Context Dependencies in Features Used to Evaluate States

Donald Mitchell

Northwestern University

Problem solvers must consider the situational factors which influence the predictiveness of features used to make judgments. In an analysis of positions from Othello games, we found that the predictive validity of many features was dependent on the stage of the game and the skill of the players. This finding supports the use of context-sensitive weights for individual features in evaluation functions -- such as the application coefficients proposed by Ackley and Berliner (1983). In addition, our research provides a method for discovering these dependencies and for testing the general validity of features.

Ackley and Berliner (1983) define two important components of game playing programs: reasoning and judgment. In their words, "reasoning...is the process of imagining the environment to be other than it is (p.3)," and "judgment... is the process of forming an interpretation of the environment with respect to a goal (p.3)." In most programs, reasoning is the search algorithm and judgment is the evaluation function. Most game-playing research has focused on the reasoning process rather than the judgment process. Wilkins (1979) and Berliner (1983) are exceptions to this rule.

Psychological studies of expert-novice differences (deGroot, 1965, Simon and Chase, 1973) have found that experts and novices use similar search processes but different evaluation functions. Because of experience, experts recognize the best moves to examine and make more accurate evaluations of the outcomes.

Because most research has focused on the reasoning process, evaluation functions are usually developed on the basis of intuitions without formal analysis of reliability or validity. Most evaluation functions follow the same general format Shannon described in 1949: they are linear combinations of individual features. Ackley and Berliner (1983) detail some of the weaknesses of this simple approach. These weaknesses include the blemish effect which is an artifact of non-continuous functions, and the boundary effect which occurs at the extreme values of a function.

Wilkins' (1979) research on chess and Ackley and Berliner's (1983) on backgammon represent two significant attempts to analyze the judgment process. Our research involved the game of Othello which is the second most popular board game in Japan. The game is played on an 8 by 8 board of uniform color. The pieces are round disks which are white on one side and black on the other. Figure 1a shows the starting position. Black always starts the game. A legal move (see figure 1) consists of placing a disk on the board so that the disk captures at least one of the opponent's disks. To capture a disk, the moving player's new disk must sandwich one or more of the opponent's disks between it and at least one of the moving player's other disks without any empty squares between them. All disks so sandwiched are captured and are turned over to reveal the mover's color. If a player does not have a legal move, then it becomes the opponent's turn. The players alternate turns until neither player has a legal move. The winner is the player with the most pieces at the end of the game. Unlike chess, the final difference in material (the number of pieces for each player) contributes to player rankings.

For games such as Othello or chess, knowledge of the important features is not sufficient to make accurate judgments. The validity of some features depends on the value of other features or on the stage of the game (e.g. early, middle, and late). A good judgment process must be sensitive to these dependencies. Ackley and Berliner (1983) used the term application coefficient to refer to weighting functions which correct for these interactions between features and context. Our research provides empirical evidence that these weighting functions should be sensitive to not only the stage of game but also the skill of the players.

Procedure

There were four major steps in our study of feature importance: 1) the identification of relevant features; 2) the use of the features to evaluate positions from games between experts and games between novices; 3) the development of an "omniscient" evaluation for each position (the external criterion); and 4) the correlation of the individual features with the external criterion.

We collected 29 features from three Othello programs and from articles by Othello experts. The programs are Odin by Peter Frey, Iago by Paul Rosenbloom (1981), and Brand by Anders Kierulf (1982). Jonathan Cerf, who was the national and world champion Othello player, said of these programs, "In my opinion the top programs...are now equal (if not superior) to the best human players. (p.16, 1981)". Because of the skill of these programs, their features seemed appropriate to this research.

We applied the 29 evaluation features to positions from 135 expert games and 131 novice games. Three positions were used from each game: early (16 pieces on the board), middle (32 pieces), and late (48 pieces). After deleting duplicate positions, there were 401 expert and 393 novice positions.

We considered three sources for the independent, omniscient evaluation of each position: 1) human experts' opinions, 2) program evaluations based on a lookahead search of sufficient depth and knowledge that it could beat most experts (the best search being a complete end-game search), and 3) the actual outcome of each game. Because of difficulties in collecting experts' opinions, the first approach was not used. We did use the actual game outcome and two search estimates -- Odin's 8 ply search evaluation and a complete, end-game search applied to the late position from each game (16 ply).

