

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

Implementing Web Data Mining Framework for Developing E- Business Solution

Roger P. David¹, Jaykumar S. Karnewar²

P.G. Student, Department of Computer Science & Engineering, JCOET, Yavatmal, Maharashtra, India¹

Assistant Professor, Department of Computer Science & Engineering, JCOET, Yavatmal, Maharashtra, India²

ABSTRACT : Big Data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data so the personalized and helped in meeting better requirements of users. Growing data memory capacity as fuelled the development of data management and processing. The developers tried to build taxonomy of web knowledge so that it can be used in a better ways and succeeded in developing sound framework. However, data mining is relevant considering the deliberately demanding change in diversity of the web, information acquisition retrieving and extracting of interestingness contents which are web based has become a very complex task. Web mining framework makes highly attempts to make view of gathering useful relevant data by reducing the complexity field in the data which received from the interaction of users from the web which is determinant feature of web data mining, and this is of our best information of web knowledge, we analyse the challenging issues in the big data revolution, this projects helps in drawing a comprehensive overall virtual framework and summarizes the corresponding techniques and methods to analyse in knowledge framework namely for e-business.

KEYWORDS: web knowledge; framework; data mining; relevant.

I. INTRODUCTION

Data mining is the means of process which self-moves the extraction of foretelling information, which discovers the interesting knowledge from large or big amounts of data which is stored in various databases, data warehouses and other knowledge acquired through various repositories [2]. The enormous amount of data seems to be good substitute and has helped in achieving these drastic goals [1]. Big Data starts with large-volume, various autonomous sources with distributed and decentralized control and finds to explore complex and evolving relationships among data. Big data concerns about large-volume of data, it becomes complex and growing data sets with multiple, autonomous sources. The use of knowledge acquisition in various fields of human day to day life has gave a push to lead the huge bulk of data storage in various formats like records, documents, images, sound recordings, many new type of available data formats [5]. Web data mining is a part of Big data mining has been hugely used in the recent last decades which made ways to break up a whole in a part of huge collections of knowing facts and unknowing actual facts [1], and these facts are being used in applying applications of mining in current web which dramatically based on databases which ruled over it [3]. It is beneficial to use data mining which is an application based on database to determine the patterns and relationships which is unknown so it helps to manage big data problems when it is used in Web, hence it is necessary to develop data mining in web framework [2] to manage big data related problems.

Web Knowledge mining in big data important action of acquiring which is a multi-level used field which includes, machine learning, natural language processing methods, statistical means, small and huge databases, recollecting of knowledge gathering, multimedia, etc. which is a most important and obvious part in web mining [5]. We know that web offers an unprecedented opportunity and difficulties to knowledge mining which can be studied further in near future [2]. This paper helps in giving an exact explanation about using various ways to deal with the problem of identifying relevant information [5], from the page content structure of web page and to introduce algorithms in the structure of web hyperlink taxonomy and prominent use of huge data is acquired in web mining.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

II. RELATED WORK

The structure of mining of data is the revolution of a field with the methods of various fields, which includes database management systems, Statistics, Artificial Intelligence [5]. The new era of data mining retrospective applications was imagined in the year probably in 1980's basically by scientifically based technique tools focused on single tasking methods which was important in that period [5], but as the revolution started the trend started changing and in recent early day's data mining trends has went under the future development of World Wide Web and when its usage will grow [3], it will continue to generate ever more n more content, structure, and usage data it will required large work space [2], and the value of Web mining will keep increasing desperately and at a fast pace. Research development needs to be done in initiating the right blocks of Web metrics, and their procedures in mining issues, extracting process models from the usage data, understanding how different parts of the process model will have drastic impact various on Web metrics which will sound good, with respect that how the process models is retrospective changing in response to the various changes that are making change in excitement to the user [4], who is developing Web mining techniques which will help to improve various other aspects of Web services, using various methods to initiate false means, frauds and intrusion detection in web mining.

The unacceptable performance was initially encountered by various internet companies such as Amazon, Google, Face book and Twitter, but they are now becoming a hurdle for other companies and which desire to contribute meaningful real-time service. A framework for mining has getting setter in Web data streams using of means dynamic learning in that respective period of time required, it gives the effect of likewise measures on the mining process and converting the mined excellent example in the disagreeable single passing requirement [2]. "Implementing Web Service Framework Research of Data Mining in E-business", it helps in promoting the actual knowledge that can be used in web data mining [1]. It will help only in supporting only that data that can be used in e-business. It makes serious concentration in mining the domain based data items. It doesn't help in supporting complex format of the data [2]. "Agent Based Framework for Semantic Web Content Mining", this kind of work focuses on giving a prominent agent-based framework for mining semantic web contents making use of clustering techniques in actual framework[2]. Clustering will help provide user with query relevant cluster of web contents, which will better satisfy user requirement and it provide support in that kind of actual situation usage of web surfing time [2]. It can take the hidden semantic web contents using clustering methodologies. The main issue is that it is very difficult to handle complex based data [2].

