage にまで遡る。もちろんそのときにはプログラムが作られたわけではない。しかし、彼は、1840年代の自著の中で、チェスのようなゲームをプレーする純粋に機械的な方法について述べ、先読みの概念に相当するものを示している。本格的にコンピュータチェスについて考え始めるのは、Shannon と Turing が相次いでコンピュータチェスの論文を発表した 1950 年ごろである。驚くべきことに、そのときにはもう、Min-Max 戦略や局面評価関数などの基本的概念が示されている<sup>15)</sup>。また、Turing の考えにより、指し手をシミュレートした対戦が行われている。

1956 年ごろになると、6×6 のミニチュア盤でのプログラムが動作し、初心者との対戦が行われている。1958 年ごろには初めて完全なチェスプログラムが作られた。1967 年には、Greenblat のプログラム MACKHACK VI のレーティング (チェスの強さの値。後述) が 1640 になった。このプログラムで、現在常識となっているゲーム木探索のさまざまなアイデアが出現している。たとえば、探索手の並べ換え (先に良い手が見つかる効率が良い)、ポーン構造 (ポーンの配列が、良い評価特徴になる)、同一局面情報の活用 (同じ局面を再度調べないようにする) などである。

1970 年になると、第 1 回北アメリカ・コンピュータチェス選手権がニューヨークで開催された。このときは、Slate らの CHESS 3.0 が優勝した。その後、この選手権は毎年行われ、それを目的として多くのコンピュータチェスがしのぎを削ることになる。1974 年からは、世界選手権も開催されるようになり、1977 年ごろには、CHESS X. X (CHESS 3.0 の後継) がほぼ最強のプログラムとして君臨した。このころまでには、主にアルゴリズムやデータ構造などのプログラミングの

アイデアが強さを左右していた。

1980 年になると、Thompson と Condon による BELLE が登場する。これは、PDP 11 と特別なハードウェアを組み合わせたもので、速度を重視したプログラムが中心になってくる。つまり、可能な手をすべて高速に読む技術の競争になってきたのである。1983 年ごろには、Hyatt, Gower, Nelson による CRAY BLITZ が最強になる。これは名前のおり、スーパーコンピュータ CRAY 上で動く、その後、1985 年ごろから Berliner らによる HITECH、およびその少し後から、Hsu らによる DEEP THOUGHT (前身は CHIPTEST) が出現する。BELLE, HITECH, CRAY BLITZ の戦国時代を経て、1989 年以降は、DEEP THOUGHT が水をあけてトップに立っている。マイクロコンピュータ (パソコンや市販の小さいチェスマシン) の分野では、MEPHISTO がここ数年では最強である。

#### 2.2 コンピュータチェスの強さ

ここでは、コンピュータチェスの強さについて述べる。

チェスの強さを表現する尺度として、ELO システムというものが考えられており、一般にレーティングと呼ばれる。これは、対戦するごとに変化する数値で、十分に多くの対戦を行うと正確な強さを示す値に落ち着く。米国チェス連盟のレーティングによると、チェスの指し方を知っている程度の人のレーティングは約 800 である (国際レーティングは、米国のものから 100 を引いた値である)。レーティングが 2200 になると、米国のマスターというタイトルが自動的に与えられる。世界のグランドマスターはたいてい 2550 以上で、世界最強といわれる Gary Kasparov は 2900 である。

これに対して、コンピュータチェスの強さは表-1 のように年々改善されてきた<sup>16)</sup>。

表-1 コンピュータチェスのレーティングの改善 (文献 16) のものを修正)

年	プログラム	レーティング
1967	MACKHACK VI	1640
1974	CHESS 4.0	1730
1977	CHESS 4.5	2271
1980	CHESS 4.7	2168
1981	BELLE	2258
1983	CRAY BLITZ	2363
1985	HITECH	2530
1988	DEEP THOUGHT	2745

\* 編集: 長尾 確 (ソニーコンピュータサイエンス研究所)

これらの値に直線を当てはめてみると、

$$49.2 \times (\text{西暦年} - 1900) - 1697$$

という式になるという。そして、これを外挿すると、1994年には Kasparov に十分勝るとされるレーティングである 2911 を越える。1990年ごろには、まだ DEEP THOUGHT は Kasparov と対戦して負けているが、その敗因の多くは、プログラムのバグや定石の整備があまりされていない点であった。

