

LIU Zhi-qing, DOU Qing

Automatic pattern acquisition from game records in GO

CLC number TP391

Document A

Article ID xxxx-xxxx (2006) xx-xxxx-xx

Abstract Computer programs of GO, because of enormous complexities of the game, are typically constructed using a knowledge-based approach with heuristics and pattern matching. In this approach, quantity, quality and consistency of patterns used in computer programs of GO determine, to a large extent, the strengths of the programs. This paper presents an effective method to automatically acquire comprehensive GO patterns from large collections of game records. Statistical usages of the patterns ensure consistency and quality of the patterns, which, in turn, can help improve the strengths of computer GO programs. Additionally, statistical usages of patterns from different sources of game records clearly show subtle and significant discrepancies among various types of GO players, and clarify certain myths in the playing of GO.

Keywords computer games, automatic knowledge acquisition, spatial patterns recognition

1 Introduction

Computer games has been a part of the core of artificial intelligence since it became a field of study; and the game of GO is one of its grand challenges [19][16][1]. GO is a zero-sum, perfect-information board game of two players, who play the game by alternately placing stones of her color (black or white) on an empty crossing on a board of the game, which consists of a 19-by-19 grid. Lines on the game board are customarily ordered from edges of the board and are such named, e.g., first lines, second lines, etc. We, following the tradition of Chess, shall refer to a stone placed on the game board as a move.

The game of GO is a territorial game, and the player securing more territory at the end of the game wins. All alive chains eventually secure certain territories, which may or may not include the stones in the chains themselves depending on

whether the Chinese rules or the Japanese rules are used. Differences between the two sets of rules are generally minimal; both assume no-trivial knowledge of the game when territories are counted, because dead chains of stones are just recognized as such without being actually captured. This complication makes the game of GO highly different from other games such as Chess where concrete winning targets exist and can be achieved by an individual move that, for example, captures the king in Chess.

A game record of GO is, in addition to some supplemental information such as identification of the players and the result, just a sequence of moves, each of which is denoted as a two-dimension coordinate representing the position of the game board on which a stone is placed by the move. In general, the stone of a move can be placed at any empty crossing on the game board, subject to certain rules of the game such as the rule of no-suicide and the rule of ko.

Computer GO has been studied since 1969, when the first paper on this subject was published by Zobrist [21]. Many important progresses on computer GO have since been made, and some of the most influential works include [21][3][4][5][12][20]. However, All these works are still elusive with respect to the ultimate goal of defeating top-level human players in GO. The reasons are well recognized: Key computer techniques for successful play of other games such as Chess do not apply to GO. One of the key techniques is an accurate static evaluation of the game board integrated with an efficient search of the game tree. However, this technique is almost useless in computer GO because of the following two complications of the game: 1) GO has a significantly larger search space than other games due to its much larger branching factor, and 2) an accurate and static evaluation of the game board is not tractable [17]. Consequently, researchers typically construct computer GO players using different approaches, e.g., a knowledge-based one using heuristics and pattern matching.

A pattern in GO is a well established sequence of moves that can be accepted by both players, and GO patterns are frequently seen in professional games. It is believed in the GO community that patterns, which define relationships among moves, are of much more importance than the individual moves themselves [14]. In other words, it is relationships

Received date: 2006-9-4

LIU Zhi-qing (✉), DOU Qing

BUPT-JIUDING Computer GO Research Institute, School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Email: zhiqing.liu@gmail.com

among individual moves that effectively determine skills of GO playing. Effective recognition of patterns and their competent use are crucial for both human players as well as computer programs of the game. To this end, dictionaries of patterns at different stages of the game have been compiled, and pattern databases of various forms are used in almost all competent computer GO players.

The primary contributions of this paper are follows: 1) We define GO patterns as spatial relationships between moves within a fixed area on the game board; 2) This spatial definition allows GO patterns to be acquired automatically from game records and urgency of patterns determined statistically from their usage data; and 3) Statistical usages of patterns may differentiate different types of game records from which the patterns are acquired. These subsequently may help improve quality, quantity, and consistency of patterns used in computer GO programs, and clarify certain myths in the playing of GO.

We organize rest of this paper as follows: Section II reviews related works on pattern acquisition in the game of GO. Section III presents our spatial definition of GO patterns and discusses its advantages over traditional pattern definitions. Section IV presents patterns and their statistical usages acquired automatically from four different sources of game records: professional, top-level amateur, computer program, and random game records. Section V discusses subtle and substantial discrepancies among the different sets of patterns extracted and clarifies some myths in the playing of GO. Section VI concludes this paper with a summary of results and a discussion of future works.

