

棋譜解説を目的とした囲碁棋譜からの着手情報の抽出

中 村 貞 吾†

囲碁の棋譜解説を目的として、棋譜から解説の対象となる着手情報の抽出を試みる。まず、棋譜中の各々の着手に対して、着点の位置、直前の着手からの相対位置、着点の周辺配石の特徴を付与する。そして、着手系列の n -gram を用いて石の流れを重視した次着手予測を行ない、解説対象となる候補着手を抽出する。

Extraction of Move Information for Commentary from Go Game Records

TEIGO NAKAMURA†

We show a method for extracting moves for commentary from Go game records. We use n -gram model to take into account move sequences producing the current position.

1. はじめに

対戦型思考ゲームをプレイする際には、プレイヤーは相手の着手からその意図や戦略を見出しそれに応じて着手を行なっている。囲碁が別名「手談」とも呼ばれているように、対局は着手を通じたコミュニケーションであると考えられる。我々は、思考ゲームの中でも特に囲碁を対象として、ゲームと言語との関わりを明らかにすることを最終の目標とした研究を行なっている。

このような背景のもとに、我々は従来より、囲碁の棋譜を個々の着手が符号化されたテキスト（棋譜テキスト）であるとみなして、棋譜テキストに自然言語処理の分野で利用されている技術を適用して知識獲得を行なってきた^{3)~6)}。囲碁は、チェスや将棋とちがって個々の石に先験的な役割が定まっていないため、プレイヤーは石の配置や着手の履歴の情報から石群の役割を識別する必要がある。本論文では、囲碁の棋譜解説を目的として、棋譜から解説の対象となる着手情報の抽出を試みる。

関連研究として、コンピュータ将棋の読み筋や評価値に基づいた将棋棋譜の自動解説の研究⁷⁾がある。プロ棋士に匹敵する実力を有するコンピュータ将棋では、その形勢判断や長手順の読み筋が棋譜解説に十分に利用できることが示されている。

一方、コンピュータ囲碁の分野では、近年、モンテカルロ木探索を用いた対局囲碁プログラム（モンテカルロ碁）の実力が向上している。そこで、棋譜中の個々の着手の良否や形勢をこれらの対局囲碁プログラムを用いて評価し、その評価値や読み筋を棋譜解説に利用することも考えられる。しかし、一般にモンテカルロ碁では、全局的な勝率に基づく局面評価が行なわれているため、形勢が良い場合にはヌルイ手が優先されたり、形勢が悪くなると自滅の道に進むなど着手が安定しない。人間同士の対局棋譜の解説を考えたときには、モンテカルロ碁の評価値は、全局的な形

勢判断の目的としては有用であるが、その局面に至る着手の履歴を考慮した上での解説対象となる着手の抽出にはこれだけでは不十分である。そこで、過去の棋譜から獲得した配石状況や手順情報を利用して各々の着手の打たれやすさを評価し、その結果を利用することにする。まず、棋譜中の各々の着手に対して、着点の位置、直前の着手からの相対位置、着点の周辺配石の情報を付与する。そして、着手系列の n -gram 統計を利用して次着手予測を行ない、着手の打たれやすさに基づいて候補着手を抽出する。これにより、石の流れなど、全局的な形勢とは異なる観点からの着手の評価を行なう。

2. 着手情報

着手された石自体の役割を次の3つの側面から眺めて特徴づける。

- (1) 着点の盤面上の絶対位置
囲碁では、1線、2線、3線、... など、盤端からの距離によって着手の位置が特徴づけられることが多い。そこで、座標 (x, y) の着点に対して、 $\min\{10 - |x - 10|, 10 - |y - 10|\}$ の値を付与する。
- (2) 直前の相手方の着手と現在の着点との位置関係（相対位置）
直前の相手方の着手の座標を (px, py) とするとき、 $(\min\{|x - px|, |y - py|\}, \max\{|x - px|, |y - py|\})$ のペアの値を付与する。
- (3) 着点の周囲の状況
Stern らのテンプレート²⁾ をもとにして、図1に示す9種類の形状のテンプレートを用意した。そして、各テンプレートの範囲内の配石状況を Zobrist ハッシュを用いて符号化する。

