



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

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

2010年3月23日(火)



研究の背景と目的

- 囲碁の別名：「手談」
 - プレイヤは相手の着手からその意図や戦略を見出し，それに応えて着手する
 - 対局は着手を通じたコミュニケーション
- 棋譜：個々の着手が符号化されたテキスト (棋譜テキスト)
 - テキスト処理技術の適用



研究の背景と目的

- 囲碁の別名：「手談」
 - プレイヤは相手の着手からその意図や戦略を見出し，それに応えて着手する
 - 対局は着手を通じたコミュニケーション
- 棋譜：個々の着手が符号化されたテキスト (棋譜テキスト)
 - テキスト処理技術の適用

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行なう



- 将棋棋譜の自動解説
 - プロ棋士に匹敵する実力を有するコンピュータ将棋の読み筋や評価値を利用 [金子 2009]
- 囲碁では？
 - 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) の実力は向上



- 将棋棋譜の自動解説
 - プロ棋士に匹敵する実力を有するコンピュータ将棋の読み筋や評価値を利用 [金子 2009]
- 囲碁では？
 - 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) の実力は向上
 - 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
 - 形勢が良い場合にはヌルイ手が優先
 - 形勢が悪くなると自滅の道に進む



- 将棋棋譜の自動解説
 - プロ棋士に匹敵する実力を有するコンピュータ将棋の読み筋や評価値を利用 [金子 2009]
- 囲碁では？
 - 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) の実力は向上
 - 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
 - 形勢が良い場合にはヌルイ手が優先
 - 形勢が悪くなると自滅の道に進む
 - 人間同士の対局棋譜の解説を考えると
 - 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
 - だが，これだけでは不十分



- 解説の対象
 - トッププロの対局 (タイトル戦)
 - ...
 - 初級者同士の対局
- 解説の形態
 - リアルタイム解説
 - 事後の解説, 手直し, 感想戦など
- 与えられた局面で
 - 局面の状況の説明
 - 最善手 (進行) の予想
 - 次善手の説明
 - 悪手の説明



- 石の流れなど，全局的な形勢とは異なる観点からの解説対象着手の抽出
- 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
 - 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
 - 着手系列の n -gram 統計を利用して次着手予測
 - 着手の打たれやすさに基づいて候補着手を抽出

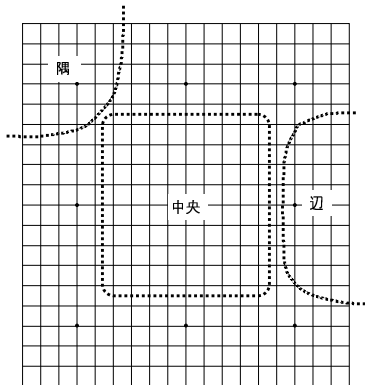


着手情報

① 着点の絶対位置

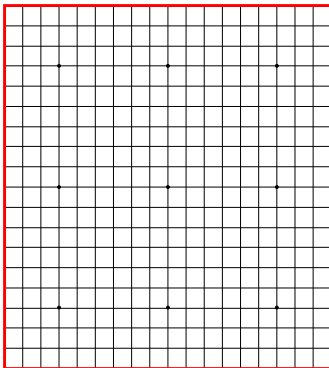
- 盤上の位置に対する囲碁用語

- 星, 三々, 小目, 高目, 目ハズシ, 天元, 隅, 辺, 中央



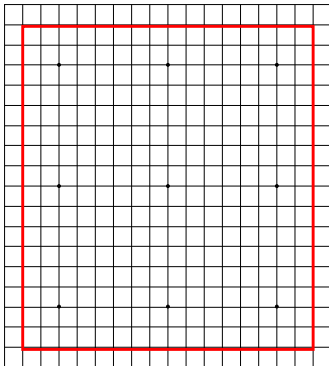


- ① 着点の絶対位置
 - 盤上の位置に対する囲碁用語
 - 一線，二線，三線，四線，...



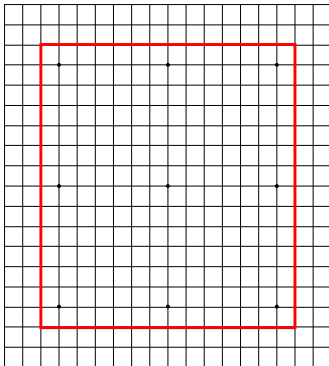


- ① 着点の絶対位置
 - 盤上の位置に対する囲碁用語
 - 一線，二線，三線，四線，...



