

Bayesian Classifier Web Object Filters for User Navigation with Validation Cross Point

A. K. Santra

Phd, Dean

CARE School of Computer Applications
Trichy – 620 009

S. Jayasudha

Research Scholar Bharathiar University
Coimbatore – 638401

ABSTRACT

Web caching and pre fetching plays a significant role among the users and websites in reducing the response time of user requests and saving the network bandwidth. The websites which are most likely visited are stored under web cache and web pre-fetch to retrieve them later shortly. Thus web caching and web pre-fetch are one of the most thriving resolutions for enhancing the performance of the web-based systems. However, the complexity in formatting the ideal Web objects that will be re-visited in the prospect is still a trouble faced by an existing conventional cluster based caching techniques. In this work, Bayesian classifier is used for web object filtering process which enhances the performance of caching techniques. Apart from that, the validation process is also being taken place to improve the user navigation process under web based systems. The validation is made of user identity on web product navigation is done with User-profile relativity to national identification number. Web services enable the product vendor regarding the demand of their products by specific valid users. The validation is done with both the existing cluster based web object filters and the proposed classifier based web object filters and compares the results with it. An experimental evaluation is done with the several log files to estimate the performance of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation in terms of efficiency of classification, validation, performance rate and compared with an existing cluster based web object filters and identified the better web caching technique.

Keywords

Web cache, Web pre-fetch, BNN classifier, and validation process.

1. INTRODUCTION

Web proxy caching acts as a solution role in enhancing the web performance by maintaining web objects probable to be utilized in the prospect in the proxy server. Thus, the Web proxy caching assists in dropping user supposed latency, shrinking network bandwidth procedure, and improving loads on the beginning servers. Since the distributed space to the cache is restricted, the space must be exploited sensibly. Consequently, an intellectual method is necessary to supervise the Web cache content proficiently. The cache substitution is the core or heart of the Web caching; therefore, the proposal of proficient cache substitution algorithms is critical for caching mechanisms realization. The most widespread Web caching methods are not proficient sufficient and may endure from a cache contamination problem because they believe just one factor and disregard other factors that have contact on the effectiveness of Web proxy caching.

The cache pollution determines that a cache comprises of objects that are not commonly visited. This is the origin a

decrease of the successful cache size and concerns depressingly on performance of the Web proxy caching. Many Web proxy caching strategies endeavored to merge some factors which can manipulate the performance of Web proxy caching for building resolution of caching. Nevertheless, this is not a simple task since one factor in a scrupulous situation may be more significant in other environments. So far, the complexity in shaping which perfect Web objects will be re-visited is motionless a main confront faced by the existing Web proxy caching techniques. In other words, which Web objects should be cached or restored to construct the best use of accessible cache space, progress hit rates, diminish network traffic, and lessen loads on the innovative server.

In Web proxy server, Web proxy logs file can be measured as inclusive and prior information of potential accesses. An accessibility of Web proxy logs files that can be subjugated as training data is the major inspiration for developing machine learning techniques in adopting intelligent Web caching approaches. The second inspiration is, while Web environment alters and informs quickly and constantly, an efficient and adaptive method is mandatory in Web environment.

Classification in data mining is a diverse method than clustering. Classification is comparable to clustering in that it also fragment customer records into different segments called classes. But not like clustering, a classification examination entails that the end-user/analyst recognize ahead of time how classes are distinct. For example, classes can be definite to symbolize the possibility that a customer defaults on a loan. It is essential that each confirmation in the dataset used to construct the classifier previously have a value for the attribute utilized to classify classes. Since each record has a value for the attribute utilized to identify the classes, and since the end-user chooses on the attribute to utilize, classification is much less investigative than clustering. The purpose of a classifier is not to discover the data to determine motivating segments, but slightly to choose how new records should be confidential.

In this paper, we propose an efficient classifier based Web Object Filters from Web Pre-fetching and Web caching scheme which evaluates the web user navigation patterns and user references of product search. The validation is also being taken place based on user identification number and compared the results among the techniques.

