Improving Opening Book Performance
Through Modeling of Chess Opponents

Steven Walczak

University of South Florida
4202 E. Fowler Ave., CIS 1040
Tampa, FL 33620

(813) 974-6768   phone
walczak@bsn.usf.edu   email

Abstract

Opening books play a vital part in the performance of current chess programs. Use of the opening book places the chess program in a position that is already several moves into the game and can be positionally inferior. A method for acquiring opening knowledge about a specific opponent is presented. The method analyzes the historic performance of the opponent to find the opening sequences that are known to the opponent. Knowledge about the opening preferences of an opponent affords a strategic advantage to a chess program. The performance of this method is demonstrated and analyzed. Current chess programs that utilize knowledge about the opening repertoire of an opponent will be able to decrease the size of their opening books and can develop a game strategy from the start of the chess game instead of the beginning of the middle game.

INTRODUCTION

Current computer chess machines implement opening books to guide the opening moves of chess games. These books consume system space and are purely tactical in nature, suggesting moves that follow a particular sequence that is recorded in the opening book. The use of books does not permit the development of strategy until the book is no longer in use. This means that computer chess systems cannot utilize heuristic domain knowledge for strategic planning until after the opening moves have already occurred, which can be as late as the tenth or even twelfth move of a game for complicated opening sequences.

White (1990) advocates reducing the size of the opening book as much as possible. This will free computer system space and generally increase the speed of opening move selection by

chess algorithms that use opening books. White presents a method for using abstraction to take advantage of the redundancies found in typical chess openings. The abstraction method of White will reduce the size of an opening book which for the first five moves alone has a size of over 28 million combinations, but the reduction is only marginal.

Studying the past performance of a specific opponent enables the opening book to be tailored to recognize the specific openings known and practiced by that opponent. Additional openings that represent strategies to be used against the opponent can also be included in the opening book. As an example, prior to the World Championship Match against Spassky, Fischer only varied from the King's Pawn opening, e4 or PK4, three times as white during tournament play (Evans, 1970). By including only those openings exhibited by an opponent in previous matches, the size of opening books can be significantly reduced while the efficiency of the corresponding book will be proportionately increased.

The use of an opening book that is tailored to match the prior performance of a specific opponent permits the development of strategy from the start of the game. John Nunn has demonstrated the strategic advantages gained by playing opening systems that are not familiar to an opponent (Nunn & Griffiths, 1987). In a game against G. Anthony, Nunn intentionally avoids the Pelikan variation of the Sicilian because of Anthony's detailed knowledge of and previous successes against this specific variation. Nunn's book gives other examples of selecting opening lines of play to avoid an opponent's strengths with specific openings. The opponent specific opening book detailed in this paper permits the selection of opening lines that have not been previously demonstrated by an opponent. The use of this strategy in selecting opening lines of play which is similar to the method used by human chess competitors will place the chess computer in an advantageous position.

LEARNING OPENING SEQUENCES OF AN OPPONENT

Human competitors in game domains rely heavily on knowledge about an opponent's prior performance to develop strategies of play. Baseball and football coaches use scouting reports and films of their adversaries to devise counter-strategies

that will give their team a strategic advantage (Aliston & Weiskopf, 1984 and Riggins & Winter, 1984). The trait of studying opponent's is common to the game of chess as well. World Chess Champions Alekhine, Botvinnik, and Tal have all stated that they perform a detailed analysis of the playing ability and style of opponents prior to matches (Horowitz, 1973 and Schonberg, 1973). Each of these World Chess Champions was able to utilize knowledge that was learned while analyzing the prior performance of their opponent to defeat that opponent in tournament play.

For the research described in this paper, only games which have a won or lost result are used to analyze an opponent's opening repertoire. This is because a won or lost result permits further analysis concerning the re-use of particular openings by an opponent, whereas drawn games are less conclusive. In the 1948 Hague-Moscow World Championship Tournament, Botvinnik played twelve games which had a win/loss result. In the first five moves alone, Botvinnik displayed ten different opening combinations. Using the straight forward approach of recording every move for the openings played, with wildcard matches allowed for the black player's move when Botvinnik was playing the white pieces, 100 individual moves would need to be stored. Applying White's method (White 1990), five moves that are duplicated in the first position can be eliminated as well as five more that are duplicated in the second position. Eliminating all redundant moves still requires storing 84 individual moves.

