A Proposed Framework Using Neural Network in Web Mining for Improving the Performance of E-Learning System

Dar Masroof Amin¹, Atul Garg²

¹Research Scholar, MMICT&BMB (MCA) Maharishi Markandeshwar University, Mullana (Ambala), 133207, India
²Associate Professor, MMICT&BMB (MCA) Maharishi Markandeshwar University, Mullana (Ambala), 133207, India

Abstract: Web Based learning has evolved as powerful tool for implementing educational requirements. It is a type of learning that is created by combining the electronic data and making it available on the Web. During last decade Web based learning expanded exponentially because of drastic increase in hardware software technology. At the same time the parallel growth in neural networking systems provides us tools for making systems intelligent. This paper will provide a framework that will incorporate neural networking learning methods to traditional e-learning architectures.

Keywords: E-Learning System, Cloud Based Architecture, Knowledge Discovery in Databases (KDD), Artificial Neural Network (ANN), Infrastructure as Service (IAAS), Platform as Service (PAAS), Software as Service (SAAS), Content Delivery Network (CDN), Cloud Based E-Learning.

1. Introduction

E-learning is generally conceptualized as electronically delivered learning and mostly this can includes all learning products delivered by computers through Intranet, Internet, satellite or other remote technologies. There is no clear and explicit definition of the concept of e-Learning. Brandon Hall defines e-learning as instruction delivered electronically completely by a Web browser, through the Internet or an Intranet or through general storage devices and may be likely other multimedia platforms [5]. The general understanding about the growth of e-learning relates exclusively to Web-based training or learning products delivered via a Web browser over an Internet. Morgan Keegan’s[2] investment analysis team deems e-learning as a technology encourages investors to consider the “e” in e-learning to represent “effective” [1] that leverages because of distributive power of Internet. E-learning is sometimes classified as asynchronous or synchronous and these terms refer to the extent to which a course is bound by place and/or time [34]. Definitions in the research literature are partially exclusive and sometimes contradictory and there are few common terms used consistently [2]. It is sheer difficult to differentiate the term “e-Learning” from terms such as online learning, virtual learning, network learning, online learning, multimedia-based learning, Web-based learning, Internet-enabled learning and similar terms [3]. E-Learning is often seen as learning where the Internet and the Web play an important role. The term is also employed in a broader sense, as learning where any electronic technology is used, but it excludes aspects that might fit under “distance learning”, but are not electronic, such as books or notes etc. Sometimes e-learning is referred as the intentional use of networked information and communication technology in teaching and learning. There terminologies used for e-learning like virtual learning, distributed learning etc. fundamentally all refer to the type of educational process which utilizes information and communication[1],[2]&[3]. But on closer analysis it will be clear that these labels refer to slightly different educational processes and can’t be used synonymously with the e-learning. These various types or modalities of e-learning activity are represented in Fig (1).

E-Learning Modalities are further divided in two parts. Each part is explained below:

1.1 Individualized Self-Spaced

1.1.1 Online
In this type of learning the computational resources such databases or related content via an Internet or Intranet.

1.1.2 Offline
In this type of learning individual computational resources are used without Internet.
1.2 Group Based

1.2.1 Synchronous
It is a kind of real-time environment in which a group of people are together sitting via an Intranet or Internet.

1.2.2 Asynchronous
There is a time delay during exchanges among groups of learners who are working over an Intranet or Internet.

In section 2, related literature of renowned researchers in the related field is discussed, section 3 provides details about dimensions of e-learning systems, section 4 explains the related work done in network architectures. The section 5 provides working and flow of traditional web mining system. In section 6 proposed frameworks and its working is described and in section 7 concludes the purpose of this research script.

2. Literature Review
E-learning has been liberalized in stages and the previous decade saw a drastic growth in the systems. Based on satellite communication through videoconferencing it became possible in 1998 to give classes interactively and simultaneously. Then in 2001 interactive classes has been started by using the Internet and other information technologies, although such classes do not have to be run simultaneously.


Obara in [12] provided the specialized architecture of ICT at University level learning systems that will enhance the quality of services provide to the individual users during heterogeneous learning process. Advanced Learning Infrastructure Consortium (ed.) provided analysis about future e-learning systems [13]. The researcher also explained that it would increase dynamic nature of e-learning system.