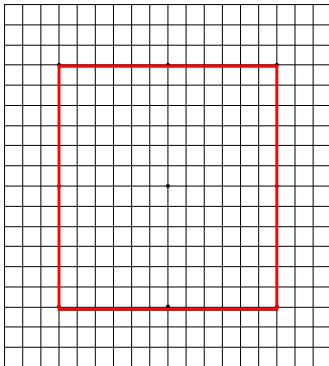


- ① 着点の絶対位置
 - 盤上の位置に対する囲碁用語
 - 一線，二線，三線，四線，...



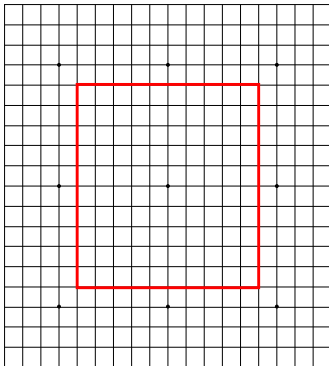


- ① 着点の絶対位置
 - 盤上の位置に対する囲碁用語
 - 一線，二線，三線，四線，...



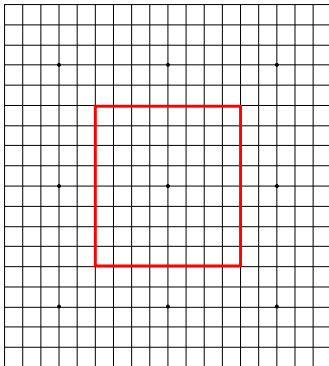


- ① 着点の絶対位置
 - 盤上の位置に対する囲碁用語
 - 一線，二線，三線，四線，...





- ① 着点の絶対位置
 - 盤上の位置に対する囲碁用語
 - 一線，二線，三線，四線，...



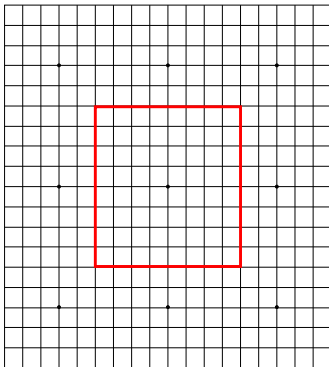


① 着点の絶対位置

- 盤端からの距離に応じた特徴

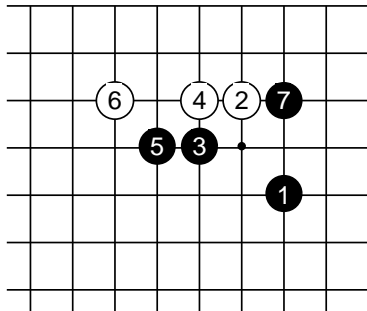
座標 (x, y) の着点に対して,

$\min\{10 - |x - 10|, 10 - |y - 10|\}$ の値の符号





② 直前着手からの相対位置

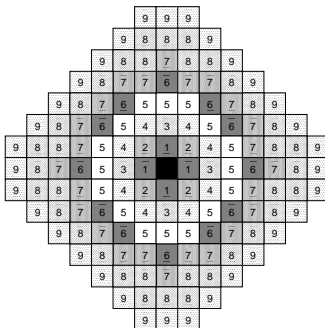


- 現在の着点 (x, y)
- 直前の相手方の着点 (px, py)

$(\min\{|x - px|, |y - py|\}, \max\{|x - px|, |y - py|\})$ の値の符号



- ③ 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]

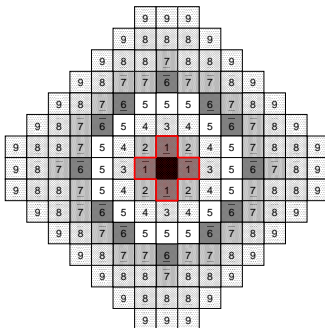




着手情報

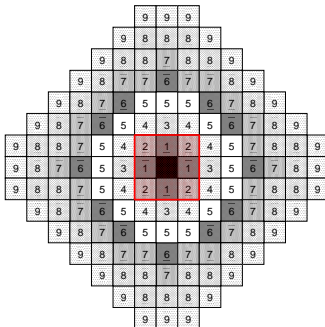
③ 着点の周辺の配石状況

- 着点を中心としたテンプレート [Stern 2006]