このように、実質的に 40 年ほどの歴史を経て、コンピュータチェスは最強の人間に匹敵する強さになったのである。

### 2.3 コンピュータチェスの技術

コンピュータチェスの技術について、ごく一部だけを述べる。その中心は、 $\alpha\beta$  枝芻に代表されるゲーム木探索である。これは、個々のゲームからは独立した技術であるが、コンピュータチェスの研究と密接に結びついて進歩した。この周辺に、さまざまなチェス特有の技術が生まれてきている。

終盤に関して、残り駒の種類による終盤データベースを用いた手法がある。これは、ゲーム木では、処理しえないもので、可能な局面のデータベースを参照することにより、それらが勝ちか負けかをすべて求める。ある残り駒のパターンは引き分けと思われていたのが、実は一方が勝つというのが機械的に証明された例もある。そのための手順が非常に長かったので、チェスの引き分けルールが変更された。

ハードウェアについても、さまざまなものがある。たとえば、DEEP THOUGHT は 3 台のプロセッサで並列的にゲーム木を探索していた。また、指し手生成器 (move generator) を専用の LSI で作ってある。こうしたハードウェアの性能の高さが DEEP THOUGHT の強さを形づくっている。

### 2.4 この章のおわりに

コンピュータチェスが最強の人間に勝つまでには、あと若干の年数が必要だろう。Kasparov が対戦するとすれば、当然死にものぐるいで対策を立ててくるであろう。もちろん、人間が負けるのは時間の問題であるが、人間側も進歩するのである。

その後、コンピュータチェスの歴史が終わると

考える人もいるが、そうとは言えないと思う。人間と対戦する必要はもはやなくなり、さらに強いものを目指して、マシン同士の戦いが続くのである。また、コンピュータチェスの延長線上には他のゲームの問題がある。たとえば、コンピュータ将棋にとって、コンピュータチェスはあまりにも大きな先輩であり、学ぶべきことは非常に多い。

本章で述べた歴史に関して、さらにくわしいことは文献 16) に書かれている。また、コンピュータチェスに関する研究で、時代を画したような論文が文献 15) に再録されている。コンピュータチェスを深く追求しようとされる方は、ICCA (国際コンピュータチェス協会) に加盟するとよいだろう (宛先は文献 16) の中にある)。

(平成 4 年 12 月 9 日受付)

## 3. 人間対コンピュータ (その 2)

### —将棋— (松原 仁)

#### 3.1 チェスと比べた将棋のむずかしさ

将棋とチェスは起源が同じとされ、実際よく似たゲームである。しかし、

- 将棋は  $9 \times 9$  の盤面でチェスは  $8 \times 8$  の盤面である (将棋は駒の総数が 40 でチェスは 32 である)。

- 将棋は 8 種類の駒でチェスは 6 種類の駒である。

- 将棋は相手から取った駒を持ち駒として再利用できるがチェスにはできない。

- 将棋では玉 (および金) 以外の全ての駒が敵陣にはいれれば成れるのに対して、チェスではポーンのみが成れるだけで、しかも敵陣の最も奥まではいらなければならない。

という違いがある。これらの違い、主に持ち駒制度があるかないかによって、将棋とチェスは情報処理的な性質が非常に異なっている。チェスは終盤になると駒の総数が少なく可能な指し手の平均が数十であるが、将棋は駒の総数がかわらず可能な指し手の平均が数百になる。ゲームとしての場合の数はチェスが 10 の 120 乗に対して将棋は 10 の 200 乗にも及ぶ。さらに、チェスは駒の損得という要因だけで形勢の判断ができるのに対し、将棋は駒の損得、駒の働き、手番などさまざまな要因の組み合わせとして評価しなくてはならない。評価の計算に時間がかかるだけでなく、駒

の働きという、定量化がきわめて困難な要因を扱わなければならない。チェスに比べて将棋の局面の評価がむずかしいのは、将棋はチェスより大駒（遠くまで動ける駒）が少なく小駒（大駒以外の駒）が多いためと思われる。

したがって、チェスシステムを作るのに有効だった手法、すなわち専用マシンによる力任せの全幅探索手法はそのままでは将棋システムを強くするためには使うことができない<sup>1)</sup>。またハードウェアの進歩を待つだけでは 10 の 80 乗の差は到底克服できない。たとえば、可能な手の数の平均がチェスは 35 で将棋は 350 としよう。9 手先まで全幅探索をすることで、チェスの  $35^9$  と  $350^9$  というオーダーの差があるうえに、一つ一つの局面の評価にチェスより将棋のほうがはるかに時間を要するからである。将棋では全幅探索が非現実的なので、強くするためには何か新しい技術を生み出す必要がある。このような理由で、コンピュータ将棋を強くすることはそれ自体が楽しいだけでなく、AI としても十分に価値のある研究対象になっている。