Mednis (1990) has recently claimed that chess masters can often acquire a particular opening style from various first moves by performing transpositions. For example, a chess computer decides to play a Slav Defense responding to 1. c4 with c6, planning for the continuation of 2. d4 d5. The white player however, prefers to play a Caro-Kann opening and thus returns 2. e4. Because of the problem of transpositions, a statistical analysis of the opening repertoire of a specific adversary is saved instead of the individual moves involved in the various transpositions. This statistical analysis permits the research methodology described to reduce the total number of moves saved for the twelve games of Botvinnik mentioned above, from 84 (for White's method) to twenty-five individual moves, a seventy percent savings.

The knowledge base of opening moves, or opening book, for a particular opponent will contain the character representation of each move and some additional information as shown in Table 1. The additional information included with each move is: the color of the pieces being played, the number of times the move has been executed in opening play, the earliest move (in game turns) that this move has been made, the latest turn this move has been made, and the statistical mean of all turns for which the move has been used. A header block of information that gives the total number of games and the number of games for the white player that have been analyzed is included in each knowledge base. For the knowledge base of Botvinnik openings observed at the Hague-Moscow Tournament the numbers for total games and

games as the white player are twelve and seven respectively. By storing the integer values in the table as binary data, the space required for each move is roughly equivalent to the space required by White's algorithm for each pair of moves. The total space required by our method still affords a forty percent savings in system space.

Table 1: Player specific, opening move KB.

Move	Color	Frequency	Start	Finish	Average Turn
d4	White	6 games	1	1	1.00
c4	White	6 games	2	3	2.33
Nf3	White	4 games	2	3	2.25
Nc3	White	6 games	3	4	3.33
e3	White	5 games	4	5	4.40
a3	White	3 games	5	5	5.00
e6	Black	4 games	1	1	1.00
d5	Black	4 games	2	2	2.00
d4	Black	1 game	3	3	3.00
c5	Black	4 games	1	4	3.00

The remainder of this section is concerned with demonstrating the accuracy of the research method for identifying the opening move selections of an opponent. Once the ability to accurately predict an opponent's choice of opening moves has been shown, then combining this method with specific opening lines from a standard chess opening book that are not well known to the opponent affords computer chess programs the smallest possible opening book that will cover the openings relevant to defeating a particular opponent.

To determine which move an opponent is likely to select, a probability is assigned to each of the moves in the knowledge base of the same color pieces. Because the move choices for the black piece player are independent of those for a white piece player, the knowledge base size can be reduced even further by separating the white and black moves into two separate knowledge bases. The appropriate knowledge base can be loaded into the chess machine before the start of play. The move with the highest probability is then selected as the most likely move choice for the opponent. To account for statistically significant opening move variations employed by an opponent, if an additional opening move exists in the knowledge base with a probability higher than fifty percent, then that move will also be

recommended. When two alternative moves are recommended by the inference mechanism, the chess system must utilize some heuristic means for determining which of the two moves to use in the current game.

The probability assigned to each move choice is formed by first calculating the frequency of the move using the equation: P(move) = (Number of times observed) ÷ (Total games of that color analyzed). For example, the probability of Botvinnik choosing the d4 move as his first choice would be 85.7% = 6 ÷ 7, since the d4 move was observed in six of the seven games analyzed for Botvinnik playing the white pieces.