2. RELATED WORK

Pre-fetching and caching are the eminent approaches for improving the presentation of the Web. They have turned out to be imperative objects the Web infrastructure. The merits of these approaches have presented go up to new developed which comprises tools and service dealers that offer cache

servers. It presents caching and pre-fetching services to clients and provider of Web resources [1].

By incorporating Web caching and Web prefetching, these two strategies can harmonize each other because the Web caching method develops the sequential locality, while Web prefetching system operates the spatial region of Web objects by utilizing an algorithm presented in [2]. Numerous dissimilar classifications and clustering approaches are used in a way. Among that, knn classifier [3] is used for the retrieving an information from the user in a reliable manner. First, current studies have revealed that Web cache hit ratio (HR) and byte hit ratio (BHR) raise in a log-like method as a purpose of cache size [4] with a Better Algorithm that Optimizes Both Hit Rate and Byte Hit Rate in Internet Web Servers.

Compared with web caching, perfecting goes one step promote by expecting users' future requests and preloading the probable objects into a cache and also discussed the performance evaluation of the proxy cache [8]. When a user ultimately desires the probable objects, they are obtainable in the cache. In the past, numerous prefetching strategies have been proposed using the inventive caching algorithm as Intelligent Predictive GDSF#, (IPGDSF#) [5]. A web caching and prefetching review results are discussed in [6]. Apart from knn classifier, Bayesian classifier is also being used with medical diagnosis [7].

Web proxy caching acts as a input role in enhancing the Web presentation by charging Web objects probable to be used in the future in the proxy server. Bayesian network (BN) [9] is used to improve the presentation of conservative Web proxy caching strategies such as Least-Recently-Used (LRU) and Greedy-Dual-Size (GDS). An admission control technique [10] is also being used with the delay reduction in cache. The classification of web proxy cache [11] is partitioned with the client side web cache [12]. In this work, we have implemented a Bayesian classifier for web based user navigation system and the validation checking also been takes place.

3. PROPOSED BAYESIAN CLASSIFIER BASED WEB OBJECT FILTERS

The proposed work is reliably designed for evaluation of the user navigation pattern of web based interface system using the Bayesian classifier system and the validation process is also been done with both the users and products to check their validity among the user navigation system. The proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation is accessed under two processes. The first process is to implement a Bayesian classifier for web object filtering process. The second process is to check the validation of the user and the product based on the appropriate valid attributes. The architecture diagram of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation is shown in fig 3.1.

The first process here is to classify the web objects based on the users visit on a particular web page. Web page objects are classified by using pre-fetched and web cached contents using Bayesian classifier. The web cluster objects with resemblance recovery in successive user sessions estimate the User Navigation.

The second process describes the validation scheme which checks the validity of both user and product before it delivers into appropriate manner. The validation takes place based on assigning the valid user id for all the users whoever involved in the web based user navigation system. The valid user id is given to the user based on their viewed websites and a valid e-mail id.

From the figure (fig 3.1), it is noted that the classification of web objects is done at first by using Bayesian classifier. When the user or a group of users viewed some web pages, the web cache and web pre-fetch stores all the web pages which are frequently used by the users. For an ease of retrieval, Bayesian classifier is used for classification task which classify the web page contents based upon the users requests.

After classification, it is necessary to provide information to a distinct valid user and the same way only a valid and secure product web site is visible to the user. For this purpose, in this work, we are also performing a validation testing which checks both the user and the product are valid or not. The web site information is retrieved to the user only if the user and the product are valid. The validation is checked out by a unique identification number. The user identification number is given based on their valid e-mail id. And the product id is sponsored by the reputed organization.

3.1 Bayesian Classifier for User Navigation

In this work, Bayesian classifier is used to classify the web objects with the retrieved web page contents. A Bayesian classifier is one of the most popular learning models those supports probability estimations to discover a class of an experimental pattern. The Bayesian classifier also is known as causal probabilistic classifier, Bayesian or simply Bayes.

A Bayesian classifier is supported on the design that the function of a (natural) class is to envisage the standards of features for associates of that class. Examples are clustered in classes since they have widespread values for the features. Such classes are regularly called natural kinds. The proposal following a Bayesian classifier is that, if a representative recognizes the class, it can guess the values of the extra features. If it does not recognize the class, Bayes' rule can be utilized to expect the class given (some of) the attribute values. In a Bayesian classifier, the knowledge agent constructs a probabilistic model of the features and utilizes that model to forecast the classification.