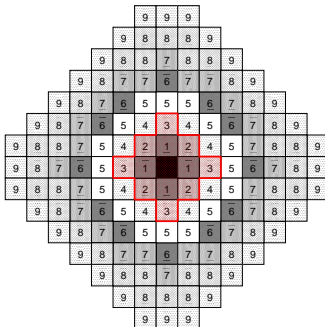
- 3 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]





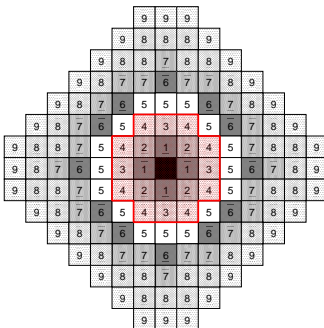
③ 着点の周辺の配石状況

- 着点を中心としたテンプレート [Stern 2006]



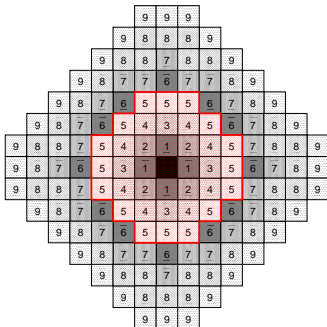


- 3 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]



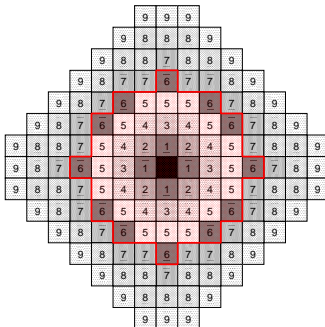


- ③ 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]



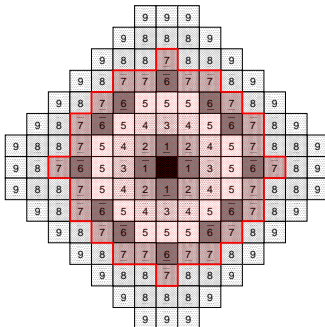


- 3 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]



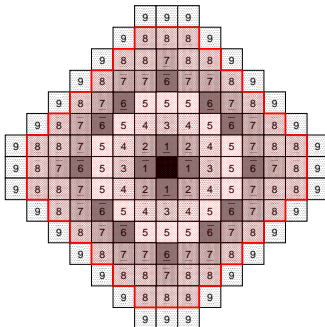


- ③ 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]



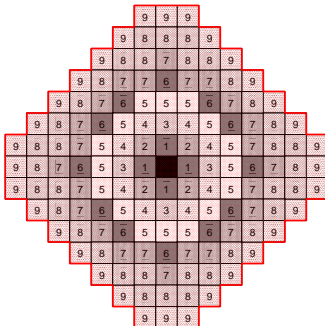


- 3 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]





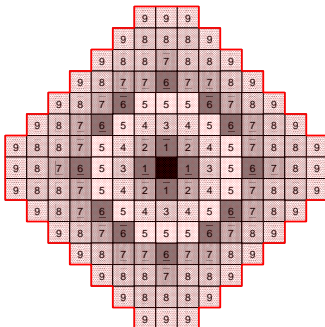
- ③ 着点の周辺の配石状況
 - 着点を中心としたテンプレート [Stern 2006]





③ 着点の周辺の配石状況

- 着点を中心としたテンプレート [Stern 2006]



- 9種類のテンプレートの範囲内の配石状況の Zobrist ハッシュ値の符号



n -gram モデル

- 棋譜中の各着手に対して前述の観点による符号化
- 着手系列 $s_{1,n} = c_1 c_2 \dots c_n$
- この記号列の n -gram 遷移確率

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

- n が大きくなると出現頻度が低下 \implies データスパースネス問題
- 確率値の補正 (スムージング)

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



着手予測

- ① 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
 - 各パターンのスコア
 - パターンが実際に使用された相対頻度によるスコア [荒木 2007]
 - 直前の着点からの距離に応じて単調減少する重み
 - 着点 m を中心とするパターン p のスコア $SC(m, p)$
- ② n -gram 確率
 - 各々の観点毎に符号列を構成
 - 観点 i に関する符号列に対する n -gram 遷移確率 P_i



