Quantitative Study of Markov Model for Prediction of User Behavior for Web Caching and Prefetching Purpose

Dharmendra T. Patel
Asst.Professor,Smt.Chandaben Mohanbhai Patel Institute of Computer Applicatios, CHARUSAT, Changa, Gujarat

Kalpesh Parikh, PhD.
Director,
Intellisense IT, Ahmedabad, Gujarat, INDIA

ABSTRACT
In modern era every organization depends on internet to conduct business and as a result of that many hidden data are available in several log files of servers; which could serve many purposes in business and that give the birth of web mining field. Web Mining could useful for many applications in business but this paper focuses on web caching and prefetching application to reduce latency while accessing internet. The common problem in organization is; in spite of sufficient internet bandwidth; sometimes users feel delay while accessing several pages. The problem could be solved out by developing predictive model based on web caching and prefetching criteria and many research have been done using Markov based predictive model to reduce access latency while using internet. This paper focuses on quantitative study of Markov based predictive model for web caching and prefetching to determine limitations of Markov Model on prediction perspectives.

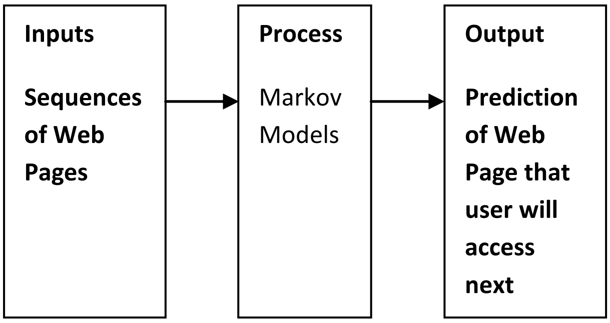
General Terms
Web Mining

Keywords
Markov Model, Web Mining, Web Caching, Web Prefetching, Access Latency

1. INTRODUCTION
World Wide Web has emerged as an information hub in current era and modern business is depended on internet so all most all organization using internet for dissemination of information and for e-commerce related transactions. Huge amount of web data has invented new research area and that is known as Web Mining [1]. Web Mining is one kind of data mining [3] technique applies on web data. Web mining research [2][4] [5] applies to many applications areas and those are very helpful for business perspectives. This paper focuses on web caching [6][7][8] and prefetching[9][10] application area of Web Mining.

In computer field, cache is well known term that is used to stores the data that can be retrieved faster when ever required. The same concept can be used for web data to predict the web page that is frequently accessed by the web user and this terminology is termed as web caching. There are numbers of researches [16][17][18][19] have been done to improve web caching mechanism but still the cache hit ratio is not improved much because web caching is only provides temporal locality[20] and to understand the behavior of web user spatial locality is required and that gave the birth of new concept and that is web prefetching. There are two main categories of web prefetching algorithms (1) Content Based and (2) History based. The scope of this paper is history based algorithms. Several authors [21][22][23][24] have given their contribution in history based web prefetching approaches. The history based algorithms predicts user behavior based on the access history of user available in several server log files [11] [12]. One popular predictive model based on history based algorithms is Markov Model [13][14][15]. This paper studies Markov Model in quantitative perspectives and identifies several critical limitations of it in prediction application point of view. Section-2 of paper deals with study of Markov Model as a predictive Model. Seciton-3 will be focus on Markov Model and Pattern Generation Process. Section-4 deals with results. Section-5 will provide conclusion of the paper.

2. MARKOV MODEL AS A PREDICTIVE MODEL
Markov model is well known stochastic model that assumes certain properties that is known as markov properties. Due to these assumptions computations are possible with the model otherwise they are complex to implement practically. The scope of Markov Model is not limited to mathematical applications but it is also applicable to computer science applications particularly in web mining research. In web mining, Markov Model has given its valuable contribution as a predictive model. Markov model is well suited for modeling and predicting a user's browsing behavior from the data available on several server log files. The input requires for predicting user browsing behavior is sequences of web pages and output is the web page that user will likely access next. Figure-1 describes input-process-output of Markov Prediction Model.



(Fig.1 Input-Process-Output of Markov Model)

Several properties of Markov Chain Process must be satisfied before any problem can be classified into it. Following are main properties of Markov Model.