2 Related works

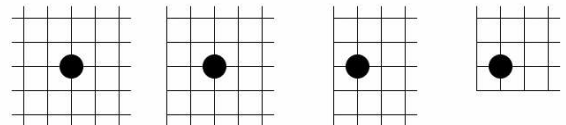
As stated in the introduction, effective recognition of patterns and their competent use are crucial for computer programs of the game. The patterns used in the computer GO players are either translated from pattern dictionaries directly, acquired through machine learning [2], generated by enumeration with urgency determined by static analysis [12][6], human expertise [10], or reinforcement learning [10], or extracted from game records statistically [18][15][9]. The statistical method of pattern extraction reported by Drake et al. [15] is similar to our approach. However, their experiments were conducted on a much smaller scale, using 874 9-by-9 game records of low-level amateur players. Nakamura's work on statistical pattern acquisition from game records [18] is also similar to our method of pattern extraction, with the exception of using n-gram statistics instead of spatial statistics as in this paper. As such, Nakamura failed to discover regularities in pattern frequency distribution because GO patterns, played on the two-dimension game board, are not strictly linear. In summary, all existing efforts of automatic acquisition of GO patterns are ineffective largely because of GO patterns are represented linearly and sequentially of moves.

Only occasionally used in computer GO, statistical methods have not been used for investigation of frequency distribution of moves and patterns in GO. Besides for pattern extraction as reviewed above, statistical methods are used mostly in Monte Carlo game evaluation and move generation in computer GO [11][8][10]. Monte Carlo evaluation in computer games was formalized in a framework by Abramson [1], in which the expect outcome of a game position was estimated to be the average of outcomes of N random games from the position. We in this paper also acquire patterns from random game records in order to conduct a comparative study on patterns and their statistical usages. To this end, we generate a large number of random games as in Monte Carlo GO.

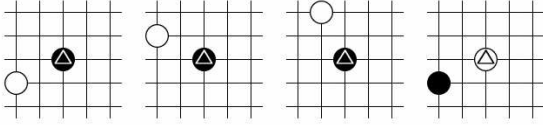
3 Spatial pattern definition

Nakamura's work on pattern acquisition from game records using n-gram statistics [18] is interesting, which is similar to acquisition of linguistic collocations such as words and word combinations from a corpus. However, it has a major deficiency. GO patterns are fundamentally different from words and word combinations in natural languages, because the game is played on a two-dimension game board. The organization of a word or word combination follows strictly the chorological order of its components because natural languages are all rigidly linear. In contrast, moves in a GO pattern are not necessarily played consecutively, and are frequently played out-of-sequence, especially at the level of professionals. Thus, the chorological order of GO moves is of only limited importance; but more important are spatial relationships among the moves. Therefore, we in this paper are concerned only with spatial relationships of the moves in patterns, and define patterns accordingly.

Although large amount of GO patterns can be found in pattern dictionaries, they are not used in our experiments because dictionary patterns have different physical properties such as shape and size, making comparisons of their statistical usages meaningless. Instead, we specify a fixed area of the game board serving as the region in which patterns are defined, and acquire patterns and their statistical usages automatically and statistically from game records. The fixed area used in our pattern definition is a 5-by-5 square, centered by the current move. When the move is close to an edge or a corner of the game board, the region for pattern definition is degenerated and its area naturally reduced to the remaining part of the region. For example, the following four areas are all valid, although some are degenerated, pattern definition regions with the most recent move (marked as the black stone) shown at the center.



Additionally, our spatial definition of GO patterns combines equivalent shapes into one single one, called the canonical form, by eliminating symmetry properties of the shapes, as in the work of [15]. Shapes are considered to be equivalent if they are the same with respect to the rotation and flipping operations. Shapes are also considered to be equivalent with respect to color switch. For example, the following four shapes are considered to be the same, with the current move marked with a triangle and shown at the center.



In the canonical form of patterns, the current move is always a stone of the black color, subject to color switch. A maximum of eight equivalent shapes are possible for a given pattern in its canonical form, among which four are obtained due to the rotation operation and the other four are due to the flipping operation. A shape is coded as a 64-bit number; all equivalent shapes are sorted as such and the shape with the smallest code is defined to be the canonical form of the equivalent shapes. In summary, we define a spatial GO pattern to be the canonical form of stones and their positions on the 5-by-5 board area centered by a current move, with all positional and color symmetric properties eliminated.