Results

We correlated each feature with the external criterion. Any position in which a feature did not apply was excluded from the feature's correlation. If there were fewer than 100 positions included in a correlation, then the number is shown in parentheses on the figures. The correlations were derived separately for each game stage and skill level. Therefore, there were six correlations for each feature representing the two skill levels and three game periods.

Overall, for experts, most features' predictiveness increased as the game progressed, but for novices, the predictiveness remained constant or decreased. Comparing the two skill levels, most features were better predictors for novices' positions than for experts' positions. The predictiveness of all features was affected by the skill of the players or the stage of the game. Five examples of different effects are discussed below.

Figure 2a presents the correlations for a feature which becomes more predictive as the game progresses. This feature is the number of moves to the squares immediately adjacent to the corner. Because the corners are the most important squares on the board and because a square can only be captured if the opponent has a piece next to it, players usually avoid moves to the squares adjacent to corners. Accordingly, most programs negatively weight these moves. Contrary to expectation, our results data show that the ability (not necessarily the action) to capture these squares is positively correlated with success and that this relationship becomes stronger as the game progresses.

Figure 2b shows a feature whose correlation increases as the game progresses and is more positive for novices than for experts. The feature is the number of edge pieces which cannot be immediately captured. Edge pieces are considered important because they cannot be captured in as many directions as the rest of the pieces. Several Othello programs positively weight this feature. As figure 2b shows, this feature positively relates to position strength only for novices' late game positions. For experts' positions early in the game, the feature is a significant, negative predictor, but as the game progresses, the correlation increases to zero. For novices, the feature is a significant, negative predictor early in the game and crosses over the zero correlation to become a significant, positive predictor late in the game. Throughout the game, the correlation is more negative for experts than for novices.

Figure 2c shows the correlations for a feature which is more positively predictive for novices than for experts. The feature is the number of pieces which can never be captured. This feature is important because the goal in Othello is to have the most pieces at the end of the game. As figure 2c shows, the number of uncapturable pieces is more predictive in novices's games than in experts' games. In games between experts, there were no uncapturable pieces until late in the game and then the correlations were not as large as those for novices' games. It is interesting to note that when uncapturable pieces were present early in the novices' game, the player who had the uncapturable pieces won the game.

Figure 3a demonstrates an interaction between player skill and game period. The feature is the total number of pieces for each player. Most novices seem to believe that maximizing pieces is a good strategy. As figure 3a shows, the number of pieces is negatively correlated with the strength of a position. At the beginning of the game, the number of pieces is only predictive for novices. In the middle of the game, the number of pieces is predictive for both groups of players, and at the end of the game, the number of pieces is predictive only for for experts.

Figure 3b demonstrates a different type of interaction between skill level and stage of game. The feature is the number of pieces occupying the edge squares immediately adjacent to an empty corner. This feature differs from the feature in figure 2a in that this feature counts pieces rather than moves, ignores the squares around occupied corners, and ignores the square diagonally adjacent to the corners. The intuition behind this feature is that these squares provide a method of attack for the opponent to capture the corner. For novices, this feature is always important, but for experts its importance decreases sharply as the game progresses. An intuitive explanation for this difference between skill levels is that an expert knows when placing a piece on one of these squares is not dangerous but a novice does not.

Conclusion

Our data demonstrate that the skill of players and the stage of game influence the ability of features to evaluate positions. A plausible explanation for this effect is that the feature definitions do not adequately account for exceptions. One method to deal with these exceptions is to use the application coefficients proposed by Ackley and Berliner (1983). The correlations provide the data necessary to determine which features need application coefficients and how to construct the coefficients. A second approach is to make the feature definitions more sophisticated by adding rules to handle exceptions. This approach requires greater knowledge of the game.

