Asian Journal of Information Technology 15 (17): 3280-3286, 2016
ISSN: 1682-3915

Markov Model Based Prediction for Effective E-Content Delivery in Cloud

G. Malini, T. Mala and A. Kannan
Department of Information Science and Technology, CEG Campus,
Anna University, 25 Chennai, India

Abstract: In the past few years, E-learning has become one of the most sought powerful twenty first century tools. There is a need to redesign the current educational system to meet the internet and network based technology enabled education using e-learning. In this study, a new approach for cloud based content delivery is proposed in which the E-learning content is stored in a distributed manner across the cloud, thereby providing easier and faster access of course materials, along with fault tolerance feature. In this model, the future content requests of the users are predicted in advance, depending on the history of requests made by the users with the help of Markov model based prediction and are provided. The predicted content, if not available in the user's cloud site storage is replicated. Through the experiments carried out in this research, it has been shown that the system is scalable for large number of users and the prediction deployed achieves minimum access time and response time of the users.

Key words: Learning object, SCORM, cloud computing, HMM prediction, content delivery

INTRODUCTION

E-learning systems that encompass online learning, employee training courses and e-books have been grabbing global limelight, becoming part and parcel of the modern technology. They provide learners the comfort of studying at any time and at any location they want. The teaching materials in different E-learning systems are usually defined in specific data formats and hence, the sharing of teaching materials among these systems is a daunting and difficult task (Tseng and Chen, 2015). As a result, the task of creating teaching materials is quite expensive. So, with an objective of solving this problem of teaching material format, several standard formats, including SCORM, IMS, LOM, AICC etc. have been proposed by international organizations (Ochoa and Duval, 2009). These standard formats have made it possible to share, reuse and recombine teaching materials in different learning management systems. Recently, at SCORM, ADL unveiled plans for the Content Object Repository Discovery and Resolution Architecture (CORDRA) which is designed to serve as a reference model and is driven by an identified need for contextualized learning object discovery.

The distribution of E-learning resources facilitates passing of the teaching experiences of instructors and provides various E-learning resources for the targeted users and consequently, the targeted users refer to learners, instructors, domain experts and so on (Alabbadi, 2011). The trend involving targeted users, E-learning resources and learning services in Internet is on an upswing. However, on the flipside, this trend has thrown up new E-learning research issues such as management, provision and saving of E-learning resources. Here comes cloud computing environment which provides an appropriate solution to these issues (Chaudhary and Saxena, 2015). It makes sure that users use the computing resources on demand and pay money according to their usage on a metering pattern similar to that for water and electricity consumption. Hence, a new business model is born, where computing resources are the main services. The next session gives an over view of the survey of literature which has been done exhaustively for developing ideas from the previous researcher of people.

In this study, a new Markov Model based prediction method is proposed to know the learners requirements in advance and to provide suitable E-learning contents to the learner. This research has been carried out in the cloud platform in order to increase the scalability and availability. The main advantage of the proposed work is that it provides relevant contents with high reliability and consumes less time for retrieval of suitable contents.

Literature review: There are many researcher (Wu *et al*., 2011; Tseng and Chen, 2015; Chaudhary and Saxena, 2015; Sampson *et al*., 2012; Senthilnayaki *et al*., 2015) on E-learning which are present in the literature. Among

Corresponding Author: G. Malini, Department of Information Science and Technology, CEG Campus, Anna University, 25 Chennai, India

them, learning object model is a new paradigm which has been developed for disseminating knowledge and increasing interoperability through metadata (Sampson *et al*., 2012). In their model, each learning object has its sole and independent function that can be integrated into larger components and provides reusability. An ontology based framework for storing and retrieving the E-learning contents is proposed by Senthilnayaki *et al.* (2015).

