ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 11, November 2016

Segmentation of User Task Behaviour by Using Neural Network

Arti Dwivedi(Mtech scholar) , Asst. Prof. Umesh Lilhore

Abstract— **This paper, introduces "Segmentation of User's Task" to understand user search behavior. Today the massive use of the web has drastically increased the traffic in the network together with the load that web servers need to manage. Sometime user leaves the sites because of its high latency time, so this becomes one important subject for research. Web mining is considered to store the information in a specific format known as weblog. This important feature has lot of implementation such as query recommendation, query log analysis, and many other techniques for improvement. As the query classification is the main motive of this paper where one can generate the class of the query for providing better response of the query. This paper proposes a model where proper knowledge from the ontology and the web usage of the web feature vector are created for training the Error back propagation neural network. This model will increase the accuracy of the classification, so that web server response will be small. Here overall execution time will be reduced because the trained neural network runs in constant time.**

Index Terms— **Information Extraction, web log, web query ranking, web mining.**

I. INTRODUCTION

As the internet users are increasing each day, the requirement of the web world is becoming quite high. In order to increase the transparency and quickness in the work large amount of work depends on this digital network. This attracts many researchers for improving the performance of the network and reduces the latency time of the internet, so that things get simpler and fast for the daily users. Here hardware part is the way of optimizing the network but in parallel software also need to be updated. This paper focuses on optimizing the web power by learning the user behavior for reducing the latency time of searching the required matter of specific interest. As websites are very important source of information for almost all kind of requirements, so these requirements of people attract number of people to provide various services. But targeting the correct customer is basic requirement of the service or business. Research in this area has the objectives of helping e-commerce businesses in their decision making, assisting in the design of good Web sites and assisting the user while navigating the Web.

Web Mining Features:

Web Structure Mining: This feature develops the website page structure where links between the pages shows relation between pages. Web structure mining helps in finding the similar pages where relation between the websites also looks closer. Here importance of this feature in web mining is quite low as compared to other features.

Web Content Mining:

- Web content mining explains the non-manual search of data resources. So available online data in form of text, image is termed as content of the website.
- Mining for multimedia data come under the content of mining work. Here some kind of good quality of information crawls from the source for gathering information form the internet.
- This mining helps in two aspects first is information retrieval and other is data structuring from unstructured to semi-structure.

Web Usage Mining: This involves user activity in the work. Here it stores some kind of patterns or movement of the user on the browser or website. In abstract the potential strategic aims in each domain into mining goal as: prediction of the user's behavior within the site, comparison between expected and actual Web site usage, adjustment of the Web site to the

Manuscript received Nov, 2016.
Arti Dwivedi(Mtech student) *: Department of Computer Science & Engineering Mtech Co-ordinator NRI Institute of Information Science & Technology Bhopal India.*
Asst. Prof. Umesh Lilhore, *Department of Computer Science & Engineering Mtech Co-ordinator NRI Institute of Information Science & Technology Bhopal India.*

interests of its users. There are no definite distinctions between the Web usage mining and other two categories.

II. TECHNIQUES OF CLUSTERING

Partitional Clustering: The partitional clustering algorithms use a feature vector matrix1 and produce the clusters by optimizing a criterion function. Such criterion functions are as follows: Maximize the sum of the average pairwise cosine similarities between the query's assigned to a cluster, minimize the cosine similarity of each cluster centroid to the centroid of the entire collection etc. [6] compared eight criterion functions and concluded that the selection of a criterion function can affect the clustering solution and that the overall quality depends on the degree to which they can correctly operate when the dataset contains clusters of different densities and the degree to which they can produce balanced clusters. The most common partitional clustering algorithm is k-means, which relies on the idea that the center of the cluster, called centroid, can be a good representation of the cluster.

Hierarchical Clustering: Hierarchical clustering algorithms produce a sequence of nested partitions. Usually the similarity between each pair of query's is stored in a **nxn** similarity matrix. At each stage, the algorithm either merges two clusters (agglomerative methods) or splits a cluster in two (divisive methods). The result of the clustering can be displayed in a tree-like structure, called a dendrogram, with one cluster at the top containing all the query of the collection and many clusters at the bottom with one query each. By choosing the appropriate level of the dendrogram we get a partitioning into as many clusters as we wish. The dendrogram is a useful representation when considering retrieval from a clustered set of querys, since it indicates the paths that the retrieval precess may follow [7].

Graph based clustering: In this case the query's to be clustered can be viewed as a set of nodes and the edges between the nodes represent the relationship between them. In [8] edges bare a weight, which denotes the strength of that relationship. Graph based algorithms rely on graph partitioning, that is, they identify the clusters by cutting edges from the graph such that the edge-cut, i.e. the sum of the weights of the edges that are cut, is minimized. Since each edge in the graph represents the similarity between the query's, by cutting the edges with the minimum sum of weights the algorithm minimizes the similarity between query's in different clusters. The basic idea is that the weights of the edges in the same cluster will be greater than the weights of the edges across the clusters. Hence, the resulting cluster will contain highly related querys.