3.2 詰め将棋システム

詰め将棋は、王手の連続で相手玉を詰めるという将棋から派生した人工的な（一人）パズルであり、むずかしさだけでなく新規性や美しさなどによって評価される一種の芸術作品でもある。王手の連続という規則のために場合の数が限定されるので、(指し) 将棋よりもコンピュータに向いている。詰め将棋に関してはこの数年で急速にコンピュータの能力が進歩した。たとえば野下の作成した詰め将棋解答システム t1<sup>6)</sup> は、図-1 の詰め将棋をワークステーション上で 30 秒あまりで解いた\*。

これは人間の専門家並みの速さである。t1 以

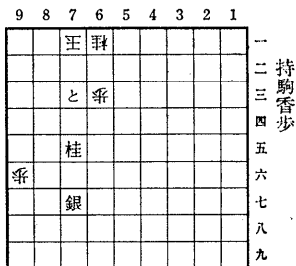


図-1 詰め将棋の例

降もアルゴリズムは進歩し、さらにはハードウェア能力も進歩しているので、現在すでに数十手詰めまでの詰め将棋の解答能力は人間の専門家のそれを越えたと言ってよい (t1 の新版 t2 は図-1 を 10 秒で解くまでに進歩している)<sup>2)</sup>。すなわち、詰め将棋を速く解く勝負においてコンピュータは人間に勝つ。詰め手数が数十手から百手を越えるような詰め将棋を解くことに興味は移りつつある。また、詰め将棋の問題は（おおざっぱに言って）答は一通りでなければならない。答が 2 通り以上存在するものは「余詰め」問題と呼ばれ、問題として不完全なので棄却される。ある問題が余詰めかどうかの判定はこれまで人間が手作業で行っていたが、探索木を綿密に調べなければならないので、手間がかかるうえにミスも多かった。最近ではコンピュータでチェックすることが可能になり、以前は完全とみなされていた有名な問題が余詰めと判明したりしている<sup>3)</sup>。近い将来には、コンピュータに詰め将棋の問題を作らせる試みも出てこよう。これはコンピュータによる芸術の創造という観点から眺めても面白い。

3.3 いまの将棋システム

この 4、5 年の間、強い将棋システムを作る試みが盛んになされるようになってきた<sup>3)</sup>。パソコン向けの商用のよく売れるソフトウェアを作ることが主な原動力となっている。将棋は海外での知名度がいま一つなので、もっぱら国内で開発が進んでいる。現在の実力はアマ 2、3 級程度である。将棋を真面目に勉強した人間にはまだ勝てないが、初心者よりは強く、初級者といい勝負である。時につれて着実に実力が上がっているため、近いうちにアマ初段レベルに到達するのは確実である。ただし、それ以上実力があっても商用ソフトとしてよく売れるとは思えないので、今後名人を目指すには商売とは違う研究としての動機が求められるよう。

コンピュータ将棋協会 (CSA: Computer Shogi Association) という非営利目的のグループがあり、そこで毎年コンピュータ将棋の選手権を開いている<sup>7),8)</sup>。観戦している人間が熱くなって助言をしても、コンピュータはどこ吹く風である。そこでここ数年上位に居るのは、森田将棋、極、永世名人、柿木将棋、といった商用ソフト (の改良版) である。

\* 門脇芳雄氏の報告に基づく。正解手順は本章の最後に記す。

						王	王	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							
						馬	馬	馬							

図-2 森田将棋 vs. 柿木将棋


図-3 永世名人 vs. 極

図-2は91年のコンピュータ将棋選手権での森田将棋と柿木将棋の対戦の一局画面（図は見やすさを考えて実際とは先手後手を逆にしてある）である。

すでにこちら側の森田将棋が優勢であるが、ここで森田将棋は▲4三歩という手を指し、柿木将棋が飛車を逃げたので、森田将棋の完封勝ちとなった。不利なときにはいちかばちかの勝負手を指すべきであり、柿木将棋としては同飛としたほうがまだ勝負の綾があったと思われる。将棋システムは不利な局面でも評価値が最善の手を指すが、人間のように敵が間違えやすい手を指すといった工夫が必要であろう。図-3は同じ選手権での先手永世名人対後手極の対戦の一局画面であるが、ここで先手は▲7七飛という目を疑うような手を指して序盤から劣勢に陥ってしまった\*。