(1) There are finite numbers of possible states in problem.
(2) All States of problem are both collectively exhaustive (at least one of the events must occur) and mutual exclusive (cannot occur at the same time).
(3) The transition probability depends only on the current state of system.
(4) The long run probability of being in particular state will be constant over time.
(5) The transition probability of moving to alternative states in next time period, given a state in the current time period must sum to 1.

Markov Model are represented by three main parameters < **Actions, States, Transition >**

Where **Actions =** Set of all possible actions that can be performed by the user

States= Set of all possible states for which Markov Model built

Transition = Transition probability Matrix (TPM) | States | * | Actions | where each entry **Tij** corresponds to the probability of performing the action **j** when process in state **i**.

Based on transition probability matrix prediction of next state is done from initial state. Markov Models are available in from of different chains. In First Order Markov Chain prediction of next action is done by looking at only last action performed by user. In Second Order Markov Chain prediction of next action is done by looking at last two actions performed by user. Similar way in next order chain number of last actions can be increased. Markov model is best suited for applications of prediction so most of all research related to prediction uses Markov Model as a prediction model.

3. MARKOV MODEL AND PATTERN GENERATION PROCESS

Generally the framework of any prediction model consists of three main steps:

(1) Cleaning Process: - This step removes unnecessary records from log file depends on application perspective.

(2) User and Session Identification Process: - Based on cleaned data, this process identifies unique user and associated session information.

(3) Pattern Generation Process: - Based on Unique user and sessions information, this phase generates patterns by which behavior of users can be predicted.

In Markov Model, pattern generation process is based on session identification process. There are numbers of sessionization strategies [25] and sessions are formed based on those strategies. Generally combination of IP address, agent and sessionization heuristics technique is used to determine sessions from raw log file because no additional technology is required to determine sessions. Once sessions are formed; that is inputted in Markov Model to generate pattern as an output. Pattern is generated based on category of Markov Chains. Markov chain can be described as a set of states where state set consists of different possible states of an application like s1,s2,s3……sn. The process of pattern generation will start from one of these states and moves successively from one state to another with a probability denoted by ay probability let us say pij; and this probability does not depends on previous state of current state known as a transition probability. Thus the pattern generation process using Markov Chain depends on transition probability. The length of sequence of pages generated as an output depends on specific category of Markov Chain. In First Order Markov Chain prediction of next page is done by looking at only last action performed by a user. In the following order of Markov Chains; the number of last actions can be increased to predict sequence of web pages.

To generate patterns using Markov Model; following steps are to be performed irrespective of the category of Markov Chain.

(1) Raw Log file is cleaned by removing unnecessary entries from it based on application.

(2) From pre-processed raw log file; generate sessions based on any suitable sessionization strategy.

(3) From number of sessions; generate occurrence matrix which determines occurrence of particular page from current set of pages.

(4) Generate transition probability matrix based on current state.

(5) From transition matrix identify maximum probability of each raw to determine prediction of next page.

4. EXPERIMENTAL RESULTS

Assume that following numbers of sessions (refer Table-1) are identified from raw log file. Analysis of number of sessions indicates that there are 15 unique states of identified sessions.

To generate patterns from above table-1, one simulation tool is designed in Microsoft Excel. Following are results of that simulation tool based on first and second order Markov

Chains. Table-2 describes the occurrence matrix of first order Markov Chain. Table-3 contains the values of transition probability matrix of first order Markov Chain. Similar way Table-4 and 5 are dedicated to second order Markov Chain. Figure-2 describes the space complexity of different orders Markov Chains.

The experimental results indicates that the space complexity from one order markov chain to next order markov chain exceed a lot to predict next page. In first order markov chain only 15 * 15 matrix is required to predict next page while in second order markov chain 74 * 15 matrix is required means 5 times space complexity is increased. Another thing is coverage may reduce from one order Markov Model to Next Order. Second Order Markov Chain transition probability matrix contains many rows with all zeroes.