Some of the features we analyzed tried to account for exceptions. For most of these features, the extra rules resulted in problems with boundary and blemish effects. In addition, because most of the added sophistication is based on intuitions, the simpler features were often better predictors than the complex ones. For example, in addition to the feature depicted in figure 3b, we had a feature which looked for over 30 types of patterns which might occur around a corner to estimate how likely it was that a player would capture the corner. This feature was less effective than the simpler count of occupied edge squares next to empty corners (figure 3b).

We are not trying to imply that features should not be made more sensitive to exceptions by adding sophistication, but to point out the difficulty of properly accounting for all the situational variables which can influence the validity of a feature. Using correlations provides the information needed to determine which features should be made more sophisticated and for testing the changes. Even if a feature cannot be made sensitive to exceptions, application coefficients can improve the average predictiveness of the feature.

References

Ackley, D. & Hans Berliner. The QBKG System: Knowledge Representation for Producing and Explaining Judgments, Carnegie-Mellon Univ., Department of Computer Science, CMU-CS-83-116, March, 1983.

Cerf, Jonathan. Machine vs. Machine, Othello Quarterly, 1981, 3(1), 12-16.

deGroot, Adriaan D. Thought and Choice in Chess, Mouton: The Hague, 1965.

Hasegawa, Goro. How to Win at Othello, Jove Publications: New York, 1977.

Kierulf, Anders. Brand 2.3 -- an Othello program, Informatik ETH Zurich, 1982.

Rosenbloom, P. A World-Championship-Level Othello Program, CMU-CS-81-137, Carnegie Mellon University, Department of Computer Science, August 1981.

Shannon, Claude. Programming a Computer for Playing Chess, Philosophical Magazine, 1950, 41, 256-275.

Simon, H. & W. Chase. Skill in Chess, American Scientist, 1973, 61(4), 394-403.

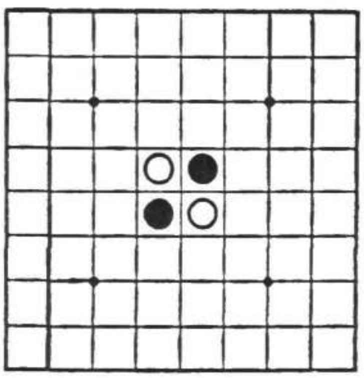
Wilkins, D. Using Patterns and Plans to Solve Problems and Control Search, Stanford Artificial Intelligence Laboratory Memo AIM-329, July 1979.

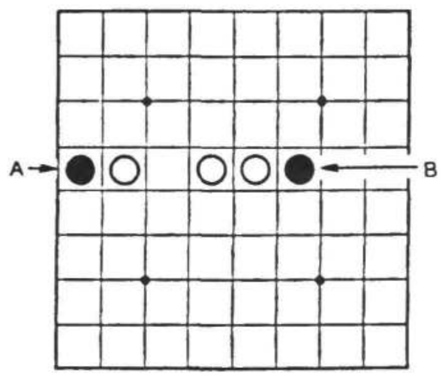
Figure Captions

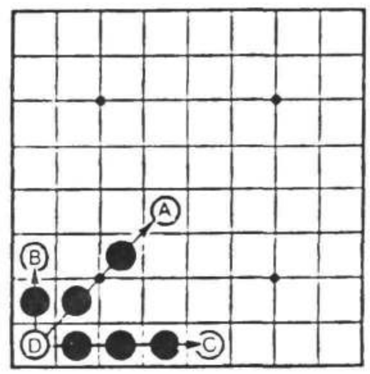
Figure 1. a) The starting position in Othello, b) an example of a legal move -- a move to D (the corner) would capture all the squares which the arrows pass through, c) an example of an illegal move -- black cannot move to A because there are empty squares between A and B. (Note: Pieces are left out of b and c to make the figures easier to read. These positions are not possible in a game. From, Hasegawa, 1977)

