

# Survey On Techniques And Features Of Dynamic Webpage Recommendation

**Dr. Yachana Bhawsar**

Lecturer

Government Women's Polytechnic College, Bhopal (M.P.)

---

**Abstract**— Website content attracts people to visit on page but retention on site depends on its search and relevancy. Many of sites are working on dynamic page recommendation based on user behavior. Some of sites provide query based search to understand the user requirement and provide better solutions. This paper has provided a understanding of web page recommendation research area, its importance. Work proposed by scholars was also detailed with proposed methods. Paper has summarized some of techniques of dynamic page prediction with their pros and cons. As mining need some features before analysis or finding patterns, so paper has list some of important web mining features as well. Paper has finds evaluation parameters of webpage recommendation parameter algorithms.

**Keywords:** Data mining, Feature Extraction, Page Recommendation, Web mining.

## **I. Introduction**

People now live in a data-driven age where practically every element of everyday life is data-driven and capable of digital monitoring and support. We produce a wide range of data kinds whether we shop, work, socialize, communicate, unwind, or travel [1, 2]. These include numerical variables, photos, videos, music, and text. Data science is an interdisciplinary discipline that combines computer science, statistics, and mathematics. It uses systems, algorithms, and scientific approaches to extract insights from organized and unstructured data. Improving decision-making processes is its main goal.

Online search algorithms, which form the core of the learning paradigm, have made a variety of online resources indispensable for learners in recent times. For students looking for internet resources or information, search engines like Google, Yahoo, Bing, CC Search, Search Encrypt, OneSearch, and Wiki.com are invaluable [3]. These search engines provide quick access to any topic's internet resources and provide results in text, picture, and video forms. As a result, students are able to look up material in text, video, and picture formats, and this trend of looking up information online is only getting stronger every year [4].

Recommender systems have become an important data science application in this area [5]. These systems use cutting-edge analytics and machine learning techniques to sort through enormous amounts of data and

---

give users useful information. This helps users make decisions about food, clothes, housing, travel, employment, education, and entertainment, among other daily life aspects.

In the field of data analysis and computers, artificial intelligence (AI), particularly machine learning (ML), has grown significantly, allowing applications to function intelligently [6]. The prediction model has the ability to suggest particular websites or web pages that have pertinent data on them. Web navigation prediction can help search engines and other web-based applications provide more relevant search results and suggestions in addition to improving the user's browsing experience. These systems are able to better match users with the material they need by analyzing user online browsing patterns and their expected destinations. This can thus result in a more effective and efficient online search process.

The remainder of this paper is structured into four sections. In the next section, paper has summarized papers of webpage recommendations done. Further techniques were compared on the basis of positive outcomes of techniques with limitations. In next section paper has list features of the web mining. Finally evaluation parameters were list in the paper for comparing different techniques.

## **II. Related Work**

Honey Jindal et al. [7] emphasize the identification and elimination of noisy information from user-navigated web pages. Pruning user-browsed sessions by removing noisy webpages and their relations can aid in developing high-performing prediction models with fewer errors. Four pruning models, namely PM3ER, PM3EN, PM3GI, and PM3EP, have been proposed to minimize prediction errors by eliminating noisy web pages and their relations within a model comprising diverse navigation patterns.

A unique Autoencoder model based on high-order structure is proposed by Tao Yi et al. [8] in order to overcome the drawbacks of the current link prediction techniques. They have developed an efficient motif adjacency matrix learning technique that makes it possible to create many motif adjacency matrices and extract high-order structures from directed networks. To address the link prediction problem in directed networks, they expand the Graph Autoencoder (GAE) and Variational Autoencoder (VAE) frameworks.

A novel QoS prediction technique based on a heterogeneous graph attention network is introduced by S. Lv et al. [9]. They build an attributed user-service network using user location data. In this network, node- and semantic-level attentions are covered by a hierarchical attention machine (HGN2HIA). Semantic-level attention evaluates the relative relevance of distinct meta-paths, whereas node-level attention determines the significance of a node in relation to its neighbors based on meta-paths. With this method, user embeddings for QoS prediction are produced in a hierarchical fashion.

L. Tang et al. [10] present a deep learning-based framework for detecting phishing websites. They implement this framework as a browser plug-in capable of real-time phishing risk assessment when users visit web pages. The service combines multiple strategies, including whitelist filtering, blacklist interception, and machine learning (ML) prediction, to enhance accuracy and reduce false alarms.