Session Number	**Sequence of Web Pages**
1	p1,p2,p5,p7,p8,p9,p10
2	p5,p6,p8,p9,p12,p15,p2,p5
3	p2,p3,p4,p5,p6,p7,p9,p10,p11,p12,p15,p14,p13
4	p1,p2,p4,p6,p8,p9,p10,p12,p14,p15,p3,p9,p8,p6
5	p1,p5,p7,p9,p11,p12,p13,p14,p15,p2,p3,p14
6	p2,p3,p8,p7,p9,p4,p6,p10,p11,p12,p13,p15
7	p1,p2,p3,p4,p6,p9,p11,p12,p14,p15,p8
8	p5,p7,p6,p5,p2,p1,p5,p6,p9,p10,p12,p14,p11,p10,p9
9	p3,p6,p4,p5,p6,p7,p9,p10,p11,p12,p15,p14,p13,p11,p10
10	p1,p2,p5,p7,p8,p9,p10,p12,p13,p14,p15,p10
11	p1,p2,p4,p6,p8,p9,p10,p12,p14,p15,p3,p9,p8,p6,p8,p9,p10
12	p1,p3,p4,p5,p8,p9,p10,p12,p11,p15,p3,p9,p8,p6,p10
13	p1,p3,p6,p9,p11,p12,p13,p14,p15,p2,p3,p14,p10,p12
14	p5,p6,p8,p9,p12,p15,p2,p5, p1,p2,p5,p7,p8,p9,p10
15	p1,p2,p4,p6,p8,p9,p10,p12,p14,p15,p3,p2,p1
16	p1,p5,p6,p2,p3,p8,p7,p9,p4,p6,p10,p11,p12,p13,p15
17	p1,p2,p4,p6,p8,p9,p10,p12,p14,p15,p3,p9,p8,p6,p7,p10
18	p7,p8,p9,p10,p2,p3,p4,p5,p6,p7,p9,p10,p11,p12,p15,p14,p13
19	p1,p5,p7,p9,p11,p12,p13,p14,p15,p2,p3,p14,p15,p8,p6
20	p5,p7,p6,p5,p2,p1,p5,p6,p9,p10,p12,p14,p11,p10,p9,p13,p5
21	p4,p3,p2,p1,p9,p10,p12,p6
22	p1,p2,p14,p15,p4,p5,p6,p8
23	p2,p3,p4,p5,p6,p8,p1,p11,p12
24	p1,p2,p4,p8,p9,p12,p13,p14,p1,p5,p4
25	p6,p8,p2,p1,p3,p4,p5,p7,p8,p9,p10,p11,p12,p13

(Table-1 Sessions Table)

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**
P1	0	10	3	0	6	0	0	0	1	0	1	0	0	0	0
P2	5	0	9	5	5	0	0	0	0	0	0	0	0	1	0
P3	0	2	0	6	0	2	0	2	4	0	0	0	0	3	0
P4	0	0	1	0	7	7	0	1	0	0	0	0	0	0	0
P5	1	2	0	1	0	10	8	1	0	0	0	0	0	0	0
P6	0	1	0	1	2	0	4	10	4	3	0	0	0	0	0
P7	0	0	0	0	0	2	0	5	7	1	0	0	0	0	0
P8	1	1	0	0	0	5	2	0	14	0	0	0	0	0	0
P9	0	0	0	2	0	0	0	4	0	17	4	3	1	0	0
P10	0	1	0	0	0	0	0	0	2	0	6	10	0	0	0
P11	0	0	0	0	0	0	0	0	0	3	0	11	0	0	1
P12	0	0	0	0	0	1	0	0	0	0	1	0	8	7	5
P13	0	0	0	0	1	0	0	0	0	0	1	0	0	5	2
P14	1	0	0	0	0	0	0	0	0	1	2	0	3	0	11
P15	0	5	5	1	0	0	0	2	0	1	0	0	0	3	0

(**Table-2 Occurrence Matrix of First Order Markov Chain**)