3. n -gram モデル

棋譜中の各着手に対して前節で述べた符号化を行ない、得られた着手系列を $s_{1,n} = c_1 c_2 \dots c_n$ とする。この記号列に対して、 n -gram 確率は記号列の相対頻度を用いて次の式で与えることができる。

† 九州工業大学大学院 情報工学研究院 知能情報工学研究系
Department of Artificial Intelligence, Kyushu Institute of Technology
E-mail: teigo@ai.kyutech.ac.jp
URL: <http://www.dumbo.ai.kyutech.ac.jp/~teigo/>

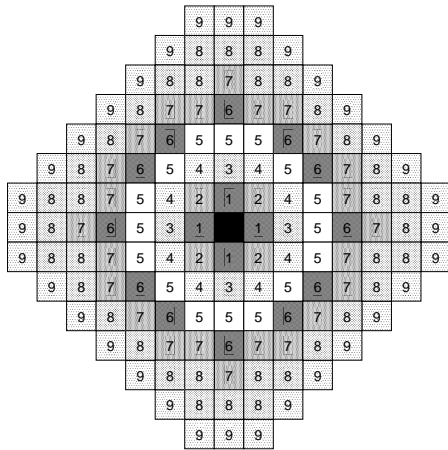


図 1 使用した 9 種類のテンプレート

$$P(c_n | s_{1,n-1}) = \frac{f(s_{1,n})}{f(s_{1,n-1})} \quad ; \quad f(x) \text{ は } x \text{ の出現頻度}$$

一般に n が大きくなると、データスパースネスの問題が起きるため、確率値を補正するスムージングを行なう必要がある。本研究では、次式の線形補完を用いた。

$$P(c_n | s_{1,n-1}) = \sum_{i=1}^{n-1} \lambda_i P(c_n | s_{n-i,n-1})$$

4. 着手予測

着手予測は、テンプレートにより獲得したパターンを用いたスコアリングに n -gram 確率を併用する形で行なう。各パターンのスコアは、荒木ら¹⁾ が用いているパターンが実際に使用された相対頻度によるスコアと直前の着点からの距離に対して単調減少する重みを乗じて算出しておく。着点 m を中心とするパターン p のスコアを $SC(p, m)$ と書く。予測対象の局面 (N 手目) に対して、その局面に至る着手符号列を $s_{1,N-1}$ とするとき、次の評価値 $E(m)$ が最大となる着点 m を予想着手とする。

$$E(m) = \prod_i P_i(m | s_{1,N-1}) \times \sum_p SC(m, p)$$

5. 実験

棋譜データとしては、日本棋院「棋譜データ集 96」CD-ROM を使用した。この中から、ランダムに選んだ 20,000 局の棋譜を学習用棋譜とし、これとは別の 1,000 局の棋譜をテスト用棋譜として用意した。

学習用棋譜に対して 3 節で述べた各々の観点による着手情報を付与して棋譜テキストを構成し、 n -gram を抽出した。抽出された n -gram の統計情報を表 1 に示す。

図 2 にテスト用棋譜の累積正解率のグラフを示す。1 位正解率は 34.5%，5 位以内の正解率が 61.0%，10 位以内の正解率が 71.5% となり、荒木らの結果¹⁾ をわずかながら上回る正解率を達成できた。図 3 は、10 手の範囲で区切った区間毎の累積正解率のグラフで、図 4 は、 n -gram 確率を用いずにテンプレートのスコアのみを用いた場合のグラフである。これを比較すると、 n -gram 確率を併用することにより、序盤の着手予測の精度が向上したことがわかる。

