

## Webpage Recommendation by Calculating Average Visiting Time on Webpages

Sejal D<sup>a</sup>, Kamalakant T<sup>a</sup>, Dinesh Anvekar<sup>b</sup>, Venugopal K R<sup>a</sup>, S S Iyengar<sup>c</sup>, L M Patnaik<sup>d</sup>

<sup>a</sup>Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore-560001, sej\_nim@yahoo.co.in

<sup>b</sup>Alpha College of Engineering, Bangalore

<sup>c</sup>Florida International University, USA

<sup>d</sup>Indian Institute of Science, Bangalore, India

Predicting users' future requests based on previously visited pages is important for web page recommendation, reduction of latency and on-line advertising. These applications compromise with prediction accuracy and modelling complexity. We propose a method for webpage Recommendation(WNPWR) which creates and generates a classifier based on sessions as training examples. As sessions are used as training examples, they are created by calculating the average time on visiting web pages rather than traditional method which uses 30 minutes as default timeout. This paper uses standard benchmark datasets to analyse and compare our framework with two-tier prediction framework. Simulation results show that our generated classifier framework WNPWR outperforms two-tier prediction framework in prediction accuracy and time.

**Keywords :** Web Navigation, Web Prediction, Webpage Recommendation.

### 1. INTRODUCTION

The possibility of visiting a web page by a user based on the history of earlier accessed web pages is known as web prediction. Prediction of web user's behaviour is critical in web mining to enhance the performance of the search engine. The organisation of the web is modelled as a graph, where each node represents a web-portal, and the edge represents user's navigation. Distribution of all web pages visited can be calculated and utilized in reweighting and reranking results. The information provided by the navigation path is of prime importance than the query given by the user. Web cache performance of search engine can be improved by storing predicted pages in the cache.

Behavioural targeting is a key issue of predicting future behaviour of web users. Behavioural targeting is a technique to improve efficiency of advertising by online website publishers and advertisers by extracting knowledge

of web browsing behaviour of users. Behaviour targeting selects advertisement to display with the help of web-browsing behaviour of users. The user analysis approach is the centre of interest in on-line advertising and properly targeted advertisements generate more consumer interest.

Predicting web user's shopping behaviour has an important role in product recommendation. Product recommendations is the dynamic shopping recommendations across mobile, email and web channels. It depends on each customer's past and current purchases behaviours. It also helps in website optimization, improve conversions and increases revenue by making related product recommendations to the customers.

*Motivation:* The web log records the navigational behaviour of the user. Pre-processing of the raw data is required before giving these data as input to prediction model. Pre-

with 30 minutes time-out. It is observed in  $NASA_{Medium}$  dataset that the actual user sessions are more than 30 minutes, hence these sessions has misclassified classifiers and accuracy is decreased. In WAR, sessions are created by calculating the average time of visiting web pages, hence accuracy increases with increase in training datasets.

Figure 2 presents a comparison between TPFramework and WRA in terms of prediction time by using different training dataset percentage on both datasets. It is observed from Figure 2 that prediction time decreases with increases in training datasets and decrease in testing datasets. The average prediction time for WRA is decreased by 35.0350% and 15.6042% for  $NASA_{Low}$  and  $NASA_{Medium}$  datasets respectively in comparison to TPFramework.

The major difference between our proposed framework and Two-tier framework is methodology of session identification (in our framework Method 2 which is discussed in user and session identification section). From table 6, it can be observed that method 2 has generated more number of sessions compared to method 1 which is used as training dataset. In method 2, the average visiting time depends on the visiting webpages, so sessions are mapped to a realistic value than when a single constant value is used. The WRA results in high prediction accuracy in comparison with TPFramework an account of these reasons. Even the average session length is small in method 2 compared to method 1, which results in lower prediction time. It is also observed that session generation time with method 2 is more compared to method 1.

## 6. CONCLUSIONS

In this work, we have proposed Webpage Recommendation Algorithm which creates and generates a classifier based on sessions as training examples. Sessions are created by calculating average time on visiting web pages, which maps to the realistic better value than

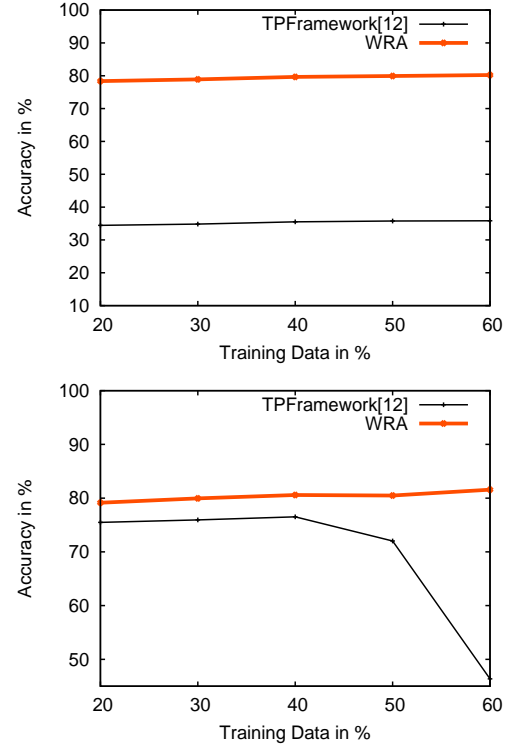


Figure 1. Prediction Accuracy Comparison between TPFramework and WRA for  $NASA_{Low}$  and  $NASA_{Medium}$  Datasets