Gharibshah, Z et al. [11] propose two deep learning-based frameworks, LSTMcp and LSTMip, for user click prediction and user interest modeling in the context of advertising. Their goal is to accurately predict user

click probabilities and specific Ad campaign preferences. They achieve this by collecting page information displayed to users and using LSTM networks to learn latent features representing user interests.

Afolabi et al. [12] introduce a semantic web content mining approach for recommender systems in online shopping. This approach involves two phases: semantic preprocessing of textual data using developed and existing ontologies and recommendation generation using the Naïve Bayes algorithm.

Hashemi et al. [13] explore web page classification methods and their application scenarios, employing machine learning algorithms to categorize web pages based on features extracted from text and HTML tags. Their findings highlight the importance of HTML tags in identifying web page topics using keyword frequencies, an aspect that is often overlooked in classification methods.

### III. Techniques of Page Recommendation

Year	Techniques	Pros	Limitations
2014 [14]	Task Trailing and Query Segmentation	Enhance page suggestion by analyzing the user's search history from their browser or the work they've completed recently.	Requires the user's consent before starting any task trail sequences.
2015 [15]	Aprior Algorithm	Use site logs to analyze user activity; personal privacy is unaffected.	The process of generating patterns from datasets takes time.
2016 [16]	Deep Neural Network Dimension Reduction Method	Make advantage of item and user features in order to forecast. Describe a technique for data cleaning to improve the accuracy of the pattern.	Low prediction accuracy
2017 [17]	Recurrent Neural Network	RNN learning of user behavior (Browser History).	Only the user's current behavior determines the work.
2019 [18]	Particle Agent Swarm Optimization	Make use of the content and web log features.	Using PSO for content feature selection lowers work accuracy

2020 [19]	Improved Triangle Similarity	User product rating for product / Page Recommendation	Other features of web was not utilizes, so new product ranking is difficult
2020 [20]	Latent Dirichlet Allocation	Product recommendation / page recommendation based on user ratings.	Web logs and content features are not used since it is not always possible to obtain information about user actions.

#### IV. Features of Web Recommendation

##### Web Content Mining

It is the concept to search any particular data online and also involves the use of mining web data contents. Multimedia data mining is one of the parts of content mining that mine the high-level information from these multimedia sources.

**Term Frequency:** It gives the count of the category of words in documents. So this TF is the occurrence of words of a similar category in an article or document.[15]

Document Term Frequency: Gives the number of documents that contain any particular term.

**IDF:** or inverse document frequency shows the ability to provide information of words in a document by categorizing it common or rare. It is the value of a logarithmically inverse fraction of the total documents that contain any word.

$$IDF(t) = \log \left( \frac{N}{n} \right)$$

In which n= total number of documents that contain in dataset and n is the number of times that term t appears in the document.

**TF-IDF [19]** (Term Frequency-inverse Document Frequency): weight the terms based on inverse document frequency. It simply means the more the term is common in all the documents the less that term is important and so will be weighted less.

$$\text{TFIDF}(t) = \text{TF}_t * \log\left(\frac{N}{n_t}\right)$$

**TF-IDF-CF:** TF-IDF was having some shortcomings and so this new parameter was introduced to determine class characteristics, and this class was called frequency by authors and it calculates the term frequency in documents belonging to a particular class.

$$\text{TFIDFCF}(t) = \log(\text{TF}_t + 1) * \log\left(\frac{N + 1}{n_t}\right) * \frac{n_{c,t}}{N_c}$$

Frequency of appearance of  $t$  in any document and  $N_c$  is the total number of documents within same class  $c$  document.

### Web Usage Mining

This is the process of discovering useful values based on user interactions while using the web. The goal of this mining is to predict the user behavior and adjust the website based on their interest [21]. There are no definite distinctions between the other two categories and this web usage mining.

**Markov Model:** It involves series of numbers to develop a  $K$ th order Markov model. Numeric data set such as the weblog page were used to get these patterns.

**Regression:** As per need several linear and logistic features were obtained from numeric data. To find any feature from temporal data regression is used.

### V. Evaluation Parameter

**Precision** of a transaction is provided as the ratio of the number of web pages appropriately predicted and the overall amount of web pages predicted.

$$\text{Precision} = \text{Approximate\_Correct\_pages} / \text{All\_predictions}$$

**Coverage** is calculated as the ratio of the amount of web pages appropriately predicted and the overall amount of web pages visited by the user.

$$\text{Coverage} = \text{Approximate\_Correct\_pages} / \text{All\_Visited\_Pages}$$

**M-metric** is utilized with the intention of obtaining a single evaluation measure, and it is defined in this manner

$$\text{Precision} = (2 * \text{Precision} * \text{Coverage}) / (\text{Precision} + \text{Coverage})$$

### Execution Time

Total Time taken for the execution of the algorithm in the prediction of the page base on the different size of dataset.