着手予測

- 予測対象の局面： N 手目
- その局面に至る着手符号列 $s_{1,N-1} = c_1 c_2 \dots c_{N-1}$
- 次の評価値 $E(m)$ が最大となる着点 m を予想着手とする

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

P_i : 観点 i に関する符号列の n -gram 遷移確率
 $SC(m, p)$: パターン p による着点 m のスコア



- 棋譜データ：日本棋院「棋譜データ集 96」CD-ROM
 - 学習用棋譜：ランダムに選んだ 20,000 局

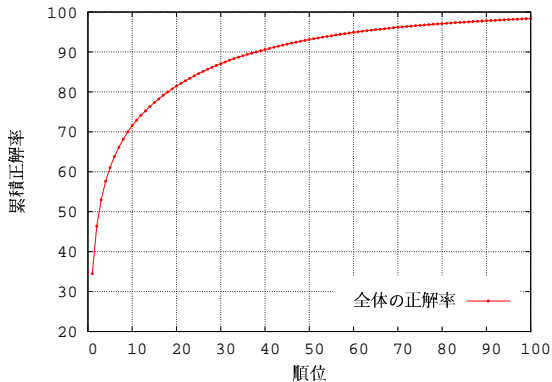
種類 \ n -gram	1-gram	2-gram	3-gram	4-gram	5-gram
テンプレート 1	45	1,759	31,185	127,262	168,210
テンプレート 2	1,021	110,640	140,255	107,635	79,288
テンプレート 3	50,111	128,230	84,045	59,031	42,913
テンプレート 4	122,945	83,134	54,434	40,116	31,595
テンプレート 5	91,867	45,656	33,878	28,327	25,115
テンプレート 6	63,486	34,480	28,159	25,410	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