Neural Network based Clustering: The Kohonen's Self-Organizing feature Maps (SOM) is a widely used unsupervised neural network model. It consists of two layers: the input layer with n input nodes, which correspond to the n query's, and an output layer with k output nodes, which correspond to k decision regions (i.e. clusters) [9, 13]. The input units receive the input data and propagate them onto the output units. Each of the k output units is assigned a weight vector. During each learning step, a query from the collection is associated with the output node, which has the most similar weight vector. The weight vector of that 'winner' node is then adapted in such a way that it will become even more similar to the vector that represents that query, i.e. the weight vector of the output node 'moves closer' to the feature vector of the query. This process runs iteratively until there are no more changes in the weight vectors of the output nodes. The output of the algorithm is the arrangement of the input query in a 2-dimensional space in such a way that the similarity between the input query's is mirrored in terms of topographic distance between the k decision regions.

Link-based clustering

Text-based clustering approaches were developed for use in small, static and homogeneous collections of query's. On the contrary, the www is a huge collection of heterogeneous and interconnected web pages. Moreover, the web pages have

additional information attached to them (web query metadata, hyperlinks) that can be very useful to clustering. According to [10], 'the link structure of a hypermedia environment can be a rich source of information about the content of the environment'. The link-based query clustering approaches take into account information extracted by the link structure of the collection. The underlying idea is that when two queries are connected via a link there exists a semantic relationship between them, which can be the basis for the partitioning of the collection into clusters.

III. Literature Survey

Hai Dong, Farookh Hussain and Elizabeth Chang in [1] proposed Web Query Classification technique which depends on web distance normalization. In this architecture middle categorized queries are send to the target class by normalizing and mapping the web queries. By defining the frequency, position and position frequency categories are ranked into three class. In the system Taxonomy-Bridging Algorithm is used to map target category. The Open Directory Project (ODP) is used to build an ODP-based classifier. This taxonomy is then mapped to the target categories using Taxonomy-Bridging Algorithm. Thus, the post-retrieval query is first classified into the ODP taxonomy, and the classifications are then mapped into the target categories for web query.

Classification of web query to the user intendant query is major task for any information retrieval system. MyoMyo ThanNaing [2] proposed "Query Classification Algorithm". To classify the web query input by the user into the user intended categories, MyoMyo ThanNaing use the domain ontology. Ontology is useful in matching of retrieve category to target category. User query are extracted in Domain terms are used as input to the query classification algorithm. Matched terms of each domain term are extracted in further sub category. Compute the probability for matched categories. Then all querys are ranked by their probability and displays to the user's desk.

Ernesto William De Luca and Andreas Nürnberger [3] proposed method of web query classification using sense folder. In this method the user query is separated in small terms. These small terms are matched with target categories using ontology. Ontology is specifically the set of rules. Word vectors (prototypes) are used to create semantic category. Then Search results are indexed by using sense folder.

At last retrieved queries are displayed to the user desk. Suha S. Oleiwi, Azman Yasin [4] proposed method of web query classification using Ontology and classification. All are retrieve queries indexed according to their probability. Probability depends on how often the queries are searched on web by user.

Another study proposed an algorithm named Query-Query Semantic Based Similarity Algorithm (QQSSA). This algorithm works on a new approach it filters and breaks the long query into small words and filters all possible prepositions, conjunction, article, special characters and other sentence delimiters from the query. And then expand the query into logically similar word to form the collection of similar words. Construct the Hyponym Tree for query1 and query2 etc. And based upon some distance measure he classifies the query.

Another approach is Classification methodology by S. loelyn Rose, K R Chandran and M Nithya [5]. The classification methodology can be fragmented into the following phases. Feature Extraction, and Mapping intermediate categories to target categories The features extracted in the first phase are mapped onto various target categories in this second phase by Direct Mapping, Glossary Mapping, Wordnet Mapping, Semantic Similarity Measure.

IV. Objective

- Increase page prediction accuracy.
- Decrease execution time of the prediction algorithm.
- Decrease space complexity for the prediction algorithm.
- Features selection is not website dependent.

V. Problem Formulation

In [11] entire work is focused on user search behavior clustering where two algorithms proposed first the SPread method (QC-SP) and second the Query Clustering using Bounded SPread method (QC-BSP) where clustering time of the first algorithm is high as compared to the second. Here they simply compare the key words of the query with the previous one. In second algorithm QC-BSP bounded time is used for clustering, so less number of comparison is use for clustering.

This approach is named as Query Clustering using Weighted Connected Component of a Graph (QC-WCC) and outperformed other popular clustering algorithms like Query Flow Graph, K-means, and DBScan in task clustering, as indicated in [13]. One major shortcoming of QC-WCC is its high time complexity of constructing the graph and extracting connected component, which is $O(N)^2$ where N is the average number of queries of a session and k is the dimension of features. Considering the massive volume of search logs some sessions could be very long (depending on the time threshold of session segmentation), the overall time complexity for the entire search logs is intolerable.