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**
P1	0	0.47619	0.142857	0	0.285714	0	0	0	0.047619	0	0.047619	0	0	0	0
P2	0.2	0	0.36	0.2	0.2	0	0	0	0	0	0	0	0	0.04	0
P3	0	0.105263	0	0.315789	0	0.105263	0	0.105263	0.210526	0	0	0	0	0.157895	0
P4	0	0	0.0625	0	0.4375	0.4375	0	0.0625	0	0	0	0	0	0	0
P5	0.043478	0.086957	0	0.043478	0	0.434783	0.347826	0.043478	0	0	0	0	0	0	0
P6	0	0.04	0	0.04	0.08	0	0.16	0.4	0.16	0.12	0	0	0	0	0
P7	0	0	0	0	0	0.133333	0	0.333333	0.466667	0.066667	0	0	0	0	0
P8	0.043478	0.043478	0	0	0	0.217391	0.086957	0	0.608696	0	0	0	0	0	0
P9	0	0	0	0.064516	0	0	0	0.129032	0	0.548387	0.129032	0.096774	0.032258	0	0
P10	0	0.052632	0	0	0	0	0	0	0.105263	0	0.315789	0.526316	0	0	0

P11	0	0	0	0	0	0	0	0	0	0.2	0	0.733333	0	0	0.066667
P12	0	0	0	0	0	0.045455	0	0	0	0	0.045455	0	0.363636	0.318182	0.227273
P13	0	0	0	0	0.111111	0	0	0	0	0	0.111111	0	0	0.555556	0.222222
P14	0.055556	0	0	0	0	0	0	0	0	0.055556	0.111111	0	0.166667	0	0.611111
P15	0	0.294118	0.294118	0.058824	0	0	0	0.117647	0	0.058824	0	0	0	0.176471	0

(**Table-3 Transition Probability Matrix of First Order Markov Chain**)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
P1,P2	0	0	1	5	3	0	0	0	0	0	0	0	0	1	0
P1,P3	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0
P1,P5	0	0	0	1	0	3	2	0	0	0	0	0	0	0	0
P1,P9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
P1,P11	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P2,P1	0	0	1	0	2	0	0	0	1	0	0	0	0	0	0
P2,P3	0	0	0	4	0	0	0	2	0	0	0	0	0	3	0
P2,P4	0	0	0	0	0	4	0	1	0	0	0	0	0	0	0
P2,P5	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0
P2,P14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
P3,P2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P3,P4	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0
P3,P6	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
P3,P8	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
P3,P9	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
P3,P14	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
P4,P3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
P4,P5	0	0	0	0	0	1	1	5	0	0	0	0	0	0	0
P4,P6	0	0	0	0	0	0	0	4	1	2	0	0	0	0	0
P4,P8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P5,P1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

P5,P2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P5,P4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P5,P6	0	1	0	0	0	0	2	4	1	0	0	0	0	0	0
P5,P7	0	0	0	0	0	2	0	2	2	0	0	0	0	0	0
P5,P8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P6,P2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P6,P4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P6,P5	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
P6.P7	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0
P6,P8	1	1	0	0	0	0	0	0	7	0	0	0	0	0	0
P6,P9	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0
P6,P10	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
P7,P6	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
P7,P8	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
P7,P9	0	0	0	2	0	0	0	0	0	3	2	0	0	0	0
P7,P10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P8,P1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
P8,P2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P8,P6	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0
P8,P7	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
P8,P9	0	0	0	0	0	0	0	0	0	10	0	3	0	0	0
P9,P4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
P9,P8	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
P9,P10	0	1	0	0	0	0	0	0	0	0	4	9	0	0	0
P9,P11	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
P9,P12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2
P9,P13	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P10,P2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P10,P9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
P10,P11	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0

P10,12	0	0	0	0	0	0	0	0	0	0	1	0	1	6	0
P11,P10	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
P11,P12	0	0	0	0	0	0	0	0	0	0	0	0	6	1	3
P11,P15	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P12,P6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P12,P11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
P12,P13	0	0	0	0	0	0	0	0	0	0	0	0	0	4	2
P12,P14	0	0	0	0	0	0	0	0	0	0	2		0	0	4
P12,P15	0	2	0	0	0	0	0	0	0	0	0	0	0	3	0
P13,P5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P13,P11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
P13,P14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
P13,P15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P14,P1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P14,P10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P14,P11	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
P14,P13	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
P14,P15	0	3	4	0	0	0	0	2	0	1	0	0	0	0	0
P15,P2	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0
P15,P3	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0
P15,P4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P15,P8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
P15,P14	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0