いまの将棋システムは1局のうちに数回はとんでもない悪手を指す。全数探索ができないので有力な一部の手だけを読むのだが、その際にたまたま当然の手を見逃してしまうからである。プログラムの一種のバグであるが、これをいかに減らすかが強くするための課題である。

3.4 囲碁システム

囲碁は将棋と並んで日本で盛んなゲームであ

る。将棋がもっぱら国内だけでプレイされているのと異なり、囲碁は中国、韓国、台湾などでもかなりプレイされている。他に類似のゲームが存在しないために、西欧でも多くの愛好者がいる。囲碁システムの研究も日本ではなく西欧で始まったほどである。囲碁は盤面の広さが19×19もあり、将棋よりもさらに探索空間が広い。場合の数は10の300乗にも及ぶ。局面の評価もどのような要素を組み合わせればよいかすら分かっていない。石の死活の判断もむずかしい。囲碁と将棋の差は将棋とチェスの差よりさらに大きいので、囲碁システムを強くするのは将棋よりもはるかに大変である。

囲碁システムの研究も最近では国際的に活発になり<sup>4),5)</sup>、コンピュータ囲碁の世界選手権が毎年開かれている<sup>10)</sup>。現在最も強いものはオランダのGoliathである\*が、その実力はまだアマチュアの7, 8級程度である。囲碁のプロの最高のタイトルである本因坊に勝つにはまだかなりかかるであろう。

3.5 この章のおわりに

チェスが研究対象としての役割を終えつつあり、次の対象として将棋や囲碁が適当になってきた。詰め将棋システムは人間を越えつつある。将棋システムはまだ名人の実力にはほど遠いが、早ければ十数年のうちに遅くとも二、三十年後には名人に勝つシステムが実現できるであろう<sup>9)</sup>。その実現のために最大のネックとなるのは、ゲームを題材とする研究に対する、日本における「冷たい雰囲気」である。これまで将棋や囲碁のシステム化がチェスに比べて遅れていたのは、将棋や囲碁がチェスに比べて場合の数が大きいという技術的理由もさることながら、「冷たい雰囲気」で研究が満足にできなかったという社会的理由が大きい。日本人の名人が外国製の将棋システムに負けるという二重に辛い事態を迎えないためにも、貴重な日本文化である将棋や囲碁のシステム化研究に対する世間の暖かい目を期待したい。日本が世界に対して研究の独自性を発揮できるまたとないチャンスである。これから研究をスタートさせても、いまならまだ十分にトップクラスへ追いつくことができる。自分が名人になるのは無理だが、名人を作るのは可能かもしれない。

\* これはたまたまの悪手の例としてあげたものであって、永世名人は最も強い部類のソフトである。

\* 92年の選手権では番狂わせで負けたが、実力は群を抜いている。

## 付 録

## 図-1の正解手順

■7二歩□8一玉■8五香□8四桂合■同香□8  
 三金合■同香不成□9二玉■8二香成□9三玉■  
 8三成桂□9四玉■8四金□9五玉■8七桂  
 まで15手詰め

(平成4年11月30日受付)

## 4. コンピュータは人間をだませるか？

—チューリング・テストをめぐる—

(大沢英一)

## 4.1 チューリング・テストとは

計算機科学の基礎に大きな足跡を残した偉大な数学者 Alan M. Turing は、1950年に出版した小論文<sup>19)</sup>において、計算機の知能を判定するためのテスト（一般にチューリング・テストと呼ばれる）を提案している。そのテストは、模倣ゲームと呼ばれる3人のプレーヤによって行われるゲームに基づく。3人のプレーヤのうち、1人は男性、1人は女性、そして残りの1人は質問者である。質問者は他の2名（回答者）から離れた部屋にいて、端末などにより回答者と自然言語により会話ができる。このゲームにおける質問者の目標は、2人の回答者のうちどちらが男性で、どちらが女性であるかを決定することである。実際にこのようなゲームを人間どうして行った場合、質問者はある割合で間違った答えを出すであろう。そこで、このゲームを計算機の知能をテストするという目的に利用するために、ゲームの設定を変えて、たとえば回答者の男性を計算機に置き換えてみる。このとき、質問者が性別の判定を間違える割合が人間どうして行ったゲームの場合とほとんど変わらなければ、その計算機は（人間をだませるくらいの）知能を持っていると結論付けて良いのではないか、というのが Turing の目論みである。このゲームは計算機の知能のテストとして適切なのであろうか？ また、より原点に戻って知能とは何なのか？ 知能の程度はどのように測ることができるのか？ これらの点について、この論文が発表された直後から現在に至るまで哲学者や計算機科学者の間で盛んに議論が行われてきている。