Malik *et al*. (2016) proposed a new Access Control Model for Data Stored on Cloud Computing to enhance the security. Many researchers proposed techniques for building Content Delivery Networks (CDN) using cloud environment which is more suitable for E-learning (Chen *et al*., 2012; Lin *et al*., 2011; Wang *et al*., 2011). In addition, a few authors worked on methodologies for efficiently storing and retrieving E-learning contents in cloud (Chaudhary and Saxena, 2015; Malini and Mala, 2013; Alabbadi, 2011; Phankokkruad, 2012). Deborah *et al*. (2014) provided a survey of important works on Learning styles assessment and theoretical origin in an E-learning scenario. In addition, new metrics and cloud based E-learning methods have been proposed by various researchers in the past (Deniel Ani *et al*., 2015).

Markov Model is an important technique which is used for classification and prediction (Mabroukeh and Ezeife, 2009) with semantic content analysis. Awad and Khalil (2012) proposed a new model for the Prediction of User's Web-Browsing Behavior using Markov model and hence can be used for predicting the learners interest in E-learning environment. Malini *et al*. (2013) proposed intelligent techniques using semantic analysis for multi-lingual education using the cloud computing platform. Their research is more suitable for accessing the contents in Tamil and English with mobility features. However, all these metrics focused on.

MATERIALS AND METHODS

Proposed framework for E-learning system on cloud

Content creation component: The overall architecture diagram of the system is shown in the Fig. 1. The first step is to create learning objects, which are the basis of E-learning content and which will be used in our educational cloud setup. The learning objects are created according to the SCORM standards. SCORM is an acronym of Sharable Content Object Reference Model. The SCORM is a collection and harmonization of specifications and standards that define the interrelationship of content objects, data models and protocols in such a way that objects are made sharable across systems, thus promoting reusability and interoperability of learning content across learning management systems.

Information extractor: This information is extracted using the term-frequency inverse document frequency statistic. This clusters all documents and extracts the frequently occurring words which can thus be used as keywords. Keyword extraction is an important technique for document retrieval, webpage retrieval, document clustering and text mining and so on. These keywords are used during search and retrieval of the learning objects.

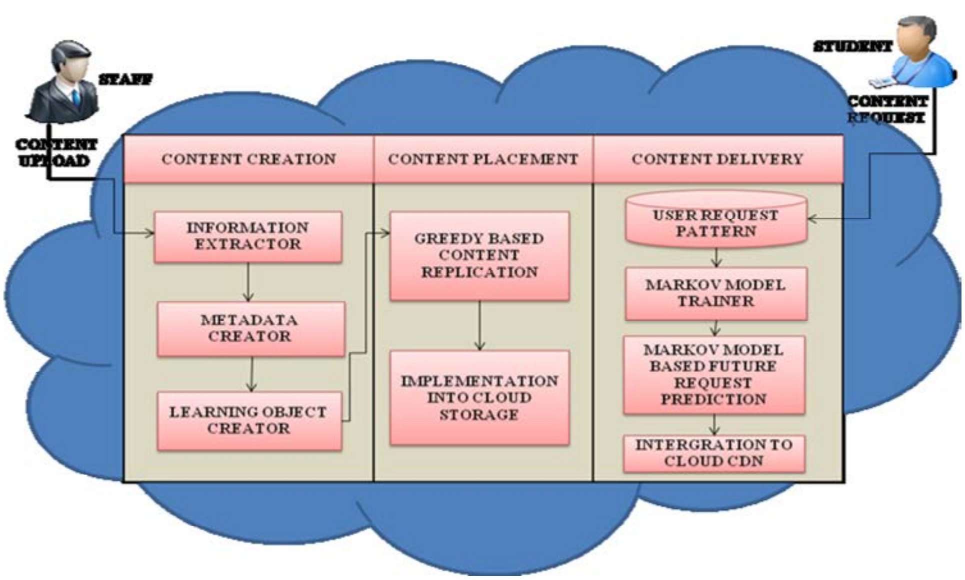

Fig. 1: Overall architecture of E-learning system on cloud

Metadata creator: Defining data with information and making it content is a process similar to the operations performed in everyday situations. Metadata is data about data, which defines the human aspect of content. Metadata first appeared on the web when the immense amount of data over the internet became impossible to process or to even understand. Every SCORM content package contains an imsmanifest.xml file in its root. The metadata, which is a part of the learning object, requires details such as title, keywords, description, author, date of the E-learning content.