After the frequency has been calculated then each probability is adjusted to account for the distance in moves that the proposed move is away from the actual game turn. Move choices that have a statistical mean for the average game turn greater than 2.5 turns away from the current game turn are not evaluated. For example, the second move of Botvinnik (assuming that the d4 move has already been made) is between the c4 and Nf3 moves in the knowledge base. The initial frequency probabilities of the respective moves are 85.7% and 57.1%, but the mean game turns for use of the move are 2.33 and 2.25. Therefore, the base probabilities are adjusted to account for the difference in mean time of usage and the actual game turn to produce probabilities of 69.45% and 50.89% respectively for the c4 and Nf3 move choices. The adjustment for non-exact game turn matches is determined using the pseudocode shown in the box below. To demonstrate the usage of the code:

69.45% = 85.7% - ((2.33 - 2) * .25) - (2.33 - 2.25)

```
DIFF = (ABS(current turn - mean turn) * 0.25 + ABS(mean turn -
       MIN(mean turn of all potential moves)))
IF ( ABS(current turn - mean turn) >= 1.0)
   THEN DIFF = DIFF * 2
P(move) = P(move) - DIFF

// Note: the phrase "mean turn" refers to the average game turn of all
//       observed occurrences for a particular move.
// ABS is the absolute value function
// MIN is the minimum value function
```

The ability to select moves which have occurred at different times/turns of a game other than the current game turn permits transpositions to be predicted when a particular move has been frequently used in prior games at times which are relatively close to the current game turn being predicted. The combination of the rule which does not examine moves that have a distance greater than 2.5 from the current turn and the pseudocode shown above prevent a move that have always been used at a precise point (say the first or fifth move) from affecting predictions about the likely move choice of an opponent at times of the game which are relatively distant from the mean usage time for the move.

The choice of the number five to be the length of the opening games studied was made to capture high quality

information. Most chess masters are consistent in the repetition of opening moves up to the fifth game turn. The actual time at which a chess player leaves his known book ranges from the fifth to the tenth move of the game. To prevent moves that are not representative of the standard openings known to an opponent, the initial knowledge base implementation of the research described in this paper only analyzed the first five moves of each game. Further research was performed to test the effect of extending the statistical analysis of an opponent's move choices to a length of ten game turns and this research is described below.

RESULTS OF THE STATISTICAL ANALYSIS OF AN OPPONENT'S MOVE CHOICES

The twelve games played by Botvinnik in the Hague-Moscow Tournament were analyzed by the program IAM, Inductive Adversary Modeler (Walczak, 1992), to test the effectiveness of the statistical approach for predicting an opponent's movement choices based solely on the previous play exhibited by the opponent. A knowledge base of twenty-five moves, eleven for Botvinnik playing the white pieces and fourteen for Botvinnik as the black piece player, was created by IAM (the first ten moves are displayed in Table 1). Additionally, IAM knew that the knowledge base was constructed from twelve total games, for seven of which Botvinnik was the white player.

A simulation of the first win/loss result game from Botvinnik's 1951 defense of the world championship against Bronstein was performed to evaluate IAM's effectiveness in predicting an opponent's opening moves using the statistical knowledge base of opponent specific opening moves. Botvinnik played the white pieces in this game. Prior to each move to be made by Botvinnik, IAM selected the move from its knowledge base that it believed Botvinnik would play next. The results are shown below in Table 2. The probability numbers associated with each move are also displayed in the table. Basing the predictions of the move choices an opponent will make on the historical performance of the opponent, produces an eighty percent accuracy rate of predictions. Also, the one prediction that was not satisfied had a confidence level (probability) of less than fifty percent . Simulations against other games played by Botvinnik provided similar results.

Next, the opponent specific opening move knowledge base was tested against a game in which the opponent, Botvinnik, was playing the black pieces. The game used for the simulation was the next game of the match against Bronstein. The knowledge base of opponent specific openings performed poorly against Botvinnik as the black player, only predicting forty percent of the first five moves correctly. This result follows naturally from the psychological observation that as the black player, an opponent is required to react to the movement choices made by the white piece player.

