Energy Stability in Cloud for Web Page Ranking

Sutirtha Kumar Guha, Anirban Kundu and Rana Dattagupta

Abstract We are going to propose a new approach to implement predictive Web page ranking in cloud environment. User query is transferred to the server side of cloud based search engine. Searching and ranking procedures have been executed on the server side and final output is sent to the client as a response. In this paper, prediction task has been performed based on user behaviour. Available data, resources, and functional support systems have been processed using schedulers. Session, duration, and other user relevant data are combined with the processed information for prediction. Different Web based data repositories are considered as cloud structure for energy stabilization.

Keywords Web page ranking · Predictive ranking · Data cloud · Resource cloud · Support cloud · Information bag controller · Cloud session · Energy stability measurement

S.K. Guha (✉)
Seacom Engineering College, Howrah, West Bengal, India
e-mail: sutirthaguha@gmail.com

A. Kundu
Netaji Subhash Engineering College, Kolkata, West Bengal, India
e-mail: anik76in@gmail.com

R. Dattagupta
Jadavpur University, Kolkata, West Bengal, India
e-mail: ranadattagupta@yahoo.com

S.K. Guha · A. Kundu
Innovation Research Lab, Howrah, West Bengal, India

S.C. Satapathy et al. (eds.), *Information Systems Design and Intelligent Applications*, Advances in Intelligent Systems and Computing 434,
DOI 10.1007/978-81-322-2752-6_61

1 Introduction

1.1 Overview

Web is considered as an important system to acquire and gather important information in our daily life. Web area is congested due to ever increasing dependability over Web. Introduction of search engine is considered as an important tool for indexing and tagging that growing Web database. Different Web pages are connected with the search engine. Matched URLs are displayed based on user query. Matched Web pages are ranked based on different algorithms. More accurate and fast ranking is achieved by predictive ranking approach. Deterministic ranking or searching is achieved based on previous user data.

1.2 Related Work

A method is introduced in [1] to predict the ranking position of a Web page. A set of successive past top-k rankings are presumed and evolution of Web pages is analysed in terms of ranking trend sequences used for Markov Models training [1]. An algorithm is proposed in [2] for conducting web page access prediction. The use of page rank algorithm is extended to predict next page with several navigational attributes [2]. Ranking in cloud environment is proposed in [3]. The ranking operation is performed based on the data as mentioned in [4]. Cloud classification is analysed by accessibility on the specific cloud in [5]. Cloud classifier is analyzed structurally in [6]. SaaS (Software-as-a-Service) is introduced for user specific application using heterogeneous distributed system [7]. K-Erlang distribution method is implemented in cloud computing to enhance the performance [8]. Different self organizational characteristics of intelligent cloud are discussed in [9]. Load balancing in cloud is proposed in [10].

Rest of the paper is organized as follows: Proposed system design is depicted in Sect. 2 along with detailed analytical studies. Experimental results have been depicted in Sect. 3. Section 4 concludes the paper.

2 Proposed Work

Prediction based ranking is introduced in this paper to achieve efficient searching and ranking in cloud environment. Operational activities of proposed system are shown in Fig. 1.

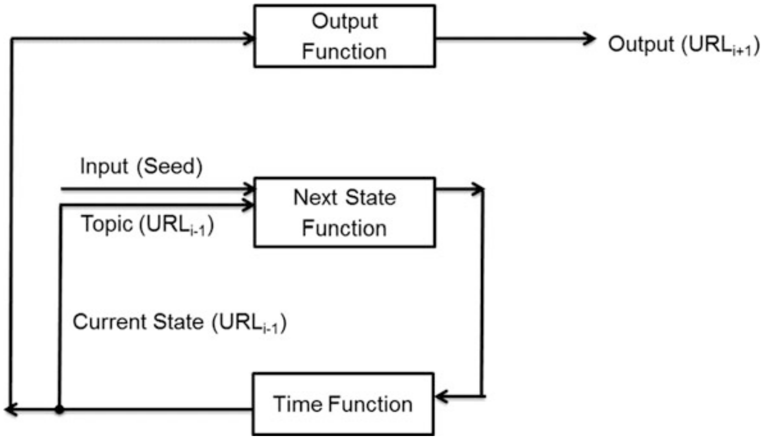


Fig. 1 State diagram of proposed system

Input from user and previous state (URL_{i-1}) are combined and processed through ‘Next State Function’. This processed information is sent to ‘Time Function’. ‘Time Function’ sends the information to ‘Output Function’ to produce predicted output (URL_{i+1}) and ‘Next State Function’ as URL_i. Operational flow-chart of our proposed system is shown in Fig. 2.