- テキスト用棋譜：上とは別の 1,000 局の棋譜



着手予測の正解率

- テスト用棋譜の累積正解率
 - 1位正解率：34.5%
 - 5位以内の正解率：61.0%
 - 10位以内の正解率：71.5%

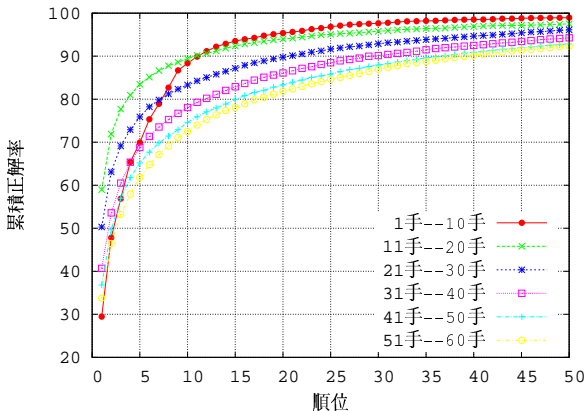




着手予測の正解率

● テスト用棋譜の累積正解率

- 10手の範囲で区切った区間毎の累積正解率

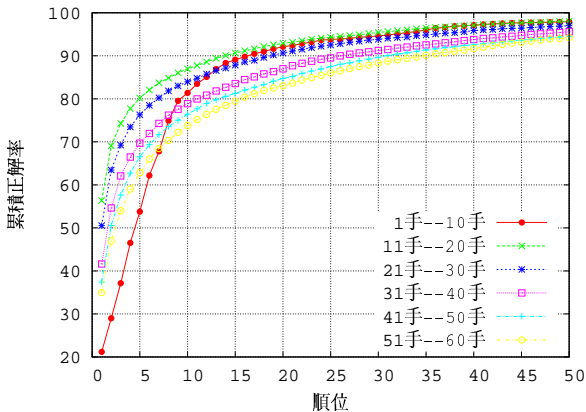




着手予測の正解率

● テスト用棋譜の累積正解率

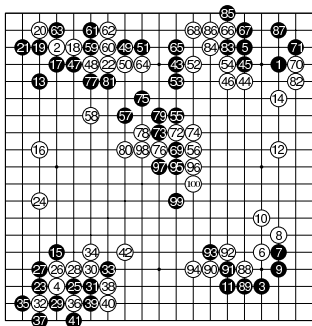
- n -gram 確率を用いずにテンプレートのスコアのみを用いた場合



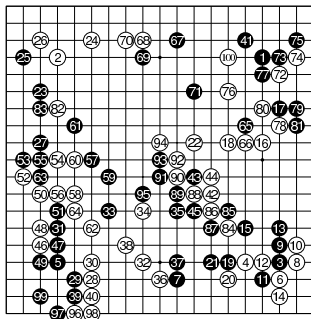


解説本の棋譜

- 打碁観賞シリーズ (林海峯，日本棋院，2004)：6 棋譜
- 名局細解 (林海峯，誠文堂新光社，2004)：4 棋譜
- 名局細解 (加藤正夫，誠文堂新光社，1987)：4 棋譜



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
 学習データに含まれる

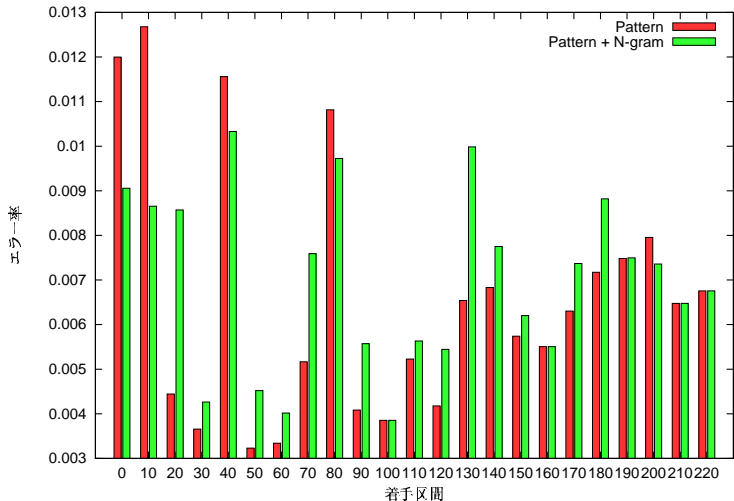


棋譜 2：第 14 回世界選手権富士通杯
 (黒) 林海峯九段-(白) 劉昌赫九段
 学習データ中ではない



棋譜1のエラー率

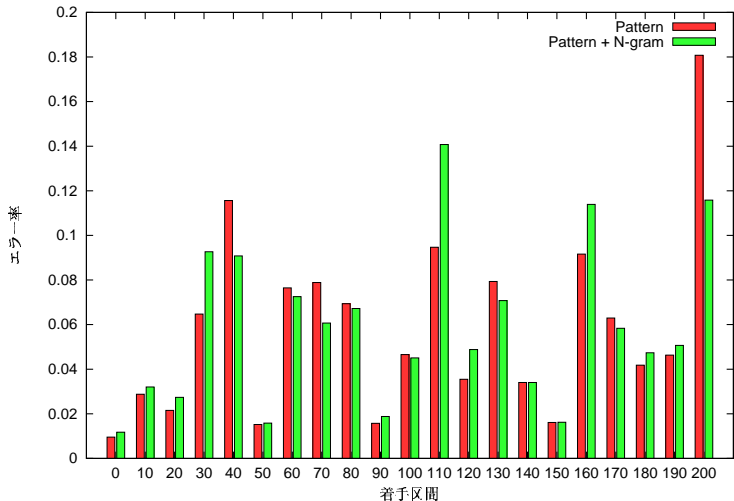
着手区間毎のエラー率





棋譜2のエラー率

着手区間毎のエラー率





棋譜 1 と棋譜 2 の予測結果

- 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
- 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
- 棋譜 2 では序盤でも予測精度が若干低下
 - 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

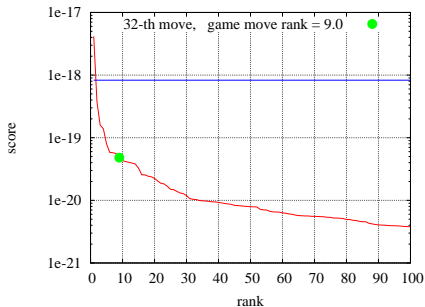
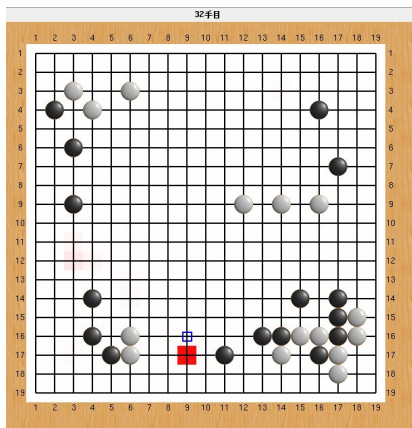