4 Automatic pattern acquisition

Given the spatial pattern definition as described above, GO patterns and their statistical usages can be acquired automatically from game records. This section discusses an efficient algorithm for automatic pattern acquisition and presents high-frequency patterns acquired from four collections of game records from various sources. The four game record collections are the following:

1. A collection of 9447 game records of professional players, which is provided by Yu, Ping, the Chinese professional six-Dan player, who has used the game collection for a study that is totally unrelated to the works reported in this paper;
2. A collection of 10000 game records played online at the KGS server by top-level amateur players, with either one of the players (or both) is amateur seven-Dan or stronger, or both are amateur six-Dan;
3. A collection of 1370 games records constructed by self-play of GNUGO 3.6, a top-level computer GO program; and
4. A collection of 3553 games records constructed by self-play of a random computer GO program, which plays all legal moves randomly with equal probability, except those moves that may destroy one's own eyes.

Given the pattern definition, each move in a game specifies a, possibly degenerated, 5-by-5 region in which a pattern occurs. Therefore patterns and their statistical usages are acquired from game records by the following algorithm, whose complexity is linear with respect to the number of moves in the game record collection:

```

for each game record R in the game record collection loop
  for each move M in R loop
    Play M on the game board;
    Obtain the 5-by-5 region R centered by M;
    Rotate and flip R into its canonical form;
    if R is not in our pattern database then
      Add R into the pattern database;
      Set frequency number of R to be 0;
    end if;
    Increase the frequency number of R by 1;
  end loop;
end loop;

```

The patterns and their statistical usages are maintained in a database table, in which each entry represents a unique pattern and, for performance purposes, each pattern is represented by a 64-bit number.

Running this algorithm on the four collections of game records allows us to acquire four sets of patterns and their statistical usages, one pattern set for each of the collections of game records. The four pattern sets have 980771, 769238, 164260, and 474117 unique pattern entries respectively. Figure 1, 2, 3, and 4 show the top twenty patterns with highest frequency of occurrence from the four pattern sets respectively.

The patterns such acquired have certain properties that are worth to be noted. First of all, the patterns are of a large quantity. As an example of the patterns from the professional game records, roughly 100 unique patterns are acquired from each of the game records, although as the number of game records increases, the number of unique patterns acquired from each game record will gradually decrease. Secondly, statistical usages of the patterns from a large collection of game records ensure that the patterns acquired are consistent and of a good quality. The frequency of occurrence a pattern indicates, to a large extent, its relative urgency and usefulness, because the higher the frequency of occurrence a pattern, the more useful and the more urgent the pattern is. Unlike hand-tuned urgency of patterns that are often subject to interpretations of the individuals who perform the task, the statistical usages of pattern are objective, reflecting a collective view of its usage. The patterns are thus consistent and of a high quality and can be used to help improve the strength of computer GO programs, as in the case of [10], in which 3-by-3 patterns are enumerated and their urgency values are either tuned manually by a human expert or acquired automatically by computer with reinforcement learning.

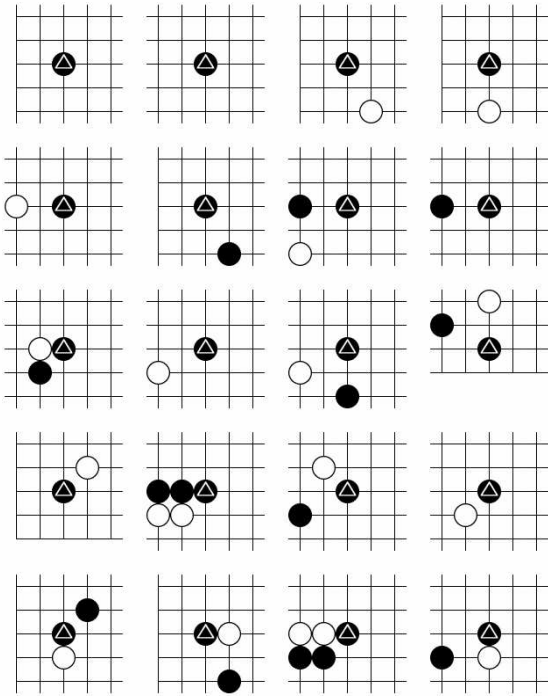


Fig. 1 Top twenty high-frequency patterns in the professional game records.