An extension to the knowledge base was then added to account for the responsive nature of play required from the black

Table 2: Results of IAM predicting Botvinnik's opening play. (Botvinnik vs. Bronstein, 1951 World Championship)

Game Turn	Actual Move Made by Botvinnik	Move Predicted for Botvinnik by IAM	Probability of move
1	d4	d4	85.7%
2	c4	c4	52.3%
3	Nc3	Nc3	77.8%
4	e3	e3	70.4%
5	Bd3	a3	42.9%

Table 3: IAM predictions for Botvinnik playing the black pieces.

Game Turn	Move Made by Botvinnik	Original Prediction without Response Code	New Prediction with Response Code
1	c5	e6	c5
2	Nc6	Nc6	Nc6
3	cxd4	e6	cxd4
4	Nf6	Nf6	Nf6
5	d6	e5	d6
Prediction Accuracy		40%	100%

piece player. This was accomplished by modifying the analysis algorithm with "response code" that records the previous move made by the white player for each of the moves made by the opponent as the black player. The size of the knowledge base is increased proportionately, because a single move such as d5 (advancing the queen pawn two ranks) will be considered as two separate moves if it follows two different moves made by the white player. Now, predictions for an opponent playing the black pieces are first selected from the moves that have previously followed the corresponding white piece move just made in the current game. Using this technique, the size of the knowledge base is increased, for the twelve games of Botvinnik from the Hague-Moscow tournament, to thirty-three individual moves. The entire increase in size corresponds to the recording of black piece moves considered as unique choices for movement. The number of black piece moves recorded in the knowledge base went from fourteen to twenty-two moves. The new size of the knowledge base still affords nearly a twenty percent savings over other methods.

The new opponent specific opening move knowledge base was tested against the same game as the previous knowledge base. The results displayed in Table 3 show that the new knowledge base which emulates the responsive behavior of the chess opponent accurately predicts one hundred percent of the opponent's opening move choices. Other game simulations with Botvinnik playing the black pieces have provided similar results with prediction accuracy ranging from sixty to one hundred percent.

Finally, the knowledge base was augmented to see if any benefit would be gained by analyzing the first ten moves, instead of just the first five moves, that were made by a specific opponent in previous games. Botvinnik demonstrates a wide variety of move choices after the fifth move of the game. The opponent specific opening move knowledge base without response code held seventy-two individual moves that had been made by Botvinnik in previous games. This is triple the size for the five move

knowledge base instead of double, demonstrating the greater availability of move choices executed by the opponent. The simulation of both games for Botvinnik (as the white and black piece player) did not show any improvement from the larger knowledge base with the same moves being accurately predicted, but now the percentage is reduced to forty and fifty percent since the number of moves being predicted has doubled. Further game simulations have shown that doubling the number of moves analyzed will only produce a ten to fifteen percent maximum increase in the number of moves that are predicted accurately.

A more detailed study was undertaken using the responsive code. The results are shown in Table 4, for 69 championship games played by Botvinnik and for the first fifteen games of the 1990-91 Championship between Kasparov and Karpov. An overall prediction percentage of 83.5 percent was obtained for the Kasparov and Karpov players. Furthermore, it is of interest to note that the knowledge base of moves for each player playing the white pieces contained eight total moves (Kasparov primarily used a Ruy Lopez opening and Karpov primarily used a double queen's pawn opening) and the knowledge base of moves for Kasparov playing black contained eight unique moves (for the King's Indian and Gruenfeld defenses) stored in eleven different responsive pairs, while the knowledge base for Karpov contained nine unique moves (for the Morphy and Tchigorin defenses) in eleven responsive pairs. Furthermore, fifty percent of the prediction errors for Kasparov playing the black pieces occurred when Kasparov switched from playing the King's Indian Defense (used in his first three games as black) to playing a Gruenfeld Defense.

USE OF OPPONENT HISTORIES

The collection of the opening repertoire known to an adversary and the statistical analysis of these openings will permit current chess programs to develop strategies from the start of the game. The previous section has demonstrated the ability of this methodology to accurately predict the opening sequence of moves

Table 4: Results of using opponent histories to predict openings.

