World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:4, No:12, 2010

Analysis of Sequence Moves in Successful Chess Openings Using Data Mining with Association Rules

R.M.Rani

***Abstract*—**Chess is one of the indoor games, which improves the level of human confidence, concentration, planning skills and knowledge. The main objective of this paper is to help the chess players to improve their chess openings using data mining techniques. Budding Chess Players usually do practices by analyzing various existing openings. When they analyze and correlate thousands of openings it becomes tedious and complex for them. The work done in this paper is to analyze the best lines of Blackmar-Diemer Gambit(BDG) which opens with White D4... using data mining analysis. It is carried out on the collection of winning games by applying association rules. The first step of this analysis is assigning variables to each different sequence moves. In the second step, the sequence association rules were generated to calculate support and confidence factor which help us to find the best subsequence chess moves that may lead to winning position.

Keywords—Blackmar-Diemer Gambit(BDG), Confidence, sequence Association Rules, Support.

I. INTRODUCTION

THE rapid development of computer technology, especially increases the memory capacities and decreases the costs of storage media has led businesses to store huge amounts of external and internal information in large databases at low cost. Mining useful information and helpful knowledge from these large databases has thus evolved into an important research area. Among them association rule mining has been one of the most popular data-mining subjects, which can be simply defined as finding interesting rules from large collections of data. Association rule mining has a wide range of applicability such as Market basket analysis, Medical diagnosis/ research, Website navigation analysis. Association rules are used to identify relationships among a set of items in a database.[3] Applying Association rules on chess database is one of the new research area . There may be millions of variations of different possible combinations in the existing chess database. Chess is a game of skill between two players. It is played on a chessboard of 64 squares(8*8) coloured alternately[12]. Chess theory is broadly divided into three major sections which are the openings, the middle-game, and the end-game respectively[11]. In this paper the Association rule is applied on chess database openings of Blackmar-Diemer Gambit or BDG which is a new application area. A gambit is an opening that involves a sacrifice of material such as pawn or piece, in order to achieve concrete advantages in the position[14]. Generally there are hundreds of various Gambits in chess openings. BDG is a sharp line for players who enjoy an attacking style of chess. The idea of this opening is to open F file for a swift and brutal attack.

This opening can be especially devastating weapon for blitz players because black can get taken by surprise and one wrong move will lead to a quick checkmate.

The BDG is a chess opening characterized by the moves1. d4 d5 2. e4 dxe4 3. Nc3, to be followed by f3 on White's fourth move. This gambit is considered an aggressive opening. It arose as a development of the earlier Blackmar Gambit, named after Armand Blackmar, was the first player to publish analysis on the opening in the chess literature[1].

Diemer was born in 1908 in the German town Radolfzell, in Baden. At a young age he was a passionate chess player, and Until 1956 his greatest success was a first place in the blitz championship of Baden. In his best period he could be considered a mediocre master. He was the prophet of relentless aggression in chess and he wrote "Play the Blackmar-Diemer gambit and mate will come by itself!"[2]. After many years of analysis, Diemer wrote a book on the opening in the late 1950s, titled 'Towards Mate From The First Move!'. Black can skip the BDG in several ways, the members of the BDG-community have developed related gambits.[2]

- 1.d4 Nf6 2.f3 d5 (c5 may lead to a kind of Benoni) 3.e4 dxe4 4.Nc3 simply transposes.
- 1.d4 Nf6 2.Nc3 d5 3.e4 Nxe4 is called the Hübsch Gambit.
- 1.d4 d5 2.e4 c6 (Caro-Kann) 3.Nc3 dxe4 4.f3 is invented by Philip Stuart Milner-Barry in 1932 and 4.Bc4 Nf6 (or Bf5) 5.f3 by Heinrich Von Hennig in 1920. So these sequences are older than Diemer's idea.
- 1.d4 d5 2.e4 e6 (French) 3.Be3 is the Alapin-Diemer Gambit; sometimes White plays the typical f2-f3 a bit later[9].
- 1.d4 d5 2.e4 e6 3.Nc3 dxe4 4.f3 and 3...Nf6 4.Bg5 dxe4 5.f3 are very rare.
- 1.d4 d5 2.e4 e6 3.Nc3 Bb4 4.a3 Bxc3+ 5.bxc3 dxe4 6.f3 is the Winckelmann-Reimer Gambit.
- d4 d5 2.e4 Nc6 (Nimzowitsch Defence) 3.Nc3 dxe4 4.d5 may be followed by 5.f3 or 5.f4.[10]