## 4.2 第1回 Loebner 賞大会

さて、Turing は前述の小論文において、「50年以内にはギガ・オーダーのストレージをもつ計算機

が利用可能になり、模倣ゲームに計算機を参加させて5分程度の会話の後に結論を下すような場合、質問者（人間）が30%以上の割合で判定を誤るように計算機を知的にプログラムすることが可能になるであろう」と予測した。はたしてこの予測は正しかったのであろうか？ この問いに答えることを目的とした第1回 Loebner 賞大会（以下、第1回大会と略す）が、1991年にアメリカ合衆国ボストン市で開催された。以下では、その大会について Epstein による報告<sup>12)</sup>を基に簡単な紹介を行う。

## 4.2.1 大会委員会による事前の議論

Loebner 賞大会委員会は大大会を開催するに当たって、チューリング・テストを実施するうえでのさまざまな問題点について議論した。一連の議論を通して、第1回大会で採用するテストではオリジナルのチューリング・テストからの変更を余儀なくされた。前述のチューリング・テストでは、1人の質問者に対して1人の回答者と一つのプログラムという設定であったが、大会では、このような会話状況の設定をとることをあきらめた。その理由は、複数のプログラムが参加してきた場合、各プログラムとペアを組む回答者のレベルを公正に揃えることができないであろうという予想による。もし回答者にばらつきがあった場合、それにより相対的に不利になったり有利になるプログラムが出てきてしまう。よって第1回大会では10個の端末を用意し、そのうちの複数の端末は人間により、また残りの端末はプログラムにより操作される仕掛けをとった。審査員である質問者には、少なくとも2個以上の端末は人間により、また、少なくとも2個以上の端末はプログラムにより操作されていることが事前に伝えられた。10人の審査員は端末を通して約15分間相手と会話を行い、端末を操作している相手の知能のレベルを評価することとした（オリジナルのチューリング・テストのように相手の性別を判定することが目標ではない）。

審査員による評価の方法も問題であった。プログラムの会話内容が人間と区別できないと言うためには、どのような評価方法が適切なのであろうか？ この点に関しては、Turing の原論文でも具体的な手法について述べられていない。実際に第1回大会で採用された評価方法とは次に述べるよ

うに比較的単純なものであった。審査員は、各端末による会話内容がどのくらい人間らしいかにより、端末を順位付けする。全審査員による順位の平均をとって、最も高い順位を得たプログラムを優勝とすることにした。また各審査員は、自己の順位評価において、どの順位までが人間であり、どこからがプログラムであるかという基準線を示すことを求められた。このような評価方法により、もしいずれかのプログラムの平均順位が、ある人間の平均順位よりも高ければ、そのプログラムはこの現代版変形チューリング・テストをパスしたと結論付けてもよいのではないだろうか、と委員会は考えた。

また、端末への表示をバッファリングするかどうかの問題となった。つまり、人間がタイプしたときのようなたどたどしい感じで端末に文字を表示することをプログラムに許すかどうかである。委員会のメンバである Weizenbaum (自然言語対話システム ELIZA<sup>21)</sup>の開発者)などは、プログラムにそのような振舞いをさせることは容易であり、結果には無関係であろうと予想した。しかしながら、委員会はこの問題についての結論を出すことはできず、最終的には、プログラムがどのように文字出力を制御するかはプログラマの自由とした。よって、文字の出力はバースト・モード(出力をバッファリングしておいて一挙に表示)、チャット・モード(一文字ごとに表示)のどちらでもよく、また、ミス・スペルの訂正やバック・スペースの使用なども許される。

さらに委員会は、現在の自然言語処理の技術水準からして、プログラムが自由主題会話を取り扱うことは困難であろうと判断した。この問題に対処するために、端末ごとに会話の主題を限定することとした。各端末でどのような主題の会話ができるか分かるように、端末の横に主題を明記した紙が貼られた。回答者も与えられた主題以外の事柄を話すことは禁じられた。プログラムが会話する主題に関してはプログラマの自由とした。実際に選ばれた主題は、ドライ・マティーニ、気まぐれな会話、女性の洋服、シェークスピアの演劇、などであった。

