

囲碁棋譜からの順序パターンとその共起知識の獲得

梶山 貴司 中村 貞吾

九州工業大学大学院 情報工学研究科

{kajiyama, teigo}@dumbo.ai.kyutech.ac.jp

概要

囲碁は状態空間が広くまた静的な局面評価が難しいため、パターン知識の使用が不可欠である。本研究では複数の手順列に対し、その順序を表した情報を順序パターンとし、棋譜データベース中における順序パターンの共起頻度とあわせて知識として獲得する。1局の棋譜テキストにおいて非連続な複数の着手列の中にも相関があり、その知識を得ることは次着手予測や棋譜解析に有用である。本論文の手法では、囲碁棋譜データベースから定石などの定型手順を獲得し、2つの定型手順の共起頻度を計数して、長い順序パターンの共起頻度を代数計算で獲得する。そして総手数500万手の棋譜データベースから順序パターンを獲得する実験の結果を示す。

Acquisition of Sequence Patterns and their Co-occurrence Relations from Game Records of Go

Takashi KAJIYAMA Teigo NAKAMURA

Kyushu Institute of Technology

E-mail: {kajiyama, teigo}@dumbo.ai.kyutech.ac.jp

Abstract

Move sequence patterns can be used to reduce a lot of search and speed up the research. We acquire non-continuous move sequences with their correlations based on appearance frequency from game records of Go. In this paper, we describe how to acquire non-continuous sequence and experimental result of acquisition from "Kifu Database 96" which contains 5 million moves.

1 はじめに

囲碁はチェスや将棋などと比べると探索空間が非常に広く、また静的な局面評価が難しいため、定石や手筋、石の形といったパターン知識の利用は不可欠なものとなる。かつて、このようなパターンは定石書から入手したり、固定窓の範囲内で静的な形パターンを獲得する手法がとられていたが、近年では、パターンを範囲非限定、大規模かつ自動的に獲得する試みがなされている [5][8]。我々もパターンの範囲を固定せず、その順序を含意したパターンの知識を獲得する研究をおこなってきた。われわれは、定石や手筋といった定型手順と自然言語テキストの定型表現の持つ特徴の類似性に着目している。これまで、棋譜に記述してある一連の着手を、時系列に沿って記述したテキストととらえ、その中で定型的な手順の出現を獲得することを試みた [6][7]。結果として、連続着手の頻度情報から多くの定型的な手順を獲得することができ、囲碁プログラミングにおいて棋譜テキストをベースにしたアプローチが有効である根拠を得た。しかし、定型手順は過去に打たれた着点の情報を持たず、静的な局面に対する候補手の列挙に適用するとき、直前の着手しか評価の対象にならない点が問題であった。そこで、今回は連続ではない2つの着手列が1局の棋譜の中にどの順序で出現するかを調査し、一般に複数の着手列がどのように共起するかという情報をパターン知識として獲得を試みる。次節以降では、順序パターンとその共起知識を示し、その獲得方法を述べる。

2 順序パターンとその共起度の獲得

観戦記などに記載される譜面を見ると、黒石や白石の着点にその番号が振られている。これはどの順番で打たれたかを示す情報であるが、この情報がある場合とない場合で次に打たれるであろう着手を予測するとき、前者のほうが早く列挙できると考えるのが自然である。その根拠は

- 次着手は直前の着手の近傍にある可能性が高い。
- 直前の着手が即時にわかる。
- 人間は着手の流れ、「ここに打てばここに打つ」応酬の知識が意識の有無に関わらずある。