when a single constant value is used. Each session is mapped to one or more generated classifiers. Simulations are performed on NASA dataset and are compared with two-tier prediction framework. Our algorithm outperforms two-tier prediction framework by providing high prediction accuracy with reduced prediction time. Further, we plan to extend this paper for online webpage recommendations for mobile applications.

## REFERENCES

1. Faten Khali, Jiuyong Li, and Hua Wang. Integrating Recommendation Models for Improved Web Page Prediction Accuracy. *ACSC '08 : In the Proceedings of the 31<sup>st</sup> Australasian Conference on Computer Science*, pages 91–100, 2008.
2. Jose Borges and Mark Levene. Evaluating Variable-Length Markov Chain Models for

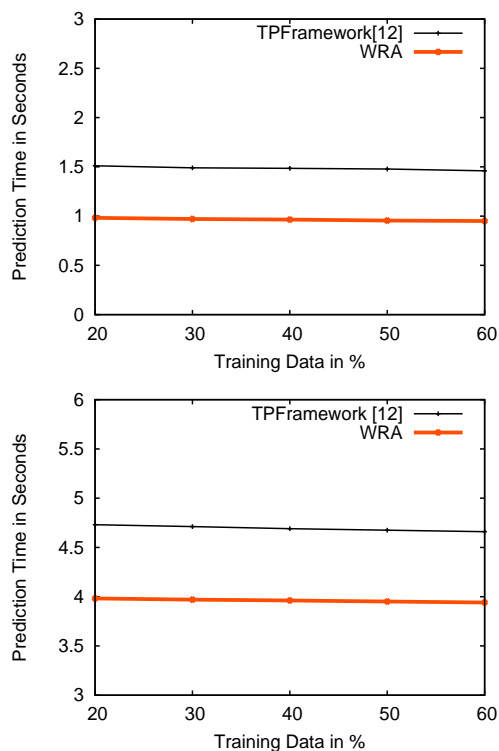


Figure 2. Prediction time Comparison between TPFramework and WRA for  $NASA_{Low}$  and  $NASA_{Medium}$  Datasets

- Analysis of User Web Navigation Sessions. *IEEE Transactions on Knowledge and Data Engineering*, 19(4):441–452, April 2007.
3. Costantinos Dimopoulos, Christos Makris, Yannis Panagis, Evangelos Theodoridis, and Athanasios Tsakalidis. A Web Page Usage Prediction Scheme using Sequence Indexing and Clustering Techniques. *Journal on Data and Knowledge Engineering*, 69(4):371–382, April 2010.
  4. Chu-Hui Lee, Yu-lung Lo, and Yu-Hsiang Fu. A Novel Prediction Model based on Hierarchical Characteristics of Web Site. *International Journal of Expert Systems with Applications*, 38(4):3422–3430, April 2011.
  5. Abdelghani Guerbas, Omar Addam, Omar Zaarour, Mohamad Nagi, Ahmad Elhajj, Mick Ridley, and Reda Alhajj. Effective Web Log Mining and Online Navigational Pattern Prediction. *Journal on Knowledge Based Systems*, 49(2):50–62, September 2013.
  6. Mukund Deshpande and George Karypis. Selective Markov Models for Predicting Web-Page Accesses. *ACM Transaction on Internet Technology*, 4(2):163–184, 2004.
  7. Mamoun A Awad and Latifur R Khan. Web Navigation Prediction using Multiple Evidence Combination and Domain Knowledge. *IEEE Transactions on System, Man and Cybernetics-Part A:Systems and Humans*, 37(6):1054–1062, November 2007.
  8. Krzysztof Dembczyński, Wojciech Kotłowski, and Marcin Sydow. Effective Prediction of Web User Behavior with User-Level Models. *Journal on Fundamental Informatica*, 89(3):189–206, 2008.
  9. Chao Liu, Ryen W White, and Susan Dumais. Understanding Web Browsing Behaviors through Weibull Analysis of Dwell Time. *SIGIR '10 : In the Proceedings of 33<sup>rd</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 379–386, 2010.
  10. Siriporn Chiphlee, Witcha Chiphlee, Naomie Salim, and Mohd Salihin Bin Ngadiman. Using Hybrid Markov Model for Web Access Prediction. *Journal of Information Technology*, 3(3):86–91, 2012.
  11. Thwe Phyu. Proposed Approach for Web Page Access Prediction using Popularity and Similarity based PageRank Algorithm. *International Journal of Scientific and Technology Research*, 2(3):240–246, 2013.
  12. 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*, 42(4):1131–1142, August 2012.
  13. Frederick Kin Hing Phoa and Juana Sanchez. Modelling the Browsing Behavior of World Wide Web Users. *Open Journal of Statistics*, 3(2):145–154, 2013.
  14. Ryen W White, Peter Bailey, and Liwei Chen. Predicting User Interests from Contextual Information. *In the Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 363–370, 2009.
  15. Ruma Dutta, Anirban Kundu, and Debajyoti Mukhopadhyay. Clustering-based Web Page Prediction. *International Journal on Knowledge and Web Intelligence*, 2(4):257–271, 2011.
  16. Robert Cooley, Bamshad Mobasher, and