評価値の分布

- 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



棋譜 2 の局面 (32 手目)



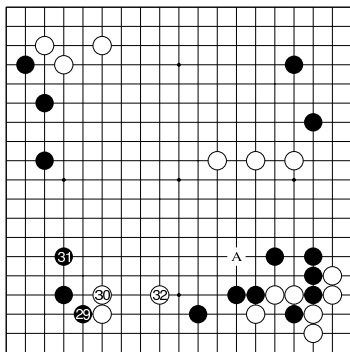


棋譜 2 の解説記事 (第 6 譜 : 本譜)

● 6 譜 (黒 29 ~ 白 32) : 高いヒラキ

黒 29、31 は、白の根拠を奪って攻めようという手。攻めることが右下の黒の補強にも役立ちます。

白は 32 と高く開き、中央進出を視野に入れていきます。石が上へ行けば、白 A がねらいやすくなるのです。





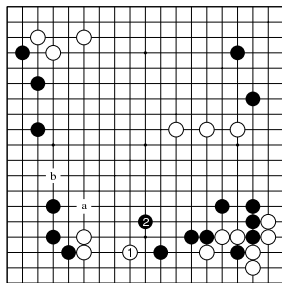
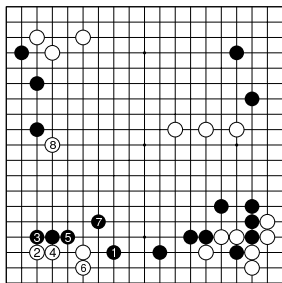
棋譜 2 の解説記事 (第 6 譜 : 変化図)

● カタツキ

譜の黒 2 9 で 1 のハサミでは白 2 と三々に入れ、黒 7 と封鎖して左辺を模様化しても、白 8 のカタツキがぴったりの消しになります。

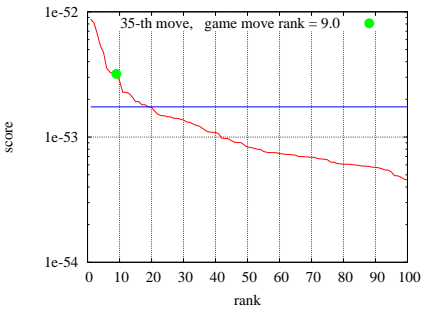
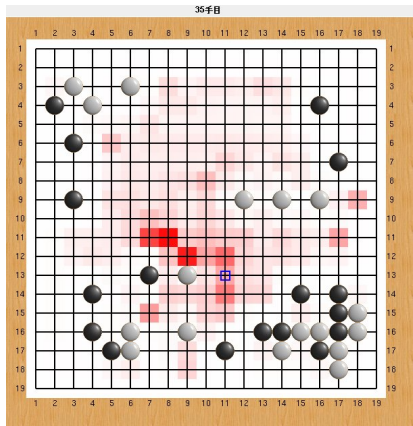
● 攻めの調子

譜の白 3 2 で 1 の低いヒラキなら、黒 2 と攻めて局勢展開の主導権を握ることは容易でしょう。白 a なら黒 b で左辺がそのまま確定します。





棋譜2の局面(35手目)



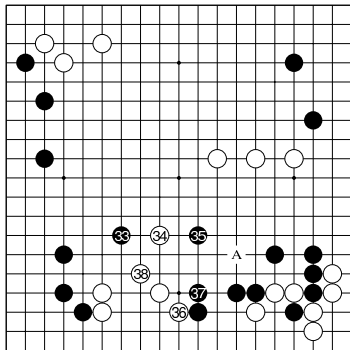


棋譜2の解説記事(第7譜：本譜)