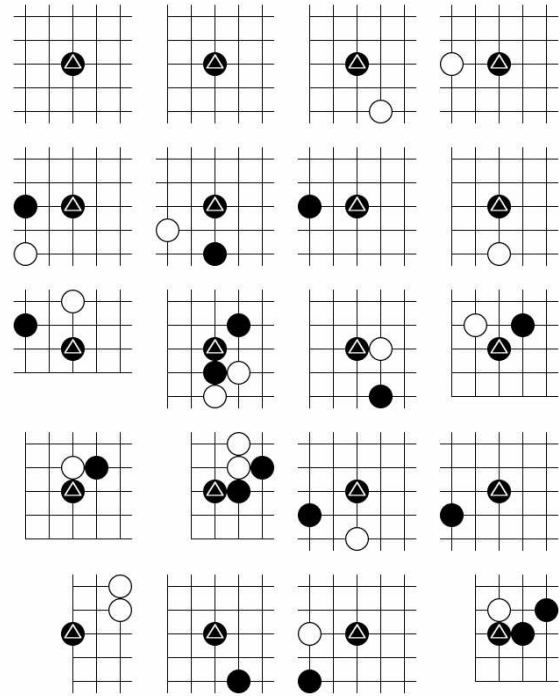


Fig. 3 Top twenty high-frequency patterns in the GNUGO game records.

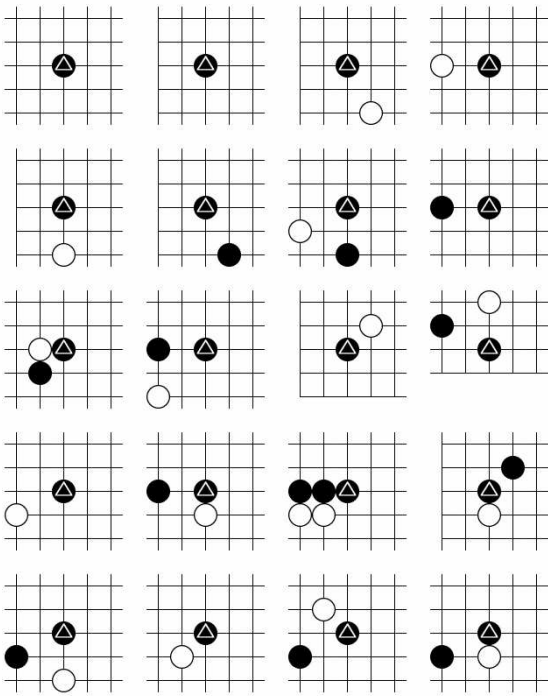


Fig. 2 Top twenty high-frequency patterns in the top-level amateur game records.

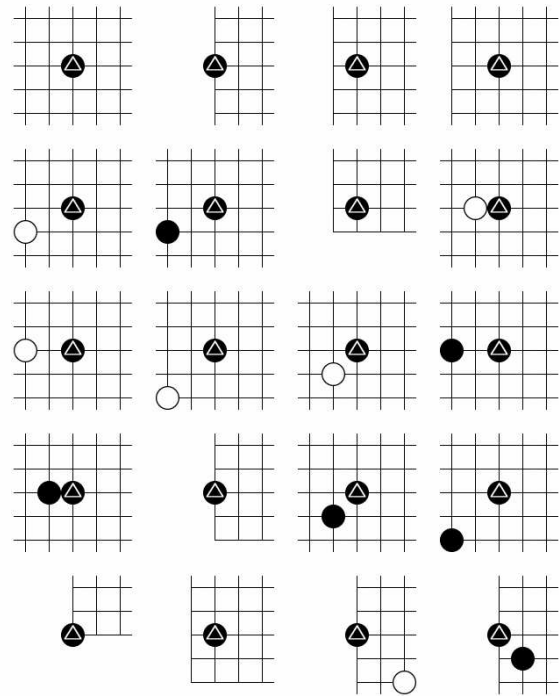


Fig. 4 Top twenty high-frequency patterns in the random game records.