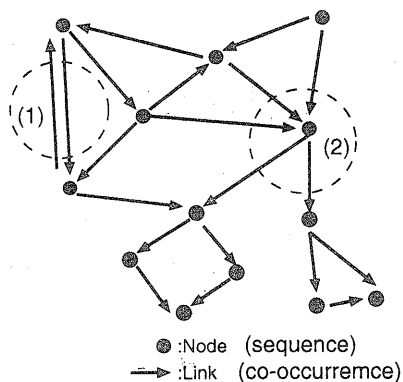


図 1: 手順とその共起関係を表すのグラフの例

ということが挙がる。人間のプレイヤーにとっては次着手の決定の際に直前の数手の着手の情報を利用している。また、次着手の候補手を決定する際、直前の着手を中心としてある程度の範囲の石の形も手掛かりとしている。このように1局のゲームにおいて過去の手順を参照していることは頻繁にある。複数の手順同士が1局の中でどう関わっているかを調査し、知識として獲得することは有益である。

ここで、手順とその共起の知識をグラフ構造で表現してみよう。1手以上のあらゆる手順をノードとし、棋譜データベース中に共起する2手順、つまり2つのノードについて、出現順に有向枝で結ぶ。そうするとたとえば図1のような有向グラフが生成されるであろう。図1の場合を例にとると、着手前後の存在(1)や、重要な分岐点(2)の存在が浮かびあがってくる。このような知識は、着手生成の手掛かりになるほか、図2のように共起関係から1局の棋譜グラフ構造の生成にも使える。

このようなグラフと同等な知識をテキストベースの形式で表現し、複数の手順が打たれた順番の情報を持つパターン、すなわち順序パターンを提案し、獲得することを目標に議論を行う。

2.1 順序パターン

n 手順列の順序パターンとは、1局の棋譜中における n 個の手順の出現順を列挙したものである。ここで、手順列とは1手以上の連続した着点の列のことである。

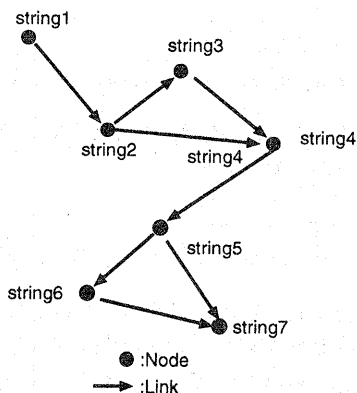


図 2: 1局に出現した手順列の共起グラフ

n 手順列の順序パターンは以下のように記述する.

$$STRING_1 : STRING_2 : \dots : STRING_n$$

$STRING_i$ は手順列であり,

$$STRING_1, STRING_2, \dots, STRING_n$$

の順に手順列が出現することを示す. また, コロンは両側の手順の非連続を許すものである. コロンを挟む両手順は重複しない. なお, 各手順列は石の色と着点の組を時系列に沿って列挙した文字列で, 以下のように記述する.

$$STRING_i :=$$

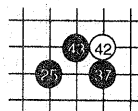
$$side_1(pos_1), side_2(pos_2), \dots, side_{k_i}(pos_{k_i})$$

$$side_i := \{SAME|DIFF\}$$

$pos := [x, y]$, $[x, y]$ は着点の座標を表す符号.

例えば, 3つの手順列からなる図3上部のパターンを同図下部のように記述する.

順序パターンは棋譜データベースの棋譜中に出現する頻度の情報をもつ. $n \geq 2$ のとき n 手順列の順序パターンの場合は特に共起度と呼ぶことにする. 定石の部分列などの断片的な手順がないと仮定したとき, 良いパターンとは頻度の高いパターンである. この仮定を満たすためには, 順序パターンの獲得の際は断片的な手順を排除する必要がある.



SAME(-1, -1)
: SAME(1, -1)
: DIFF(1, 0), SAME(0, 0)

図 3: 順序パターンの記述例

2.2 順序パターンの獲得

一般に, n 手順列の順序パターンを統計的手法のみで獲得するのでは大変手間がかかる. そこで, はじめに 1 手順列で構成する順序パターンを獲得する. つぎに 2 つの 1 手順列 S_1, S_2 をひと組として 2 個の手順列の順序パターンを獲得する. さらにそれをもとに一般の n 個の手順列の順序パターンを獲得してゆく.