2 節で述べた各々の観点毎に符号列を構成する。観点を i で表わし、その符号列に対する n -gram 確率を P_i と書くことにする。

表 1 n -gram の総数

n -gram 種類	1-gram	2-gram	3-gram	4-gram	5-gram
テンプレート 1	45	1,759	31,185	127,262	168,210
テンプレート 2	1,021	110,640	140,255	107,635	79,288
テンプレート 3	50,111	128,230	84,045	59,031	42,913
テンプレート 4	122,945	83,134	54,434	40,116	31,595
テンプレート 5	91,867	45,656	33,878	28,327	25,115
テンプレート 6	63,486	34,480	28,159	25,410	23,327
テンプレート 7	41,813	26,615	23,557	22,031	20,818
テンプレート 8	31,532	21,870	20,083	19,091	18,286
テンプレート 9	27,436	19,707	18,496	17,776	17,101
絶対位置	10	99	972	8,730	53,210
相対位置	205	21,010	118,517	176,699	159,817

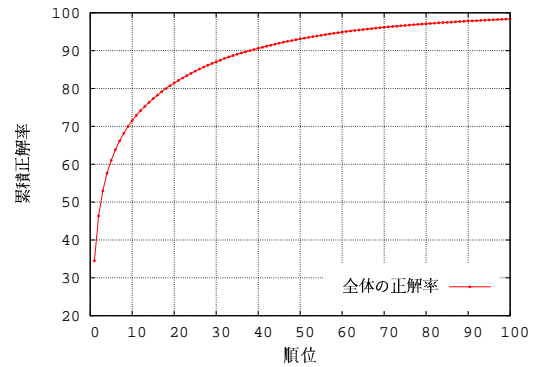


図 2 全体の正解率

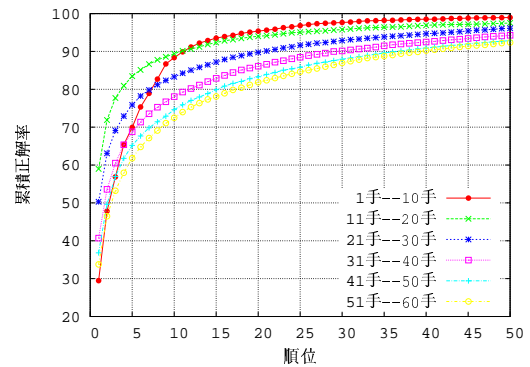


図 3 着手区間毎の正解率

次に、棋譜の解説が行なわれている書籍（「打碁観賞シリーズ」⁸⁾ と「名局細解」^{9),10)}）中の棋譜に対して着手予測を行なった。その中から 2 つの棋譜（図 5、図 6）を選び結果を概観する。なお、図 5 の棋譜 1 は学習用棋譜に含まれている棋譜で、図 6 の棋譜 2 は学習用棋譜には含まれていない棋譜である。

n -gram を使用することにより、序盤の着手予測の精度は向上したが、全体的にみると、正解率はテンプレートのスコアのみを用いた場合に比べて若干低下した。図 7 と図 8 に、着手区間毎のエラー率の比較を示す。棋譜 1 は序盤は定型どおりに進んでいるので、 n -gram の効果が顕著に見られるが、棋譜 2 では序盤でも予測精度の若干の低下がみられた。これは、序盤から典型的な定型から外れた進行になっていることが原因であると考えられる。

棋譜の解説を考えた場合、着手予測の精度だけでなく、実戦では打たれなかった着手についての評価とその分布の状況も重要な情報となる。図 9 は、棋譜 2 の 32 手目の局

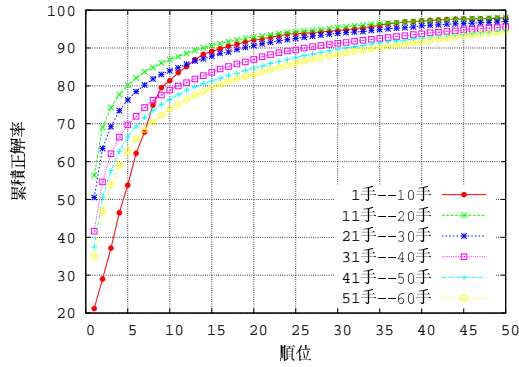


図 4 着手区間毎の正解率 (n-gram 確率なしの場合)