The researchers in [11], [12], [13] & [14] provided an analysis that there is scarce number of system architectures that are actually capable of adapting the change of e-learning system architectures. In [16] Carver et al., 1999 provided how to enhance student learning through hypermedia courseware and incorporation of student learning styles. Gilbert & Han in [15] explained the process of adapting instructions to accommodate learning style. In [17], Paredes & Rodriguez proposed ways for considering learning styles in adaptive Web-based system; Stash & Brau in [18] discussed how systems learning materials was often presented to best fit the learning style of learner, which is usually evaluated through a predefined questionnaire. For long questionnaires learners were not able to stand by so long thus, results become unreliable.

Villaverde in [15] describes an approach to the problem of mapping student’s actions within e-learning environments into learning styles that will accordingly become patterns for future decisions. Researcher in [15] also described an automatic approach based on feed-forward neural networks to infer the learning styles of students and have selected the back-propagation algorithm to train the ANN described in their work. The author [15] also explained a neural network architecture that learns the associations between students’ actions in e-learning environments.

The author in [16] provided a two folded way of learning in which first fold makes it imperceptible to students by an automatic mechanism for style recognition facilitates the gathering of information about learning preferences and the second is algorithm that uses the recent history of system usage to recognize changes in learning styles. Researcher [16] also suggested that recognition mechanism can be used in adaptive e-learning environments to help in the detection of students, learning styles and thus, can easily adapt the contents of courses that are presented to them. Research can also be extended to consider further input actions available in particular e-learning systems or domains. In the next section various dimensions of e-learning systems are discussed.

3. Dimensions of E-Learning
The process of designing and developing e-learning products, which can include courses, seminars, workshops, online learning portals, chat sessions/discussion groups and more, involves a careful mixture of personnel resources, hardware and software specifications and applications, standards for interactivity and media and design parameters based on user capabilities. The organizations that are in a process to building their own products should consider these resources essential to the deployment and production of e-learning as well. These components are explained further below. From technological point of view, e-Learning is somewhat closely related to Internet-based learning, online learning, Web-based learning and computer–based learning:

3.1 Internet Based Learning
The Web is only one of the Internet services that use HTML, URL and browsers. Some other services of Internet including Web are e-mail, file transfer facilities etc.

3.2 Online Learning
Learning may take place via any electronic medium. It is not automatically connected to a network.

3.3 Web Based Learning
Online learning could be organized through any network. Thus, Internet-based learning is only a subset of online learning.
### 3.4 Computer Based Learning

Learning includes computer-based learning that is not network-based. Such as Offline Software Packages, flash drives etc. As a result, e-Learning includes both non-network-based (online learning, Internet-based learning, and Web-based learning) and network-based learning or computer-based learning. The concept of e-Learning is employed in a broader sense, as learning that takes place via a combination of face-to-face and e-Learning. A mixture of face-to-face and e-Learning is known as blended e-Learning or hybrid [20]. The general idea about this is provided in the Fig (2).

Cloud based e-learning architecture is classified into five layers [26 & 27] and are defined below.

#### 4.2.1 Hardware Resource Layer

It is bottom most layer in the cloud service middleware which handles all the necessary computing hardware resources like CPU and its physical memory. [27].

#### 4.2.2 Software Resource Layer

Middleware and operating systems is used to create this layer [27].

#### 4.2.3 Resource Management Layer

It plays an important role to get loose coupling of hardware and software resources.

#### 4.2.4 Service Layer

Service layer is categorized into three levels namely Infrastructure as Service (IaaS), Platform as Service (PAAS), Software as Service (SAAS). These services help the cloud clients to use cloud resources [28].

#### 4.2.5 Business Application Layer

The main difference from this layer and from all other layers in the cloud based e-learning architecture as it acts as basic business logic of e-learning and frames the expansion of groups of components for e-learning [28], [29] & [30].

### 4.3 Content Delivery Networks(CDN)

Content Delivery Networks (CDNs) have emerged as solution to Internet Service Degradation and evolved further
to overcome the inherent restrictions of the Internet in terms of user perceived Quality of Service (QoS) during accessing Web content. CDN replicates content from the main server to cache servers, scattered over the globe, for delivering content to end-users in a timely and reliable manner from nearby optimal surrogates [39]. A typical content delivery environment is and where the replicated Web server clusters are located at the edge of the network to which the end-users are connected. A customer or content provider can sign up with a CDN provider for service and have its content placed on the content servers. The end-users can interact with the CDN by either specifying the content or service [40]&[42].