(Table-4 Occurrence Matrix of Second Order Markov Chain)

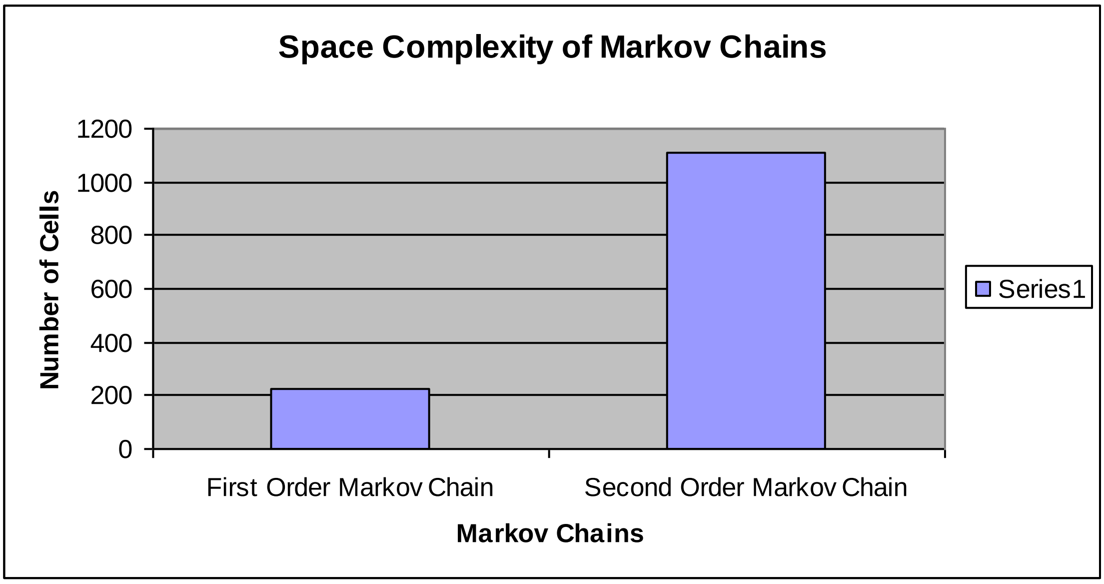
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
P1,P2	0	0	0.1	1	0	0	0	0	0	0	0	0	0	0.1	0
P1,P3	0	0	0	1	0	0.33	0	0	0	0	0	0	0	0	0
P1,P5	0	0	0	0	0	0.5	0.33	0	0	0	0	0	0	0	0
P1,P9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
P1,P11	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P2,P1	0	0	0.3	0	1	0	0	0	0.3	0	0	0	0	0	0

P2,P3	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0
P2,P4	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0
P2,P5	0.3	0	0	0	0	0	0.75	0	0	0	0	0	0	0	0
P2,P14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
P3,P2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P3,P4	0	0	0	0	1	0.17	0	0	0	0	0	0	0	0	0
P3,P6	0	0	0	1	0	0	0	0	0.5	0	0	0	0	0	0
P3,P8	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
P3,P9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
P3,P14	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0.5
P4,P3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
P4,P5	0	0	0	0	0	0.14	0.14	1	0	0	0	0	0	0	0
P4,P6	0	0	0	0	0	0	0	1	0.1	0.3	0	0	0	0	0
P4,P8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P5,P1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
P5,P2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P5,P4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P5,P6	0	0.1	0	0	0	0	0.25	1	0.1	0	0	0	0	0	0
P5,P7	0	0	0	0	0	0.33	0	0	0.3	0	0	0	0	0	0
P5,P8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P6,P2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P6,P4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P6,P5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
P6.P7	0	0	0	0	0	0	0	0	0.8	0.3	0	0	0	0	0
P6,P8	0.1	0.1	0	0	0	0	0	0	0.8	0	0	0	0	0	0
P6,P9	0	0	0	0	0	0	0	0	0	0.5	0.5	0	0	0	0
P6,P10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
P7,P6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P7,P8	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P7,P9	0	0	0	0	0	0	0	0	0	0.4	0.3	0	0	0	0
P7,P10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P8,P1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