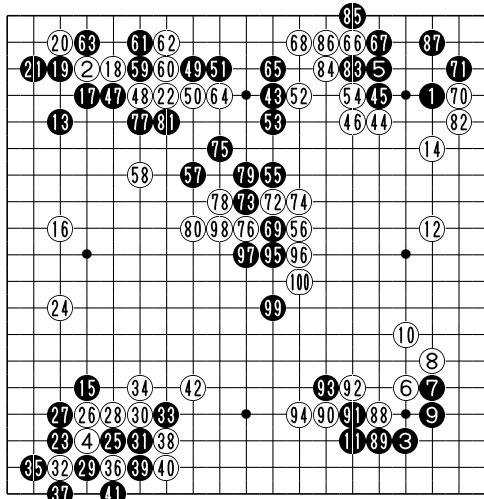


図 5 棋譜 1: 第 4 期旧名人戦七番勝負第 2 局 (黒) 林海峯八段-(白) 坂田栄男名人 (100 手まで)

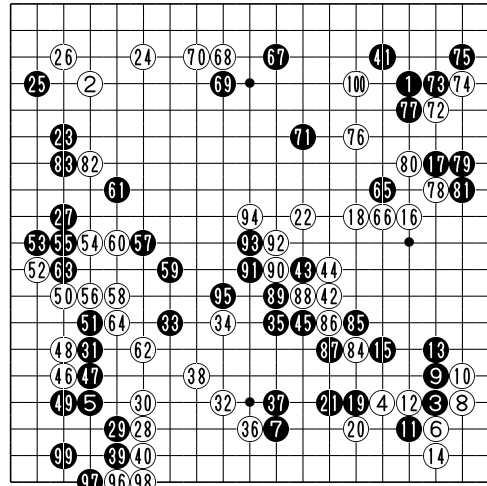


図 6 棋譜 2: 第 14 回世界選手権富士通杯 (黒) 林海峯九段-(白) 劉昌赫九段 (100 手まで)

面である。右図中の青のラインは、最大評価値の 10% の値を表わす。この局面では、1 位に予測された着手の評価値が突出している。この局面に対して、解説本文中では「白は 32 と高くヒラキ、中央進出を視野に入れていいます。」とあり、着手予想の状況と同様な解説が行なわれている。図 10 は、棋譜 2 の 35 手目の局面である。着手予測の順位は前図と同様の 9 位であるが、上位に予測された着手の予測値が突出しておらず、手が広い局面であることがわかる。n-gram 確率の使用が正解手以外の着手の評価値の分布に与えた影響を見るために、最大評価値の 10% 以内の評価値を持つ候補着手数で着手区間毎に比較したものが図 11 と図 12 である。n-gram 確率の使用により、次善手の着手数が減少していることがわかる (棋譜全体では、およそ 3/4 に減少した)。

6. おわりに

パターンと n-gram を用いた着手予測を行ない、その評価値の分布の状況が着手の解説に利用できる可能性を示した。今後は、対局囲碁プログラムによる形勢判断の結果との併用、および着手に対する実際の解説文の内容との関係や変化手順の解説の分析が課題となる。

謝辞

本研究は科研費 (21500144 「言語モデルと数理モデルを用いた思考ゲームの記述に関する研究」) の助成を受けた

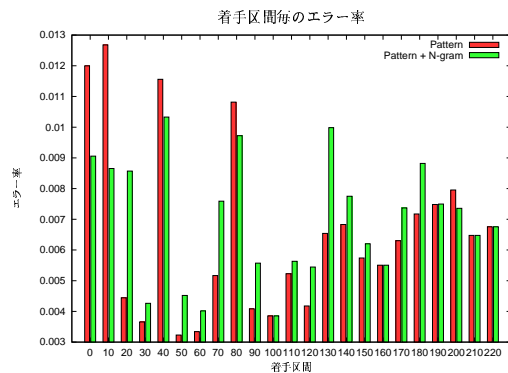


図 7 着手区間毎のエラー率 (棋譜 1)