Learning object creator: The SCORM learning object is a Package Interface File (PIF). It represents a unit of usable and reusable content. This may be part of a course that has instructional relevance outside a course organization and can be delivered independently, as an entire course or as a collection of course.

Content placement: This component builds Content Distribution Network, which makes use of the cloud for storing the created learning objects. It is aiming to minimize the storage cost at all cloud sites which are simulated with the help of various approaches in order to direct the content to the required data centre and thereby balancing load.

Greedy based content replication: The greedy based content replication is based on two factors, location and storage. Location information is dependent on the upload and downloads costs and the storage information is dependent on the ram space and external storage space needed. Given the total number of learning objects, their size, available number of cloud sites and number of hosts in each site and storage capacity and cost for storage of each site, this approach decides the content and the site where the content is to be placed. Since, this approach tries to minimize the cost incurred on storing the content at each site, this is called the "Greedy based approach" for content placement.

Two datacenters are created initially with varying storage costs, ram and storage size configurations. Here, the configuration of virtual machines required for storing the learning objects is set according to the size of the objects. If the objects are smaller in size, smaller VMs are needed and vice-versa. In our simulation, objects having a size of less than or equal to a threshold value of size requires a RAM storage of 1024 MB and objects having a size of greater than that requires a RAM storage of 2048 MB (Algorithum A).

Alogrithum A; Set the vm configuration requirements according to object size:

```
for each object_i
        if(size_i<=1000)
                ram_reqd=1024
        //create vm_small
                vm_size=size_i
        else
                ram_reqd=2048
        //create vm_large
                vm_size=size_i
```

Sort the data centers according to the storage cost. Next step is to decide in which site and host the object to be stored. Since, in a storage cloud setup where Storage as a Service is provided and one pays for what he stores, we aim to minimize the storage of e-content at each site. Since each site has different costs per memory and storage, we look at the storing the objects at the site, which has minimum storage cost (Algorithm B).

Algorithm B; date center:

```
For each datacenter d_j
   List<id_j >= min_cost(m_j + o_j ) //m_j =cost per memory,o_j = cost per storage
      remove d_j
  //Sorting the datacenters according to the costs of each d_j
```

Check storage availability for storing learning objects. The storage size of each host in each site is checked for the availability before storing objects of given size (Algorithm C).

Algorithm C; date center dj:

```
for each datacenter d_j in List <id_j >
        for each host h in d_j
                for each object l
        if(storage_h > size_l )
                if( ram_h > ram_l )
                        allocate vm to host h
                        store object l in host h, datacenter d_j
                next object
        else next host
    go to next datacenter
```

Implementation into cloud storage: The learning object content is then stored into the storage site of the cloud. An external SAN storage device is mounted into the cloud platform. The virtual machines, that are instantiated via the eucalyptus cloud platform are bound with SAN storage as when they are instantiated. Thus this serves as a storage site for the individual users which is provided by the cloud.

Content delivery component: This component is aimed at developing an application portal to upload the learning objects into a centralised storage and replicating them, reducing the access time of the requests for learning content by the students. To achieve the minimum access time, prediction of the future content request from the history of requests made by the student is done with the help of the Markov model. Hidden Markov Models will be incorporated to improve the delivery of the content based on individual user requirements. Since, the students' preference for learning content keeps on changing as the e-learning course content is delivered, it is important to keep track of these changing preferences in a real time environment. The delivery of course content is then adapted in a way which is most suited to the current learning content preference for student.

Collection of users requests for E-content: The students' requests are collected and the keywords of the metadata file are compared. Then identified corresponding metadata files and learning object files are stored in a database which acts as the training data for prediction of future content requests of the users. As and when the student makes requests for content, they are stored in what is called a history of requests.

