

ゲームプログラミングにおける長期的戦略

谷口 和友

東京工業大学 総合理工学研究科 知能システム科学専攻

コンピュータ将棋において、駒の取り合いのような評価値の変化が激しい手順に比べ駒組みのような穏やかで長い一連の手順は探索するのが難しい。本研究では長期的な戦略の知識を用いてこのような一連の手順を探索する手法を提案し、この手法を用いて実験システムを作成した。本手法では、知識として各駒に目標位置を与え、そこへ最短手数で向かう手を計算しこれを候補手とする。この候補手の中から Mini-Max 原理に基づく探索により指し手を決定する。この手法を用いた実験システムにおいて、手数が長く最後まで先読みするのが難しい棒銀の探索を実現することができた。また攻めの戦略である棒銀同様、守りの戦略である穴熊も実現することができた。

Long-term Strategies in Game Programming

Kazutomo Taniguchi

Tokyo Institute of Technology

We propose a hybrid search algorithm in computer shogi, which integrates knowledge-based strategies and search technique. Since the algorithm is guided by rules representing expert knowledge and most computing powers are allocated to a small number of promising moves, in-depth analysis of those moves is possible. A candidate move is generated in four steps: "state judgment," "strategy decision," "goal selection," and "candidate generation." A goal is a place on a shogi board, which is a target of pieces. For each goal position, candidate sequences of moves, that is, the shortest routes to the goal are generated. As a result, long and consistent sequences of moves such as the Bougin and Anaguma strategies can be generated. Our framework is effective in describing both offense and defense strategies.

1 はじめに

コンピュータ将棋では必要となる探索空間がチェスよりはるかに大きく、チェスで成功した全幅探索に近い方法は現実的でない。このため候補手を生成する際の選択的手法が必要となる。

コンピュータ将棋における探索では、静的評価関数による局面の評価を基にした、Mini-Max 原理を用いるのが主流である。数手後の見込みを考慮した局面の価値は評価しにくく、いわゆる「大局観」のようなものを評価するのは難しい。このため一般的に局面における駒の損得や位置、玉の堅さなどを考慮する。また探索空間が限られていることから、深さ限界を定めそこで探索を打ち切り評価を行う。このため深さ限界によって選ばれる手順が変化する可能性がある。

駒の取り合いのような手順では、局面の評価値が一手ごとに激しく変化する。この場合、変化の激しい手順の途中で探索を打ち切ると不適当な手順が選ばれる可能性があるため、取り合いの終わった局面で評価値を計算したい。探索空間が限られているため部分的に深く探索する必要があるが、Singular Extension[1]は他の手順よりも断然高い評価値の手順で探索を延長している。このような駒の取り合いのような手順では、評価値の激しい変化から探索を延長する手法を使うことができる。

しかし、駒組みのような手順では、局面の評価値が一手ごとに緩やかに変化する。駒組みの「一連の手順」の途中で探索を打ち切りたくないが、組み終わるまでの手数がかかるにも関わらず、組み終わらないと評価値の変化が影響しない。この場合、Singular Extensionのように評価値の緩やかな変化から探索

を延長する手法が使えない。

特に将棋では一時的な駒の損得より攻めや守りの戦略が重要であり、戦略に基づく首尾一貫した手順の途中で打ち切らずどのようにして深く探索するかが問題となる。人間は、大まかな目標を立てそれに向かう手順、いわゆる戦略に従う手順に絞り、それが実現できるかどうかを検証することが知られている [2]。この人間的なモデルを工学的なモデルとして応用することで、深く探索できない問題点を解決することができる。と考える。

そのためのアプローチとして、長期的戦略の知識を用いて目標へ向かう手順を生成し、候補手を絞ることで深く探索する手法を提案する。戦略の知識によって選択される目標を盤上の位置で記述し、目標位置へ最短手数で向かう手を候補手として生成する。その候補手の中から Mini-Max 原理に基づく探索を用いて指し手を決定する。

2 先行研究

コンピュータチェスにおける知識ベースシステム PARADISE [3] は、人間のマスターが読むような重要な候補手だけを生成し、探索する局面数を少なくすることを目指した。これは知識ベースを用いてプランを生成し、プランから指し手を決定する。