## VI. Conclusions

This paper has described the requirement of web page prediction to increase the retention of web user on site and improve sale / service quality of site. This paper has summarized various research work done by scholars in field of site web page prediction. Some of important techniques and features of web mining were also list in the paper with their importance. It was obtained from survey that focus on user search query may improve the prediction performance of website. Use with weblog pattern feature and web content feature increase the result accuracy of prediction as well.

## References

1. H.-J. Tsuei, W.-H. Tsai, F.-T. Pan, and G.-H. Tzeng, "Improving search engine optimization (SEO) by using hybrid modified MCDM models," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 1–16, 2020.
2. Z. Hadjilambrou, M. Kleanthous, G. Antoniou, A. Portero, and Y. Sazeides, "Comprehensive characterization of an open source document search engine," *ACM Transactions on Architecture and Code Optimization*, vol. 16, no. 2, pp. 1–21, 2019.
3. AUsta, I. S. Altingovde, R. Ozcan, and O. Ulusoy, "Learning to rank for educational search engines," *IEEE Transactions on Learning Technologies*, vol. 14, no. 2, pp. 211–225, 2021.
4. T. Denoeux, *Methods for Building Belief Functions*, University of Technology of Compiègne, Compiègne, France, pp. 1–67, October 2019.
5. N. N. Akram and V. Ilango, "Intelligent Web Mining Techniques using Semantic Web," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 1-7.
6. APradeep, "Web Mining: Opportunities, Challenges, and Future Directions," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-6.
7. Honey Jindal, Neetu Sardana, Ankit Vidyarthi, Deepak Gupta, Mufti Mahmud. "Mining user's navigation structure by filtering impurity nodes for generating relevant predictions", *International Journal of Cognitive Computing in Engineering*, Volume 4, 2023, Pages 248-258.
8. Tao Yi, Shanfan Zhang, Zhan Bu, Jinwei Du, Changjian Fang. "Link prediction based on higher-order structure extraction and autoencoder learning in directed networks", *Knowledge-Based Systems*, Volume 241, 2022.
9. S. Lv, F. Yi, P. He and C. Zeng, "QoS Prediction of Web Services Based on a Two-Level Heterogeneous Graph Attention Network," in *IEEE Access*, vol. 10, pp. 1871-1880, 2022, doi: 10.1109/ACCESS.2021.3138127.
10. L. Tang and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," in *IEEE Access*, vol. 10, pp. 1509-1521, 2022, doi: 10.1109/ACCESS.2021.3137636.

11. Gharibshah, Z., Zhu, X., Hainline, A. et al. Deep Learning for User Interest and Response Prediction in Online Display Advertising. *Data Sci. Eng.* 5, 12–26 (2020).
12. Afolabi, Ibukun Tolulope, et al. "Semantic Web mining for Content-Based Online Shopping Recommender Systems." *IJIT* vol.15, no.4 2019: pp.41-56.
13. Hashemi, M. Web page classification: A survey of perspectives, gaps, and future directions. *Multimed. Tools Appl.* 2020, 79, 11921–11945.
14. 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.
15. A. Ladekar, P. Pawar, D. Raikar and J. Chaudhari, "Web Log Based Analysis of User's Browsing Behavior", *IJCSIT - International Journal of Computer Science and Information Technologies*, Vol. 6 (2), 2015.
16. Yang Song, Ali Mamdouh Elkahky, and Xiaodong He. 2016. Multi-rate deep learning for temporal recommendation. In *SIGIR*. 909–912.
17. Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based News Recommendation for Millions of Users. In *SIGKDD*.
18. R. Manikandan. "A novel approach on Particle Agent Swarm Optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective". Springer Science+Business Media, LLC, part of Springer Nature 10 January 2019.
19. MINH-DUC NGUYEN, AND YOON-SIK CHO. "A Hybrid Generative Model for Online User Behavior Prediction". *IEEE Access*, volume 8, January 7, 2020.
20. Arta Iftikhar, Mustansar Ali Ghazanfar, Mubbashir Ayub, Zahid Mehmood, And Muazzam Maqsood. "An Improved Product Recommendation Method for Collaborative Filtering". *IEEE Access*, volume 8, July 20, 2020.
21. P.G. Om Prakash, and A. Jaya. "WS-BD-Based Two-Level Match: Interesting Sequential Patterns and Bayesian Fuzzy Clustering for Predicting the Web Pages from Weblogs". *The Computer Journal*, Volume 63, Issue 1, 2020.