Scientific Journal of Impact Factor(SJIF): 3.134
e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 4, April -2015

A review paper on different pattern classification techniques based on web usage mining with Neural Network

Parekh Pranav M.[1], Shah Ankur N. [2]

[1]CSE Department, PIT - Limda, parekh.pranav1@gmail.com
[2] CSE Department, PIT - Limda, ankur11586@gmail.com

Abstract —*Area of data mining includes data preprocessing, data classification, cluster analysis, Association etc. The traffic on World Wide Web is increasing day by day and large amount of data generated due to user's interaction with web sites. Web mining is the application of data mining techniques which includes web usage mining, web content mining and the third one is web structure mining. For the pattern classification in web mining different algorithms are used. This paper includes different pattern classifications techniques which are used for web usage mining*

Keywords- *Web usage mining, Naïve bayes, Support Vector machine, Pattern classification, Neural Network*.

I. INTRODUCTION

Web mining is the application of data mining techniques which is use to discover new patterns from the web. Web mining is categorized in three ways: web content mining, second is web structure mining and third is web usage mining. In the area of web mining different techniques are used for classify patterns from the web usage data. They are classified on the basis of recall, Time taken, accuracy and precision. In this paper we are going to discuss about different algorithm for pattern classification also we discuss about neural network.

II. INTRODUCTION OF WEB USAGE MINING

Web usage mining is part of the Web mining in data mining. From the server logs meaningful information is mining by the process of web mining. From web usage mining it shows that what user is actually looking for on internet. This information can be used in several ways like e-commerce, website personalization, user future request prediction, improvement of website etc [1]. The main task of web usage data is to capture web browsing data of the users from specified websites. The Web mining process can be divided in three parts (i) preprocessing, (ii) pattern discovery and last (iii) pattern analysis. In web usage mining for pattern classification the web log data is to be form in clean log after preprocessing phase.

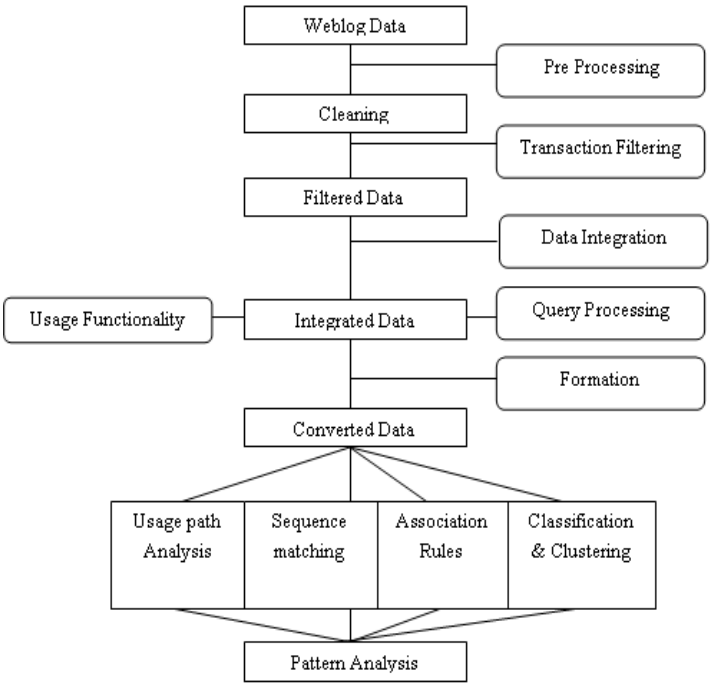

Figure 1 Architecture of Web usage mining

Most important phases of web usage mining are the reconstruction of user sessions by using heuristic techniques and classify useful patterns from these sessions by using patterns classification techniques. Different processes have been done on web log data before applying the classification algorithm on it [7]. Data cleaning and session identification is done on the web log data and then various techniques of web usage pattern discovery are applied in order to detect interesting and useful pattern [7].

III. CLASSIFICATION TECHNIQUES

A. Decision Tree Classifiers

This technique for classification is simple for classification. Here the classifier is in the form of tree structure [1].It classify instance by starting of the root of the tree and it moves at downward till the leaf node [4]. If in practice decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Limitation of decision tree classifier is more computation is required because of at each node, each candidate splitting field must be sorted before its best split can be found.

B. Rule Based classification

Rule-based classifiers use a set of IF-THEN rules for classification. The expression of rule is: IF condition THEN conclusion. The IF part of the rule is called antecedent or precondition and THEN part of rule is called as rule consequent. One or more attributes can be there in precondition part or the rule base classifiers. The consequent part consist class prediction [3].

C. Naive Bayesian classification

In this technique bayes' theorem is applied with the strong independent assumptions. Also it is it simple probabilistic classifier algorithm. In this algorithm joint probability of words and categories used to estimate the probabilities of categories given in document. The naive part of the algorithm is assumption of the word independence [1].

D. Support Vector Machine

It is a technique in which supervised learning model is applied with associated learning algorithms so it will analyze data and recognize patterns for classification and regression analysis [1]. Support vector machine constructs hyperplane or set of hyperplanes in high dimensional space so it can be used for regression, classification and for other tasks [5]. The hyper plane which has the largest distance to the nearest training data point of any class will achieve a good separation, the larger the margin the lower the generalization error of the classifier [1].