PARADISE はまず、局面を静的に解析し約 200 個のプロダクションルールを用いて探索を導くためのプランを組み立てる。プランは、指し手のリストで表されている。プランの組み立てでは、始めに初期プランを生成し、その後サブプランを加えていく。次に、組み立てたプランを選択しそれに従って指し手を決定する。プランの選択は最良優先アルゴリズムで行われる。

この過程は人間に近い読み方であると言える。このシステムの実行結果は非常によく、パフォーマンスは知識の数に影響されるため知識を増やすことでシステムはさらによくなる。しかし知識を記述するために多量の作業を要することが非常に大きな問題点であった。

コンピュータ将棋における SPEAR [4] は、駒の配置や駒の利きなどの位置的特徴の集合で表されるパターンを認識することで、候補手を生成する。生成された候補手の中から、Mini-Max 原理に基づく探索 ($\alpha\beta$ 枝刈りを含む反復深化法) を用いて指し手を決定する。

このシステムは約 6000 個のパターンを持ち、1局

面当り 8.2 個の候補手を生成する。与えた全ての局面に対し 81% の局面で、生成された候補手がプロの棋譜に現れる指し手を含んでいた。位置で表されたパターンは記述しやすい表現であると言える。

本研究の手法は知識を用いて目標位置へ向かう手順を生成し探索をその手順に絞る。目標位置へ最短手数で向かう手を求める単純な方法は多くの (無駄だと思われる) 候補手を生成するが、探索を用いてこのような手を削ることで知識の記述を容易にしている。このシステムは知識ベースを Mini-Max 探索で補うハイブリッドシステムである。ハードウェアの進歩により、知識だけで解決するのは難しい問題に対して探索を加える手段が可能となっている。

3 アルゴリズム

本研究のシステムは、知識を用いて候補手を生成し、探索を用いて指し手を決定する。システムの概観を図 1 に示す。

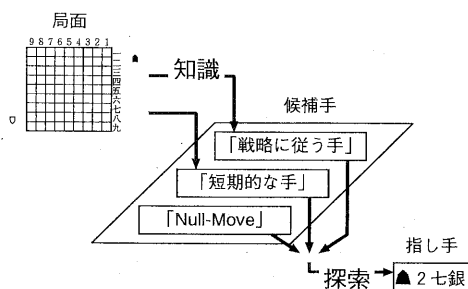


図 1: システムの概観

3.1 探索を用いた指し手決定

現在のコンピュータ将棋プログラムでは、反復深化法を用いた探索が主流である。これは深さ限界を徐々に増やししながら Mini-Max 法による探索を繰り返し行うものである。本研究で作成した実験システムでは、探索に $\alpha\beta$ 枝刈りを含む Mini-Max 法を用いた。

実験システムは、候補手として「戦略に従う手」「短期的な手」「Null-Move」の 3 種類を生成する。現段階では、“歩以外の敵の駒を取る手”、“敵の駒の利きにある味方の駒を移動させる (逃げる) 手”を「短期的な手」として生成している。また候補手以

外の“何らかの指し手”を表すために「Null-Move」を生成している。これは単純に一手パスをし局面の手番を相手に渡すだけの手である。「戦略に従う手」の生成については、以下で説明する。

3.2 知識を用いた候補手生成

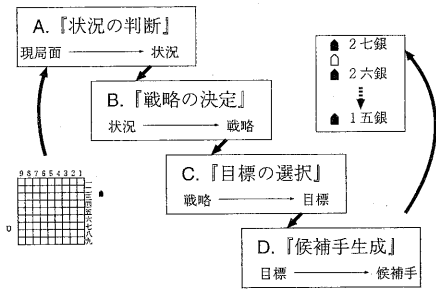


図 2: 「戦略に従う候補手」生成の枠組

「戦略に従う手」を生成するための枠組を図 2 に示す。知識の追加に伴う複雑化に対応するため、知識を問題の段階に分割してある。この構造は『状況の判断』『戦略の決定』『目標の選択』『候補手の生成』の 4 つの部分から成る。以下では、次章の実験で現れる局面図 4 における「戦略に従う手」の生成過程を述べる。