VII. PROPOSED MODEL

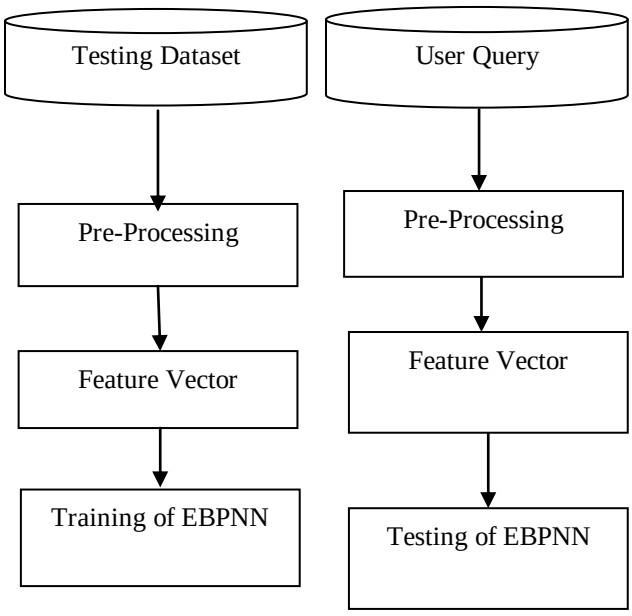

Fig. 1 The Proposed query segmentation model.

VIII. EXPECTED OUTCOME

As the query classification is the main motive of this work where one can generate the class of the query it belong for giving better response of the query. With the proper knowledge from the ontology and the web usage of the web feature vector are created for training the Error back propagation neural network.. By the use of EBPNN classification the query gets efficient and will be less time consuming. This work will increase the accuracy of the classification so the web server responsed will be small. Here overall execution time will be reduce as well because trained neural network run in constant time.

IX. CONCLUSIONS

As the user satisfaction plays an important role in information retrieval. Query recommendation is one of the best method for helping users to satisfy the users information need by suggesting queries related to current users need by maintaining query log processing files, by using past historical navigation patterns, by updating the records of query processing so that by using dynamic and static log data, also by using clicked snippets, and so on. This paper helps to review some of these query recommendation techniques with the limitations and advantages. There is no general method developed till now which rank the pages efficiently, as different webs has different requirement.

REFERENCES

[1] An Ontology-based Webpage Classification Approach for the Knowledge Grid Environment by Hai Dong, Farookh Hussain and Elizabeth Chang, 2009 Fifth International Conference on Semantics, Knowledge and Grid (IEEE-2009).

[2]Ontology-Based Web Query Classification For Research Paper Searching , By Myomyo Thannaing, International Journal Of Innovations In Engineering And Technology (Ijiet) , Vol. 2 Issue 1 February 2013.

[3] Ontology-Based Semantic Online Classification Of Querys: Supporting Users In Searching The Web By Ernesto William De Luca And Andreas Nürnberger, Ijcst, 2012.

[4] Web Query Classification To Multi Categories Based On Ontology By Suha S. Oleiwi, Azman Yasin, International Journal Of Digital Content Technology And Its Applications(Jdcta) Volume7, Number13, Sep 2013.

[5] S.Lovelyn Rose, K.R.Chandran, M.Nithya An Efficient Approach To Web Query Classification Using State Space Trees., Issn :2229-4333, International Journal Of Computer Science And Technology (Ijcst), June-2011.

[6] Zhao, Y., Karypis, G. 2001. Criterion Functions For Query Clustering:Experiments And Analysis. Technical Report #01-40. University Of Minnesota, Computer Science Department. Minneapolis, Mn (Http://Wwwusers. Cs.Umn.Edu/~Karypis/Publications/Ir.Html)

[7] Zhao, Y., Karypis, G. 2002. Evaluation Of Hierarchical Clustering Algorithms For Query Datasets, Acm Press, 16:515-524.

[8] San San Tint1 And May Yi Aung. "Web Graph Clustering Using Hyperlink Structure ".Advanced Computational Intelligence: An International Journal (Acii), Vol.1, No.2, October 2014

[9] Khan, M. S., & Khor, S. W. (2004). Web Query Clustering Using A Hybrid Neural Network. Applied Soft Computing, 4(4), 423-432. 17

[10] Kleinberg, J. 1997. " Web Usage Mining For Enhancing Search Result Delivery And Helping Users To Find Interesting Web Content",I Acm Sigir Conf. Research And Development In Information Retrival (Sigir '13), Pp. 765-769,2013.

[11] Mamoun A. Awad And Issa Khalil "Prediction Of User's Web-Browsing Behavior: Application Of Markov Model". Ieee Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics, Vol. 42, No. 4, August 2012.

[12] Thi Thanh Sang Nguyen, Hai Yan Lu, Jie Lu " Web-Page Recommendation Based On Web Usage And Domain Knowledge" 1041-4347/13/$31.00 © 2013 Ieee.

[13] Zhen Liao, Yang Song, Yalou Huang, Li-Wei He, And Qi He. "Task Trail: An Effective Segmentation Of User Search Behavior" . Ieee Transactions On Knowledge And Data Engineering, Vol. 26, No. 12, December 2014.