P8,P2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P8,P6	0	0	0	0	0	0	0.33	0	0	0.3	0	0	0	0	0
P8,P7	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P8,P9	0	0	0	0	0	0	0	0	0	0.8	0	0.23	0	0	0
P9,P4	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
P9,P8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
P9,P10	0	0.1	0	0	0	0	0	0	0	0	0.3	0.64	0	0	0
P9,P11	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P9,P12	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0.7
P9,P13	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P10,P2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P10,P9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
P10,P11	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P10,12	0	0	0	0	0	0	0	0	0	0	0.1	0	0.1	0.8	0
P11,P10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P11,P12	0	0	0	0	0	0	0	0	0	0	0	0	0.6	0.1	0.3
P11,P15	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
P12,P6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P12,P11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
P12,P13	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7	0.3
P12,P14	0	0	0	0	0	0	0	0	0	0	0.3	0	0	0	0.7
P12,P15	0	0.4	0	0	0	0	0	0	0	0	0	0	0	0.6	0
P13,P5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P13,P11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
P13,P14	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0.7
P13,P15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P14,P1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P14,P10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P14,P11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
P14,P13	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
P14,P15	0	0.3	0.4	0	0	0	0	0	0	0.1	0	0	0	0	0
P15,P2	0	0	0.5	0	1	0	0	0	0	0	0	0	0	0	0

P15,P3	0	0.3	0	0	0	0	0	0	0.8	0	0	0	0	0	0
P15,P4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P15,P8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
P15,P14	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

(Table -5 Transition Probability Matrix of Second Order Markov Chain)



(Fig. 2 Space Complexity Chart)

5. CONCLUSION

Markov Model is a well known model for web page prediction in field of web prefetching but the quantitative study of it derives many limitations of it. The quantitative study analyzes many issues of Markov Model in terms of predictions. Following are the main problems with Markov Model.

1. State Complexity of Markov Model increases in every order. The results of study derives facts that in first order markov chain only 15 * 15 means 225 cells space is required to process 15 unique states while second oder markov chain requires 74 * 15 means 1110 cells space to predict next page means space complexity increases approximately 5 times.

2. Prediction of entire sequence is very difficult tasks because higher order markov chain is required.

3. Some times probability of number of pages may equal so creates ambiguity in prediction of next page.

4. Coverage may reduce from one order Markov Model to Next Order. Second Order Markov Chain transition probability matrix contains many rows with all zeroes.

Thus the quantitative study indicates that Markov Model is not fitted as a prediction model in web prefetching. For prediction some data mining techniques should be considered for effective and simple way of prediction.

6. REFERENCES

[1] Cooley, R; Mobasher, B. ; Srivastava, J, "Web mining: information and pattern discovery on the World Wide Web " , Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference, Pages-558-567.

[2] Ramakrishna, M.T. ; Gowdar, L.K. ; Havanur, M.S. ; Swamy, B.P.M, "Web Mining: Key Accomplishments, Applications and Future Directions" Data Storage and Data Engineering (DSDE), 2010 International IEEE Conference, Pages- 187 – 191.

[3] Ming-Syan Chen; Jiawei Han; Philip S. Yu, " Data Mining : An Overview from a Database Perspectives", IEE Transactions on knowledge and data engineering, Vol-8, December-1996, Pages-866-883.

[4] Zhang Haiyang, "The Research of Web Mining in E-Commerce", Management and Service Science (MASS), 2011 IEE International Conference, Pages-1-4.

[5] Raymand Kosala; Hendril Blokeel, " Web Mining Research : A Survey", ACM SIGKDD Explorations, July 2000, Vol-2, Issue-1, Pages-1-4.

[6] H.T. Chen, "*Pre-fetching and Re-fetching in Web caching systems: Algorithms and Simulation"* Master Thesis, TRENT UNIVESITY, Peterborough, Ontario, Canada(2008)

[7] T.Chen, "Obtaining the optimal cache document replacement policy for the caching system of an EC Website", *European Journal of Operational Research*.181 (2),(2007), pp. 828. Amsterdam.