『状況の判断』 この部分は、『状況』の抽象的記述と具体的記述との対で構成される。例えば、

- 「敵の囲いの端が弱い」

という抽象的記述に対し、

```

...
if ((x_num(後, 玉, 1, 3) == 1) かつ
    (x_num(後, 金, 1, 2) == 0) かつ
    (x_num(後, 銀, 1, 2) <= 1))
then True

/* 1筋から2筋に先手の飛が1枚 */
/* 1筋から3筋に先手の銀が1枚 */
if ((x_num(先, 飛, 1, 2) == 1) かつ
    (x_num(先, 銀, 1, 3) == 1))
then True
...

```

のような具体的記述 (パターン) を用意してある。『現局面』のパターン認識を行い、用意してある『状

況』の中から局面に合う具体的記述を見つけ、その対となっている抽象的記述を選択する。

『戦略の決定』 この部分は、『状況』と『戦略』との対で構成される。対は

```

if 「敵の囲いの端が弱い」
then 「棒銀で端を攻める」

```

のような if-then ルールで表してあり、『状況の判断』部分で得られる『状況』(の抽象的記述) に対応した『戦略』を決定する。

『目標の選択』 この部分は、『戦略の決定』部分で得られる『戦略』から『目標』を導く。『目標』は盤上の目標位置で表され、駒ごとに与えられる。if-then ルールにより、抽象性の高い『戦略』から抽象性の低い『戦略』が導かれ、抽象性の低い『戦略』から『目標』が導かれる。例えば、

- 「棒銀で端を攻める」

のような抽象性の高い戦略が導かれ、

- 「味方の駒を端の歩が向かい合っている位置 (1五) に集中させる」

のような抽象性の低い『戦略』が導かれる。そして、それらから

```

飛 2八 ⇒ 1五
銀 3八 ⇒ 1五
歩 1六 ⇒ 1五

```

など、複数の『目標』位置が導かれる。

『候補手の生成』 この部分は、『目標』位置へ最短手数で向かう手を生成する。それぞれの駒が盤上の『目標』位置へ最短手数で移動するための経路を計算し、それらの経路を通る手を『候補手』としてそれぞれ生成する。例えば、

```

銀 3八 ⇒ 1五
の『目標』からは、
銀 3八 ⇒ 2七 ⇒ 2六 ⇒ 1五

```

のような経路が計算され、▲2七銀 が指し手の『候補手』となる。この場合、経路は 1 通りしかないが、仮に『目標』が

```

銀 3八 ⇒ 2四

```

ならば、計算される経路は

銀 3八 → 2七 → 2六 → 1五 → 2四
 銀 3八 → 2七 → 2六 → 3五 → 2四
 銀 3八 → 2七 → 3六 → 3五 → 2四

のようになる。

3.3 本手法の適用例

以下では、次章の実験で現れる局面を用いてアルゴリズムの適用例を述べる。

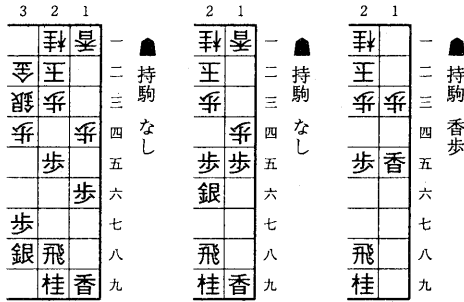


図 3: 棒銀に現れる局面

[左] 棒銀の前半 [中] 棒銀の中盤 [右] 棒銀の後半

棒銀の前半 図3[左]の局面では、前節で述べた手順により「戦略に従う手」が候補手として生成される。この局面では▲2七銀、▲1五歩などが生成されるが、これとは別に生成される「短期的な手」「Null-Move」が候補手として追加される。ただしこの局面では駒がぶつかっておらず、「短期的な手」は生成されない。

この局面で生成される『候補手』によって、探索木に子局面が作られる。子局面でも同様に『候補手』によって、さらにその子局面が作られる。囲いの端が弱い『状況』のパターンが局面に合っている間、子局面は作り続けられるため、