2.2.1 1 手順列の順序パターンの獲得

1 手順列の順序パターンとは, 1 手以上の連続着手である. たとえば定石や手筋などの定型手順は 1 手順列の順序パターンである. これは図1で言えばノードに相当する. 1 手順列の順序パターンを獲得する手法には n -gram 統計を用いた定型手順の獲得法 [6][7] がある. これらの手法により単位性のある手順を獲得することができる. 2 手順列の順序パターンの獲得は 2 つの 1 手順列の順序パターンの組を調査する.

2.2.2 2 手順列の順序パターンの獲得

2 つの手順列 $string_1, string_2$ の順序パターン $string_1 : string_2$ を獲得する作業はグラフ構造における有向枝を獲得することに相当する. まず 2 つ



図 4: $string_1 : string_2$ の出現を表すグラフ

の 1 手順列の順序パターン S_1, S_2 を組とする. ただし, S_1 および S_2 は互いに重複をしない. 次に S_1, S_2 が棋譜データベース D の全棋譜のうち何局同時に出現

現するかの情報として共起度を獲得し、共起度が正である順序パターン $S_1 : S_2$ を獲得する。 S_1 と S_2 が S_1, S_2 の順で出現する共起度 $co(S_1, S_2)$ は下式のとおりである。

$$co(S_1, S_2) = \sum_{g \in D} |S_1, S_2|_g$$

ここで、 $|x_1, x_2|_g$ は1局の棋譜テキスト g 中に手順列 x_1 および手順列 x_2 がこの順で出現する頻度である。たとえば着点の列 " c_1, c_2, c_1, c_2, c_2' " がデータベース上にあった場合は、順序パターン $c_1 : c_2$ の共起度は5である。

2.2.3 n 手順列の順序パタンの獲得

$n \geq 3$ の n 手順列についての獲得の際、全ての手順の組合せについて共起度を調査するのは調査空間の広さが非常に大きくなるため現実的ではない。そこで、2手順順序パタンのときに用いた共起度をもとに、代数計算により疑似的に共起度を与えることで順序パタンの存在を調査する。

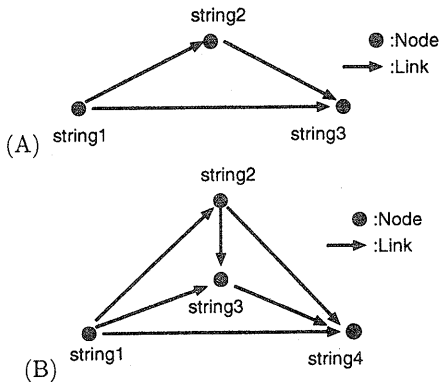


図 5: 「密な」グラフ

1局の記譜テキストを表す共起グラフ G を生成することを考えたとき、 s_1, s_2, s_3 の3手順がこの順で出現する必要条件は図 5(A) のような部分グラフが G 中に出現することである。これは3本の有向枝、すなわち3組の手順列がいずれもデータベース中に共起していて $co(s_1, s_2)co(s_1, s_3)co(s_2, s_3) > 0$ であることを指している。 s_1, s_2, s_3 の3手順がこの順で出現する共起度 $co(s_1, s_2, s_3)$ は、任意の2手順間の共

起度の最小値 $minfreq(s_1, s_2, s_3)$ 以下である。

$$minfreq(s_1, s_2, s_3) = \min\{co(s_1, s_2), co(s_2, s_3), co(s_1, s_3)\}$$

そこで1以下の正数の係数 $\alpha(s_1, s_2, s_3)$ を乗じた値を3手順の共起度とする。

$$co(s_1, s_2, s_3) = \alpha(s_1, s_2, s_3) \cdot minfreq(s_1, s_2, s_3)$$

係数 $\alpha(s_1, s_2, s_3)$ は、関連の低い組は0に近く、関連の高い手順の組は1に近い値をとる。手順 s_1 の頻度が $co(s_1, s_2)$ に等しい、すなわち s_1 のあとに必ず s_2 が生じるのであれば、最大値の1を得る。

$$\alpha(s_1, s_2, s_3) = 1 - \prod_{s_i, s_j \in S} (1 - \frac{co(s_i, s_j)}{|s_i|}) (1 - \frac{co(s_i, s_j)}{|s_j|})$$