The inner part of characteristic action of web data mining are applying various methods and algorithms which will help in order to discover and extract excellent examples or various of stored data [6]. From the previous few decades data mining and knowledge based discovery applications have gained wealthy focus due to its suggestive meaning in decision making and it has become a most essential component in various organizations and sectors [6]. The field of web data mining have been succeeded and helped in extracting into new areas of human life with various constitutes and went forward in the fields of Statistics, Databases, Machine based Learning, excellent example Reorganization, Artificial Intelligence and Computation based capabilities etc.

Data mining helps in processing the retrieval and extracting for telling acquired of knowledge data, the information can be regain and retain by identifying information from various type of huge database systems associated with it [5]. As the use of internet is growing globally web community also does various scientific research and implements for the user. This technology has a wide used for reviewing huge amount of collected data, and is applied to a different number of domains of various websites on the web [3]. Web mining technique can be widely classified into three domains: content, structure and usage mining. The actual study of three domains is described below.

2.1 Classification of Web Knowledge Mining Framework:

Knowledge Mining helps in accessing is a large amount of data information. But the use of this technique may create Multidisciplinary issues, advantages and possible opportunities field which is used for data mining when it is used on web [2].

2.1.1 Web Structure Mining:

Web structure mining is a technique through which it becomes easy for analysing the data and gives explanation of the links which is being used between different websites and web pages [2]. This technique focuses mainly on building

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

web creepers [4]. The topology of this arrangement mines it and works on hyperlinks. It is the main part for the discovering the data.

2.1.2 Web Content Mining:

It is a technique which gives descriptions and analysis and extracts information from the contents on web [2]. It focuses involving techniques for classifying, for making summary and clustering of the web contents. It may provide useful methods and interesting excellent examples about the needs for user and contribution in the behaviour of user.

2.1.3 Web Usage Mining:

It mainly focuses on digging the use of contents of web from the blocks to carry on web servers, cookies blocks, server application logs etc [2]. It helps the user, when they move from one page to another. Thus, it is more helpful for providing similarities and association between contents of web [3]. Through the use of this content it is easy to build the framework of web data mining. So the researcher focused on these three contents to study the framework of architecture.

III. PROPOSED WORK

As we know big data mining is a key technology in e-business help to collect user information, web mining can be used to analyse user data [3]. For an intelligent learning database system [52] to handle Big Data, the essential key is to go up to the large quantity of data and provide treatments for the characteristics featured by the algorithms. The improving of data management systems is increasingly breaking based on application domains i.e. relational data, associated data, stream data, clustered data, etc. So my proposed work puts forward a better way to manage Big data in data mining to retrieve extract information while using through web network. It also provides a business and technological overview of web data mining and can enforce and redefine user-entrepreneur renew relationship. My proposed framework techniques in web server will help to improve data retrieval to analyse user behaviour and manage large amount of data. Following figure shows the detail information of Web Mining Architecture in e-business.

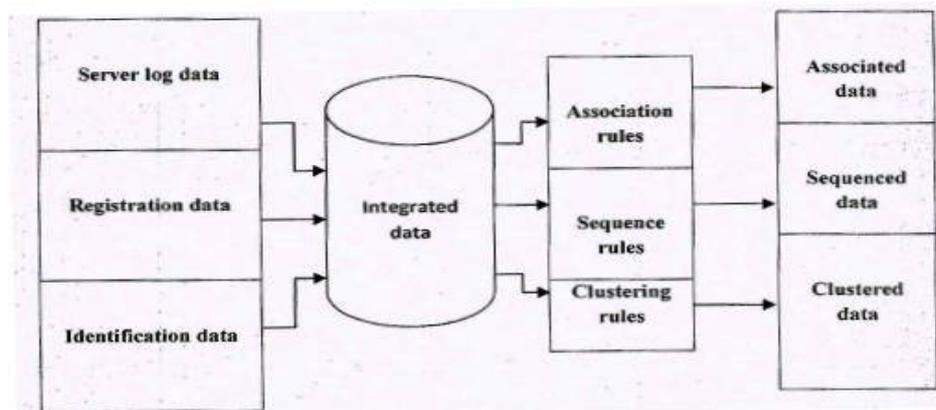


Fig 1: Architecture of Web Mining

4.1 Working of Web Mining Architecture:

When the user accesses a particular website the following work is done initially on the usage mining of data, which helps in digging the contents of user on web and blocks unreliable cookies, non-trusted pages. The user can move from one page to page.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