銀 3八 → 2七 → 2六 → 1五

のような経路を通る手順が探索される。

棒銀の中盤 図3[中]は駒交換が起こりそうな局面である。この局面で、後手は「短期的な手」と「Null-Move」を候補手として生成する。この局面では後手の駒に目標を与えていないため後手の「戦略に従う

手」は生成されないが、「敵の駒を取る手」□1五歩が「短期的な手」として生成される。

棒銀の後半 図3[右]は図3[中]以降、1五での駒交換が起こった後の局面である。この局面では図3[左]に用いた後手の囲いの端が弱い『状況』を判断するための自陣のパターンと異なっている。よって「棒銀」で端を攻める『戦略』は選ばれない。

この局面では別のパターンより、端(1三)に駒を集中させる『戦略』が選ばれる。各駒の『目標』位置として

飛 2八 → 1三
 角 6八 → 1三
 香 1五 → 1三
 香 持駒 → 1三

が与えられる。

この場合『目標』位置へ向かう▲1八飛、▲1三角成、▲1三角不成、▲1三香成、▲1三香不成、...などの手を「戦略に従う手」として生成し、プログラムは候補手を探索木に加えていく。

4 実験

本章では、様々な状況における長期戦略の特徴を調べるために、全幅探索との比較実験を行った。実験で Mini-Max 探索に用いた静的評価関数は、駒の損得と駒の位置関係を重視する単純なものである。Mini-Max 探索には $\alpha\beta$ 枝刈りを加えた。

4.1 攻めの戦略



図 4: 先手が棒銀を開始する局面

長期的な攻めの戦略を用いた例として、棒銀の開始局面図4について先読みを行った。

この局面は後手の「矢倉囲い」の端(1筋)が弱く、先手が「棒銀戦法」で右銀を犠牲に端を攻め破れば先手が有利になる局面である。人間ならば将棋を習い立ての初心者が直線的に読みを進めることのできる局面である。しかし既存の将棋プログラムでは、探索の深さが足りないため端を破る局面まで探索するのは難しい[5]。この局面の探索において、前半の取り合い部分で、銀を捨てて大きな駒損となる先手は不利な局面だと判断する。しかし後半の取り合い部分で、先手が端を破る局面まで探索すれば評価値が高くなり有利な局面であることが分かる。

図4の局面で前章で説明した知識を与え、候補手を生成した。この局面において実験プログラムは、最大15手の探索で次の手順、**▲2七銀** **△Null** **▲2六銀** **△Null** **▲1五歩** **△同歩** **▲同銀** **△同香** **▲同香** **△1三歩** **▲1八飛** **△Null** **▲1三香成** **△同桂** **▲同角成** を選択した。

コンピュータ将棋の試合には持ち時間があらかじめ決められているこのため、数分以内に探索を行わなければならない。CPU PentiumIII/500MHzのコンピュータ上で、実験プログラムは上記の15手を121秒で選択している。これは現実的な時間であると言える。

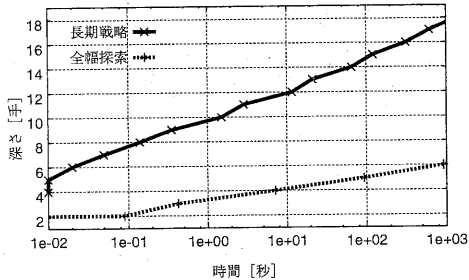


図5: 図4での探索手数の比較

探索深さの比較 図5は探索時間に対する探索可能な手数の変化を示している。全幅探索の場合、現実的な時間では5手程度の探索しかできない。探索における深さ限界を1手増やすだけで探索の時間は爆発的に大きくなってしまふ。一般的な静的評価関数では、棒銀の仕掛け前の局面から先読みをしないでその後の評価値を計算することが難しい。このため長い手順を探索する必要があるが、全幅探索では現