	White Moves Analyzed	White Moves Predicted	Black Moves Analyzed	Black Moves Predicted	Total
Botvinnik	195	158 (81%)	150	108 (72%)	77%
Kasparov	35	29 (83%)	40	34 (85%)	84%
Karpov	40	33 (82%)	35	29 (83%)	83%

that will be used by a specific opponent. The primary use of the historical analysis of an opponent's opening repertoire should be to emulate human chess masters by selecting opening sequences from an alternate book source that have not been demonstrated by the opponent. The use of openings that are unfamiliar to an opponent affords a definite strategic advantage to the chess program. The ability to predict the move choices of an opponent permits the chess program to strategically analyze the opening portion of the game prior to the actual game. This prior analysis can be used to gain the best positional and strategic position possible.

Furthermore, the overall size of opening books can be significantly reduced. A size savings of twenty to forty percent is realized by using the opponent specific opening move knowledge base. This savings in required systems space will permit current chess programs to execute quicker and help to reduce the time deficits that plague some of the current chess programs competing in human tournaments.

The possibility of an opponent misleading the knowledge base does exist. This deception on the part of chess competitors can be reduced to a minimum by only examining the history of the opponent in tournament level play, including the game records of the current tournament in which the chess program is participating. Using historical evidence from a situation where the opponent has a stake in the outcome provides high quality information about the opponent's true opening preferences.

REFERENCES

Aliston, W. & Weiskopf, D. (1984). The Complete Baseball Handbook, Allyn and Bacon, Boston.

Evans, L. (1970). Chess Catechism, Simon and Schuster, New York.

Horowitz A. (1973). The World Chess Championship A History, Macmillan, New York.

Mednis, E. (1990). Move Orders in the Opening: The Modern Master's Tool. Chess Life, 45(6), 14-16.

Nunn, J. & Griffiths, P. (1987). Secrets of Grandmaster Play, Macmillan, New York.

Reinfeld, F. (1987). The Complete Chess-Player, Prentice Hall, New York.

Riggins, J. & Winter, J. (1984). Gameplan The Language and Strategy of Pro Football, Santa Barbara Press, Santa Barbara, CA.

Schonberg, H. (1973). Grandmasters of Chess, J.B. Lippincott Co., Philadelphia.

Walczak, S. (1992). Pattern-Based Tactical Planning. International Journal of Pattern Recognition and Artificial Intelligence, 6(5), 955-988.

White, J. (1990). The Amateur's Book-Opening Routine. ICCA Journal, 13, 22-26.

APPENDIX

This appendix contains an explanation of the move notation used in the article as well as a glossary of terms.

Algebraic notation is used throughout this article. In algebraic notation, each column of the chess board is designated by a letter (a through h) and each row is designated by a number, with number 1 being the row on which the white king starts and row 8 the starting row of the black king. Hence, e2 is the space occupied by the pawn immediately in front of the white king. Moves are given by using a letter to designate the piece being moved (e.g., K for king and N for the knight) followed by the square the piece will occupy at the end of the move (e.g., Be2 indicates the bishop on the white diagonal will move to the square in front of the white king's starting square). If no piece letter is given, then the piece is assumed to be a pawn.

The following list of terms used in the text of the article refer to specific classical chess opening sequences (Reinfeld, 1987): Pelikan variation, Sicilian, Caro-Kann, Slav Defense, Ruy Lopez, Gruenfeld, King's Indian, Morphy, and Tchigorin.

Glossary of terms used in the article:

move distance — The difference between the current game turn and the mean game turn for a move. For example, the standard opening sequence for a player might be d4, c4, Nc3 (moves one to three respectively), but the player chooses to play the c4 move first. Since c4 is the expected move for the second game turn, a new move must be selected and both the d4 and Nc3 moves are normally played one game turn distance (game turns 1 and 3 respectively) from the current game turn (2), so each would have a distance of one.

transposition — A change in the standard order of a move sequence. For example, the first two moves of the Slav defense played by a black player are normally the moves d5 followed by c6. A transposition would be the moves c6 followed by d5, resulting in the same board position even though the move sequence has changed.