In classification tasks, BN depends on probability estimations, called a posterior probability, to assign a class to an observed pattern. The posterior probability [9] is determined using the Bayes theorem as shown as

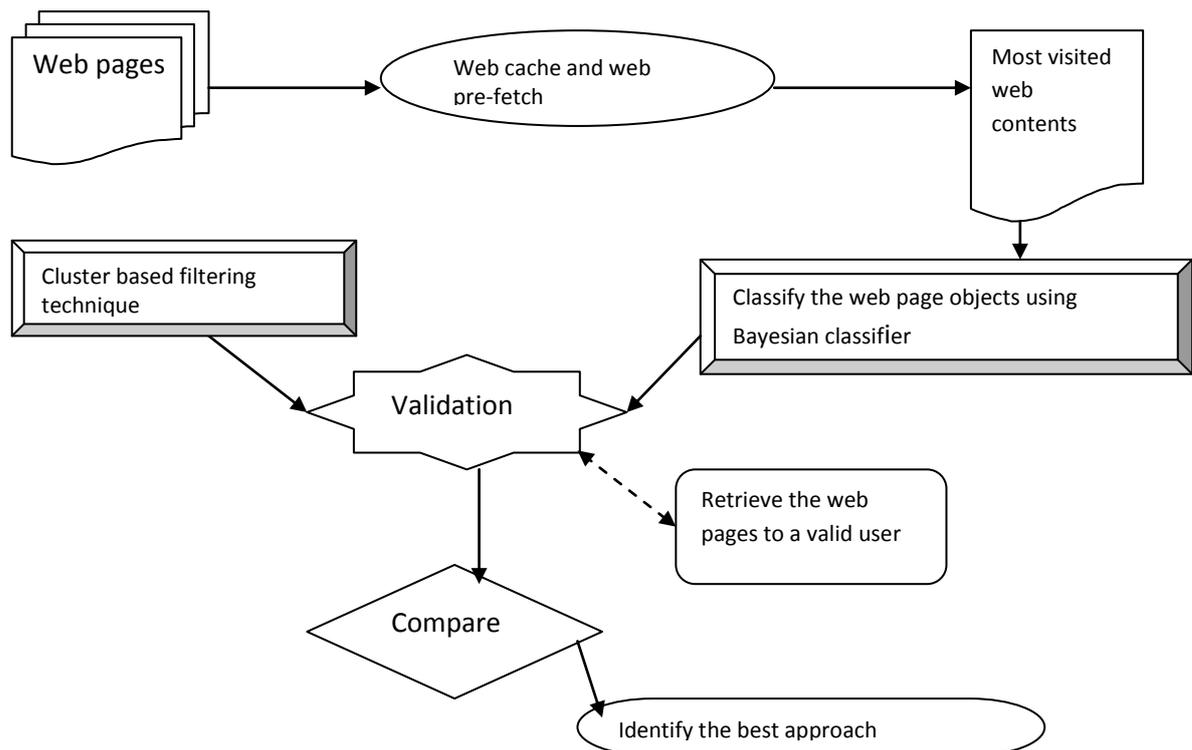


Fig 3.1 Architecture diagram of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation

$$P(C_i / x) = \frac{P(x / c_i)P(C_i)}{P(x)} \dots\dots\dots \text{eqn 1}$$

The Bayes classifier computes a posterior possibility, $P(C_i / x)$, that an experimental pattern x fits in to class c_i , from a priori possibility of this class, $P(C_i)$, and the provisional possibilities $P(x / c_i)$, which are the possibilities of discovering this pattern in class c_i . $P(x)$ is the probability that a pattern x is there all through the data. $P(x)$ can be resulted from the total probability theorem as

$$P(x) = \sum_{i=1}^N P(x / C_i)P(C_i) \dots\dots\dots \text{eqn 2}$$

Using eqn 1 and eqn 2, the web objects are retrieved from the web cache and pre-fetch (web contents). After identifying the web objects class patterns, it is necessary to calculate the frequency of web objects occurring in the user web navigation system. The frequency factor is significant pointer for forecasting the revisiting of Web objects. In the proposed Bayesian classifier for user navigation, it is enhanced by

incorporating the accumulative scores or possibilities $W(g)$ that object g will be recalled in the future depending on BN classifier as shown as

$$k(g) = L + w(g) * \frac{C(g)}{S(g)} \dots\dots\dots \text{eqn 3}$$

Where
 g- Web object
 K (g) – key value for web object
 L - Aging factor
 C (g) - cost of enticing object g from the server into the cache
 S (g) - size of object g