實的に不可能である。しかし棒銀のように目標位置を決めやすいものは、戦略に従う手を探索するこの手法が効果的である。

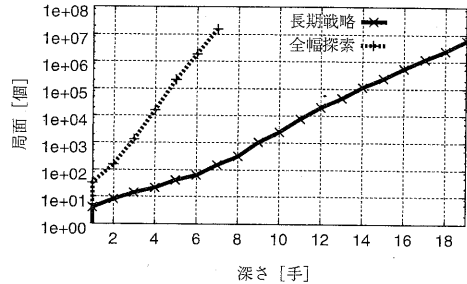


図6: 図4での深さ限界に対する探索局面数の変化

探索局面数の比較 局面図4における深さ限界に対する探索局面数の変化を図6に示す。全幅探索では探索局面数が増えすぎて深く探索できない。全幅探索による生成候補手の数が1局面当たり30手程度、長期戦略による生成候補手の数が1局面当たり4手程度であり、これが深い探索を可能にしている。

4.2 互いに攻め合う戦略



図7: 先後同型で互いに棒銀へ向かう局面

局面図4における先読みでは、敵の戦略を考えず、味方の戦略だけを考慮した。局面図7では後手にも攻めの戦略を与え先読みを行った。先後同型のためお互いに、同じ戦略に従う候補手を生成し探索する。図8に探索時間に対する探索可能な手数の変化を示した。

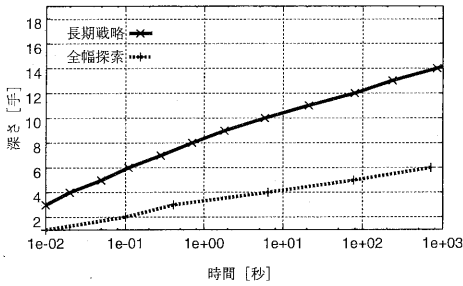


図 8: 図 7 での探索手数の比較

この局面では敵の「戦略に従う候補手」も探索に含まれるため、いわゆる勝手読み比べ候補手が増大する。最大 13 手の探索に 233 秒かかっている。選択された手順は、▲2七銀 △Null ▲2六銀 △Null ▲1五歩 △同歩 ▲同銀 △同香 ▲同香 △Null ▲1八飛 △9五歩 ▲1三香成 となった。この手順からは攻め合いのように見えないが、後手の棒銀も探索木に含まれている。この局面で、後手の棒銀は先手の棒銀より一手遅く最終的に評価が高くならなかったため、後手は棒銀を選択していない。

4.3 守りの戦略



図 9: 穴熊に入る局面

長期的な守りの戦略を用いた例として、穴熊の開始局面図 9 について探索を行った。この局面は、先手が「穴熊囲い」を目標に駒組みし、後手がそれを防ぐために攻め駒を前進させる局面である。局面図 9 において実験プログラムは、最大 9 手の探索を 18 秒で行い次の手順、▲8八玉 △7三桂 ▲9九玉 △6四歩 ▲6六銀 △6五歩 ▲7七銀 △8五桂

▲8八銀 を選択した。この局面でプログラムは次の目標を各駒に与える。先手の駒の目標はそれぞれ以下の通り。

- 玉 7八 ⇒ 9九
- 金 6九 ⇒ 7九
- 金 4九 ⇒ 7八
- 銀 5七 ⇒ 7七
- 銀 7九 ⇒ 8八

後手の駒の目標はそれぞれ以下の通り。

- 角 3三 ⇒ 9九
- 銀 4三 ⇒ 7六
- 桂 8一 ⇒ 7七
- 歩 6三 ⇒ 6七
- 歩 4四 ⇒ 4五

この目標は、先手は穴熊囲いの形を目指しており、後手は角筋を通し角筋に入る敵の駒を攻める形を目指している。



図 10: 穴熊に囲う局面

類似局面での適用 図 9 の類似局面図 10 について探索を行った。この局面では最大 9 手の探索を 49 秒で行い次の手順、▲6六銀 △5四銀 ▲8八銀 △4五歩 ▲7七銀引 △6五歩 ▲8六銀 △Null ▲7九金 を選択した。この局面でプログラムが各駒に与える目標は、上の目標とはほぼ同じである。戦略を用いるこの手法が同じような目的を持つ類似局面にも適用できていると言える。