4.1.1. Server Log Data:

For log file accessing of web server records the interaction information of visitors or users access. This web server log file, which contains many records, is used to collect or record users' access to this website, including Server log (user IP, server name, URL, time taken by client or user browsing this website, Cookie identification number), Error log (lost connection, authorization failure, timeout, etc.), Cookie (user state and access path), it also includes cleaning of data [9]. It is blocks further transaction when error is detected on a website.

4.1.2. User Registration Information data:

When user access to an e-commerce base site, it inputs and delivers some information to the server via web page [3], including registration information and exchange information. Analysis of user registration information helps you to analyse user behaviour pattern and formulate the corresponding e-business policies aiming at specific users these data is transferred to the place where whole data is integrated [4].

4.1.3. Transaction Data:

It saves user information; it also contains information about goods and order table. All the transaction is done here, the order information table will have a new record to record the user's purchase information every time a customer or user completes a transaction. All these work will be done in background database of e-business. This information mining is of great advantage to analyse behaviour of users' interest.

4.1.4. Integrated data:

It is a part of web content mining and the most important block in the web mining framework [4]. All the information about the contents, Documents available for the customer is saved here are in the form of database query language. It also keeps records of visitor's access, interaction information of user, user's registration information, and links of various customers is integrated in the form of data [3]. There is huge amount of data is stored and it is in heterogeneous form.

4.1.5. Classification of data:

The actual data mining process is done in structure mining. When particular form of data is requested by user, it becomes necessary to classify the information from the integrated data. The Transformation of data into its form is done and to fetch the particular data for a user and satisfy his request. Here the agent of framework which is a software determines the data on the basis of the characteristics [3], similarities and occurrences and sorted in particular techniques which is mined later and made available for the user or client.

Now the data is formatted in different data mining techniques such as Path Analysis, Association rule, Sequence Pattern and Clustering rules [4]. These are done on the basis of discovering a pattern on a basis of relationship between items in the same transaction for Association Rule. The clustering is obtained by making meaningful or useful cluster of object which has similar characteristics. The sequential pattern is obtained by discovering or identifying similar pattern, trends in transaction of data over a business period. It also analyses the path of an object which is occurring on a particular path.

4.1.5.1. Association Rule:

When there is a relationship between items in the same transaction a pattern called Association is discovered. Association rules is a valuable tool that has been used widely in various fields for extracting the information from data [2]. The task of association rules mining has been described, formally, let D be a database of sales transactions, and let $I = \{i_1, i_2, \dots, i_m\}$ be a set of binary literals called items. A transaction D in C contains a set of items called items, such that $D \subseteq I$. The items present in an item set are called length of an Item sets that have a length l are denoted by l -item sets. Each item set is associated with a statistical threshold named support. The support of the item set is number of transactions in C that contain [2] the item set. An association rule is an expression $P \Rightarrow Q$, where $P, Q \subseteq I$ are two sets of items and $P \cap Q = \emptyset$. X is called the antecedent, and Q is called the consequent of the association rule. An association rule $P \Rightarrow Q$ has a measure of goodness named confidence, which can be defined as, the probability a transaction contains P given that it contains Q , and is given as $\text{support}(PQ) / \text{support}(P)$.

Example of a Sequenced data is as follows:

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

When the data is extracted from the database using association rule method we get associated data. Associated data is a data which is obtained by applying association rule on items which have same relationship in a same transaction. This data is separated from the data which is present in the database. The information or data which is occurring in a transaction is extracted from the integrated data through which it can be made available to the user.

1. To obtain the associated data it finds the frequent item sets, i.e., the sets of items that have the least minimum support. Association rule depends upon strength and confidence. The support determines how often a rule is applicable to a given data set, while confidence determines how frequently items appear in transaction.
2. Use frequent item sets to generate associate data that satisfy minimum support and minimum confidence.