The work done in this paper is to analyze the frequently occurred successful sequence moves by considering the first 12-moves of Blackmar-Diemer Gambit BDG openings with the help of Data mining Techniques. By applying Association rules, the analyze done in this paper concludes which sequence moves be the best opening moves which may leads to winning position. The first step of this analysis is to collect the winning games of Blackmar-Diemer Gambit that opens with White D4 from the chess data base. Then the independent variables are assigned for each different moves of different games. The second step is to generate the sequence association rules to calculate support factor and confidence factor which helps us to find the best moves that may lead to winning position. The rest of this paper is organized as follows. In section (2) we give a formal definition of association rules and the problem definition. Section (3)shows the analysis of chess moves. Experimental results are shown in section (4) and Section (5) contains conclusions.

II. Association Rule Problem

Association rules is one of the data mining techniques which are used to discover interesting correlation in large database. Mostly the Association rules are applied to categorical data. The classical definition of Association rules is for a given a set of items $I = \{I_1, I_2, \ldots I_m\}$ and a database of transactions $D = \{t_1, t_2, t_3 \ldots t_n\}$ where $t_i = \{I_{i1}, I_{i2}, \ldots I_{i,k}\}$ and $I_{ij} \in I$, an association rule is an implication of the form $X \Rightarrow Y$ where $X, Y \subset I$ are sets of items called item sets and $X \cap Y = \phi$. The definition of Support(s) for an association rule $X \Rightarrow Y$ is the percentage of transaction in the database that contains $X \cup Y$. The confidence or strength (α) for an association rule $X \Rightarrow Y$ is the ratio of the number of transactions that contain $X \cup Y$ to the number of transactions that contain X[6],[7]. The two values support (s) and confidence(α) are used for selection of association rules. Actually the confidence measures the strength of the rule and whereas support measures how often it should occur in the database[3]. Let us consider the first 12-sequence moves of BDG winning 150 games are collected from the various chess websites[15],[16],[17],chess Academy and chess books. Upon analyzing the subsequence moves of BDG openings[18],[19], the subsequence moves can be defined as Let $O = O_1, O_2, \ldots O_m$ be a set of chess Openings and one sequence $M = \{m_1, m_2, \ldots,\}$ is subsequence moves of openings ,$1 \le j \le m$ openings such that $m_i \in O_j$. Let us consider that first 12 moves as sequence moves of openings denoted by $m_1, m_2, \ldots m_{12}$. The temporal features are included by adding move numbers as an attribute names and independent variables are assigned for each moves of BDG openings. Table I shows the first 12 moves numbers as an attribute names and its values are corresponding chess moves of 150 games with different variations

```
Algorithm(a)
Input
 D //Database consists of  n openings
Oj //Each openings in D as shown in Table I
X //First 5 moves of white opening BDG
Output
Generates the result as shown in the Table II

AlgInitialization
X← < d4 d5> < e4 dxe4>< Nc3 Nf6>< f3exf3>< Nxf3
g6> //First 5 moves
k=1;
For  each Oj ∈  D do
     Where  j  varies from 1 to n

Begin
     For each  moves mi ∈  Oj do
              Where i varies from 6 to 12
       read (Move-value);
    If  (Move-value) already represented by the
      sequence array ak, then locate the
      corresponding ak and assign it to  respective
      move number of the opening Oj.
    else
       Assign ak to the respective move number of the
       Opening Oj.
k=k+1;
End
```

Algorithm(a) assigns the first 5 moves to the variable 'X' which are common sequence to all games of BDG and AlgInitialization reads the value of remaining moves from sixth to twelfth of each games which are assigned to the independent variables. The output of this algorithm is shown in the Table II.

III. Analysis Of Chess Moves

Support and confidence are the two parameters used to set up association rules in the process of producing association rules. Given a set of openings for each opening, the support of a sequence s(S) is the percentage of total openings whose opening sequence contains S. The confidence(α) for a sequence association rule $X \Rightarrow Y$ is the ratio of the number of openings that contain both sequences X and Y to the number that contain X.

Given a minimum support threshold, a sequence is said to be large or frequent if its support exceeds this threshold. A large (frequent) sequence is a sequence whose number of occurrences is above a threshold, s. Algorithms such as Apriori, sampling are used to generate frequent subsequence moves from the database.[4],[5]. The Apriori algorithm searches for large itemsets during its initial database pass and uses it result as the seed for discovering other large data sets during subsequent passes.[8]

Applying Apriori algorithm all frequent sequence moves are found generating the association rule is straight forward. The following algorithm(b) generates efficient association rules satisfying support and confidence which help us to find the best subsequence moves.