このように、(1)オリジナルのチューリング・テストの主旨になるべく沿うように、そしてまた、(2)現時点の自然言語処理技術に基づくプロ

グラムの稚拙さをカバーするように、第1回大会では会話状況設定にさまざまな制約をおくこととなった。

### 4.3 第1回大会の様子と結果

第1回 Loebner 賞大会は、1991年11月8日、アメリカ合衆国ボストン市の計算機博物館に200人の観客を集めて開催された。技術的な問題から、10個の端末のうち2個が使用不能となったため、8個の端末を用いることとなった。そのうちの2個は人間(回答者と呼ぶ)により操作され、残りの6個は北米大陸の各地に散らばっているプログラムからモデムを通して操作された。なお前述のように、審査員には、8個の端末のうち少なくとも2個以上は人間により操作されていることが伝えられた。

新聞広告により公募した10人の審査員は、7人が女性で3人が男性、そして年齢的には19歳から48歳までの人たちであった。これらの審査員は計算機科学や人工知能とは無関係な人たちであるが、全員英語には堪能である。

大会で1位と2位の順位を得たのは、回答者として参加した2名の女性であった。また、事前の予想どおり、多くのプログラムはほとんどの会話の局面でうまい応答ができなかった。応答のまずいプログラムに共通にみられたことは、審査員の一連の質問に対して連続して無回答を出し、また、特定の返答(たとえば“Really can't say. Oh well.”とか、“don't know”など)を繰り返すことであった。

大会で3位の順位を得た(つまり、プログラムの中では1位となった)のは、Joseph Weintraubによるプログラム(主題は「気まぐれな会話」)であった。このプログラムの平均順位は上位2名の人間の平均順位とそう離れてはいなかった。驚くことに、10名の審査員のうち5名はそのプログラムに割り当てられた端末を操作しているのは人間であると判断した。また、ある審査員の判定においては、そのプログラムの順位のほうが2名の回答者のそれよりも高かった。Weintraubのプログラムは、他のプログラムよりも実質的にうまく振舞った。また、このプログラムは人間のタイピングのリズムと文字訂正などをほぼ完璧にシミュレートした。Weintraubによるプログラムとある審査員との会話の一部を以下に紹介する

(Cambridge Center for Behavioral Studies の Robert Epstein 氏の好意による). *Judge 6* は第6番目の審査員の意味で, *Terminal 5* が Weintraub のプログラムの応答である. 応答の時刻が明記されている. スクリプトにはタイプ・ミスやモデムの送信エラーがそのまま保存されている.

01: 32: 25 *Judge 6*

Hello, earth person!

01: 33: 03 *Terminal 5*

Hello there judge, are you ready to have some fun? How ya doin, judge? Hello-how's the weather up there today?

01: 33: 35 *Judge 6*

Dark clouds with occasional bursts of sunshine (or rather camera flashes).

01: 34: 45 *Terminal 5*

The trouble with the rat race is that even if you win you're still a rat. Whimsy ca be flimsy, right? How important is that to you?

01: 35: 26 *Judge 6*

can't say I understand the question...but the whimsy of poet is one thing and the whimsy of most of us is another.

01: 36: 51 *Terminal 5*

How do yo uknow you can t say you understand the question...but the whimsy of a poet is one thing and the whimsy of most of us is another.

01: 37: 37 *Judge 6*

Are you repeating me just for laughs?

01: 38: 04 *Terminal 5*

I'm sorry, I do sometimes repeat myself...perhaps it's because i can't my electr] nic organizer.

このプログラムの特徴は, Weizenbaum の ELIZA のように, 相手の発話の内容を自分の応答に反映することである. また, 会話になんらかの行き違いを生じた場合には, 人間のおかすようなタイプ・ミスをシミュレートしながら非常に上手に謝罪して回復している.

大会の結果でもう一つ興味深いことがある. 2名の回答者のうちの1人が選んだ主題は「シェークスピアの演劇」であった. 10名の審査員のうち3人は, その回答者に割り当てられた端末を操作しているのはプログラムである, と判断した. ち

なみに, その端末の出力はバッファ・モードであった. 実は, この回答者はシェークスピアの専門家であった. そして彼女は, 会話の途中でしばしば, シェークスピアの作品から長い文章を一言一句そのまま正確に引用したのである. 何人かの審査員は, それを人間技とは思えないと判断したようである.

#### 4.4 考察

第1回大会で Weintraub のプログラムが優勝した理由は何なのであろうか? Epstein は以下の二つが主な理由ではないかと予想している<sup>12)</sup>.

