A Fuzzy Integrated neural Network approach for web page prediction

Nishant Jakhar

Asstt. Prof., CSE Dept, Indus Degree College

ABSTRACT

In an organization, different users performs same kind of web page requests and the request patterns. To provide the effective page prediction in this work a three stage model is provided. In first stage of this model, the fuzzy adaptive method is defined to analyze the session adaptive frequency. The high level page pruning is done to remove the less frequent used pages. In second stage, the page pattern is analyzed using fuzzy association page analysis. Based on these most frequent patterns and less common patterns are identified. In final stage of this model, the neural network is applied to perform page prediction. In this paper, the model is proposed to provide effective page prediction.

Keywords: Page Prediction, Neural Network, Fuzzy logic

I.   INTRODUCTION

WWW is the web repository having billions of website and web pages exist. But, the access of the pages or sites within an organization is predictable based on the common requirements of users. The pages are generally accessed in same patterns and sometimes on same sites. But if each time, the pages are loaded from the web repository, it takes time as well as memory. It slow down the browsing speed over an organization. To improve the web content access, some content driven usage mining can be applied. The usage mining is about to predict the usage behavior of user and relatively identify the next possible web page visit. Once the page is predicted, the next work is to load it in the memory before the user access. This kind of page caching before user access for a web page is called web page prediction. The prediction model is defined to reduce the access time so that the computing power will be improved and the network bandwidth will be utilized effectively.

A) Web Cache

Caching in web pages provides the temporary storage to the web documents or the web pages so that the offline web access will be provided as well as the demand driven instant access will be achieved. The condition specific page access can be defined to control the caching to that the access hit will be improved. The cost effective and performance effective is required for which some machine learning algorithms are defined. The objective of these algorithms is to provide the utilization of available limited cache in terms of effective page replacement and the rule specific page access. While performing the caching, the temporal measures are also defined to bound the access within the time limits. The location driven, time driven, application driven and organization driven caching can be defined to provide effective web page access. The browsing experience and contents are analyzed to improve the caching behavior. The bandwidth utilization and the resource utilization is the primary requirement of web caching. Web caching is either maintained either by the main server, proxy server or the client side. The application specific caching is also done to provide more directed caching.
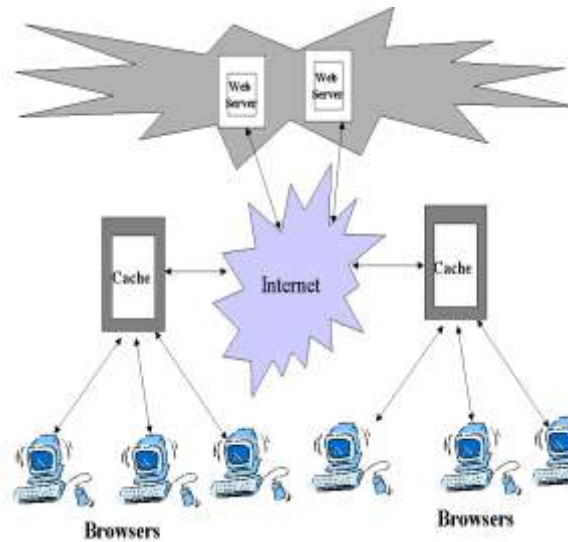
Figure 1: Web Caching Model

Here figure 1 is showing the basic caching model. The web repository is here shown in the form of multiple web servers. The internet access with specific cache is defined to perform page access. The content specific access as well as traffic driven morning was also defined to provide the adaptive caching of web pages.

Caching is basically defined and controlled by some middle server or the controller that act as proxy between the client and the server. The network gateway control is here defined with larger storage, features and the performance. The performance driven mapping with the network was provided to achieve pre-caching to the web pages. The internet connection, caching and the relative retrieval was provided to improve the performance vector. The cache hierarchy was provided with neighbor specification to track the web documents. The cache, driven access able to provide the services to multiple client in parallel. The caching is here defined to achieve the instruction driven access and to provide the resource tracking and cache access in effective way.

B) Neural Network

Neural network is the data computational model defined to establish the relationship between the complex input and output. It is the intelligent method provides the learning driven knowledge processing. The neural network uses the weight driven connectivity to transform the complex input to predicted output. The linear model is defined with both linear and non-linear relationship specification. The traditional linear modeling provides the data driven characterization so that effective prediction will be performed. The network based layered model is shown here in figure 2.
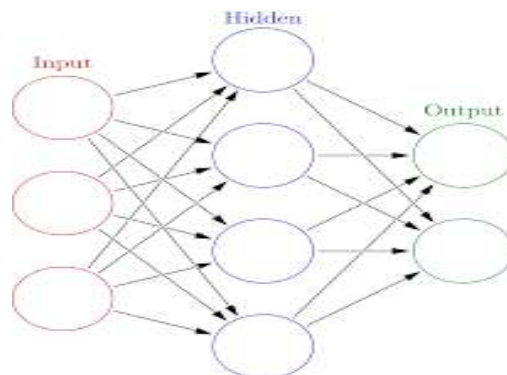

Figure 2: Layered Neural Network Model

Here figure 2 is showing the layered model with specification of 3 different layers. In first layer, the complex input is taken as input. To transforms the input to output, the weights are applied over the input. The hidden layer is defined on carry these weights and to transform the input to the output.