Algorithm(b)
Input:
D //Database of BDG openings
I // frequent subsequence moves in L
L //set of frequent subsequence moves
s //Support
α //Confidence
Output
R //Association rules satisfying α and s
ARGen algorithm
$R = \phi$;
For each subsequence of moves $m_i \in M$ do
For each x is a subset of mi such that $x \neq \phi$ do
Begin
If support(m_i) ≥ threshold value then
R=RU{x==> (m_i-x)};
End

Let 'M' be a set of frequent subsequence moves and 'm_i' be a frequent subsequence moves in 'M' as given in Table III

TABLE III
FREQUENT SUBSEQUENCE MOVES

I_i	Frequent subsequence moves
I_1	a15,a16,a17
I_2	a30,a31,a32,a33,a34
I_3	a30,a31,a58
I_4	a30,a31,a58,a59
:	:
I_n	a30,a31,a268,a279,a280

Association rules are generated with support and confidence factor as shown in Table IV

TABLE IV
SUPPORT AND CONFIDENCE FOR SOME ASSOCIATION RULES

I_i	Association rules R	s(%)	α(%)
I_1	a15←a16,a17	2	100
I_2	a30←a31,a32,a33,a34	4	5
I_3	a30←a31,a58	26	29
I_4	a30←a31,a58,a59	8	9
:	:	:	:
I_n	a30←a31,a268,a279,a280	3	3

IV. EXPERIMENTAL RESULTS

According to the algorithm(b) and with threshold values, 3 subsequence moves, 4 subsequence moves, 5 subsequence moves are analyzed with the help of Association rules as shown in the Table VI.

TABLE V
THRESHOLD VALUES OF SUPPORT AND CONFIDENCE

Moves	Support threshold	Confidence threshold
3subsequence	26%	11%
4subsequence	5%	7%
5subsequence	5%	5%

Regarding the chess game it is very difficult to find that the same sequence occurs in more than one game. A few percentages of the subsequences are found by using association rules which helps the chess player to analyze without any difficulty. Table VI shows sequence moves that occurs frequently which has greater support and confidence than its threshold values.

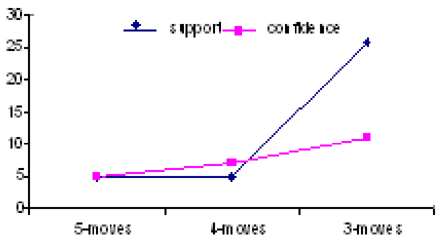

Fig.1 Comparison of Percentage of sequence moves

With reference to the Table VI and Figure 1, the result shows that the maximum percentage of support and confidence is for 3-sequence moves and it concludes that the following 8 moves may be the best successful line as a result of data mining analysis with 150 games of BDG.

1.d4 d5 2.e4 dxe4 3.Nc3 Nf6 4.f3 exf3 5.Nxf3 g6 6.Bc4 Bg7 7.o-o o-o 8.Qe1 Bf5.

It was concluded that after the eighth move, white side has an advantage because the number of pieces developed on white side is more and rook is in semi-opened File and Queen is in opened file so that the black pieces can be attacked easily.

V. CONCLUSIONS

In this paper, we have considered new emerging area by collecting the successful chess opening database for discovering the sequence moves of Blackmar -Diemer Gambit(BDG). Association rule is one of the data mining technique which has been applied for finding best sequence moves and the result was shown in the Table V and Figure 1. It concludes that out of 150 different BDG games ,26% of white winning games have followed the 3-sub sequence moves (6.Bc4 Bg7 7.o-o o-o 8.Qe1 Bf5) and won the game. The future enhancements of this paper is that the same data mining technique can be applied for different gambits to discover various subsequence moves that enable the chess player to analyze the game.

ACKNOWLEDGEMENT

The author would like to thank Mr.R.M.Ramesh chess enthusiast and King Chess Foundation (Chess Academy) for providing chess data set used in this article. The author would also like to thank Mr.T.Jagadeesan and Mr.U.Jayavelan chess coaches of King Chess Foundation for sharing their expertise.