5. Traditional Web Usage Mining

The techniques and processing steps in Web Usage mining are similar to the data mining process. Just difference in Web mining and data mining is that in Web mining data will come from server log files and at the initial level in data will come from various data bases and warehouses[36].

6. Collection of Information

There are mainly three sources from where Web mining applications can gather data.

(1) Web servers (2) Proxy servers (3) Web clients

The Web servers is largest source of Web data, which provides us with huge mass of data can be available there. In Web servers, data is generally presented in extended log format, standard common log format and LogML [36]&[37]. For example generally ECLF (Extended common log format) is used in Web servers.

6.1 Process of Knowledge Discovery in Databases (KDD) for Web Mining:

This process of mining is useful to extract information from a large collection of data. This is a widely used data mining technique which includes data preparation, selecting, cleansing and including previous knowledge in data sets and providing interpreted results on basis of observed results see in fig(5).

Steps of KDD process:
(a) Domain: Learning the application domain relevant prior knowledge and goals of application
(b) Data Selection: The step creates a target data set.
(c) Data cleaning and preprocessing: It is the process of detecting and correcting inaccurate records from a database. Generally, term refers to identifying incorrect, incomplete, irrelevant, inaccurate, etc. parts of the data and then replacing, modifying, or deleting this faulty data.
(d) Data reduction and transformation: This step finds useful features, dimensionality/variable reduction and invariant representation.
(e) Data mining: The general data mining operation takes care of summarization, classification, regression, association, clustering and choosing mining algorithm(s)
(f) Data mining: search for patterns of interest
(g) Pattern evaluation and knowledge presentation: Transformation, visualization, removing redundant patterns, etc.
(h) Use of discovered knowledge

General working Flow of Traditional Web Mining process is give in fig (6)

Figure 5: KKД Process

Figure 6: Working flow of Traditional Web Mining
Web Browser: A client side application used to access the Internet.

Web Server: Server contains the web pages that are requested from web server.

Log Files: Log files keep the records about the various types of access done by a user on Internet.

User Session: User Session is activity by a user with Unique IP address during a specified period of time.

Rule: A measurement of usability of content accessed by a web site.

Interest Rule: This specifies the accurate and necessary non-standard content to be accessed from the mining process.

6.2 Various Issues with Traditional Usage Mining With Centralized Servers:

Scalability: Traditional systems where confined to their basic architecture. There were not slots to expand the systems.

Reliability: These systems cannot hold the pressure of exponential growth and were not to fault tolerant.

Slow: Due to congestion the systems were responding slowly to the user requests.

Inferior Streaming Quality: For heterogeneous data like animations, videos the system was not able to make it real time due to inferior Streaming.

Caching: On the server, data is stored in cache hierarchy. There is possibility of mismatching of data in the local cache access patterns and web server log records [37]. E.g. user has visited page hierarchy but due to data in caching server has recorded log differently as second time access of each page would directly been from cache. So the second entry of page is missed from log [37]. So, it cannot be predicted that log that every time correct data. Therefore, caching is a big issue for accessing Web data. Atul & Anil [44] proposed portable extended cache memory to reduce Web traffic and to improve performance. But, the basic problem with this technique is security due to cache memory is applied at the client end.

CGI Data: Common Gateway Interface (CGI) is used to pass variables and user entered data to respective server. It has a functionality to hide the value and username pairs from URI. That makes it difficult to check the User in Web Usage Data.

Session Identification: Finding and Tracking the session usage duration and creation especially when parallel login with same account through different machines makes identification complex.

Dynamicity of Pages: The content of dynamic pages is changed continuously for a same user with a fixed interval. So, minor change in content makes vary the log data as result.

Transaction uniqueness: Identifying unique users is an issue and their unique transactions as same account multiplicity is available.

7. Proposed Work

Neural Networks generates internal changes to adapt behavior in response to the environment. The success of E-learning systems mostly depend on whether these systems can provide an adaptive system with an adaptive interface to the learners depending upon factors like their professional background, learning abilities, learning goals etc. These features can be incorporated only if e-learning systems will be able to classify the learners. Neural networks can be effectively used for classification.

E-Learning presented statistic data regarding the assessment activity for each student during that period of an online course. Data must be divided in separate databases for training and testing of the classifying indicators. The three types of databases used are main database, Training database and testing database. The way in which the database is divided is important, as it represents an improvement criterion for the classifying indicator. With training of the classifying indicator shows that bigger the data better will be classifying results.

