

Using Cloudflare Model Approach to Forecast Web Page Caching

Dr. Neelam Ruhil¹ Dr. Rajesh Kumar²

¹Associate Professor, DPG Degree College

²Director technology, GlobalLogic India Pvt. Ltd

Abstract—The World Wide Web is a large repository of information. There are lot of users who regularly accesses this information source, this is simple to invent certain patterns to access resources on web. Web assumption has been implemented in the past for static content. With the increasing Internet traffic and Web content, the Web assumption models are very famous. Data mining methodology categorizes the modules of clients based on their attributes and assumes future activity without allowing instant inferences and interactivity. Here some practices like information retrieval and assumption by partial matching can be used in combination with prediction modelling to increase validity and performance. There is always some scope to improve the web page access based on user requirement. One of the methods given by web is page pre-fetching which means to make available the web page to the user before the user request. In this current work we are presenting an intellectual method created on the history of web page visit used for web page prediction. A three-level approach is proposed in which we Cloudflare model is combined with AWS approach and association mining.

I. INTRODUCTION

Prediction model is a very well-known machine learning method and it is differed from the way that data mining does with Data history. Data mining methods categorizes the modules of clients by their attributes and assume future activity without allowing instant inferences and interactivity. Here some practices like information retrieval and assumption by partial matching can be used in combination with prediction modelling to increase validity and performance. Unlike other models, a Cloudflare Model is mostly challenging because of the several situations that hold the self-motivated Web in terms of user activities and the content that changes

regularly. So, to design an assumption model use the prediction probabilistic idea.

1.1 Web Cache

A Web cache is a procedure used for the temporary storage and caching of Web documents like HTML pages and images in order to reduce server load, bandwidth usage and perceived lag. It stores the reproduction of documents passing through the web and later it satisfies the request if certain circumstances are met [3]. A Cache is a repository that stores data transparently so that upcoming requests for that data can be served faster.

1.1.1 Web Caches Working

1. If the response's headers tell the cache not to keep the information, it won't.
2. If some request is secure or authenticated i.e., HTTPS, it won't be cached.
3. A cached is able to send to a client without checking with the origin server, this representation is considered fresh if:
 - a. It contains an expiry time or some age-controlling header set and is still within the fresh period.
 - b. If the representation of cache has seen recently and was adapted previously. Fresh demonstrations are aided directly from the cache, without examining the origin server.
4. In case if the representation is old, the source server will be requested to validate it.
5. In circumstances for example, when it is not connected to a network, a cache serves stale responses without examining the origin server.

1.1.2 Importance of Web cache

1. Reduced Cost of Internet Traffic
2. Reduced latency

1.2 Data Mining

Web usage mining is a subset of Web mining operations which itself is a subset of data mining in general. The aim is to use the data and information extracted in Web systems in order to reach knowledge of the system itself. Data mining is a set operation performed on a collection of data or a subset of it so as to extract meaningful patterns on the data. Another definition is “Data mining is the semi-automatic discovery of patterns, associations, changes, anomalies, rules, and statistically significant structures and events in data”. That is, data mining attempts to extract knowledge from data. If a subset is to be used, careful and unbiased sampling algorithms should be used to avoid biased result. Data mining is different from information extraction although they are closely related. To better understand the concepts brief definitions of keywords can be given as [9]:

Data: “A class of information objects, made up of units of binary code that are intended to be stored, processed, and transmitted by digital computers”.

Information: “It is a set of facts with processing capability added, such as context, relationships to other facts about the same or related objects, implying an increased usefulness. Information provides meaning to data”.

Knowledge: “It is the summation of information into independent concepts and rules that can explain relationships or predict outcomes”.

II. IMPLIMENTATION

The presented work which is focused on three main concepts called Cloudflare Model, Clustering and the Neural Network. The Markov Model is the basic prediction algorithm, that is been improved by using the C- Means Algorithm as well as the Neural network. The neural network is used as the classification tool to derive the required results.

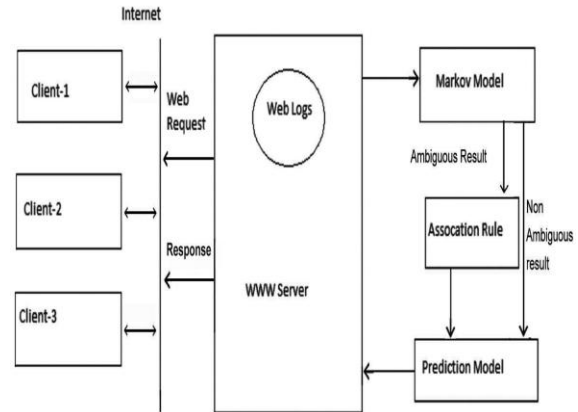


Figure 1 The integration & Association Model (IMAM)

The integration model profits from the decrease the state space complexity of the lower Markov model by using association mining in case of ambiguity. The integration model also provides the complexity of the association rules since the rules are generated only in special cases [9].