ただし、 $|x|$ は棋譜データベース中 D における手順 x の出現頻度であり、 $S = \{s_1, s_2, s_3\}$ である。

一般の n 個の手順の共起度を得る際は、係数 $\alpha(s_1, s_2, \dots, s_n)$ を、3個の手順列組のときに用いた係数 $\alpha(s_i, s_j, s_k)$ の積で得る。

$$\alpha(s_1, s_2, \dots, s_n) = \prod_{s_i, s_j, s_k \in S} \alpha(s_i, s_j, s_k)$$

ただし、 $S = \{s_1, s_2, \dots, s_n\}$ である。 n 個の手順の間に $(n-1)(n-2)/2$ 本のノードが存在するとき、 n 手順の順序パターン全体に共起関係があるとし、共起度 $co(s_1, s_2, \dots, s_n)$ を与える。

$$co(s_1, s_2, \dots, s_n) = \alpha(s_1, s_2, \dots, s_n) \cdot minfreq(s_1, \dots, s_n)$$

3 順序パターン知識獲得実験

棋譜データベースから順序パターン知識を獲得する実験を行う。まず、ノードとなる手順を獲得し、次に2手順間の共起関係の存在を調査した。そして、3手順の組のとき全節で述べた共起度の計算値とデータベースからの実測値をとり、比較した。

3.1 実験データ

日本棋院「棋譜データベース'96」から、25281局、5220039手を使用した。着点の符号は絶対座標をベースに用いた。

3.2 実験結果

まず、文献 [1] の手法を用いて n -gram 統計をおこなった。その中から、頻度 32 以上、手順長 1 または 4 以上の手順パタンのうち、隣接文字エントロピー法 [6][2] で閾値 1 以上の手順を獲得した。手順長の特徴は表 1 のとおりである。

手順長	手順数
1	361
4	1226
5	227
6	103
7	61
8	28
9	30
10	7
11	14
12	3
13	3
14	1
16	1
24	1
合計	2066

表 1: 手順長の特徴

これら 2066 手順を 1 手順順序パターンとし、前章で挙げたように 2 つの手順を組として共起度を調査し、共起関係のある順序パターンを獲得した。その例が図 9 である。順序パターンは (1) 序盤の布石とヨセの手順の組 (2) 単位性のある長手順同士の組 (3) 一つの定型手順の 2 つの部分列があった。

次に獲得した 2 手順列順序パターン $s_1 : s_2$ のうち無作為に 27 個選択した。そしてある着点 c が打たれた後に $s_1 : s_2$ が出現する順序パターン $c : s_1 : s_2$ について、疑似共起度と真の (棋譜データベースから直接計数した) 共起度との比較を行った。図 6 は全ての実測値と計算値の対応点の分布である。係数 α を乗じない minfreq と実測値の分布 (図 7) をひと目比較してみると、計算値と実測値が等しい線により密集していることが分かる。また、計算値と実測値の平均 2 乗誤差は 1.24 であった。図 9 に獲得した 3 手順列の順序パターンの中から、共起度の高いものの例を示す。