全幅探索では駒の取り合いの無いこのような手順の場合脈絡のない手順が選ばれることが多いが、戦略を用いた手法では首尾一貫した手順が選ばれている。

「目標」を決めてそれに向かう手を生成する本研

究の手法は、敵が逃げた場合に「目標」を変えなければならぬが、駒組みなど評価値の変化が穏やかな局面では「目標」は大きく動かないことが多い。このためこの手法は守りの戦略にも効果的であると言える。

4.4 長期的戦略による選択的深化

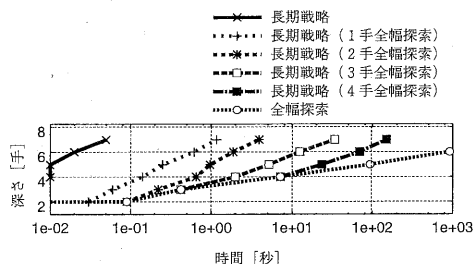


図 11: 図 4 において長期戦略と全幅探索を併用

局面図 4 において長期戦略と全幅探索を併用し探索を行った。Singular Extension は全幅探索と併用する選択的深化手法であるが、同様に長期戦略による候補手生成を選択的深化手法として用いることが有効ではないかと考え、以下の実験を行った。

図 11 は長期的戦略を用いた探索において 1 手から 4 手の全幅探索を行い、それぞれの探索時間に対する探索可能な手数の変化を調べたものである。図の「長期戦略 (1 手全幅探索)」は、探索の 1 手目の局面だけ全ての可能な候補手を生成し、それ以降の局面では長期戦略を用いて候補手を生成する。「長期戦略 (2 手全幅探索)」は、探索の 2 手目までの局面で全ての可能な候補手を生成し、それ以降の局面では長期戦略を用いて候補手を生成する。「...3 手...」、「...4 手...」も同様である。

長期戦略と全幅探索の併用は、一定時間に探索可能な手数を一気に減らしてしまい、また全幅探索の影響で首尾一貫していない手順が多く生成されるため、有効性を見つけ出すことができなかった。

4.5 主観評価

戦略を用いた探索の読み筋について、人間による主観評価をおこなった。実験は図 4 で実験システムが行った探索の探索木を被験者に読んでもらい既存のコンピュータ将棋の指し手に比べどのような読み

を行っているか解答を得たものである被験者はアマチュア 4 段の実力を持ち、さらにコンピュータ将棋に関する知識を持つ。

被験者は▲3六歩からの 3 筋の攻めも探索すべきだと述べている。長期戦略を用いた探索には 3 筋の攻めが含まれていなかった。この局面で 1 筋からの攻めの戦略を評価値が高くなるまで先読みした上で選んでいる点はいいが、3 筋からの攻めの戦略も (この場合の成功は難しいが) 読む方が望ましいとのことだった。

5 議論

5.1 先行研究との比較

PARADISE との比較から、本研究のシステムには以下の長所・短所が挙げられる。

目標位置で記述する長所 PARADISE の知識 (プロダクションルール) が、「if [条件]: 局面のパターン / then [帰結]: 適した指し手のリスト」と記述されるのに対し、本研究では、「if [条件]: 局面のパターン / then [帰結]: 適した目標位置」と記述される。(PARADISE では指し手のリストに“任意の指し手”を含むこともできるが) 本システムでは多くの指し手を 1 つの目標位置で記述できるため、帰結部の記述が容易でルール数が少なくなる。

目標位置で記述する短所 本研究の手法では、知識の帰結部を目標位置で表すことで複数の候補手を 1 つの知識で表し知識の数を少なくしているが、攻撃に対し敵が逃げたり避けたりして目標位置の変更をする場合、複数の目標位置それぞれに対応する知識を必要とする。複数のスタートからの手順を 1 つの知識で表すことはできるが、複数のゴールまでの手順を 1 つの知識で表すことができない。