Operational activities of different segments of proposed system are discussed in next sub sections.

Definition 1: **Data Cloud** Information such as text data, image data, audio data, and video data are stored in a predefined database in server side. This predefined database is addressed here as “Data Cloud”.

Definition 2: **Resource Cloud** Client request is served in server side to provide efficient and accurate performance. Different hardware and software resources are required to execute this operation. These hardware and software resources are resided in “Resource Cloud”.

Definition 3: **Support Cloud** Different support modules are resided in this part of the cloud.

Definition 4: **Information Bag Controller** Data from “Data Cloud”, “Resource Cloud” and “Support Cloud” are sent and merged to perform the job in co-operative manner. Integration and control of data from these three clouds are executed in a controller named as “Information Bag Controller”. Raw processed data is sent to the next module as web content that is a combination of required information and unwanted noise.

Definition 5: **Cloud Mining** Received data from ‘Information Bag Controller’ is processed to eliminate the noise. Processed information is then searched according to client side request and analyzed.

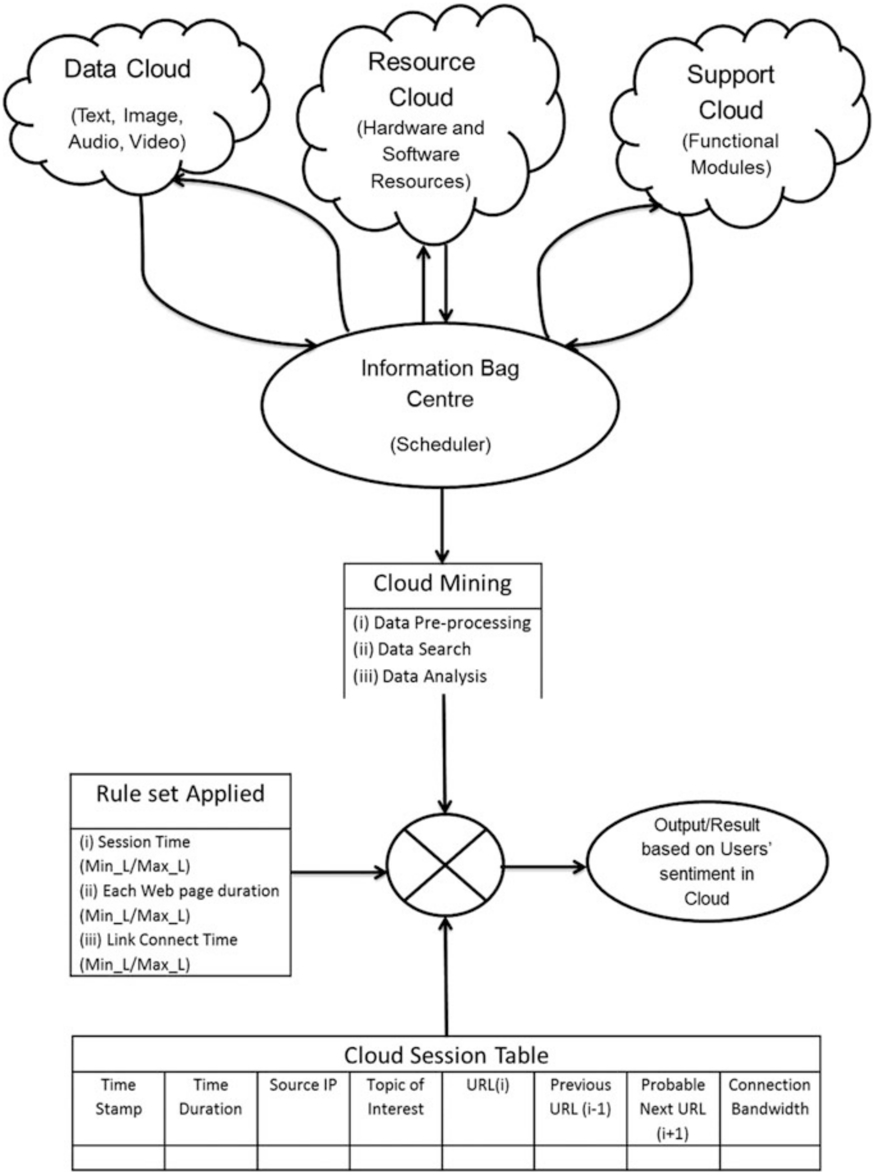


Fig. 2 Operational flowchart of proposed system

2.1 Rules Applied in Cloud

Set of rules is applied into the processed information. Major rules are as follows:

(a) Session Time (Min_L/Max_L) Visitor session on that particular URL is addresses as "Session Time". Upper and lower limit is defined as Max_L and Min_L.
(b) Each Web page duration (Min_L/Max_L) Life time of the Web page is named as "Web page duration". It ranges in between Max_L and Min_L.
