www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 6 June 2015, Page No. 12884-12888

A review on markov models and web page prediction

[1]Er. Manisha Dhull,[2]Dr.Meenakshi Sharma

1 Student of M.tech, HCTM(Kaithal),it1708117@gmail.com
2Assistant Professor in CSE Deptt. HCTM(kaithal),minnyk@gmail.com

Abstract A Markov process is a stochastic process in which the probability distribution of the current state is conditionally independent of the path of past states, a characteristic called the Markov property. Markov chain is a discrete-time stochastic process with the Markov property. In this paper we have to use the Markov model for the prediction of the different web pages based on the sequence of previously accessed page

Key words: Markov models, web pages.

1 Introduction: Markov models are used in the identification of the next page to be accessed by the web site user based on the sequence of previously accessed pages. The fundamental assumption of predictions based on Markov models is that the next state is dependent on the previous k states. The longer the k is, the more accurate the predictions are. Longer k causes the two problems. The coverage of model is limited and leaves many states uncovered and the complexity of the model becomes unmanageable. There are three modified Markov models for predicting Web page access. All k^{th} Markov model is used to tackle the problem of low coverage of a high order Markov model. For each test instance, the highest order Markov model that covers the instance is used to predict the instance. For example, build an all 4-Markov model including 1, 2, 3 and 4, for a test instance, use 4-Markov model to make prediction. Use the 3-markov model, if the 4-markov model does not contain the corresponding states and so forth [1].

Low order Markov models have higher accuracy and lower coverage than clustering. In order to overcome low coverage, all-k^{th} order Markov models have been used, the highest order is first applied to predict a next page. It cannot predict the page, then decreases the order by one until prediction is successful. This can increase the coverage, but it is associated with higher state space complexity. Clustering methods are unsupervised methods, and normally are not used for classification directly. However, proper clustering groups users' sessions with similar browsing history together, and this facilitates classification [2].

Prediction is performed on the cluster sets rather than the actual sessions. Clustering accuracy is based on the selected features for partitioning. For instance, partitioning based on semantic relationships or contents or link structure usually provides higher accuracy than partitioning based on bit vector, spent time, or frequency. However, even the semantic, contents and link structure accuracy is limited due to the unidirectional nature of the clusters and the multidirectional structure of Web pages. This involves implementing a clustering algorithm to partition Web sessions into clusters and then applying Markov model techniques based on the clusters in order to achieve better accuracy and performance of next page access prediction [2].

2 TYPES OF MARKOV MODEL

- **Frequency Pruned Markov Model**
- **Accuracy Pruned Markov Model**

2.1 Frequency Pruned Markov Model: This model describes though all k^{th} order Markov models result in low coverage, Exacerbate the problem of complexity since the states of all Markov models are added up. Note that many states have low statistically predictive reliability since their occurrence frequencies are very low. The removal of these low frequency states affects the accuracy of a Markov model. However, the number of states of the pruned Markov model will be significantly reduced [3]

2.2 Accuracy Pruned Markov Model: This model describes Frequency pruned Markov Model does not capture factors that affect the accuracy of states. A high frequent state may not present accurate prediction. When use a means to estimate the predictive accuracy of states, states with low predictive accuracy can be eliminated. One way to estimate the predictive accuracy using conditional probability is called confidence pruning. Another way to estimate the predictive accuracy is to count errors involved, called error pruning [3].When choosing the Markov model order, main aim is to determine a Markov model order that leads to high accuracy with low state space complexity. It reveals the increase of precision as the all k^{th} order Markov model increases. It shows increase of the state space complexity as the order of all k^{th} Markov model increases. Based on this information, use the all 2^{nd} order Markov model because it has better accuracy than that of the all 1^{st} order Markov model without the drawback of the state space complexity of the all 3^{rd} and all 4^{th} order Markov model [3].

3 WEB PAGE PREDICTIONS: Web page access prediction gained its importance from the ever increasing number of e-commerce and e-business. It involves personalizing, Marketing, Recommendations, helps in improving the web site structure and also guide web users in navigating through hyperlinks for accessing the information need. The most widely used techniques for discovering the patterns are Markov model, association rules and clustering, sequential patterns etc [4].

Pre-fetching and prediction is done by pre-processing of logs as it is the main requirement to provide user with best recommendations and overcomes the limitation of path completion and for pattern discover. This technique integrates the following three techniques together i.e. clustering, association rules and low-order Markov model using frequency support pruning. It achieves complete logs, better accuracy, less state space complexity and less number of rules. The predicted pages are pre-fetched and keep it in server cache which reduces the accessing time of that page and increases the web server performance [4].

The web data is heterogeneous in nature. Each session is a collection of visited Web pages by the user. Every user has a different level of browsing expertise and sessions are formed mainly haphazardly because users usually follow different paths when trying to access the same page. Clustering combines similar Web page paths or user sessions together and subsets of data are therefore more homogeneous resulting in simpler Markov model computations. By applying clustering to abstracted user sessions, it is more likely to find groups of sessions with similar pages that help increase the Markov model accuracy [2].