Like eqn 3, the Bayesian classifier will classify the object patterns based on the users visit on the web page. In an existing cluster based approach, web object filters are sued for removing the redundancy of the web object presents. In the proposed scheme, the web objects are removed from the cache based on the least recently used web objects (LRU). If the user rarely visits the web page, then that web page content has been removed from the cache.

3.2 Validation Process

The validation process carries over for both the user and the web product to check whether a secure transaction is processing on. The validation process for this proposed work takes two approaches. The first approach is to check the validity of the user. The second approach is to estimate the validity of the products.

The user validity is identified based on the unique identification number. The proposed validation process assign id for the user based on the users' valid e-mail id like gmail, ymail, etc.,. The user is considered as valid user, only if he has a valid e-mail id. So, the user validation takes place based on mail id and individual user id.

The product validity checking is as same as like the user validity checking. The product validity is done based on the individual product id assigned by the authorized organization whoever sale the product in an appropriate manner.

In the proposed validation check for Bayesian classifier based web object filters for web user navigation, the validation process becomes simple and ease to check the validity, since it classified the web objects already based on the Bayesian classifier technique. After classification, a unique id is assigned to each user and product. The web page is retrieved from the domain only when the user and the product are valid. If any one among the two is invalid, the requested web page could not be displayed. Compared to an existing cluster based object filters, the proposed Bayesian classifier object based filtering provides a good outcome with an efficient validation checking. The algorithm below describes the over all Bayesian classifier process with validation crossing points.

Algorithm

Input: Web objects W

Step 1: Initialize the web objects
Step 2: For each W
Step 3: Evaluate the factor for W
Step 4: Evaluate the overall bandwidth consumption
Step 5: End for

// Web cache and web pre-fetch

Step 6: Do
Step 7: Select largest increasing factor
Step 8: Highlight it as "pre-fetched"
Step 9: While (n web objects enters)
Step 10: Highlight others as "Non-pre-fetched"
Step 11: Do
Step 12: Select from n objects having longest lifetime
Step 13: Highlight the object as "cache"
Step 14: while (N web objects enters)
Step 15: Highlight the other objects as "Non-cache"
Step 16: End do

//Bayesian classifier based user navigation

Step 17: Apply Bayesian classifier to pre-fetched and cache objects
Step 18: Classify the number of web objects based on posterior probability,
Refer eqn 1, 2, 3
Step 19: List the classified web objects
Step 20: If W met with Non-cache and Non pre-fetched
Step 21: Discard the web object from the classification
Step 22: End if

// Validation

Step 23: For each W
Step 24: If W – user
Step 25: Assign a unique identification number if he has valid email id in gmail, ymail, rediffmail etc.,. assigned as valid
Step 26: Else
Step 27: Predict the user as invalid

Step 28: End If
Step 29: If W- product
Step 30: Assign a product id by a reputed industry and checked as valid
Step 31: else
Step 32: Invalid product
Step 33: End if
Step 34: retrieve the web page contents from web cache and web pre-fetch only
If the user and the product is valid
Step 35: End

Using the above algorithm, the entire classification process takes place for web user navigation system. The validation process carries over for both the proposed BC and an existing CBWF to estimate the best approach for user navigation system.

3.3 Cluster Based Web Object Filters

The objective of Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme is to evaluate the web user navigation patterns and user references of product search. The overall process consists of five phases.

The first phase consists of initialization phase. The phase initializes the web object. Web objects are initialized where O denotes the set of web objects, h_i denotes hit value of object i and b_i denotes bandwidth object i. Let n denote the number of objects to be pre-fetched.