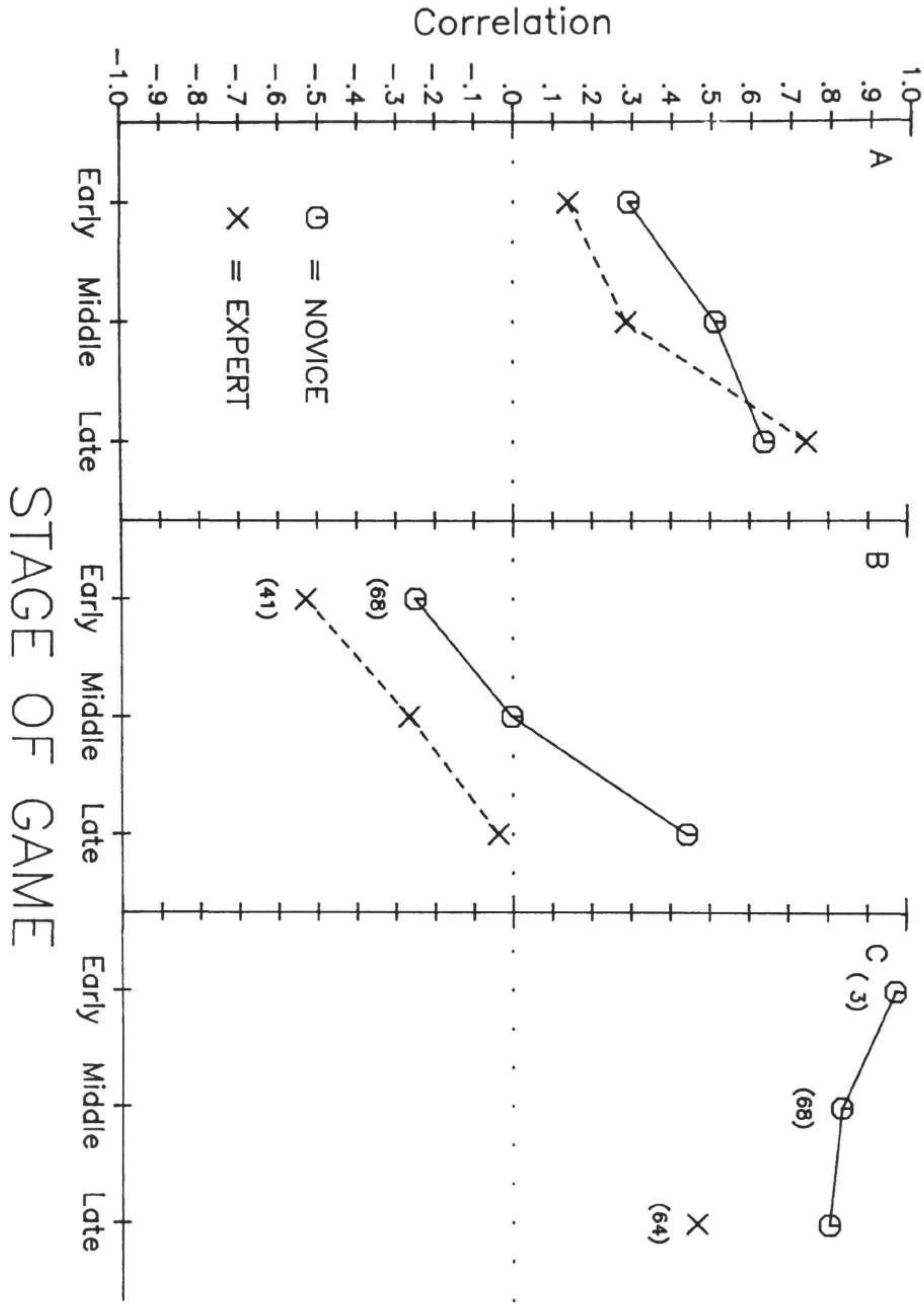
Figure 2. Correlation of evaluation features with an external measure of position strength. a) Stage of game effect -- number of legal moves to squares next to corners, b) Stage of game and skill level effect -- number of edge pieces, c) Skill level effect -- number of pieces which can never be captured.