(c) Link Connect Time (Min_L/Max_L) It is considered as the time required to connect the Web page. It resided between Max_L and Min_L.

2.2 Cloud Session Table

Different information of cloud is stored in "Cloud Session Table" for predicting user's requirements using time stamp, duration, source IP, topic of interest, URL(i), previous URL(i − 1), probable next URL(i + 1), and connection bandwidth.

2.3 Energy Stability in Cloud

Energy stability measurement (S_M) of the proposed system depends on the following factors:

C_S Cloud session;
S State;
S_t State transition;
T Time;
S_e Seed (Starting Point);
E_C Energy of cloud;

∴ Cloud Energy, E_C depends on $S_M(S_e, S, S_t, T, C_S)$
∴ Energy flow in a Cloud is considered as follows:

$$\frac{d}{dt}(E_C) = S_M(S_e, S, S_t, T, C_S) \tag{1}$$

Applying integration on both sides of Eq. (1),

$$\int \frac{d}{dt}(\mathrm{E_C})dt = \int \mathrm{S_M}(S_e, S, S_t, T, C_S)dt$$
$$\therefore \int \mathrm{E_C} = \int \mathrm{S_M}(S_e, S, S_t, T, C_S)dt \tag{2}$$

On the other hand, cloud session is represented as follows based on "Cloud Session Table" mentioned in Table 1:

$$C_S = f(T_S, T_D, S_{IP}, T_I, URL_i, URL_{i-1}, URL_{i+1}, B_W) \tag{3}$$

where,

T_S = Time stamp; T_D = Time duration; S_{IP} = Source IP; T_I = Topic of Interest; URL_i = Current URL; URL_{i-1} = Previous URL; URL_{i+1} = Next URL; B_W = Bandwidth;

Here, URL represents a Web page; and, URL_{i+1} is always probabilistic as it depends on the users' choice in real time.

Combining Eqs. (2), (3),

$$\int E_C = \int S_M(S_e, S, S_t, T, f(T_S, T_D, S_{IP}, T_I, URL_i, URL_{i-1}, URL_{i+1}, B_W))dt \tag{4}$$

$S_M(S_e, S, S_t, T, C_S)$ is realised as follows:

Data flow execution of Fig. 3 is depicted in Algorithm 1.

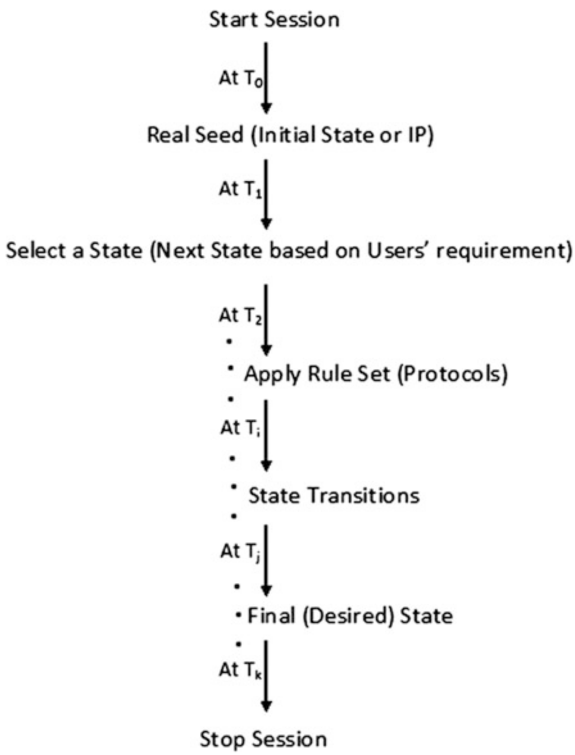
Algorithm 1: Energy_stability_measurement