The second phase consists of web pre-fetching and caching phase. During this phase, pre-fetching is performed using hit rate and bandwidth rate and caching is performed using lifetime. The overall factor is evaluated using hit bandwidth rate. The next step is to evaluate the increase factor in order to determine the pre-fetched object and non pre-fetched object. The objects having largest increase factor and the objects that have maximum life time are determined as pre-fetched and cached objects.

The third phase involves with the cluster formation. During this phase Pre-fetched and cached documents obtained from web pre-fetching and caching phase are given as input to form cluster. Clusters are forming using the Euclidean distance. Based on the minimum distance the objects are assigned and accordingly the web objects clusters are formed.

The last phase consists of filtering and evaluation. Here cluster based web objects are given as input. Similarity items are evaluated which in turn discards web pages generated by unique users and also filter uncacheable requests. The output obtained is the filtered web object.

User Navigation of web object is evaluated from the web cluster objects with similarity retrieval in subsequent user sessions using cosine based similarity. Web Object Filters are built with the interpretation of the cluster web pages related to the unique users by discarding redundant pages.

Experimental evaluation of proposed Cluster based Web Object Filters from Web Pre-fetching and Web caching scheme is carried out using the java platform. The performance is measured in terms of Objective function, Number of clusters and cluster accuracy.

4. EXPERIMENTAL EVALUATION

To evaluate the proposed Classifier based Web Pre-fetching and Web caching scheme, we conceded out experimentation that the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation evolves the classification result in terms of the values of web objects contents. The result of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation are compared with the previous cluster based web object filtering and discussed the validation process in order to present an impending on their applicability and consequence. The experimentation was behavioral supported on two separate sources of log files. The first source log file substantiates the web users' navigational behavior on an academic-oriented host (AUTH CSD Department site), while the second source log file proceedings web users' calls on a general public, most eminent server.

We conceded out experimentation depends on two log files ensuing from the primary source and referring to different time periods. More principally, the most important log file preserves records over a two-month, and its size is concerning 120 Mbytes. The next CSD log file preserves records over an eight-month period, and its size is about 480 Mbytes. The third log file, consequential from the Web server, sustains records over a one-month period, and its size is concerning 200 Mbytes. Web Pre-fetching and Web caching scheme leads the classification process and engages data filtering, which filters any log entry that is having the unnecessary web object. Thus, the beginning log entries have been considerably reduced so as to effort and be supportive for the classifying information. The performance of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation is measured in terms of

- i) Efficiency of classification,
- ii) Validation,
- iii) Performance rate

5. RESULTS AND DISCUSSION

From this work, we have seen that the classification of web objects are done efficiently based upon the user requests than the cluster based system for web based user navigation process. In the proposed scheme, the Bayesian classifier is used for the prediction of class of web objects which are retrieved from the domains. After classification technique, the

No. of web objects	Classification efficiency (%)	
	Proposed BC	Existing CBWF
10	12	5
20	20	10
30	30	16
40	45	20
50	56	25

Table 5.1 No. of web objects vs. Classification efficiency



Fig. 5.1 No. of web objects vs. Classification efficiency

validation process is done for both the proposed Bayesian classifier (BC) and an existing cluster based web object filters (CBWF) for web user navigation. The table and graph describes the performance and the best approach for web based user navigation based on the validation checking points.

The above table (table 5.1) describes the efficiency of classification based on the number of web objects retrieved by the user. The outcome of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation is compared with an existing cluster based web object filtering technique.

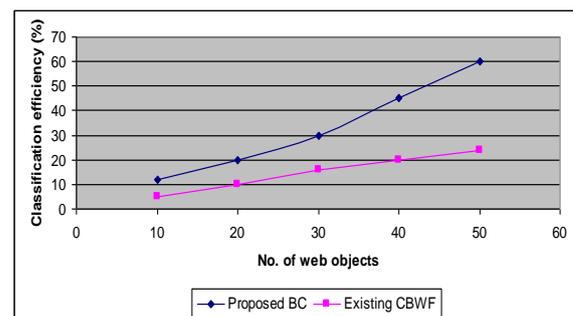


Fig 5.2 No. of web objects vs. Classification efficiency

Fig 5.2 describes the efficiency of classification based on the number of web objects visited by the user. The proposed web based user navigation scheme used Bayesian classifier where the classification of web objects is done based on the probability taken over with the users visiting page. The redundancy of the web objects are also being discarded reliably based on the obtained less objective function raised over the web page contents. The classification efficiency is measured in terms of how fast the classification of web objects is made with no redundant web objects. Compared to an existing cluster based web object filter, the proposed Bayesian classifier based user navigation system classifies well and the variance is 50% high in the proposed BC.