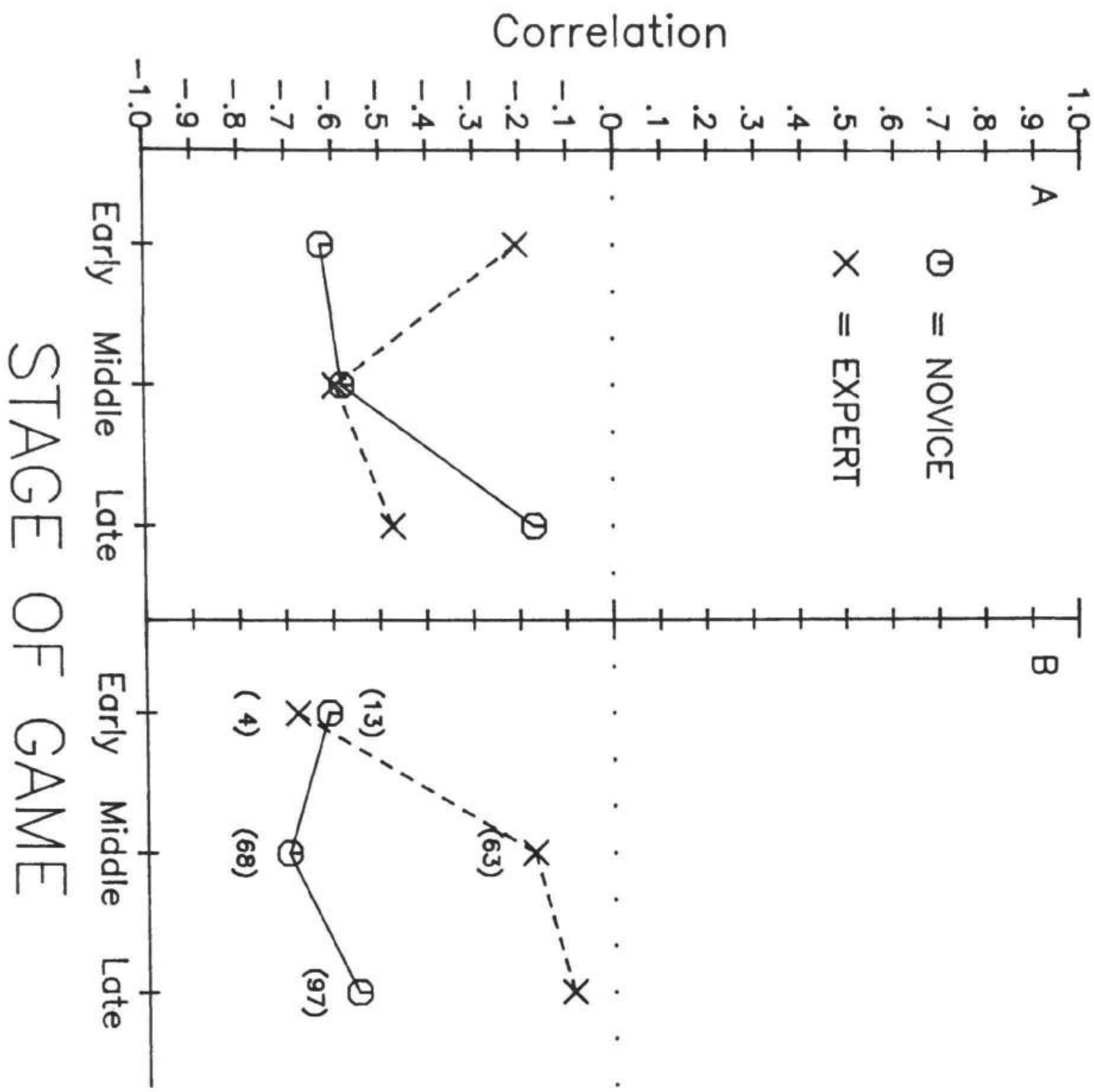
Figure 3. Correlations demonstrating interactions between skill level and stage of game. a) number of pieces for each player, b) number of pieces on edge squares next to empty corners.




A
B


A
B
D
C


Correlation
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-.8
-.9
-1.0
A
B
C
⊖ = NOVICE
X = EXPERT
(68)
(41)
(3)
(68)
(64)
Early
Middle
Late
STAGE OF GAME


Correlation
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
.0
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-.8
-.9
-1.0
A
⊖ = NOVICE
× = EXPERT
Early
Middle
Late
B
(13)
(4)
(63)
(68)
(97)
Early
Middle
Late
STAGE OF GAME