Transactional Database

Transaction Id	Items	Time
12	bread, milk	10:12
13	bread, juice, milk	12:13
14	milk, beer, bread, juice	13:22
15	bread, eggs, milk	13:26
16	beer, basket, bread, juice	15:11

Fig.2: Example of an associated data.

It's applied widely especially in market basket analysis in order to infer items from the presence of other items in the customer's shopping cart. The example of associated data is shown in the figure where the milk product is associated with other product with in every transaction.

4.1.5.2. Sequence Pattern:

The pattern analysis is one of data mining technique that finds to discover or identifies same patterns, regularly occurring events or trends in transaction data over a transaction period. A set of items that customers buy together different times in a year can be identified in previous transaction.

The rule for mining the information from the database which is in sequence pattern is as follows.

1. To find frequent items we have to scan the database. e.g. {p, q, r, s ...}
2. Always try to create a rule with only two items for every pairs of such kind of items. e.g. {p} ⇒ {q}.
3. Then, find larger rules by recursively scanning the database for adding a single item at a time to the left or Right part of each rule (left and right expansions). e.g. {p, r} ⇒ {q}, then {p, r, s} ⇒ {q}, etc.
4. Testing of each rule which is created is done to see if it is valid or not.

Example of a Sequenced data is shown in the figure, using this rule it is easy to determine the data in a sequence pattern type and can be extracted from the database. When the data is extracted from the database using association rule method we get sequenced data. Sequenced data is a data which is obtained by applying a sequence rule on data which have similar patterns.

ID	Sequences	ID	Rule	Support	Confidence
1	(a b), (c), (f), (g), (e)	1	a b c ⇒ e	0.5	1.0
2	(a d), (c), (b), (e f)	2	a ⇒ c e f	0.5	0.66
3	(a), (b), (f) (e)	3	a b ⇒ e f	0.5	1.0
4	(b), (f g)	4	b ⇒ e f	0.75	0.75
		5	a ⇒ e f	0.75	1.0
		6	c ⇒ f	0.5	1.0
		7	a ⇒ b	0.5	0.66
	

Fig.3: Example of a Sequenced data.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

The following figure shows the example of data which is present in sequence and it shows the length or degree of sequence to which it is related. This technique will extract the data present in same sequence and separated from the data present in the database. The occurrences of the data are also analysed.

4.1.5.2 Clustering Rule:

The rule that makes meaningful or useful cluster of objects which have probably same characteristics using automatic technique is known as clustering. The cluster defines the classes and puts objects in each class. Every cluster is given with a centre point known as centred. Each point is assigned to the cluster with the closes centred. Number of clusters, M, must be specified. M-means will converge for common similarity measures mentioned above. This algorithm is very important and is very simple.

- 1: Select M points as an initial centre point.
- 2: Then Repeat the initial step.
- 3: Form M cluster by assigning all points to the closest centre point.
- 4: Again compute the centre point of every cluster.

Example of a Sequenced data is as follows:

When the data is extracted from the database using clustering algorithm we get clustered data. Clustered data is a data which is obtained by applying clustering algorithm on items which have similar characteristics and occurrences.