- 参加した6つのプログラムのうち, 人間のタイピングのリズムやタイプ・ミスなどを巧妙にシミュレートしたのは Weintraub のプログラムだけである.

- Weintraub が選んだ「気まぐれな会話」のような会話主題においては, 審査員がなんらかの質問をしても, プログラムは冗談だけを言って済ませてしまうことが可能である. Weintraub のプログラムは, このような主題の性質をうまく利用しているように思える.

確かに, これらの要因は評価結果に大きな影響を与えたであろうが, 同じように重要な要因として以下で述べることも関連したのではないかと推察する.

Weintraub のプログラムが成功した理由は, ELIZA において実現されたような, 場当たりのではあるがとりあえず会話を適当に進行させるような, 特殊な技法に負うところが大きいのではないだろうか. この技法の特性は, 会話相手の発話に対して, 言い換えなどのような, とりあえず何か関連のありそうな反応だけでも示して, 会話を進行させることである. このような特性は, 人間のように資源制限を受けた (resource bounded) 知能の, ある重要な側面をうまく表現しているのかも知れない. これまでに開発されてきた自然言語対話システムの多くは, 人工知能研究が伝統的に重視してきた熟考的 (deliberative) プロセスを中心に据えている. 熟考的と対立する概念は即応的 (reactive) であるが, Weintraub のプログラムは質問者の発話に対して即応しながらも, 内容的にはなんらかの熟考を予期させるような応答を返ししながら会話を進めていたとみることができる (実際に熟考しているのかどうかは, 気まぐれな会話



という主題の性質により、うまくごまかされていると考える)。おそらく人間は、この熟考と即応のバランスをうまくとりながら、会話をすすめているのではないだろうか。知能のこのような側面に関する研究は、最近になっていくつみられるようになった<sup>11),17),20)</sup>が、いまだ十分とはいえない。今後は、自然言語対話などの研究においても、このような観点からの考察を十分に行う必要があると思われる。

なお、1992年10月号のSIGART BULLETINにはチューリング・テストに関する小特集が組まれている<sup>13),14),18)</sup>。この小特集では、チューリング・テストの適切さや計算機の知能を測るためのテストに関する要求などについて、いくつかの観点から意見が述べられている。チューリング・テストに関しては賛否両論があるが、チューリング・テストの最も重大な問題点は、それにパスしたことが何を意味するのかが適切に定義されていないことのように思われる。このように、チューリング・テストに関する議論は絶えない。しかしながら、Loebner賞大会のような知的プログラムのコンテストを開催することは、知能に関するさまざまな観点からの考察を深め、人工知能の理論、技法をさらに発展させる可能性があるということに意義を見出したい。

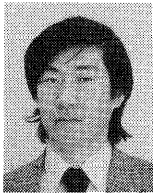
#### 4.5 この章のおわりに

Turingがチューリング・テストにパスする計算機の出現を予測した20世紀末まではまだ数年残されている。しかし、あと数年のうちにオープン・エンドなチューリング・テスト(第1回大会でおかれたようなさまざまな制約を取り除いたもの)にパスするようなプログラムを設計することは、現時点の技術水準からすると非常に困難なタスクであるように思われる。この記事をお読みになった皆さんの中で、何か良いアイデアをお持ちの方は、ぜひ一度このLoebner賞大会に参加されてみてはいかがであろうか? 大会の詳細に関する情報はCambridge Center for Behavioral Studies(住所: 11 Waterhouse Street, Cambridge, Massachusetts 02138, U. S. A., 電話: +1-617-491-9020)から得られる。

(平成5年1月5日受付)