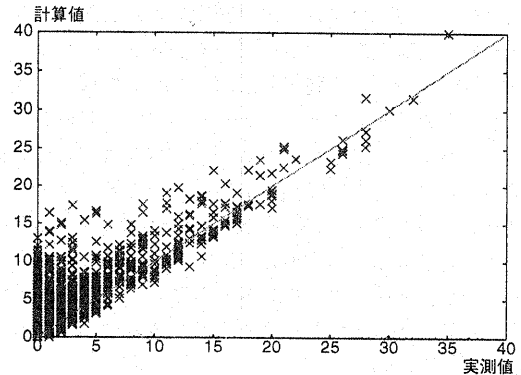


図 6: 3 手順列の順序パタンの共起度の実測値と計算値の対応点の分布

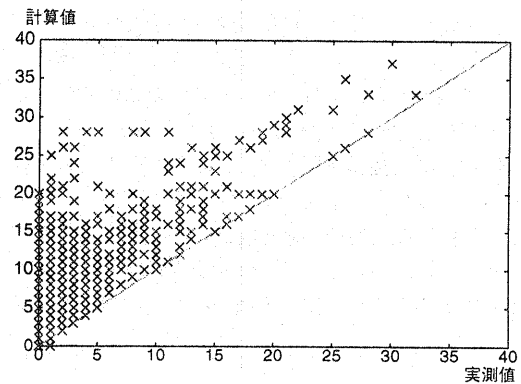


図 7: 3 手順列の順序パタンの minfreq と実測共起度の対応点の分布

3.3 考察

今回の実験の範囲では、共起度の疑似計算値と真の値の差は全体のほとんどが1未満と小さく、特に共起度の大きいときには差が小さいので、今回の実験で用いたパタンのように過去の着点との共起を調査するときは非常に有用である。また、「密な」グラフの存在を共起関係の探索の枝刈りの条件として利用することは有効な手段である。

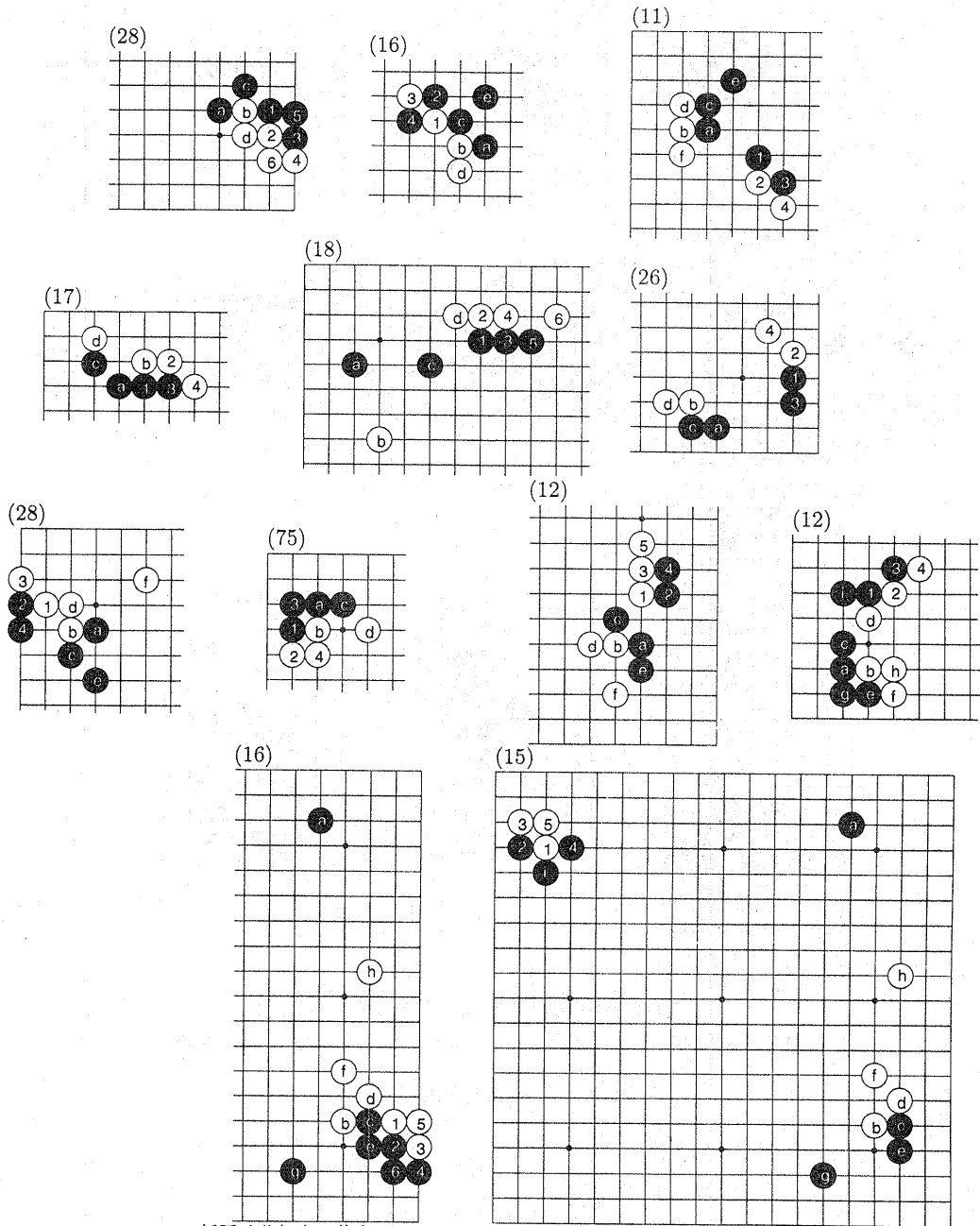
獲得した順序パタン知識は頻度または共起度があるので、次着手をシステムに予測させることができる。教師データから獲得した順序パタン知識は、手順そのものはそしてテストデータの全ての局面を出題し、パタン知識をもった次着手予想システムに候補手を列挙させる。以前、着手頻度をもとに確率文法モデルを生成し、次着手を予測させた [9]。このときは直前の着手列の頻度だけをもとに評価値を与えていたが、順序パタンを組み込むことで、過去の着手にも評価値に影響し、より人間の感覚に近い評価手続きができる。