Input: Initial State
Output: Next desired State

Step 1: Start session.
Step 2: Initial state is considered as input.
Step 3: Next state is selected based on users' requirement.
Step 4: Session time, Web page duration, Link connect time are applied on selected next state.
Step 5: State transitions are executed.
Step 6: Stop session

Data flow is started from initial state. Next probable state is selected based on users' requirement at next time stamp. Defined protocol sets are then applied on probable next state. State transitions are then executed to get desired state.

Fig. 3 Data flow chart of energy stability measurement (S_M)



Similarly, cloud session (C_S) consists of a cycle of operations initiated by particular users.

$C_S = f(T_S, T_D, S_{IP}, T_I, URL_i, URL_{i-1}, URL_{i+1}, B_W)$ is realized as follows:

Data flow execution of Fig. 4 is depicted in Algorithm 2.

Algorithm 2: Cloud_session

Input: User IP
Output: Bandwidth usage

Step 1: Start session.
Step 2: Source IP is considered as input.
Step 3: Topic of interest is searched in present Web link (URL_i)
Step 4: If, topic of interest is found in present Web link (URL_i), then go to Step 5. Otherwise, go to Step 7.
Step 5: Next Web link (URL_{i+1}) is predicted based on present Web link (URL_i) and previous Web link (URL_{i-1}).
Step 6: Total time duration is calculated and go to Step 2.
Step 7: Bandwidth usage for particular case is checked.
Step 8: Stop session.

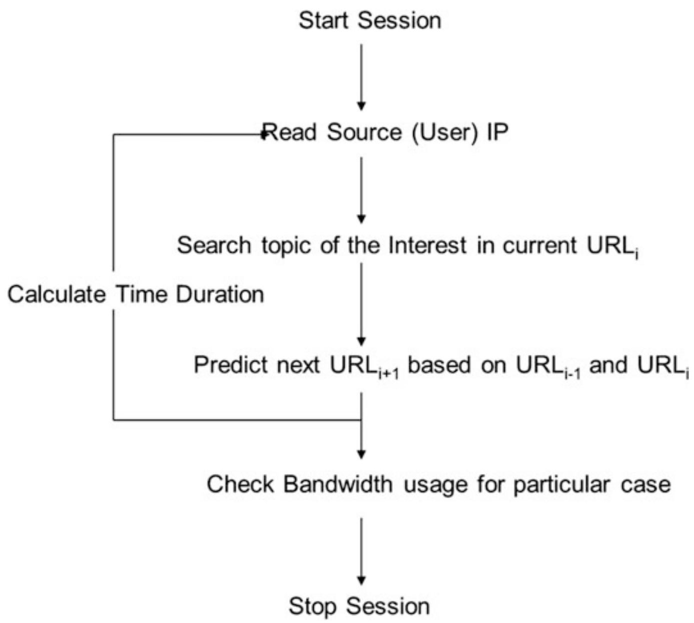


Fig. 4 Data flow chart of cloud session (C_S)

Source IP is considered as input. Matched topic is searched in present Web page (URL_i). Next Web page is predicted based on present (URL_i) and previous Web page (URL_{i-1}). This process is repeated for every matched topic in present Web page. Total bandwidth is measured and considered as cloud session.

3 Experimental Results

Different Web pages are selected at a particular time instance. Web pages from same database based on same query are selected by the same user at different time instances. Prediction success of our proposed method is shown in Fig. 5.

It is evident from the experimental data that accurate and efficient prediction is achieved by our proposed approach.