The web page prediction process has illustrated in Figure 1.1. It is clear from figure that, first of all a client request to a server for the specific web page. The server will send the URL of that page to the predictor. Then the predictor will check that specific web page, if it exists then predictor will send that page to the server and the server will immediately send that page to the client to fulfill its request. Also the predictor will send that page to the update engine which updates the data structure. The predictor uses that data structure for storing the web pages [4] [5]

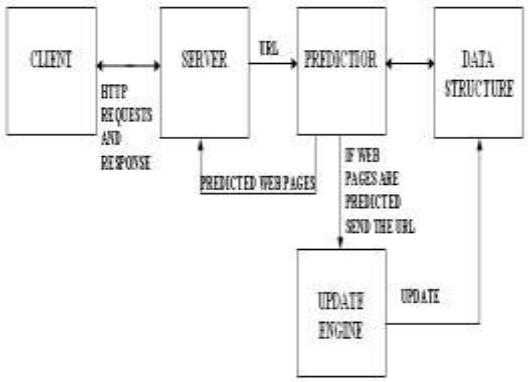

Figure 1.1: Web Page Prediction Process

4 RESEARCH METHODS

A careful investigation or inquiry especially through search for new facts in any branch of knowledge. Redman and Mory defines research as a systematized effort to gain new knowledge. Some people consider research as a movement, a movement from the known to the unknown. It is actually a voyage of discovery. Research is an academic activity and as such the term should be used in a technical sense [6].

Research is an academic activity and as such the term should be used in a technical sense. According to Clifford Woody research comprises of defining and redefining problems, formulating hypotheses or suggested solutions; collecting, organizing and evaluating data; making deductions and reaching conclusions and at last carefully testing the conclusions to determine whether the fit the formulating hypothesis.

The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered as yet

1. To gain familiarity with a phenomenon or to achieve new insights into it
2. To portray accurately the characteristics of a particular individual, situation or a group .To determine the frequently with which something occurs or with which it is associated with something else
3. To test a hypothesis of a casual relationship between variables

4.1 Applied Research The applied research is discovering, interpreting and the development of methods and systems for the advancement of human knowledge on a wide variety of scientific matters of our world and the universe The Research Processes step order may vary depending on the subject matter and researcher, the steps followed during formal research either it is basic or applied are choosing the research problem, review of related literature, collection of data, interpretation of data and preparing the research report

4.2 Basic Research The basic research is geared toward advancing our knowledge about human behavior with little concern for any immediate practical benefits that might result. The goal of the research process is to produce new knowledge which takes three main forms. First is exploratory research which structures and identifies new problems. Second is constructive research which develops solutions to a problem. Third is empirical research which tests the feasibility of solution using empirical evidence

4.3 Analytical Research In analytical research, the researcher has to use facts or information already available, and analyses these to make a critical evaluation of the material

4.4 Experimental Research Empirical research relies on experience or observation alone, often without due regard for system and theory. It is data based research, coming up with conclusion which are capable of being verified by observation or experiment. So that it is also known as experimental type of research

5 APPLICATIONS:

Web Server HTTP Request Prediction The first application of the probabilistic link prediction is HTTP request prediction. Extensive work has been done on the analysis of HTTP requests in order to enhance server performance. Most of the work involves statistical analysis of request file sizes, request patterns, and caching mechanisms. There are different methods of building a sequence prefix tree using path profiles and using the longest matched most-frequent sequence to predict the next request. Probabilistic sequence generation models such as Markov chains have not been applied to the problem of HTTP request prediction [7].

Adaptive Web Navigation The second application of the Markov chain probabilistic link predictor is system aided web navigation. Link prediction is used to build a navigation agent which suggests other sites/links would be of interest to the user based on the statistics of previous visits. In the current framework, the predicted link doesn't strictly have to be a link present in the web page currently being viewed [8].

Tour Generation The tour generator is given as input the starting URL. This generates a sequence of states using the Markov Chain process. This is returned and displayed to the client as a tour. such a tour is used to browse a web site in minimum time. [9].

Personalized Hub/Authority The notion of Hubs/Authorities is typically applied on the web graph structure. Hubs refer to web sites that are often good starting points to find information. Authority refers to web sites that contain a lot of useful information on a particular topic. The term personalized is used here to pertain to a specific set of users, or a specific type of sites [38].Personalized Hubs/Authorities extend the notion of Hubs/Authorities to focus on a specific

group of users/sites using the path traversal patterns that are collected. The Markov state transition matrix is a representation of the users traversal patterns, and can be viewed as a traversal connectivity matrix [10].

REFERENCES

[1] Jiuyong Li, **"Integrating Recommendation Models for Improved Web Page Prediction Accuracy",** Thirty-First Australasian Computer Science Conference (ACSC2008), Wollongong, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 74. Gillian Dobbie and Bernard Mans, Ed. Reproduction for academic, 2008.

[2] Dimitris Drosos, **"Web Page Rank Prediction with Markov Models"**, Database Applications and Data Mining, 2008.

[3] Mukund Deshpande, **"Selective Markov Models for Predicting Web-Page Accesses",** IEEE/WIC/ACM International Conference on Web Intelligence, 2001.

[4] Cooley R., Mobasher B., and Srivastava J., **"Grouping web page references into transactions for mining world wide web browsing patterns"**, Technical Report TR 97-021, Dept. of Computer Science, Univ. of Minnesota, Minneapolis, USA,1997.

[5] Faten Khalil, **"Integrating Markov Model with Clustering for Predicting Web Page Accesses"**, Web development and mining, 2007.

[6] C. R. Kothari, **"Research Methodology Research Methods & Techniques 2nd Edition".**

[7] S. Schechter, M. Krishnan, and M. D. Smith. **"Using path profiles to predict http requests",** In Seventh International World Wide Web Conference, 1998.

[8] J. Zhu, J. Hong, and J. G. Hughes, **"Using markov models for web site link prediction"**, HT'02, USA, pages 169–170, 2002.

[9] Chen M. S. Park J. S., and Yu P.S., **"Data mining for path traversal patterns in a web environment",** In ICDCS, pages 385-392, 1996.

[10] M. Eirinaki and M. Vazirgiannis, **"Usage-based PageRank for Web Personalization"**, In Proceedings of the 5th IEEE International Conference on Data Mining (ICDM'05), Lousiana, 2005.