[8] U. Acharjee, *Personalized and Artificial Intelligence Web Caching and Prefetching.* Master thesis, University of Ottawa,Canada(2006).

[9] B. Zhijie, G. Zhimin, and J. Yu, "A Survey of Web Prefetching", *Journal of computer research and development*, 46(2), (2009), pp. 202-210.

[10] J. Domenech, J. A. Gil, J. Sahuquillo, and A. Pont, "Using current web page structure to improve prefetching performance", *Computer Network Journal*, 54(9), (2010), 1404-1417.

[11] Technet Library,Microsoft Products,Tools, Technologies (www.technet.microsoft.com)

[12] K.R.Suneetha; Dr. R.Krishnamoorthi, " Identifying User Behavior by Analyzing Web Server Access Log File", IJCSNS International Journal of computer science and Network Security, Vol.9,April 2009, pages-327-332.

[13] Nigam, B., "Analysis of Markov model on different web Prefetching and caching schemes", Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference, Pages-1-6.

[14] Lei Shi; Yan Zhang ; Wei Lin, " Optimal Model of Web Caching and Prefetching", Proceeding of second Symposium International computer science and computational technology, China, Dec-2009,Pages-250-253.

[15] Wei-Guang Teng;Cheng-Yue Chang,;and Ming-Syan Chen, "Integrating Web Caching and Web Prefetching in Client-Side Proxies", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 16, NO. 5, MAY 2005, Pages-444-455.

[16] W. Ali, and S.M. Shamsuddin, "Intelligent Client-side Web Caching Scheme Based on Least recently Used Algorithm and Neuro-Fuzzy System", *The sixth International Symposium on Neural Networks(ISNN 2009), Lecture Notes in Computer Science (LNCS)*, Springer-Verlag Berlin Heidelberg , 5552, (2009), pp. 70–79.

[17] W. Tian, B. Choi, and V.V. Phoha,"An Adaptive Web Cache Access Predictor Using Neural Network". *Proceedings of the 15th international conference on Industrial and engineering applications of artificial intelligence and expert systems: developments in applied artificial intelligence, Lecture Notes In Computer Science(LNCS)*, Springer- Verlag London, UK 2358, (2002).450-459.

[18] A. P. Foong, Y.-H. Hu, and D. M. Heisey, "Logistic regression in an adaptive web cache", *IEEE Internet Computing*, 3, (1999), 27-36.

[19] H. ElAarag and S. Romano, "Improvement of the neural network proxy cache replacement strategy", *Proceedings of the 2009 Spring Simulation Multiconference,(SSM'09)*, San Diego, California, (2009), pp: 90.

[20] Younghyun Kim ; Sangheon Pack ; Chung Gu Kang ; Soonjoon Park , "Exploiting spatial and temporal locality for seamless vertical handover" ,IEEE Communications and Information Technology, ISCIT 2009, Pages-1078 – 1083.

[21] X. Chen and X. Zhang, "Popularity-based PPM: An effective web prefetching technique for high accuracy and low storage*", In Proceedings of the International Conference on Parallel Processing*, (2002), pp. 296-304.

[22] Ali Bayir ,Smart Miner: A New Framework for Mining Large Scale Web Usage Data, Murat,Department of Computer,Science and Engineering,University at Buffalo, USA.

[23] E. P. Markatos and C. E. Chronaki , "A Top-10 approach to prefetching on the Web ", *Proceedings of INET'98 Geneva, Switzerland*, (1998), pp. 276-290.

[24] Y. Jiang, M.Y Wu, and W. Shu, "Web prefetching : Costs , benefits and performance", *Proceedings of the 11th International World Wide Web Conference, New York, ACM*, (2002).

[25] Dharmendra Patel, Dr. Kalpesh Parikh, Atul Patel, " Sessionization –A Vital Stage in Data Preprocessing of Web Usage Mining-A Survey", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 1, Jan-Feb. 2012, pp. 327-330.

7. ACKNOWLEDGMENTS

Our thanks to Mr. Chirag Patel, Asst.Professor,CMPICA,Changa for his contribution for simulation in excel.