Number of Users	Number of Age Groups	Number of Time instances
10	02	05

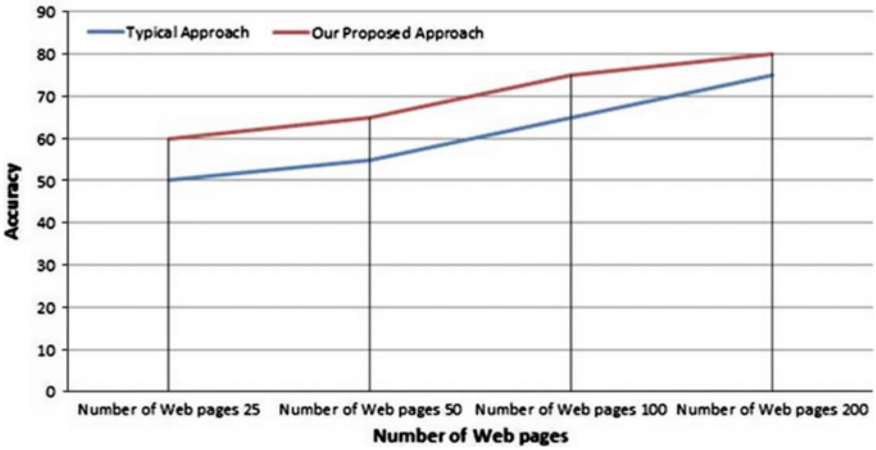


Fig. 5 Prediction success comparison based on different age groups user

4 Conclusion

In this paper, we have proposed a prediction based ranking approach in cloud. Different Web page related information such as data set, resources are considered to predict next Web page selection. Collected data are processed and analysed to produce successful prediction in cloud. Various rules are implemented on data set to achieve better outcome. User query is sent to the server to acquire most feasible result. Our proposed approach is performed on server side based on user sentiment using energy measurement. More accurate prediction is achieved as the procedure is dynamically operated based on previous user data. Hence, accuracy is increased as wider range of data is considered.

References

1. Vazirgiannis, M., Drosos, D., Senellart, P., Vlachou, A.: Web Page Rank Prediction with Markov Models. In: 17th International World Wide Web Conference, Beijing, China, April 21–25, 2015, pp. 1075–1076 (2015).
2. Thwe, P. : Proposed Approach For Web Page Access Prediction Using Popularity And Similarity Based Page Rank Algorithm. In: International Journal of Scientific & Technology Research, Volume 2, Issue 3, March, 2013, pp. 240–246 (2013).
3. Sampath, P., Ramya, D.: Performance Analysis Of Web Page Prediction With Markov Model, Association Rule Mining(Arm) And Association Rule Mining With Statistical Features (Arm-Sf). Journal of Computer Engineering (IOSR-JCE), Volume 8, Issue 5, pp. 70–74 (2013).

4. Kundu, A., Xu, G., Ji, C.: Data Specific Ranking in Cloud. International Journal of Cloud Applications and Computing (IJCAC), Vol. 4, Issue 4, pp. 32–41 (2014).
5. Kundu, A., Xu, G., Ji, C.: Analysis on Cloud Classification using Accessibility. International Journal of Cloud Applications and Computing (IJCAC), Vol.4, Issue 3, 2014, pp. 44–53 (2014).
6. Kundu, A., Xu, G., Ji, C.: Structural Analysis of Cloud Classifier. International Journal of Cloud Applications and Computing (IJCAC), Vol. 4, No. 1, pp. 63–75 (2014).
7. Kundu, A., Ji, C., Liu, R.: Software-as-a-Service using Heterogeneous Distributed System for User specific Applications. International Journal of Cloud Applications and Computing (IJCAC), Vol. 4, No. 1, pp. 15–32 (2014).
8. Banerjee, C., Kundu, A., Agarwal, A., Singh, P., Bhattacharya, S., Dattagupta, R.: Priority based K-Erlang Distribution Method in Cloud Computing. International Journal on Recent Trends in Engineering & Technology, Vol. 10, No. 1, pp. 01–11 (2014).
9. Kundu, A., Xu, G., Ji, C.: Self Organization Behavior of Intelligent Cloud. Journal of Convergence Information Technology (JCIT), Vol. 8, No. 16, pp. 39–47 (2013).
10. Kundu, A., Xu, G., Liu, R.: Efficient Load Balancing in Cloud: A Practical Implementation. International Journal of Advancements in Computing Technology (IJACT), Vol. 5, No. 12, pp. 43–54 (2013).