In brief, the new integration model results in an increase the accuracy and a decrease in state & rule complexity.

2.1 C-mean clustering

It is an important part of cluster analysis is partitioned clustering. Based on various models, several clustering procedures have been implemented and different algorithms remain to act in the work.

The paper’s significance can be briefed as:

1. Rendering a new description of the mean, a structure for separated clustering systems, called GCM i.e., General c-means Clustering Model is proposed.
2. By applying confined optimality test, the association between Partitioned clustering and Occam's razor is recognized first time in GCM.
3. A guide for developing and executing clustering algorithm is implemented based on an assumption in partitioned clustering. These decisions are confirmed by numerical experimental results.

III. RESULTS

Due to the significant growth in the amount of data the web users are facing the problems of information

Figure 6 is showing the results of Web page visit with Cloudflare Model and implementation of association rule on it. The association model shows the results respective to previous and next visited pages.

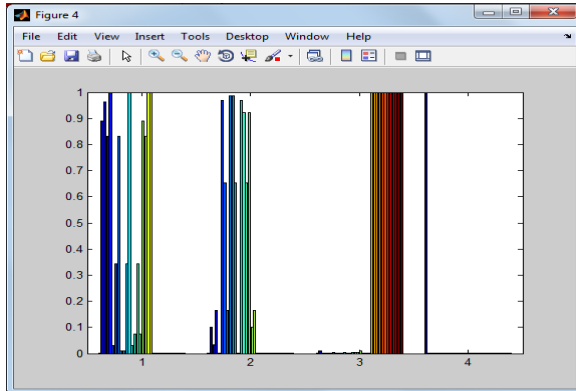


Figure 7 Clustered Cloudflare Model

Figure 7 is showing the results of clustered Markov model. All pages are divided in 4 clusters. High frequency pages are grouped in separate clusters. The cluster 4 is having the low frequency pages.

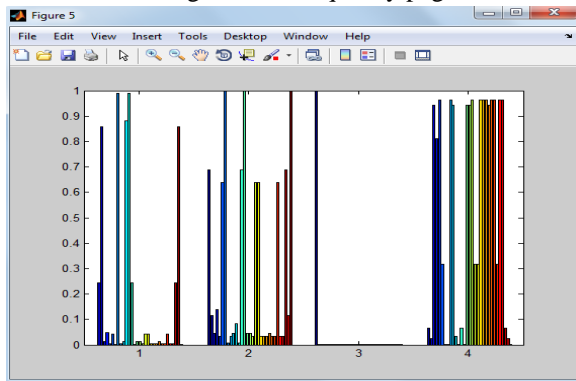


Figure 8 Web Page clustering and association using Cloudflare Model

Figure 8 is showing the results of clustered Markov model. All pages are divided in 4 clusters. High frequency pages are grouped in separate clusters. The cluster 3 is having the low frequency pages. This result is respective to the association model applied on the dataset. Here C- means clustering is applied on it.

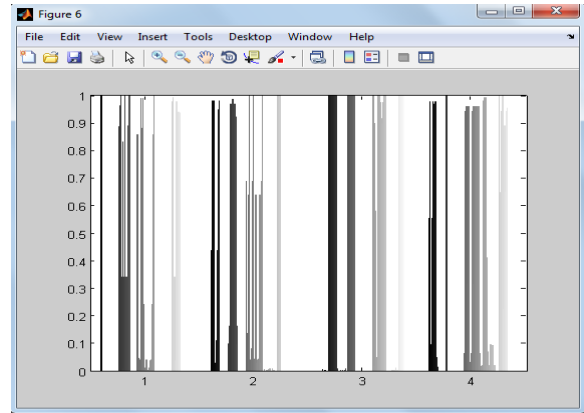


Figure 9 C-means clustering using Cloudflare Model with Web Page association

Figure 9 is showing the results of clustered Cloudflare Model. All pages are divided in 4 clusters. High frequency pages are grouped in separate clusters. The cluster 3 is having the low frequency pages. This result is respective to the association model applied on the dataset. Here C- means clustering is applied on it.

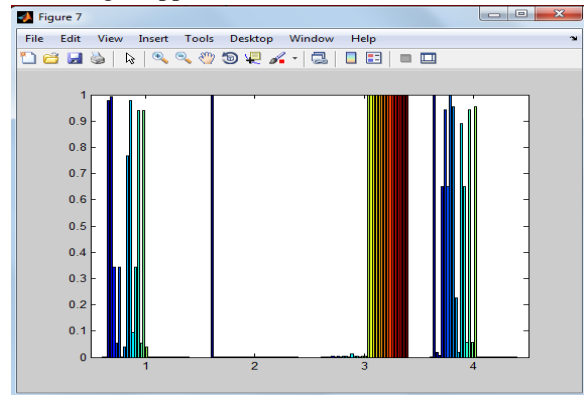


Figure 10 Clustering in C-means With Web Page associations

Figure 10 is showing the results of clustered Cloudflare Model level 2. All pages are divided in 4 clusters. High frequency pages are grouped in separate clusters. The cluster 2 is having the low frequency pages. This result is respective to the association model applied on the dataset. Here C- means clustering is applied on it.