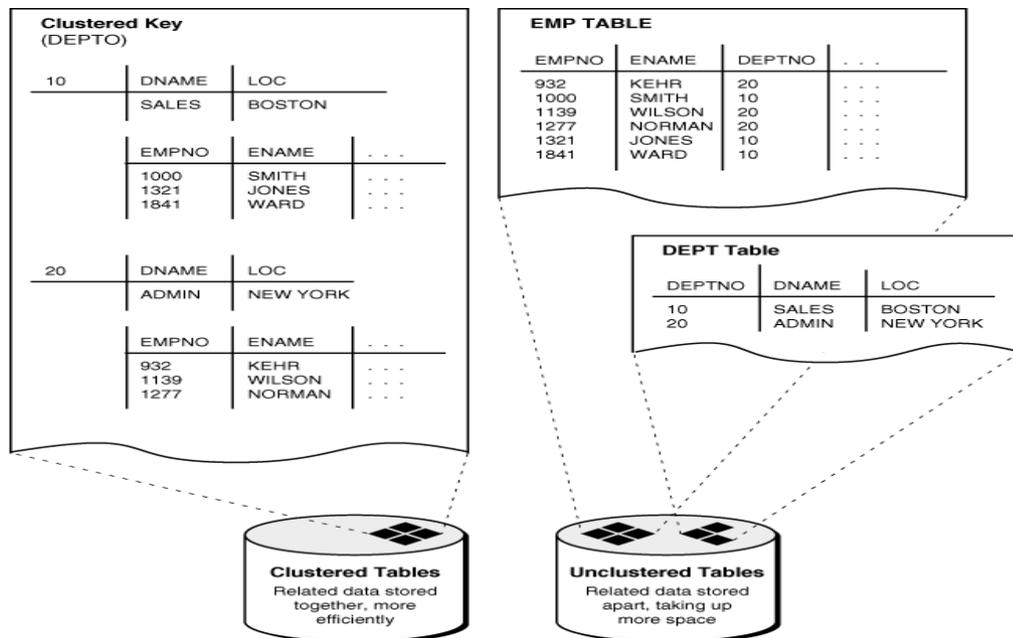


Fig.4: Example of a Clustered data.

The following figure shows the example of clustered data which will present in a database. This data is separated from the data which is present in the database which is shown in the figure. This technique helps to mine the data from the database when using in Web Data Mining Framework. By Using these framework we can mine the complex data and extract the data which the user wants. The integrity of the information remains intact on the web and it has become very easy to classify data directly into its various forms. Through the use of these webs mining architecture in e-business will help in faster method for extracting information from the data and has proper prospects for the usage of web pages and websites.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly Peer Reviewed Journal)

Vol. 5, Issue 2, February 2016

IV. EXPERIMENTAL RESULTS

Using this concept in the Web data mining frame work has given wide areas for this framework as a set in e-business and manage the huge amount of data which would be present in that database. It has put forward the different techniques for web structure mining analysis, Web content and structural mining and process analysis. It is proposed to implement a new Web knowledge mining framework used to manage and to give a probable structure for a page on web. The result of this project will be clearly indicated that how web mining (in a broad sense, Big Data Mining applied to ecommerce) is applicable to improving the services provided by e-commerce based enterprises. Specifically, we first discussed some recent approaches and techniques used in data mining. We now present some ways in which web data mining can be extended for further research. Implementing web data mining framework will provide a new approach towards big data mining technique.

V. CONCLUSION

We have implemented “Web Data Mining Framework for Developing E- Business Solution”. It can be used by different kinds of people or customers in various part of the world through web. It will help in managing large volume of data. Proper steps will ensure bright future.

REFERENCES

- [1] Hao Zhang, Gang Chen, Kian-Lee Tan, Meihui Zhang, “In-Memory Big Data Management and Processing: A Survey”, IEEE Transaction on knowledge and Engineering, Volume No. 27, Page no.1920-1947, July 2015.
- [2] Hong Liu and Jin Hua Xu, “Web Service Framework Research of Data Mining in E-business”, Advanced in Control Engineering and Information Science, Proceeding Engineering, Volume No. 15, Page no.1968-1972, May 2011.
- [3] B.Madamsamy and J.Jebmalar Tamilsevi, “General Web Knowledge Framework Mining”, International Journal of Computer Science and Engineering (IJCSSE), Volume No.4, Page No.1744-1750, 10 October 2012.
- [4] Yun Xue and Jianbin Chen, “Research and Design of Web Data Mining in Personalized E-business”, International Symposium on Web Information Systems and Applications, ISBN-978 Volume No. 3, Page No.96-99, Nanchang, China, May 22-24, 2009.
- [5] Belsare Satish and Patil Sunil, “Study and Evaluation of user’s behaviour in e-commerce Using Data Mining”, Research Journal of Recent Sciences, Volume No. I, Page No. 375-387, 25, January 2012
- [6] Venkatadari.M and Dr Lokanatha C. Reddy, “A Review on Data mining from past to the future”, International Journal of Computer Applications, Volume No.7, Page No.19-22, February 3-5, 2011.
- [7] Xindong Wu and V, “Data Mining with Big Data”, IEEE Transaction on Knowledge and Data Engineering, Volume No. 26, Page No.97-107, January 2014.
- [8] Sonal Tiwari “A Web Usage Mining Framework for Business Intelligence”, International Journal of Electronics Communication and Computer Technology (IJECCCT), Volume No.1, Issue 1, Page No. 234-245, September 2009.
- [9] R.Manjusha and R.Ramachandran, “Web Mining Framework for Security in E-commerce”, IEEE-International Conference on Recent Trends in Information Technology, Volume No.2, Page No.1043- 1048, June 3-5, 2011.
- [10] Aarti Singh, “Agent Based Framework for Semantic Web Content Mining”, International Journal of Advancements in Technology, Volume No.13, Page no.976-989, 2005.