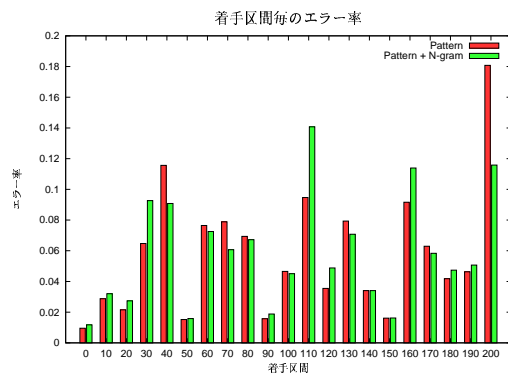


図 8 着手区間毎のエラー率 (棋譜 2)

ものである。

参考文献

- 1) Araki, N., Yoshida, K., Tsuruoka, Y. and Tsujii, J.: Move Prediction in Go with the Maximum Entropy Method, *2007 IEEE Symposium on Computational Intelligence and Games* (2007).
- 2) Stern, D., Herbrich, R. and Graepel, T.: Bayesian Pattern Ranking for Move Prediction in the Game of Go, *International Conference on Machine Learning (ICML-2006)* (2006).
- 3) 梶山貴司, 中村貞吾: 囲碁の着手記号列に対する確率

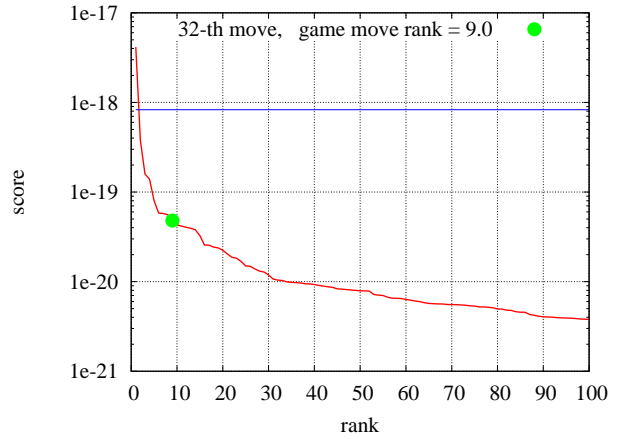
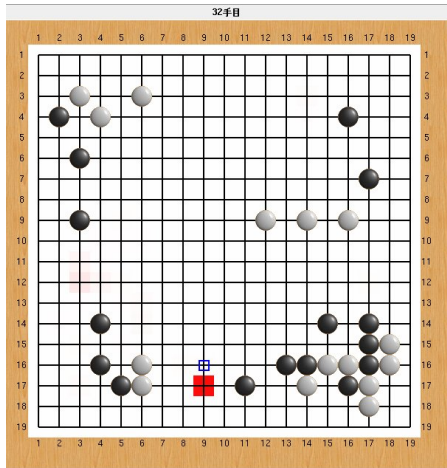


図 9 棋譜 2 の 32 手目の局面 (左) と候補手の評価値の分布 (右)