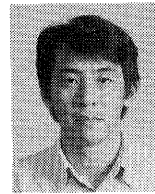
#### 参考文献

- 1) 飯田弘之, 小谷善行: エキスパートの思考をモデルとしたゲーム木探索の方式, 情報処理学会論文誌, Vol. 33, No. 11, pp. 1296-1305 (1992).
- 2) 伊藤琢巳, 野下浩平: 詰将棋を速く解く2つのプログラム, 情報処理学会34回プログラミング・シンポジウム (1993).
- 3) 小谷善行, 吉川竹四郎, 柿木義一, 森田和郎: コンピュータ将棋, サイエンス社 (1990).
- 4) 実近憲昭: 囲碁のコンピュータ化, bit, Vol. 20, No. 7, pp. 761-769 (1988).
- 5) 実近憲昭: ゲームとAI, 人工知能学会誌, Vol. 5, No. 5, pp. 527-537 (1990).
- 6) 野下浩平: 詰将棋を解くアルゴリズムについて, 電子情報通信学会コンピューテーション研究会COMP 91-58 (1991).
- 7) 松原 仁: 「CSA 第2回コンピュータ将棋選手権」観戦記, bit, Vol. 24, No. 4, pp. 400-409, 共立出版 (1992).
- 8) 松原 仁: 「CSA 第3回コンピュータ将棋選手権」観戦記, bit, 共立出版 (to appear).
- 9) 松原 仁: いざ排戦! 名人に勝つ将棋システム, 科学朝日, 1993年1月号, pp. 114-118 (1992).
- 10) 吉川竹四郎: 1992年国際コンピュータ囲碁大会報告記, bit, 共立出版 (to appear).
- 11) Bratman, M. E., Israel, D. J. and Pollack, M. E.: Plans and Resource Bounded Practical Reasoning, *Computational Intelligence*, Vol. 4, No. 4, pp. 349-355 (1988).
- 12) Epstein, R.: Can Machines Think? *AI Magazine*, Vol. 13, No. 2, pp. 80-95, Summer (1992).
- 13) Harnad, S.: The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion, *SIGART BULLETIN*, Vol. 3, No. 4, pp. 9-10 (Oct. 1992).
- 14) Johnson, W. L.: Needed: A New Test of Intelligence. *SIGART BULLETIN*, Vol. 3, No. 4, pp. 7-9 (Oct. 1992).
- 15) Levy, D. N. L.: *Computer Chess Compendium*, Springer-Verlag, New York (1989).
- 16) Levy, D. N. L. and Newborn, M.: *How Computers Play Chess*, Computer Science Press, New York, 1991. (小谷善行監訳, コンピュータ・チェス, サイエンス社, 刊行予定).
- 17) Pollack, M. E. and Ringuette, M.: Introducing the Tileworld: Experimentally Evaluating Agent Architectures, In *Proceedings of The Eighth National Conference on Artificial Intelligence*, pp. 183-189 (1990).
- 18) Shapiro, S. C.: The Turing Test and The Economist. *SIGART BULLETIN*, Vol. 3, No. 4, pp. 10-11 (Oct. 1992).
- 19) Turing, A. M.: Computing Machinery and Intelligence, *Mind*, Vol. 59, No. 236, pp. 443-460 (1950).
- 20) Walker, M. A.: Redundancy in Collaborative Dialogue, In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, volume 1, pp. 345-351 (1992).
- 21) Weizenbaum, J.: ELIZA—A Computer Program for The Study of Natural Language Communication between Man and Machine, *Communications of the ACM*, Vol. 9, No. 1, pp. 36-45 (1966).



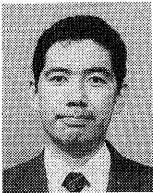
小谷 善行 (正会員)

昭和 24 年生. 昭和 46 年東京大学工学部計数工学科卒業. 昭和 52 年同大学院博士課程工学系研究科修了. 同年東京農工大学工学部数理情報工学科講師. 現在同大学電子情報工学科 (コンピュータサイエンス) 助教授. 記号処理言語を含むソフトウェア工学および知識処理に興味を持つ. 人工知能学会, 日本ソフトウェア科学会, 電子情報通信学会, 認知科学会各会員.



大沢 英一 (正会員)

1959 年生. 1982 年東京工業大学理学部数学科卒業. 1982 年ソニー(株)入社. 情報処理研究所, 総合研究所勤務. 1986~1987 年ハーバード大学大学院言語学科. 1988 年(株)ソニーコンピュータサイエンス研究所入社, 現在に至る. マルチエージェント・システム, 分散プランニング, 談話の計算モデルなどの研究に従事. 日本ソフトウェア科学会, 人工知能学会, AAAI 各会員.



松原 仁 (正会員)

1959 年生. 1981 年東京大学理学部情報科学科卒業. 1986 年同大学院工学系研究科情報工学専門課程修了. 同年電子技術総合研究所に入所. 知能情報部推論研究室ならびに協調アーキテクチャ計画室所属. 1993 年よりスタンフォード大学 CSLI 滞在研究員. 知識表現, 機械学習, コンピュータ将棋などに興味を持つ. AIUEO, 日本認知科学会, 人工知能学会, 日本ソフトウェア科学会, コンピュータ将棋協会各会員.