E. Lazy Learners

k- Nearest neighbor classifiers is method for classification and regression, which is non parametric method. In classification the input consist of the k closest training example in feature space [6]. In this the output is class membership. An object is classified by the majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors [1].

IV. NEURAL NETWORKS

An Artificial Neural network is an information processing paradigm that is inspired by the way biological nervous system works, such as processing of information by brain. ANN mimic the real life behavior of neurons and the electric message they produce between input, processing by the brain and the final output from the brain . Artificial Neural Network is configured for a specific application such as pattern reorganization or data classification, through a learning process [12].

Mainly there are three types of architecture of neural network. They are different by how neurons of neural network are linked with learning algorithm. Types are as follows:

A. Single layer feed forward network

It is a simplest type of neural network. It consists of a single layer of output nodes; here inputs are fed directly to the output via a series of weight. The some of the product s of weights and the inputs are calculated in each node, and if the value is above some threshold (0) the neurons fires and takes the activated value (1), else it takes the deactivated value (-1). Neurons with this kind of activation function are also known as artificial neurons or linear threshold units [11].

B. Multilayer feed forward network

This type of network includes multilayer of computational unit, usually they are interconnected in feed forward way. Each neuron in one layer has directly connections to the neurons of the subsequent layer [11].

C. Recurrent Network

It is a class of neural network at where connection between units forms a directly cycle. Which creates an internal state of network which allows it to exhibits dynamic temporal behavior. Un like feed forward network Recurrent neural network can use their internal memory to process arbitrary sequences of inputs. This makes them applicable to such tasks likes un-segmented connected handwriting reorganization, where they have the best known result [11].

V. COMPARISON BETWEEN TWO CLASSIFIERS

Here they use some web log data for classification. And the comparative result is shown in below table. Table 1 shows the time taken in different session for classification of data.

Table 1 Time taken in session

Time In Second		
Session	**K-mean**	**Naïve Bayesian with supervise learning**
1	20.1	18.10
2	22.15	16.23
3	28.30	17.21
4	40.25	15.12
5	42.27	14.20
6	32.52	15.23
7	25.12	19.38
8	28.03	15.29

[7]

Here this table shows the accuracy of the data. During each of the session no of test case given and the accuracy of that data also are given.

Table 2 Accuracy of Data (%)

Session	No. of test case	K-mean classify	Naïve Bayesian classify
1	255	20.39	48.57
2	170	28.47	40.28
3	128	33.60	39.25
4	85	35.30	41.91

[7]

As shown in Table 1 and Table 2 Naïve Bayesian gives more classify data. Also with the use of naïve Bayesian classification it gives more accurate data other algorithm. Because of it uses with super wise learning error also are less.

VI. CONCLUSION

This paper has attempted to provide basic idea about web usage mining and pattern classification. Also it includes the some information about neural network.

Here different algorithms are given for classification. But in different case it gives different result. For the different datasets they work differently. We have to choose one among them, which works best on our dataset. So, as per the comparison of the algorithm Naïve Bayesian classification works more accurate than other k-mean classifiers in this case as shown in table[1] and in table[2]. It also provides more accurate data within less time.

Using Neural Network techniques for pattern classification for web usage data gives more accuracy and at the same it gives less error will.

REFERENCES

Format of Reference

[1] Surpreet Dhilon, Kamljit Kaur "Comparative study of claasification algorithm for web usage mining." International journal of advanced Research in Computer Science and Software Engineering, July 2014.

[2] Anshuman Sharma "Web usage minig using neuarl network" International Journal of Reiviews in Computing, April 2012

[3] Nanhay Sharma, Achin Jain, Ram Shringar raw, "Comparison analysis of web usage mining pattern recognization techniques." Interanational Journal of Data mining & Knowledge management Process, July 2013.

[4] Jiawei han Kevin, Chen Chaun Chang "Data mining for web intelligence". November 2002.

[5] Sujata Padmakumar, DR. punithavalli, Dr. Ranjith, "A web usage mining approach to user navigation pattern and prediction in web log data" International Jounal of Science research, April 2014.

[6] Sonali Muddalwar, ShankarKawar "Applying artificial Neural Network in web usage mining". International jounal of computer science and management Research vol 1 Issue 4 November 2012.

[7] Romil V. Patel, Dheeraj Kumar Singh, "Pattern Classification based on Web usage Mining using neural Network techniques," International Journal of Computer Application(0975-8887) Volume 71-No.21,June 2013

[8] Jiawei Han, Micheline Kamber, Jain Pei, " Data Mining – Concepts and techniques" Elsevier publishers.

[9] S. Rajsekaran & G.A. Vijayalashmi Pai, "Neural Network, fuzzy logic and Genetic Algorithms"- PHI Learning Private Limited.

[10] A B M Shawakat Ali, Saleh A. Wasimi, " Data mining: Methods and Techniques." Cengage learning India Private Limited.

[11] B. Mahalakshmi, S.Kalaichelvi, "Web mining and Neural Networks" ISR National Journal of Advanced Reserch in Computer Networking, wireless and mobile communicatons. Vol.1 Issue 1. 08-may-2014.

[12] Anshuman Sharma, "Web Usage Mining Using Neural Network" International Journal of Reivews in Computing, 10 April 2012 Vol 9.