REFERENCES

[1] http://www.Blackmar-Diemer Gambit - Wikipedia, the free encyclopedia.htm
[2] http://www.chlodwig.com/Partien/Diemer/Diemer_Story.htm
[3] “DATA MINING-Introductory and Advanced Topics”, Margaret H. Dunham,pp 164-172
[4] J.Li et al., “Mining Risk Patterns in Medical Data”, in Proc. Of the eleventh ACM SIKDD international conference on knowledge discovery in data mining,pp 770-775,2005
[5] M.J.Zaki,”Mining non redundant association rules”. Data mining and Knowledge discovery Journal,pp,:223-248,2004
[6] J.Han and Kamber, Data mining, Concepts and Techniques, Morgan Kaufmann Publishers,pp 498-500,2006
[7] I.H Witten, E. Frank, Data Mining-Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers,2005
[8] M.Karolis, C.S. Pattichis, Association Rule analysis for the assesment of the risk of coronary Heart events, 31st international conference of the IEEE EMBS,September 2-6,2009,.
[9] “Chess Openings:Traps and Zaps “,American Writer, Pandolfini.
[10] “Nimzo Indian Defence Theory and Games”,K.R.Seshadri, Internal Arbiter.
[11] “Begin Chess”, Manuel Aaron,pp14-15
[12] “Begin Chess”, David B.Pritchard,pp 9-10
[13] “Your First Move”, Chess for Beginners,Alexei Sokolsky.pp 42.
[14] “Gambit Champion openings”, Eric Schiller, International Arbiter.
[15] http://www.bookuppro.com/ecopgn/d.html
[16] http://www.chessopeningsdatabase.com/3/d4,d5,/Chess-Openings-Database.htm
[17] http://www.chessvideos.tv/chess-opening-database/
[18] http://www.chessvideos.tv/chess-opening-database/search/Blackmar-Diemer-gambit
[19] http://www.chess.com/article/view/openings-for-tactical-players-blackmar-diemer-gambit

TABLE I
FIRST 12 SUBSEQUENCE MOVES OF BDG

Openings No.	I move	II move	III move	IV move	V move	VI move	VII move	VIII move	IX move	X move	XI move	XI move
1	d4 d5	e4 dxe4	Nc3 Nf6	f3exf3	Nxf3 g6	Bc4Be6	Bxe6fxe6	Qd3Nc6	Qb5 Rb8	Ne5Qd6	Bf4 a6	Nxc6
2	d4 d5	e4 dxe4	Nc3Nf6	f3exf3	Nxf3 g6	Bc4Bf5	Ne5 e6	Nxf7Kxf7	g4Be4	Nxe4Nxe4	o-o+Ke7	Re1
3	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	Bc4Bg4	Bxf7+Kxf7	Ne5+Ke8	Nxg4Nxg4	Qxg4Bg7	Be3Nc6	o-o-o-
4	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	Bc4Bg4	Bxf7Kxf7	Ne5+Kg7	Nxg4Nc6	d5 Nb4	o-o Nxg4	Qxg4
:	:	:		:	:	:	:	:	:	:	:	:
150	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	Bc4 e6	Bg5Be7	o-o Nfd7	Qd2 Nf6	Ne5 Rf8	Bh6 Rg8	Qf2

TABLE II
CODE OF CHESS SEQUENCE MOVES OF BDG

Openings No.	X←I,II ,III ,IV, V moves					VI move	VII move	VIII move	IX move	X move	XI move	XI move
1	d4 d5	e4 dxe4	Nc3 Nf6	f3exf3	Nxf3 g6	a1	a2	a3	a4	a5	a6	a7
2	d4 d5	e4 dxe4	Nc3Nf6	f3exf3	Nxf3 g6	a8	a9	a10	a11	a12	a13	a14
3	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	a15	a16	a17	a18	a19	a20	a21
4	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	a15	a16	a17	a22	a23	a24	a25
5	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	a15	a16	a17	a26	a27	a28	a29
6	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	a30	a31	a32	a33	a34	a35	a36
:	:	:		:	:	:	:	:	:	:	:	:
148	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	a30	a31	a268	a279	a285	a286	a96
149	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	a30	a31	a268	a279	a287	a288	a289
150	d4 d5	e4 dxe4	Nc3 Nf6	f3 exf3	Nxf3 g6	a30	a31	a268	a290	a291	a292	a293

TABLE VI
BEST SUBSEQUENCE MOVES OF BDG

Moves	Sub sequence moves	Chess sequence moves	s(%)	α(%)
3- subsequence	X←a30,a31,a58	X←Bc4 Bg7,o-o o-o,Qe1 Bf5	26	29
4- subsequence	X←a30,a31,a58,a59	X← Bc4 Bg7,o-o o-o, Qe1Bf5,Qh4 Bxc2	8	9
5- subsequence	X←a30,a31,a268,a279,a280	X← Bc4 Bg7,o-o o-o,Qe1Nc6,Qh4Bg4,Be3Bxf3	5	5