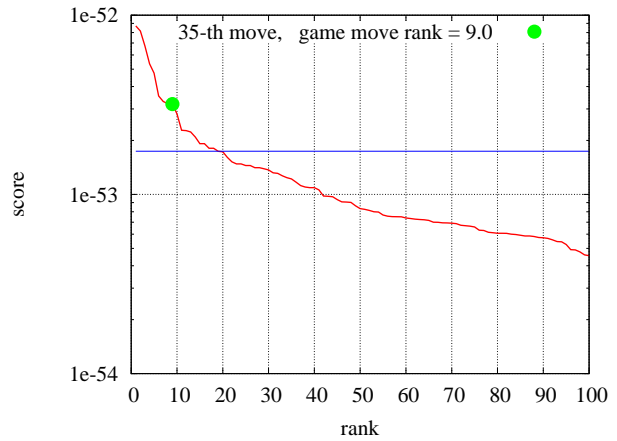
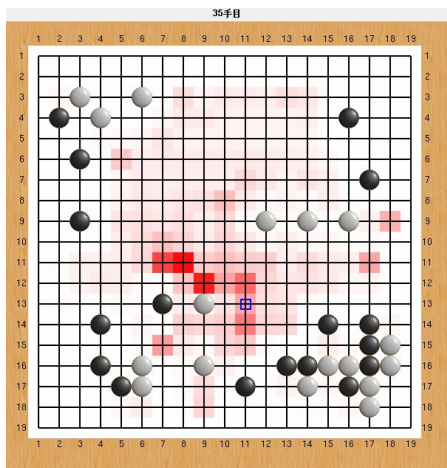


図 10 35 手目の局面 (左) と候補手の評価値の分布 (右)

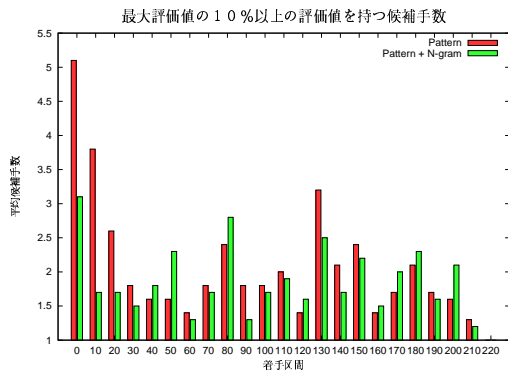


図 11 最大評価値の 10% 以上の評価値を持つ候補手数 (棋譜 1)

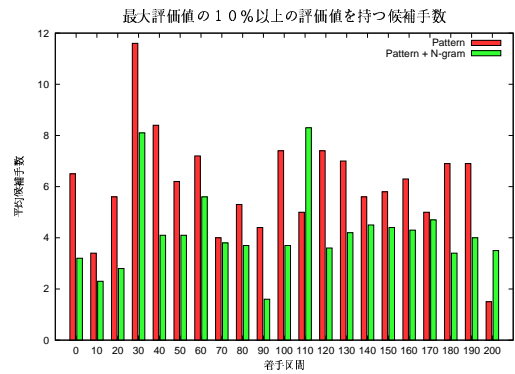


図 12 最大評価値の 10% 以上の評価値を持つ候補手数 (棋譜 2)

- 文法モデルの作成, ゲームプログラミングワークショップ, Vol.5, pp.161-168 (1999).
- 4) 中村貞吾: n-gram 統計を用いた棋譜データベースからの定型手順の獲得, ゲームプログラミングワークショップ, Vol.4, pp.96-105 (1997).
 - 5) 中村貞吾: 着手記号列の出現頻度に基づく囲碁棋譜からの定型手順獲得, 情報処理学会論文誌, Vol.43, No.10, pp.3030-3039 (2002).
 - 6) 中村貞吾, 梶山貴司: 棋譜テキストからの特徴素抽出とそれを用いた棋譜の分類, 情報処理学会ゲーム情報

- 学研究会報告, Vol.2000-GI-2-11, pp.79-86 (2000).
- 7) 金子知通: コンピュータ将棋を用いた棋譜の自動解説と評価, ゲームプログラミングワークショップ, Vol.14, pp.9-16 (2009).
 - 8) 林 海峯: 打碁観賞シリーズ 4 林海峯, 日本棋院 (2004).
 - 9) 林 海峯: 林海峯 名局細解, 誠文堂新光社 (2004).
 - 10) 加藤正夫: 加藤正夫 名局細解 1, 誠文堂新光社 (1987).