Markov model trainer: The collection of requests of the students serves as a training data. That undergoes transitions from one state to another, among a finite or countable number of possible states. The transition probabilities between the various states are calculated and the next state is the one with the maximum transition probability. The transitions probabilities between various states of Markov model are used for calculating the transition probabilities between the various states (requests).

Markov model based prediction: Given the transition probability matrix and the current request of the student, the future request based on Markov chain is obtained. The predicted content has a higher probability of being the next which may or may not be present in that user's site. In this way, the access time is minimised when the user requests the next predicted content, which in the other case has to wait till the content is replicated only at the request time. In the next, complete implementation of the system is discussed and explained elaborately.

RESULTS AND DISCUSSION

The experimental data are collected through a cloud-based E-learning system for higher education students. The underlying assumption is that each student will have different preferences for the various learning activities and content available in the content database which form the discrete Hidden Markov Model (HMM)

Table 1: Access time with and without prediction

Learning object	File size (MB)	Access time with prediction	Access time without prediction
Stack	85.5	0.259	2.938
Queue	68.4	0.232	1.873
Linked list	67.4	0.241	1.845
B &B+ Tree	34.6	0.114	1.093
Heap	21.6	0.139	0.842
Avl trees	53.1	0.252	1.782
Binary search tree	62.2	0.223	1.903
Red black tree	44.9	0.201	1.538
Sorting	29.2	0.189	1.131
Splay tree	11.6	0.059	0.337

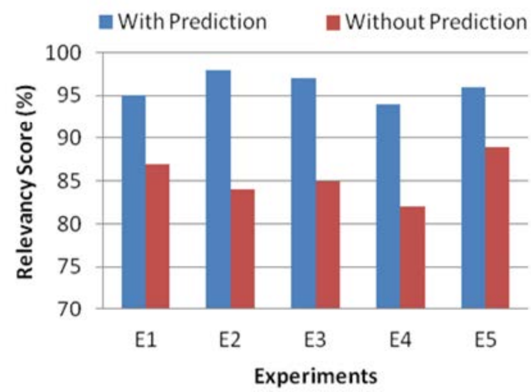

Fig. 2: The analysis of elevancy experiments

observation sequences in our experiment. With the support of our Anna University, 50 participant records are collected for 2 day. The students are ranked according to their Computer Science performance in the past one week and based on this ranking, a top ranking group of students are chosen. From each group, a number of students' requests are collected at random and their navigation sequences are used as the training data set while the others form the test data set. For each group of students under study, a corresponding HMM is constructed to model their navigation behavior. Random transition matrix initialization is employed to reach global optimal solution. Table 1 lists access time with and without prediction values.

Figure 2 shows the relevancy analysis for the E-learning contents retrieved using conventional methods and prediction based method. Figure 2, we can draw a conclusion that the accuracy is increased when the prediction technique is used, thereby replicating the content beforehand, instead of doing at the time of request.

Given the input as transition probabilities and the current request, the next request is predicted. This follows the Markov chain principle that the next state depends on the current state. The request which has maximum transition probability from the current request becomes the next request. If "stack" is the current request, the transition probabilities are as follows:

