Introduction to Integrating Web Mining With Neural Network

Er.Romil.V.Patel
PG Student, IT
Parul Institute Of Engineering & Technology,
Vadodara, Gujarat

Dheeraj Kumar Singh
Ass.Professor, IT
Parul Institute Of Engineering & Technology,
Vadodara, Gujarat

Mr.Ankur.N.Shah
PG Student,CE
C.U.Shah College of Engineering & Technology
WadhwanCity, Gujarat.

***ABSTRACT*- The World Wide Web is huge, unstructured, universal and heterogeneous. In recent years the growth of the World Wide Web exceeded all expectations. Today there are several billions of HTML documents, pictures and other multimedia files available via internet and the number is still rising. But considering the impressive variety of the web, retrieving interesting content has become a very difficult task. Web usage mining is one of the technique of web mining is very useful to discover knowledge from secondary data obtained from the interaction from users with the web. The web usage mining is very essential for effective website. In this paper we give basic idea about web mining, neural network and GNG algorithm.**

Keywords: Neural Network, Web Mining, Natural Language Processing, GNG

I. INTRODUCTION TO NEURAL NETWORK

A neural network is a machine that is design to model the way in which the brain performs a particular task or function of interest the network is usually implemented by using electronic components or is simulated in software on a digital computer. A neural network is also referred as neuron computers, connectionist networks, parallel distributed processors etc.

II. FEATURES OF NEURAL NETWORK

Following are different features of neural network.

[*A*] *Nonlinearity:* An artificial neuron can be linear or nonlinear. A neural network made up of an interconnection of nonlinear neurons. So neural network itself a nonlinear. Nonlinearity is a special kind of sense that is distributed throughout the network. Nonlinearity is a highly important property, particularly if the underlying physical mechanism responsible for generation of the input signal (e.g. speech signal) is inherently nonlinear.

[*B*] *Input-Output mapping:* A popular paradigm of learning with a teacher or supervised learning. It involves modification of the synaptic weights of a neural network. It done by applying set of labeled training samples or task examples. Each example consists of a unique input signal and a corresponding desired response. The network is presented with an example picked at random from the set. The synaptic weights of the network are modified to minimize the difference between the desired response and the actual response of the network. The training of network is repeated until it reaches to the steady state it means there is no significant change in the synaptic weights. The previously applied training examples may be reapplied during the training session but in a different order. So network learn from the examples by constructing an input output mapping for the problem.

[*C*] *Adaptivity:* Neural networks have a built in capability to adapt change in synaptic weights according to the surrounding environment.

[*D*] *Evidential response:* A neural network can be designed to provide information not only about which particular pattern to select but also about the confidence in the decision made. This information may be used to reject ambiguous patterns and improve the classification performance of the network.

[*E*] *Contextual information*: It is deal with naturally by a neural network. Knowledge is represented by vary structure and activation state of a neural network. Every neuron in the network is potentially affected by the global activity of all other neurons in the network.

[*F*] *VLSI implementability:* The massively parallel nature of neural network makes it potentially fast for the computation of certain tasks. So it is well suited for VLSI technology.

[*G*] *Uniformity of analysis and design:* In neural network all domains of application use the same notations. So it is easy to share theories and learning algorithms between different applications of neural network. It provides seamless integration of modules.

[*H*] *Fault tolerance:* A neural network implemented in hardware form, has the potential to be inherently fault tolerance. In order to make neural network fault tolerance it is necessary to train network such a way.

[*I*] *Neurobiological analogy:* The design of a neural network is motivated by analogy with the brain. It gives proof that fault tolerance is fast and powerful. Neurobiologists look to neural networks as a research tool for the interpretation of neurobiological phenomena.

III. TYPES OF NEURAL NETWORK ARCHITECTURES

There are fundamentally three different types of architecture of neural network. They are differing by how neurons of neural network are linked with learning algorithms. Following are different types.