5 Discussion

It is obvious from Figure 1 to 4 that high-frequency patterns in the four patterns sets are different. Much of the difference among the four types of GO players can be described by the order of the statistical usage of a pattern in the four pattern sets. Given a pattern P , let $O_p(P)$, $O_a(P)$, $O_g(P)$, and $O_r(P)$ denote respectively the order of the statistical usage of P in the pattern sets of professional, top-level amateur, GNUGO, and random game records. Table 1 shows, for the top twenty high-frequency professional patterns, their respective order in the top-level amateur, GNUGO, and random pattern sets.

It is obvious with knowledge of GO that the top-twenty high-frequency patterns in the professional pattern set are all important and played frequently. For example, the most frequent pattern, playing a stone on a third line of the game board with empty surroundings, is an important opening move. However, not all of the top-twenty high-frequency professional patterns occur with a high order of frequency in other pattern sets. It is clear from Table 1 that 18 of the top-twenty patterns appear on the top-twenty list of the top-level amateur pattern set, while only 11 and 6 of them appear on the top-twenty lists of the GNUGO and random pattern sets respectively. It is easy to see that, in the perspective of pattern usages, the games played by top-level amateur are similar to the games played by professionals. However, significant differences exist between the games played by top-level human players and those of computer programs. For example, two high-frequency professional patterns do not appear at all in the random games, and appear very infrequently in GNUGO games.

Even though statistical pattern usages between the professionals and the top-level amateurs are similar, discrepancies in the order of their pattern statistics reveal differences in the ways that the professionals and the top-level amateurs play. These differences, although significant, are generally so subtle that people have not been able to identify them with quantitative evidences. We detail the differences as follows, where P_k denotes the pattern in the professional pattern set whose order of statistical usage is k :

1. As indicated by the relative order between P_1 and P_2 , and between P_4 and P_5 in both the professional and the top-level amateur pattern sets, the professionals prefer playing on the third lines in opening games; while the top-level amateurs prefer the fourth lines. Clearly the professionals are more concerned with protecting territories.
2. As indicated by the relative order of P_7 in both pattern sets, the professionals feel more pressures under tsume and prefer a one-space jump in response.
3. As indicated by the relative order of P_{10} in both pattern sets, the professionals prefer a knight move to approach the opponent.

Table 1 The orders of the top twenty high-frequency patterns from the professional pattern set in the top-level amateur, GNUGO, and random pattern sets

$O_p(P)$	$O_a(P)$	$O_g(P)$	$O_r(P)$
1	2	2	2
2	1	1	4
3	3	3	66
4	5	8	51
5	4	4	9
6	6	18	78
7	10	5	86
8	8	7	12
9	9	21	223
10	13	23	5
11	7	6	95
12	12	9	674
13	11	86	3841
14	15	212	NA
15	19	24	177
16	18	101	11
17	16	49	1983
18	25	11	6633
19	23	430	NA
20	20	84	102

4. As indicated by the relative order of P_{11} in both pattern sets, the professionals do not play as urgent as the top-level amateurs the one-space jump when approached by a knight move from the opponent.
5. As indicated by the relative order of P_{13} in both pattern sets, the professionals do not play as urgent as the top-level amateurs the 3-3 point invasion for the opponent's 4-4 point.
6. As indicated by the relative order of P_{14} and P_{19} in both pattern sets, the professionals feel more urgency in such a circumstance to either extend for defense or to hane for attacking.
7. As indicated by the relative order of P_{15} and P_{16} in both pattern sets, the professionals feel an urgency to attack a single stone of the opponent with a shoulder hit.

While our approach presented in this paper is capable to acquire a large amount of high-quality patterns automatically and efficiently from GO game records, more can be done in this direction to further improve machine learning. One possible research task in the future is to study how game heuristics such as game status analysis techniques can be used such that important patterns can be acquired more effectively.

6 Conclusion

Our experiments on automatic acquisition of spatial patterns in the game of GO allow us to make a few conclusion remarks at the end of the paper. First, we can conclude that GO patterns shall be defined spatially in reflection of its two-dimension characteristics, because n-gram statistics as used in [18] can not fully acquire and classify patterns effectively, especially for professional game records. Secondly, symmetry properties must be removed in order to effectively obtain statistical usages of patterns, because, otherwise, the number of patterns will increase significantly and related patterns cannot be classified together, resulting in ineffective statistics. Thirdly, a large amount of quality and consistent patterns can be automatically acquired from game records, and the pattern acquisition algorithm is very efficient, whose complexity is linear to the number of moves in the game records. Fourthly, statistical usages of patterns indicate relative usefulness and urgency of patterns, which can be used to help improve the strength of computer programs. Additionally, discrepancies in statistical usage ordering of patterns acquired from different sources of game records can reveal subtle and important discrepancies of the different players of the game records. And finally, we believe that more can be done in the direction of distinguishing different-level of GO playing skills with their statistical usages of patterns.