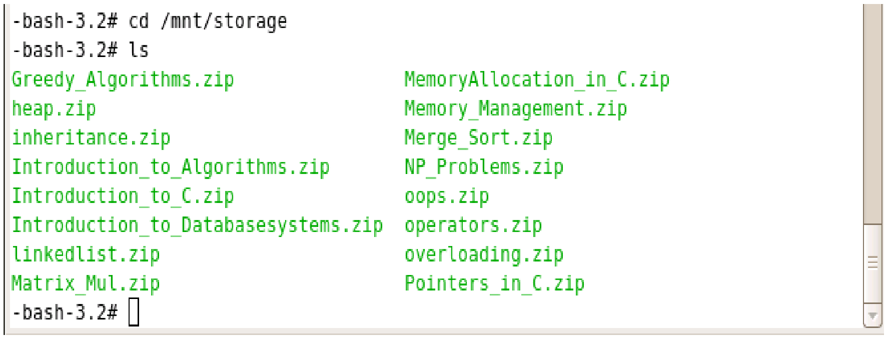

Fig. 3: List of contents before replication

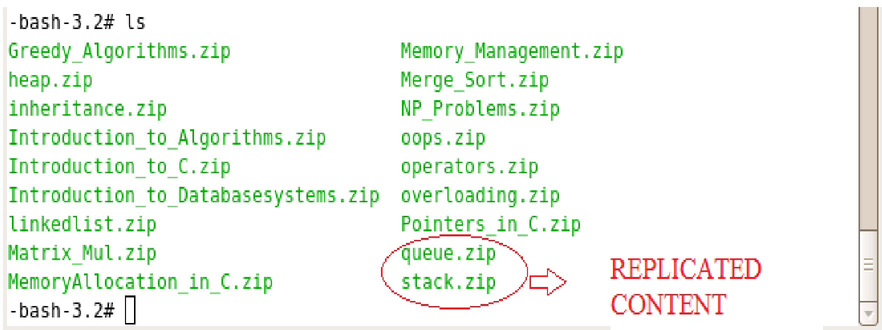

Fig. 4: List of contents after prediction and replication

- T (Stack→Queue) = Total number of transitions to state Stack to Queue = 37
- T(Stack→Heap) = Total number of transitions to state Stack to Heap = 7
- T(Stack→Array) = Total number of transitions to state Stack to Array = 3
- T(Stack) = Total number of transitions from state Stack = 50
  - Stack →Queue = 370/500 = 0.74
  - Stack→Heap = 70/500 = 0.14
  - Stack→Array = 30/500 = 0.06

"Queue" has the maximum transition probability. Hence, it is predicted as the next request of those students whose current request is "Stack". Suppose a student's requests "linked list" and the contents of storage in his site are as follows. Now after prediction, the contents of "stack" and "queue" which are probably the next requests are replicated (Fig. 3). Now, the contents are as follows.

The next request for learning object and the location is decided, based on the predicted next user requests and replicated within the site that is nearer to the user if the content is not found in the user's site (Fig. 4). This component achieves minimum access time that is the time taken for the requests to bring the content from the cloud storage is reduced as it brings the content from the user's site. As the user requests a learning object, the prediction algorithm is run dynamically at the backend, thus predicting the next possible request from the history of previous requests. The predicted content is checked whether it is already present in the user site's storage. If not, the content is replicated from the SAN storage into the user site's storage. By doing this, when the next user's request and the predicted request are the same, the access time of the content is reduced by replicating the content already in the user's site.

Content placement store different data sets of learning objects for storage has been simulated for our cloud based content delivery network. It has been observed that minimum storage access time has been achieved for the greedy based approach of storing the learning content, which selects the sites in the order of lesser storage access time, when compared we compared greedy approaches with normal approach and based on the result have been plotted as Fig. 5 to the normal approach of storage in the given order of cloud sites. The various conclusions and observations of this system and further areas of research and extensions are discussed in the next study.

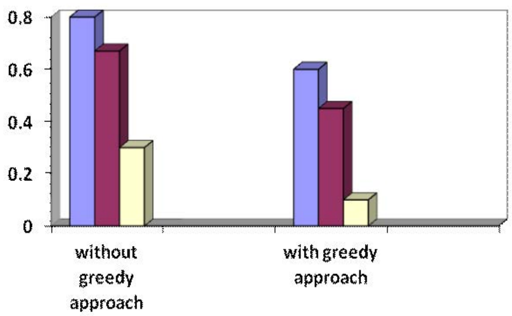

Fig. 5: Comparison of access time with and without greedy based approach

CONCLUSION

In this study, a novel student-instructor-friendly e-learning content delivery network has been developed. The salient features are its SCORM-based learning objects, providing reusability and interoperability, greedy based approach for storing them in cloud storage sites, which aims at minimizing the costs of storage, access of the e-content by the students and Markov Model based prediction which tries to predict the future requests to be made by the students and replicating that content near the user's site, thus minimizing the access time. In addition, new storage structures are used for effective data manipulation.