In this paper, an effective fuzzy integrated neural network model for optimizing the web page access. The method first analyze analyzed the web traffic based on the fuzzy rules. These rules are applied on single page and later on association page vector. The pruning is also applied to identify the most effective page access. In second stage, the neural network model is applied to perform the page prediction. In this section, the exploration to the web repository is provided along with the concept of web page access and caching. In section II, the work defined by the earlier researchers is discussed. In section III, the proposed fuzzy adaptive work model is provided. In section IV, the conclusion obtained from work is presented.

II. EXISTING WORK

In this section, the work provided by the earlier researchers is presented for web caching and web page prediction. Different machine learning algorithms defined by the researchers are discussed in this section. Author[1] has defined a work to improving the cache design model under improving the web page access. Author defined two different methods for analyzing the meta data and relative to the user location the caching was provided. The method reduced the access traffic over the web server and improved the client response vector. The cache hit ratio was observed to verify the work effectiveness. Author[2] has provided a work to improve the data retrieval over the web along with the content delivery. The novel and selective web page prefetching is defined to provide the effective web access. The optimization to the available resources was provided at moderate level so that the communication overhead is reduced. The caching and the network compatibility is also improved in this work.

The web system estimation and resource driven response observation was also considered as the effective measurement so that the improved access will be obtained. Author[3] has defined an environment specific class mapping based on the application domain. The application specific page priority assignment is done so that the effective prefetching and optimized web access will be obtained. An extensive improvement to the size, behavior and caching was provided by the author. The web document specific service access as defined to achieve the effective document tracking. Author[4] has defined the content specific availability analysis, latency analysis and the bandwidth observation. Author improved the architectural improvement in terms of cache deployment. The content security, dynamic behavior and effective functional processing was discussed by the author.

Author[5] has defined the measurement level similarity at content level so that the access improvement will be achieved. The content level pre-fetching was provided to achieve the similarity utilization and the cluster formation. The cluster specific mapping and the interest driven analysis was provided to improve the caching results. A global affinity based matrix formation and its application driven access was defined by the author to improve the service access in optimized form. Author[6] has defined resource caching based method to provide the web scalability. The server load, infrastructure and caching behavior was optimized by the author. The protocol level validations were defined to improve the caching behavior. The web caching improvement and caching was provided at content level.

Author[7] has provided a performance driven architectural with analytical specification to improve the distributed caching. The bandwidth usage, latency and the cache specific space utilization was also improved by the author. The caching level maximization along with cooperative communication in distributed environment was provided by the author. The expected feature analysis in the architectural form was provided to share the load and to achieve the distributed service access. Author[8] has defined a behavior driven analysis along with service access log so that the clustered behavior of the network will be improved. The service similarity analysis was provided at lower level so that the multilevel communication scheme is provided. The arbitrary content similarity and the session specific gain analysis was provided by the author at lower level. The observations are taken for multiple sessions so that the matrix based cluster formation will be achieved. Author[9] has defined the markov model method for prediction of web page. The performance evaluation relative to the web page was provided by the author.

Author[10] has defined a web search method for different topics and navigation with specification of interactive page access. The content structure analysis and the relevance feedback specific evaluation was provided to improve the web page access. A segmentation adaptive algorithmic approach was provided to achieve the effective page prediction. Author[11] has defined a parallel web fetching algorithm under cluster formation. The session persist method was provided along with scheduling algorithm to improve the prefetching results. Author[12] has defined the problem solution as the scalable method to process the larger web page database in segmented form. Author applied the caching scheme with performance measure using probabilistic markov model. The problem criteria specification and nonlinear featured observation was provided to improve the processing and memory requirements. The performance driven optimization was provided by the author. Author[13] has defined a work on automata driven optimization to observe the network behavior and relative data processing. The predictive markov tree modeling under accuracy map was provided to achieve the accurate class mapping.

Author[14] has defined a page ranking algorithm for web page prediction. The personalized method has analyzed the page visit frequency in the defined period and applied the factor specific analysis to predict the next possible page. Author[15] has defined a client specific method to provide effective page tracking and identification of next possible page visit for the client.