No. of users	Validation (%)	
	Proposed BC	Existing CBWF
10	30	10
20	45	15
30	60	20
40	75	24
50	80	30

Table 5.2 No. of users vs. validation

The above table (table 5.2) describes the validation checking point based on the number of users present in the web based navigation system. The outcome of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation is compared with an existing cluster based web object filtering technique.

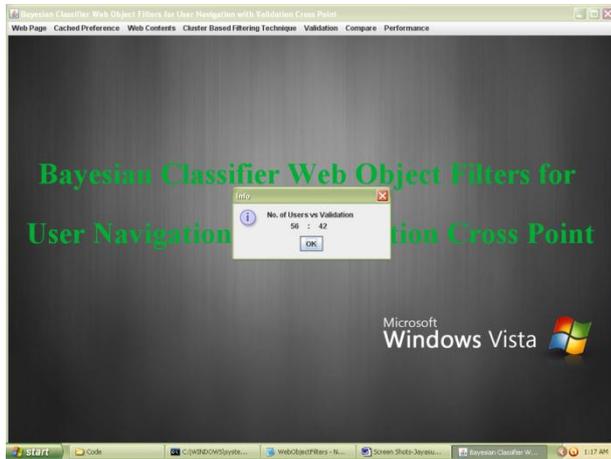


Fig. 5.3 No. of users vs. validation

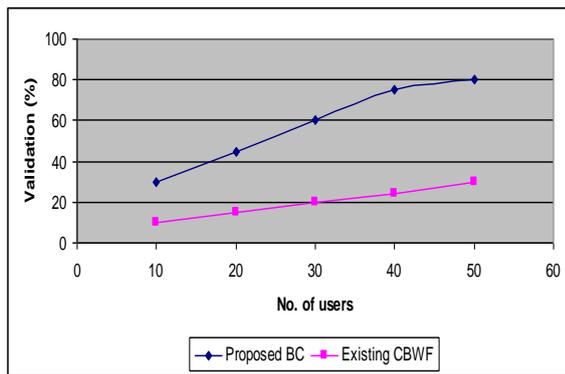


Fig. 5.4 No. of users vs. validation

Fig 5.4 depicts the validation process of the user based on the number of users involved in the retrieval process. The proposed Bayesian classifier process provides a validity of the user based on assigning the unique identification number to both the user and the product. The proposed scheme retrieves the web objects to the user only if both the user and the product have valid id. The valid id is given to the user only if he has a valid mail id. The product id is applied only if the reputed organization authorized it. So the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation provides valid web site product information to the authorized user depends upon the valid id. Compare to an existing cluster based web object filter, the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation performs a reliable validation.

The below table (table 5.3) describes the performance of the schemes based on the web objects present in the web based navigation system. The outcome of the proposed validation check for Bayesian classifier and cluster based web object filters for web user navigation is compared with an existing cluster based web object filtering technique.

Size of web objects	Redundancy rate	
	Proposed BC	Existing CBWF
10	5	12
20	7	18
30	11	24
40	12	28
50	13	32

Table 5.3 Size of web objects vs. performance rate

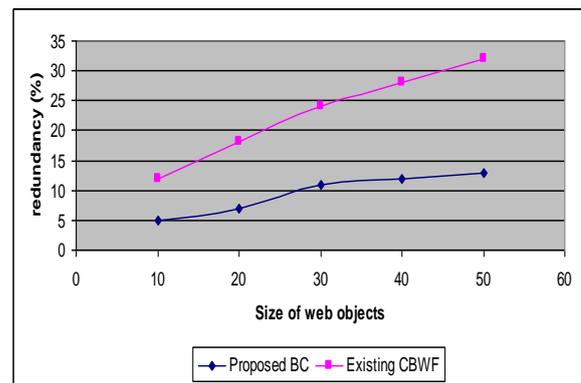


Fig 5.3 size of web objects vs. performance rate

Fig 5.3 describes the performance discussion of both the proposed BC and an existing CBWF when more number of web objects enters. Compared to an existing cluster based, the proposed Bayesian classifier for web cache and web pre-fetch provides a better classification of the objects by providing a unique identification number to each user. Finally, it is being depicted that the proposed scheme provides an efficient classification process by providing a unique identification number for an authorized transaction among the users and the web pages.