4 おわりに

囲碁における着手列の前後関係を記述する順序パタン知識について述べた。獲得したパタンの頻度である共起度の獲得には、3以上の手順列の共起度については代数計算で疑似的に与える手法を示した。順序パタンの獲得実験により、2手順列の共起関係の存在を発見し、また、3手順の順序パタンについて計算値と実測値の誤差を有意に小さくすることができた。これからは、より大規模な順序パタンの獲得と次着手予測への適用実験を行い、有用性を検証したい。

参考文献

- [1] 長尾, 森: "大規模日本語テキストの n グラム統計の作り方と語句の自動抽出", 情報処理学会言語処理研究会報告 NL 96-1, pp.1-8, 1993.
- [2] 下畑, 杉尾, 永田: "隣接文字の分散値を用いた定型表現の自動抽出", 情報処理学会自然言語研究会報告 NL-110-11, pp.71-78, 1995.
- [3] 津田: "岩波講座ソフトウェア科学-9 数値処理プログラミング", 岩波書店, pp.151-218, 1988.
- [4] 有川, 宮野: "情報処理シリーズ-9 オートマトンと計算可能性", 培風館, pp.33-66, 1986.
- [5] 小島, 植田, 永野: "生態学アナロジーを用いた囲碁パターン知識の獲得", ゲームプログラミングワークショップ'96, pp.133-140, 1996.
- [6] 中村: "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'97, pp.96-105, 1997.
- [7] 中村, 梶山: "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の獲得", 情報処理学会ゲーム情報学研究会報告 99-GI-1, pp.107-114, 1999.
- [8] 小島, 吉川: "知識と用語の相互獲得モデルとその詰碁への適用", ゲームプログラミングワークショップ'99, pp.169-176, 1999.
- [9] 梶山, 中村: "囲碁の着手符号列に対する確率文法モデルの作成" ゲームプログラミングワークショップ'99, pp.161-168, 1999.



括弧は共起度，着点 a, b, c, ... は 1, 2, 3, ... より過去の着手である。

図 8: 2 手順列の順序パタンの例