[1] Single layer feed forward network

[2] Multilayer feed forward network

[3] Recurrent network

IV. INTRODUCTION TO WEB MINING

The Application of data mining techniques to the World Wide Web referred as web mining. We make use of the web in several ways. For example, finding relevant information, discovering new knowledge from the web, personalized web page synthesis, learning about individual users etc. Web mining techniques provides a set of techniques which provide solutions to different problems. However web mining techniques are not the only tools to handle these problems. Other related techniques from different research areas such as database (DB), information retrieval (IR) and natural language processing (NLP) can also be used.

When we see web mining in terms of data mining it have three interest of operations say clustering (e.g. finding natural groupings of users, pages, etc.), associations (e.g. which URLs tend to be requested together) and sequential analysis (e.g. the order in which URLs tend to be accessed). As in most real world problems the clusters and associations in web mining do not have clear cut boundaries and often overlap considerably.

Web mining techniques can be categorized as web content mining, web structure mining and web usage mining.

[*A*] *Web content mining:* It studies the search and retrieval of information on the web. Web content mining future can be divided as web page content mining and search result mining. It has to do with the retrieval of content available on the web into more structure forms as well as its indexing for easy tracking information locations. Web content may be unstructured (plain text), semi-structure (HTML documents), or structured (extracted from databases into dynamic web pages). Such dynamic data cannot be indexed and consist what is called "the hidden web". A research area closely related to content mining is text mining.

[*B*] *Web structure mining:* It focuses on the structure of the hyperlinks (inter document structure) within the web. The goal of web structure mining is to categorized the web pages and generate the information such as the similarity and relationship between them, taking the advantage of their hyperlink topology. Then it focuses on the identification of authorities.

[*C*] *Web usage mining:* It discovers and analyzes user access patterns. The term web usage mining was introduced by Cooley et. al. in 1997 and in according with their definition: web usage mining is the automatic discovery of user access patterns from web servers. Web usage mining is the process of identifying browsing patterns by analyzing the user's navigational behavior. This information takes as input the usage data i.e. the data residing in the web server logs, recording the visits of the users to a web site.

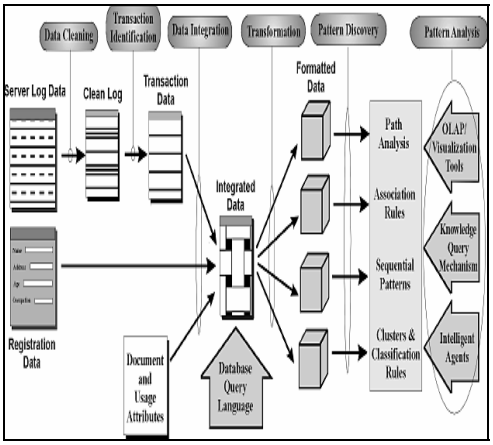

Figure: - A general architecture of web usage mining.

Examples of Application of Web Mining

[1] Web mining in foreign languages

It should be noted that the language code of Chinese words is very complicated compared to that of English. The GB code, BIG5 code and HZ code are common Chinese word codes in web documents. Before text mining, one needs to identify the code standard of the HTML documents and transform it into inner code, then use other data mining techniques to find useful knowledge and patterns.

V. ADVANTAGES OF WEB MINING

A.) Web mining is attractive for companies because of several advantages. In the most general sense it can contribute to the increase of profits, be it by actually selling more products or services, or by minimising the costs. In order to do this, marketing intelligence is required. This intelligence can focus on marketing strategies and competitive analyses or on the relationship with the customers. The different kinds of web data that are somehow related to customers will then be categorised and clustered to build detailed customer profiles. This not only helps companies to retain current customers by being able to provide more personalised services, but it also contributes in the search for potential customers. The scenario clearly illustrates this. By analyzing the web log data (usage mining), Sharon's favourite bookstores discovered that Sharon is a potential buyer. They were able to make her a tempting offer by displaying a specific banner on her browser window. The other bookstore was able to identify Sharon as a potential customer by searching the web for homepages and analyzing the data on the pages (content and structure mining). She was sent a special offer by e-mail, which would most probably match her preferences.