6. CONCLUSION

In this paper, we have presented an efficient Bayesian classifier based Web Object Filters from Web Pre-fetching and Web caching scheme to estimate the web user navigation prototypes and user references of product search. The proposed Bayesian classifier based Web Pre-fetching and Web caching scheme competently incorporates Web caching and pre-fetching contents. Classifying was done with the requested web page objects obtained from pre-fetched and web cached contents using the probability models. Web Object Filters have built with the interpretation of the classified web pages related to the distinctive users by discarding redundant pages. Classification based Web Object Filter scheme evaluates the User Navigation from the web classified objects with similarity retrieval in successive user sessions. Experimental evaluation of proposed Bayesian classifier based Web Object Filters from Web Pre-fetching and Web caching scheme is carried out using the java platform and the results shown that the proposed BC outperforms well by validating each user and product by a unique identification number.

7. REFERENCES

- [1] Teng WG, Chang CY, Chen MS (2005). Integrating Web caching and Web prefetching in client-side proxies. *IEEE Trans Parallel Distributed Syst*;16(5):444–55.
- [2] J. B. Patil et. al. (Feb 2011), “Integrating intelligent predictive caching and static prefetching in web proxy servers”, *International Journal on Computer Science and Engineering (IJCSSE)*, Vol. 3 No. 2 .
- [3] J.N. Kok et al. (2007), “IKNN: Informative K-Nearest Neighbor Pattern Classification”, (Eds.): *PKDD 2007, LNAI 4702*, pp. 248–264.
- [4] J. B. Patil and B. V. Pawar (2008) , “GDSF#, A Better Algorithm that Optimizes Both Hit Rate and Byte Hit Rate in Internet Web Servers”, *International Journal of Computer Science and Applications*, ISSN: 0972-9038, Volume 5, Number 4, Pages 1-10.
- [5] Sulaiman, S., Shamsuddin, S. M., Forkan, F., Abraham, A.(2008): Intelligent Web caching using neurocomputing and particle swarm optimization algorithm. *Second Asia International Conference on Modeling & Simulation (AICMS08)*.
- [6] Ali, W., Shamsuddin, S.M., Ismail, A.S(2011): A survey of Web caching and prefetching. *Int. J. Advance. Soft Comput. Appl.* 3(1), 18-44 .
- [7] Oliveira, L.S.C., Andreão, R.V., Sarcinelli-Filho, M.(2010): The Use of Bayesian Networks for Heart Beat Classification, in *Brain Inspired Cognitive Systems 2008*, A. Hussain, et al., Editors. 2010, SpringerNew York. p. 217-231.
- [8] Kumar, C.(2009): Performance evaluation for implementations of a network of proxy caches. *Decision Support Systems* 46(2), 492–500.
- [9] Waleed Ali, Siti Mariyam Shamsuddin et. al (2012) ., “Intelligent Bayesian Network-Based Approaches for Web Proxy Caching”, *Journal on knowledge based systems*, Volume 31, July,
- [10] Kaya, C.C., Zhang, G., Tan, Y., Mookerjee, V.S.(2009): An admission-control technique for delay reduction in proxy caching. *Decision Support Systems* 46(2), 594–603.
- [11] Ali, W., Shamsuddin, S.M., Ismail, A.S.(2011): Web proxy cache content classification based on support vector machine. *Journal of Artificial Intelligence* 4(1), 100–109
- [12] Ali Ahmed, W., Shamsuddin, S.M.(2011): Neuro-fuzzy system in partitioned client-side Web cache. *Expert Systems with Applications* 38(12),14715-14725.