III. RESEARCH METHODOLOGY

As a user generally performs the common web requests or follows some particular access pattern, there is the requirement to optimize the web page access for specific domains and pages. Web page Caching is one of such method that provides the restoration of the pages that are been frequently in use. But as the number of users increases as well as user activity increases, it becomes difficult to provide caching to all the frequent pages. As the cache memory size is limited and the visit activity is higher, there is the requirement of some intelligent method for selection of cacheable web pages as well as to replace them frequently. In this work, an intelligent method is provided for effective web page prediction. The method has first analyzed the individual use visit for multiple sessions. Later on the association specific visit analysis is performed. The mapping of this visit is here done using fuzzy logic. At higher level, the pruning is applied to remove the pages with less frequent page visit. In final stage of this method, the neural network is applied to perform the web page prediction. The algorithmic method is shown here in figure 3.
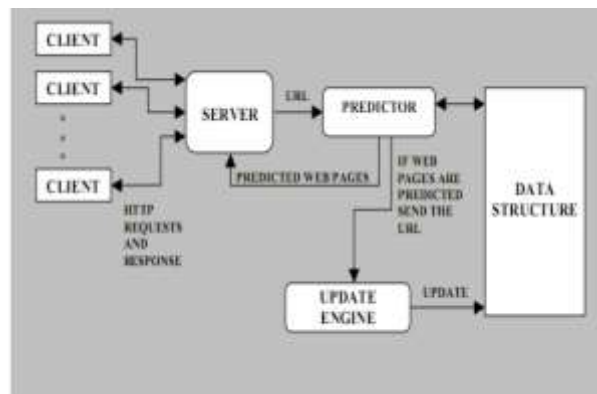

Figure 3: Web Page Prediction Model

Here figure 3 is showing the predictive model for web page visit. The model is here divided in two main stages. The model begins as the server captures the clients requests for the service access. The predictor is defined in first stage to identify the URL on server side. The predictor load the requests in the data structure and apply the main fuzzy adaptive method to perform the pruning and load the effective web pages over it. In the second stage, the updator is applied using intelligent neural based approach to analyze the page visit combinations and relatively apply the prediction. Each of the work layer is defined here under

A) Layer 1: Individual Page Analysis

Once the dataset is collected, the fuzzy logic is applied on the session pages. The user independent fuzzy logic is here applied to categorize the pages based on the frequency of visits. The high frequency, low frequency and medium frequency pages are identified. The fuzzy rule training method is applied to perform page identify the page visit. The conditional formulation is done to remove the pages with low frequency.

B) Layer 2: Association Page Analysis

In second layer, the visit pattern is analyzed based on the visit measure for two pages. The pair of two pages is here identified under the fuzzy logic. In this stage, the pruning is applied on the association analysis. The pages that does not support any pattern and not visiting in combination are removed under fuzzy rules.

C) Layer 3: Neural Updation

In third layer of this model, the neural network model is applied to perform the prediction. The page pair is identified and relatively caching of web pages is done. The distance and frequency specific analysis is defined to generate the predictive

decision. From this weight specific model effective page combination is defined. The predictive page on the page visit is identified.

CONCLUSION

In this paper, an intelligent neural based model is defined for web page prediction. The presented model is here divided in three main stages. In first stage, single page analysis is done under fuzzy approach. In second stage, the associated page analysis is done. In final stage, the neural predictive model is applied for web page prediction.

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), 1998.

[2]. Reinhard P. Klemm, " Web Companion: A Friendly Client-Side Web Prefetching Agent", IEEE Transactions on knowledge and data engineering, 1999.

[3]. Xiangping Chen," Lifetime Behavior and its Impact on Web Caching" Proceeding WIAPP IEEE, 1999.

[4]. Greg Barish and Katia Obraczka, "World Wide Web Caching: Trends and Techniques", IEEE Communications Magazine, May 2000.

[5]. Jitian Xiao," Measuring Similarity of Interests for Clustering Web-Users", IEEE, 2001.

[6]. Brian D.Davison, "A Web Caching Primer", IEEE Internet Computing, 2001.

[7]. Pablo Rodriguez, Christian Spanner, and Ernst W. Biersack, "Analysis of Web Caching Architectures: Hierarchical and Distributed Caching", IEEE/ACM Transaction on Networking, AUGUST 2001.

[8]. Jitian Xiao, "Clustering of Web Users Using Session-based Similarity Measures", Proceedings. 2001 International Conference IEEE, 2001.

[9]. Alexandros Nanopoulos, "A Data Mining Algorithm for Generalized Web Prefetching", IEEE Transactions on knowledge and data engineering, 2003.

[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 2003.

[11]. Cairong Yan, "Parallel Web Prefectching on Cluster Server", CCECE/CCGEI, IEEE, 2005.

[12]. M. Junchang, G. Zhimin, "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, 2006.

[13]. Ruma Dutta, "Offering Memory Efficiency utilizing Cellular Automata for Markov Tree based Web-page Prediction Model", 10th International Conference on Information Technology IEEE, 2007.

[14]. Yong Zhen Guo, " Personalized PageRank for Web Page Prediction Based on Access Time-Length and Frequency", International Conference on Web Intelligence, IEEE 2007.

[15]. Payal Gulati, "A Novel Approach for Determining Next Page Access", First International Conference on Emerging Trends in Engineering and Technology, IEEE, 2008.