7.1 Properties of Proposed Framework.

Reliability: The Reliability in increased the proposed system with respect to load and growth.

Self-scaling computational units: The attention subsystem is based on competitive learning enhancing pattern features but suppressing noise.

Speed: Speed of System is increased because of capability to handle efficiently and learn new patterns by ART.

Self-adjusting memory search: The system can search memory in parallel, and adaptively change its search order.

Quality: Quality of streaming will be increased to maximum limits.

The parameter that needs to be selected for architecture improvement is the number of neurons of the layer hidden between the input layer and the output. Fig (7) shows the how neural network can be incorporated in traditional Web mining flow. The results in [41] show that if the neural networks will be implemented in collecting the real content, the response time and efficiency of Web can be increased drastically.
Also by using the neural networks there will be the best configuration from the error point of view. And there will be improvement of the performances along with the growing database by retraining, thus including the new information collected from the subsequent wards.

With this approach, the working of traditional architectures will become dynamic and efficient. The main objective is to provide an optimal solution at acceptable cost by looking for an approximate solution to problems. The soft computing methodologies (involving neural networks, fuzzy sets, rough sets and genetic algorithms) hold promise in Web mining.

ART is a family of different neural architectures where, the most basic architecture is ART1 by Carpenter and Grossberg, in [43].

ART1 can recognize and learn binary patterns. ART2 class of architectures categorizes arbitrary sequences of analog input patterns [43]. The below give architecture in (Fig 8) explains the internal course of work of ART systems with traditional e-learning systems.

The proposed approach includes Web-log analysis via introducing ART structure concept of neural networking for widely distributed, huge, highly heterogeneous, interconnected, semi structured, evolving, hypertext information repository of World Wide Web. The architecture with ART models can self-organize in real time producing stable recognition while getting input patterns beyond those originally stored. The series of steps followed in the above frameworks are:

1) The front end learner is prime source of queries for accessing Web application.
2) The Interface provides the access to the data base.
3) The data from the general KDD process is fetched to the user.
4) In addition to fetching data to the general user a copy of user links is sent to the ART system.
5) First a hidden layer which is composed of a group of neurons equal to the biggest layer of the network and the performances of the classifying indicators will be assessed.
6) Gradually a neuron is added to the hidden layer and the performances of the classifying indicators are re-assessed.
7) Optimum number of neurons of the hidden layer is achieved when the error rate is minimum or has no significant fluctuations.
8) Here is now the ANN can be used where in [35] there is minimum calculated error rate and overall performance.
9) The patterns are learned in phase 4 for future use
10) The learned Web Usage pattern is then open to access through interface. Same pattern can be evaluated for further saving on KDD architecture
11) Next time if the same information is accesses again, there will be a response time by systems as it is previously adapted to the system.

The stabilization of learning and activation occurs in the attention-subsystem by matching bottom-up input activation and top-down expectation. The orienting subsystem works like a novelty detector. It controls the attention subsystem when a mismatch occurs in the attention-subsystem. By using Adaptive Resonance Theory (ART), the results achieved can be comparatively best, more specially and specifically its feature of dynamic vigilance parameter and top down expectation and bottom up input action approach. The classification will be easy for huge amount of Web logs using ART model. The classification and clustering of any type of complex log data on the basis of specific analysis such as session tracking and pattern identification can be done my ART model.

In the above given architecture the Web log files that are accessed frequently are used by the proposed architecture for learning processes. The architecture adapts the process that is used by different users in accessing heterogeneous data. In a recursive manner the system adapts itself to perceive the users ways of accessing the links on the World Wide Web.
In addition to responding to the Web user, the same information is adapted in the network. This will help in caching the same page when that will be accessed or fetched by the multiple users next time. Already learned patterns directly access their corresponding category. Using the environment as a teacher the system can adaptively ovulate intentional vigilance. Disapproval of the current recognition in the environment of the system, changes this parameter to be more vigilant.

8. Conclusions

The advances in neural networking have created drastic change in making the software systems capable of learning and prediction of future trends. In the era the researchers are getting better performance using neural networks. In this research script we applied neural network at Web mining log phase at the server end. So using neural network in Web Mining architecture of e-learning systems will enhance the performance of mining data on Web and by using adaption concept of neural networks, the e-learning systems can be more dynamic by allowing personalization of the e-learning environment to suit the needs of the learner.

References

[28] Ivan I Ivanov, “Cloud Computing in Education: The Intersection of Challenges and Opportunities.”