Acknowledgements The authors are indebt to Professor LIN Jin-tong, the president of Beijing University of Posts and Telecommunications for helping establish the BUPT-JIUDING Computer GO Research Institute, which makes this work possible. The authors thank Yu Ping for insightful discussions of the game and for providing the game records of professional players used in this paper.

References

1. Abramson B., "Expected-outcome: a general model of static evaluation", IEEE Transactions on PAMI, 12, 1990, pp. 182-193.
2. Abramson M., Harry W., "A distributed reinforcement learning approach to pattern inference in Go", In International Conference on Machine Learning Applications, Los Angeles, CA, 2003.
3. Benson D. B., "Life in the game of Go", Information Sciences, 10, pp. 17-29, Reprinted in Computer Games, D. N. L. Levy (Ed), Vol. II, Springer-Verlag, New York, N.Y. 1976, pp. 203-213.
4. Berlekamp E., Wolfe D., "Mathematical Go Endgames, Nightmares for the Professional Go Player", Ishi press international, San Jose, London, Tokyo, 1994.
5. Boon M., "Pattern Matcher for Goliath", Computer GO 13, winter 89-90.
6. Bouzy B., "Go patterns generated by retrograde analysis", In Computer Olympiads, Maastricht, 2001.
7. Bouzy B., Cazenave T., "Computer Go: an AI oriented survey", Artificial Intelligence, 132, 2001, pp. 39-103.
8. Bouzy B., Helmstetter B., "Monte Carlo Go Developments", Advances in Computer Games, Many Games, Many Challenges. J. van den Herik and H. Iida, E. Heinz (ed), 2003, pp. 159-174.
9. Bouzy B., Chaslot G., "Bayesian generation and integration of k-nearest-neighbor patterns for 19x19 Go", In IEEE 2005 Symposium on Computational Intelligence in Games, Colchester, UK, 2005. pp. 176-181.
10. Bouzy B., Chaslot G., "Monte-Carlo Go reinforcement learning experiments", In IEEE 2006 Symposium on Computational Intelligence in Games, Reno, USA, 2006.
11. Brüggmann B., "Monte Carlo Go", 1993, online at <http://www.joy.ne.jp/welcome/igs/Go/computer/mcgo.tex.Z>
12. Cazenave T., "Generation of patterns with external conditions for the game of Go", In Advances in Computer Games Conference, Paderborn, 1999.
13. Chen K., Chen Z., "Static analysis of life and death in the game of Go", Information Sciences, 121, 1999.
14. Cho C., Go: A Complete Introduction to the Game, Kiseido Publishing Co, 1997.
15. Drake P., Levenick J., Veenstra L., Venghaus A., "Pattern Matching in the Game of Go", online at https://webdisk.lclark.edu/xythoswfs/webui/_xy-13651_1-tid_jDVKYDTf
16. Müller M., "Computer Go", Artificial Intelligence, 134, 2002, pp. 145-179.
17. Müller M., "Position Evaluation in Computer Go". ICGA Journal, 25(4), 2002, pp. 219-228.
18. Nakamura T., "Acquisition of move sequence patterns from game record database using n-gram statistics", In Game Programming Workshop 97, 1997. In Japanese.
19. Schaeffer J., van den Herik J., "Games, Computers, and Artificial Intelligence", Artificial Intelligence, 134, 2002, pp. 1-7.
20. Wolf T., "Forward pruning and other heuristic search techniques in tsume go", Information Sciences, 122, 2000, pp. 59-76.
21. Zobrist A., "A model of visual organization for the game of go", Proceedings AFIPS 34, 1969, pp. 103-112.

Biography: LIU, Zhi-qing is a professor at Beijing University of Posts and Telecommunications (BUPT). Prof. Liu received a Ph.D. in computer science at New York University. Prof. LIU's current research interest is on all aspects of computer processing of the game of GO, which include pattern matching, knowledge mining and representation, and statistical natural language processing. DOU, Qing is a bachelor student at School of Software Engineering, BUPT. Mr. DOU's research interest is Artificial Intelligence.