RECOMENDATIONS

Future works in this direction can be the use of map-reduce techniques to perform parallel retrieval in order to enhance the performance further.

REFERENCES

Alabbadi, M.M., 2011. Cloud computing for education and learning: Education and learning as a service (ELaaS). Proceedings of the Conference on 14th International Interactive Collaborative Learning, September 21-23, 2011, Piestany, Slovakia, pp: 589-594.

Awad, M.A. and I. Khalil, 2012. Prediction of user's web-browsing behavior: Application of Markov model. IEEE Trans. Syst. Man Cybern. Part B: Cybern., 42: 1131-1142.

Chaudhary, J. and A. Saxena, 2015. e-Learning system by using cloud computing. Int. J. Innov. Scient. Res., 17: 246-250.

Chen, F., K. Guo, J. Lin and T.L. Porta, 2012. Intra-cloud lightning: Building CDNs in the cloud. Proceedings of the 31st Annual IEEE International Conference on Computer Communications, March 25-30, 2012, Orlando, FL., USA., pp: 433-441.

Daniel Ani, U.P., K. Taiwo, E. Ogbuju and S.E. Adewumi, 2015. Internet capabilities for effective learning and research. Int. J. Innov. Scient. Res., 13: 341-348.

Deborah, L.J., R. Baskaran and A. Kannan, 2014. Learning styles assessment and theoretical origin in an e-Learning scenario: A survey. Artif. Intell. Rev., 42: 801-819.

Lin, C.F., M.C. Leu, C.W. Chang and S.M. Yuan, 2011. The study and methods for cloud based CDN. Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, October 10-12, 2011, Beijing, China, pp: 469-475.

Mabroukeh, N.R. and C.I. Ezeife, 2009. Semantic-rich Markov models for web prefetching. Proceedings of the IEEE International Conference on Data Mining Workshops, December 6, 2009, Miami, FL., USA., pp: 465-470.

Malik, A.A., B. Hussain and S.O.H. Kazmi, 2016. Access control model for data stored on cloud computing. Int. J. Innov. Scient. Res., 20: 233-241.

Malini, G. and T. Mala, 2013. TEDU-Tamil educational cloud for anytime anywhere Tamil learning. Proceedings of the 12th International Tamil Internet Conference, August 15-18, 2013, Kuala Lumpur, Malaysia, pp: 415-422.

Malini, G., T. Mala and A. Karuppaiah, 2013. Semantic based for e-Learning content storage on cloud. Proceedings of the International Conference on Information Communication and Embedded Systems, February 21-22, 2013, Chennai, India, pp: 389-393.

Ochoa, X. and E. Duval, 2009. Quantitative analysis of learning object repositories. IEEE Trans. Learn. Technol., 2: 226-238.

Phankokkruad, M., 2012. Implement of cloud computing for e-Learning system. Proceedings of the International Conference on Computer and Information Science, Volume 1, June 12-14, 2012, Kuala Lumpur, Malaysia, pp: 7-11.

Sampson, D.G., P. Zervas and G. Chloros, 2012. Supporting the process of developing and managing LOM application profiles: The ASK-LOM-AP tool. IEEE Trans. Learn. Technol., 5: 238-250.

Senthilnayaki, B., K. Venkatalakshmi and A. Kannan, 2015. An ontology based framework for intelligent web based e-Learning. Int. J. Intell. Inform. Technol., 11: 23-39.

Tseng, C.S. and M.T. Chen, 2015. Developing library knowledge services by e-Learning. Int. J. Innov. Sci. Res., 18: 65-71.

Wang, Y., X. Wen, Y. Sun, Z. Zhao and T. Yang, 2011. The content delivery network system based on cloud storage. Proceedings of the International Conference on Network Computing and Information Security, Volume 1, May 14-15, 2011, Guilin, China, pp: 98-102.

Wu, B., K. Qian, P. Bhattacharya, M. Guo and W. Hu, 2011. Live programming learning objects on cloud. Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies, July 6-8, 2011, Athens, GA., USA., pp: 362-363.