図 3[右] の局面で、後手が△1三歩の代りに△1二歩を指すかも知れない。よって、先手は 1三を目標位置とする知識と、1二を目標位置とする知識を別々に用意する必要がある。このような場合に、知識の帰結部を目標位置で表す本研究の手法は、知識が複雑化する問題点を解決していない。

知識と探索を併用する長所 PARADISE が知識ベースであるのに対し、本システムでは Mini-Max 原理に基づく探索を併用している。PARADISE では不適切なプランを生成しなくするためにルールを

厳密に記述する必要があるが、本システムでは条件部・帰結部を厳密に記述せずに、不適切な目標位置を生成し不適切な候補手を生成しても、Mini-Max探索で削り指し手としては選択しない。

手順を組み立てられない短所 PARADISEでは手順を組み立てることができる。ある小さな戦略に別の戦略を組み合わせることで、大きな戦略を作ることができるのに対し、本システムでは戦略を組み合わせることができない。



図 12: 手順の組み立てを必要とする局面

図 12 は手順の組み立てを必要とする簡単な例題局面である。味方の持駒の金を 4 二に打つと敵の玉を詰ますことができるが、4 二の位置を敵の 3 二の金が守っている。仮に 4 二金が有効であれば 2 三飛成は考えない。このような場合 PARADISE では、プラン「4 二金」からプラン「2 三飛成 同金 4 二金」を組み立てて、2 三飛成を指すことができる。

しかし本システムで、2 三飛成を生成するためには、2 三の歩が守られていても取りに行くような手を生成するための知識を別途用意し、いつもそのような手を探索しなければならない。この方法では知識が煩雑化し探索も肥大化してしまう。探索の途中で目標を再設定し他の駒の連動させることは今後の課題となる。

5.2 当面の課題

今回の実験システムでは「短期的な手」として“敵の駒を取る手”、“敵の駒から逃げる手”のみを生成しているが、見落としが多いため改良する必要がある。また標準的な実装であるハッシュテーブルを用いていない。手順前後や Null-Move による同一局面が探索の中に非常に多く出てくが、ハッシュテーブルを使うことで改善できると考えられる。

6 おわりに

本研究ではコンピュータ将棋において「長期的戦略」を用いる手法を提案した。またこの手法を用いたシステムを作成し実験を行った。このシステムは、既存のプログラムが一連の手順を深く探索できない問題点を、戦略に基づく首尾一貫した手順に探索を絞ることで解決した。

生成候補手は目標へ向かう手順なので連続的な手となる。この手法は知識ベースのみのシステムに比べてたくさん無駄な手を生成するが、探索を用いることで解決し、記述を容易にしている。位置情報のような、単純な知識で目標を記述し、これを探索の絞り込みに用いることができた。

知識ベースには細かな特化した知識が必要となるが、位置情報で目標位置を記述することでこの問題点を緩和できた。またこれらの候補手に対してのみ探索を行うことで、既存のプログラムでは探索できない棒銀の問題を解くことができた。棒銀のような長く、かつ首尾一貫した手順を生成することは難しいが、これを解いている。また敵味方互いに戦略を用いて探索することもできた。さらに攻めの戦略である棒銀にだけでなく守りの戦略である穴熊にも適用することができた。

参考文献

- [1] Thomas Anantharaman, Murray S. Campbell, and Feng-hsiung Hsu. Singular Extensions: Adding Selectivity to Brute-Force Searching. *Artificial Intelligence*, Vol. 43, pp. 99-109, 1990.
- [2] 吉川厚, 小島琢矢. 囲碁対局プロトコルの諸相情報処理学会研究報告 99-GI-2, 2000. (掲載予定).
- [3] David Wilkins. Using Patterns and Plans in Chess. *Artificial Intelligence*, Vol. 14, pp. 165-203, 1980.
- [4] Reijer Grimbergen and Hitoshi Matsubara. Pattern Recognition for Candidate Generation in the Game of Shogi. In *Proceedings of the Workshop on Computer Games (W31) at IJCAI-97*, pp. 7-12, 1997.
- [5] 松原仁. CSA 第 9 回コンピュータ将棋選手権観戦記. *bit*, Vol. 31, No. 7, pp. 46-49, 1999.