● 7譜(黒33～白38)：ともに好形

上から攻めるにしても、黒33がいいようです。白34で進出形ですが、黒35と耳の急所に迫って薄みを強調し、合わせて白Aの守りを兼ねることが出来ます。

白も36、38と地を持って治まる好形ですが、黒に先手が回りそうです。





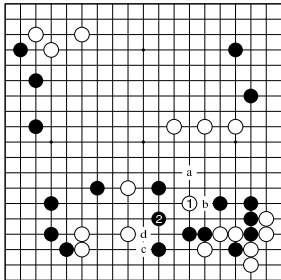
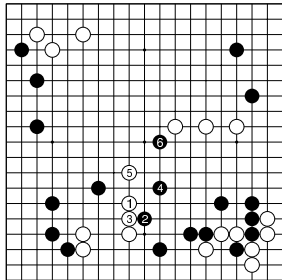
棋譜2の解説記事(第7譜:変化図)

● カラミの可能性

譜の白3 4で1のトビなら、黒2から4、5と攻めを継続しながら進出します。下辺の白はまだ治まっておらず、カラミの可能性も生じます。

● 利きは一方

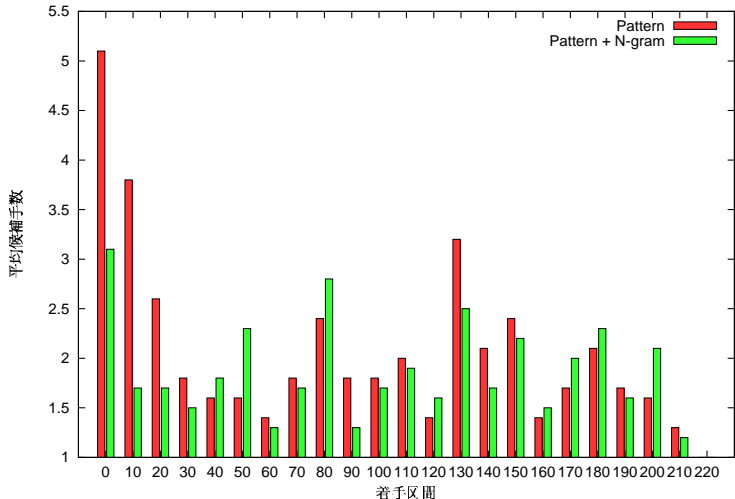
譜の白3 6で1のノゾキなら黒2の受け。この形は白a黒bと利かされても、下辺の白cには黒dと抵抗できます。





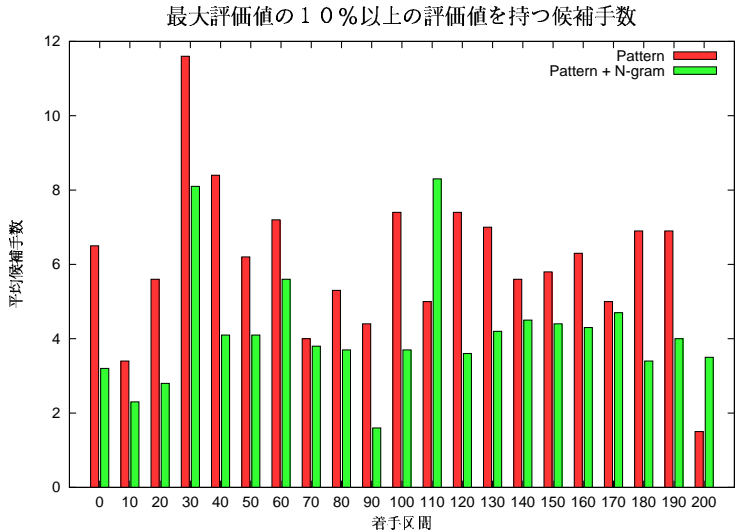
棋譜1の評価値の分布

最大評価値の10%以上の評価値を持つ候補手数





棋譜2の評価値の分布





n-gram の効果

- n -gram 確率を併用することにより
 - 序盤の着手予測の精度は向上
 - ゲーム全体の着手予測精度は若干低下
- 正解手以外の着手の評価値の分布
 - 次善手の候補着手数が増加
(棋譜全体では、およそ $\frac{3}{4}$ に減少)



まとめ

- 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
- 今後の課題
 - 対局囲碁プログラムによる形勢判断の結果との併用
 - 実際の解説文の内容との関係の分析