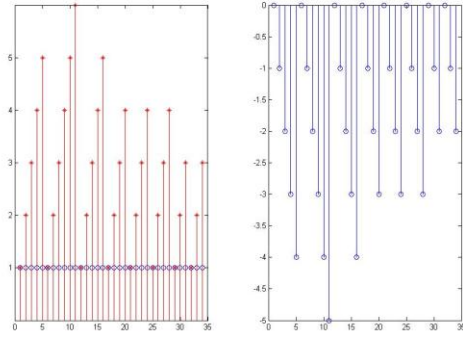


Figure 11 Analysis of Training Dataset

The available Web log dataset is divided in two parts called training dataset and the testing dataset. The selected 90% data values as the training dataset and 10% as well 30% dataset as the testing dataset. In the above figure, the Web usage analysis based on training dataset is defined. The first image in figure 11 is showing the usage of Web pages in terms of Web page usage frequency respective to the training dataset.

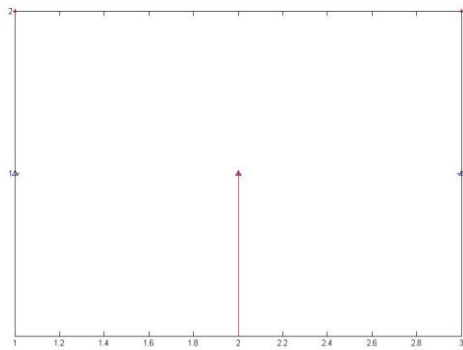


Figure 12 Analysis of Training Dataset

In the above figure, the Web usage analysis is defined based on testing dataset is defined. For this analysis, just taken the 10% of the available dataset as the training dataset. The figure 12 is showing the usage frequency respective to the testing datas.

IV. CONCLUSION

The evolution of Web application, especially in in the field of electronic commerce, thereby needs substantial interest to well understand and analyse Web data usage and imply the information for the betterment of users. Number of issues arises in the

area of Web Usage Mining. In the proposed paper, we defined numerous applications of Web Mining recommended by different authors. The paper considers the most stimulating study area that deals with the most of semantics to progress the consequences of Web Tradition Mining applications.

REFERENCES

- [1] R. Tewari, M. Dahlin, H. Vin and J. Kay, "Beyond hierarchies: design considerations for distributed caching on the Internet", In Proceedings of the 19th International Conference on Distributed Computing Systems (ICDCS), 2018.
- [2] Reinhard P. Klemm, "Web Companion: A Friendly Client-Side Web Prefetching Agent", IEEE Transactions on knowledge and data engineering, 2022.
- [3] Xiangping Chen, "Lifetime Behaviour and its Impact on Web Caching" Proceeding WIAPP IEEE, 2022.
- [4] Greg Barish and Katia Obraczka, "World Wide Web Caching: Trends and Techniques", IEEE Communications Magazine, May 2020.
- [5] Jilian Xiao," Measuring Similarity of Interests for Clustering Web-Users", IEEE 2021.
- [6] Brian Davison, "A Web Caching Primer", IEEE Internet Computing, 2021.
- [7] Pablo Rodriguez, Christian Spanner, and Ernst W. Biersack, "Analysis of Web Caching Architectures: Hierarchical and Distributed Caching", IEEE/ACM Transaction on Networking, AUGUST 2021.
- [8] Jitian Xiao, "Clustering of Web Users Using Session-based Similarity Measures", Proceedings. 2021 International Conference IEEE, 20211.
- [9] Alexandros Nanopoulos, "A Data Mining Algorithm for Generalized Web Prefetching", IEEE Transactions on knowledge and data engineering, 2023.
- [10] S. Yu, D. Cai, J.-R. Wen, and W.-Y. Ma, "Improving pseudo- relevance feedback in Web information retrieval using Web page segmentation", In Proceedings of the Twelfth International World Wide Web Conference, May 2023.

- [11] Caihong Yan, "Parallel Web Prefetching on Cluster Server", CCECE/CCGEI, IEEE, 2024.
- [12] M. Jincheng, G. Zhiming, "Finding Shared Fragments in Large Collection of Web Pages for Fragment-based Web Caching", Fifth International Symposium on Network Computing and Applications (NCA'06), IEEE, 2024.
- [13] Ruma Dutta, "Offering Memory Efficiency utilizing Cellular Automata for Markov Tree based Web-page Prediction Model", 10th International Conference on Information Technology IEEE, 2024.
- [14] Yong Zhen Guo, "Personalized PageRank for Web Page Prediction Based on Access Time-Length and Frequency", International Conference on Web Intelligence, IEEE 2024.
- [15] Payal Gulati, "A Novel Approach for Determining Next Page Access", First International Conference on Emerging Trends in Engineering and Technology, IEEE 2020.