B.) From Sharon's point of view we could say that she was pleased by the fact that the web site of her favourite bookstore displayed an interesting banner and she was not aware of any missed offers from this bookstore. And although in this scenario Sharon was a bit annoyed by the unsolicited e-mail send by the first bookstore, she might just as well have been attracted by the offer and she might even have become a customer of this bookstore. Clearly both web mining categories contribute to the general goal of gaining marketing intelligence, be it each in its own way.

VI. PROPOSED RESEARCH APPROACH

In the present work, this paper proposes the use of SOM to identify the use's habits. This kind of artificial neural network will be tried to gather the users by patterns of pages accesses. To obtain this result we need to process the Web Log files to identify users and session of users; after that with this session's user, we'll train the ANN. The selection of "Self Organization Method" is so because it isn't necessary to supervise to the training. The principle of self-organizing the neural networks of optimal complexity is considered under the unrepresentative learning set.

VII. INTRODUCTION TO GNG (GROWING NEURAL GAS) ALGORITHM

Neural gas is an artificial neural network, inspired by the self-organizing map and introduced in 1991 by Thomas Martinetz and Klaus Schulten. The neural gas is a simple algorithm for finding optimal data representations based on feature vectors. The algorithm was coined "neural gas" because of the dynamics of the feature vectors during the adaptation process, which distribute themselves like a gas within the data space. It is applied where data compression or vector quantization is an issue, for example speech recognition, image processing or pattern recognition. As a robustly converging alternative to the k-means clustering it is also used for cluster analysis.

VIII. CONCLUSION

This paper has attempted to provide a basic idea about the artificial neural network, web mining and GNG algorithm. In this paper a general overview of Web usage mining is presented. Web usage mining is used in many areas such as e-Business, e-CRM, e-Services, e-Education, e-Newspapers, e-Government, advertising, Digital Libraries, marketing, bioinformatics and so on. One of the open issues in data mining, in general and Web Mining, in particular, is the creation of intelligent tools that can assist in the interpretation of mined knowledge. Clearly, these tools need to have specific knowledge about the particular problem domain to do any more than filtering based on statistical attributes of the discovered rules or patterns.

REFERENCES

[1] Mohamad H. Hassoum "Fundamentals of artificial neural networks", PHI.

[2] Laurene Fausett "Fundamentals of neural networks architectures, algorithms and applications", pearson education.

[3] Arun K Pujari "Data Mining Techniques", Universities Press.

[4] Navin Kumar Tyagi, A.K.Solanki, and Sanjay Tyagi "An algorithmic approach to data preprocessing in web usage mining", An international journal of information technology and knowledge management, July-December 2010, Vol.2.

[5] Sonali Muddalwar and Shashank Kawar "Applying artificial neural network in web usage mining", an international journal of computer science and management research, Nov-2012, Vol.1.

[6] Vaishali A. Zilpe and Dr. Mohammad Atique "Neural network approach for web usage mining", A national conference on emerging trends in computer science and information technology (ETCSIT), 2011.

[7] Anshuman Sharma "Web usage mining using neural network", An international journal of reviews in computing, April-2012, Vol.9.

[8] T.Krishna Kishore, T.Sasi Vardhan and N.Lakshmi Narayana "Probabilistic semantic web mining using artificial neural analysis", An international journal of computing science and information security, Mar-2010, Vol.7.

[9] S.Taherizadeh and N.Moghadam "Integrating web content mining into web usage mining for finding patterns and predicting user's behaviors", An International journal of information science and management, January / June-2009, Vol.7.

[10] Naresh Barsagade "Web Usage Mining and Pattern Discovery: A Survey Paper", DEC-2003.