Jaideep Srivastava. Data Preparation for Mining World Wide Web Browsing Patterns. *Journal on Knowledge and Information Systems*, 1(1):5–32, 1999.

17. CE Dinuca and D Ciobanu. Improving the Session Identification using the Mean Time. *International Journal of Mathematical Models and Methods in Applied Sciences*, 6(2):265–272, 2012.
18. Internet Traffic Archive, [Online]. Available : <http://ita.ee.lbl.gov/html/contrib/sask-http.html>.



**Sejal D** is a research scholar in the department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. She received her Bachelor degree in Computer Science and Engineering from Gujarat University and Master degree from University Visvesvaraya College of Engineering, Bangalore University, Bangalore. Her research interests include Data Mining and Web Mining.



**KamalaKant T** received BE degree in Information Science and Engineering from Visvesvaraya Technological University, Belgaum, in 2011 and the ME degree in Computer Science and engineering, from University Visvesvaraya College of Engineering, Bangalore, Karnataka, in 2014. His research interests include data mining and web mining.



**Dinesh K Anvekar** is currently working as Dean at Alpha College of Engineering, Bangalore. He obtained his Bachelor degree from University of Visvesvaraya college of Engineering. He received his Master and Ph.D degree from Indian Institute of Technology. He received best Ph.D Thesis Award of Indian Institute of Science. He has 15 US patents issued for work done in IBM Solutions Research Center, Bell Labs, Lotus Interworks, and for Nokia Research Center, Finland.

He has filed 53 patents. He has authored one book and over 55 technical papers. He has received Invention Report Awards from No.kia Research Center, Finland, Lucent Technologies (Bell Labs) Award for paper contribution to Bell Labs Technical Journal and KAAS Young Scientist Award in Karnataka State, India.



**Venugopal K R** is currently the Principal, University Visvesvaraya College of Engineering, Bangalore University, Bangalore. He obtained his Bachelor of Engineering from University Visvesvaraya College of Engineering. He received his Masters degree in Computer Science and Automation from Indian Institute of Science Bangalore. He was awarded Ph.D. in Economics from Bangalore University and Ph.D. in Computer Science from Indian Institute of Technology, Madras. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science and Journalism. He has authored and edited 57 books on Computer Science and Economics, which include Petrodollar and the World Economy, C Aptitude, Mastering C, Microprocessor Programming, Mastering C++ and Digital Circuits and Systems etc., He has filed 101 patents. During his three decades of service at UVCE he has over 550 research papers to his credit. His research interests include Computer Networks, Wireless Sensor Networks, Parallel and Distributed Systems, Digital Signal Processing and Data Mining.



**S S Iyengar** is currently Ryder Professor, Florida International University, USA. He was Roy Paul Daniels Professor and Chairman of the Computer Science Department of Louisiana state University. He heads the Wireless Sensor Networks Laboratory and the Robotics Research Laboratory at USA. He has been involved with research in High Performance Algorithms, Data Structures, Sensor Fusion and Intelligent Systems, since receiving his Ph.D degree in 1974 from MSU, USA. He is Fellow of IEEE and ACM. He has directed over

40 Ph.D students and 100 post graduate students, many of whom are faculty of Major Universities worldwide or Scientists or Engineers at National Labs/Industries around the world. He has published more than 500 research papers and has authored/co-authored 6 books and edited 7 books. His books are published by John Wiley and Sons, CRC Press, Prentice Hall, Springer Verlag, IEEE Computer Society Press *etc.*, One of his books titled Introduction to Parallel Algorithms has been translated to Chinese.



**L M Patnaik** is currently Honorary Professor, Indian Institute of Science, Bangalore, India. He was a Vice Chancellor, Defense Institute of Advanced Technology, Pune, India and was a Professor since 1986 with the Department of CSA, Indian Institute of Science, Bangalore.

During the past 35 years of his service at the In-

stitute he has over 700 research publications in refereed International Journals and refereed International Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academies in India; Fellow of the IEEE and the Academy of Science for the Developing World. He has received twenty national and international awards; notable among them is the IEEE Technical Achievement Award for his significant contributions to High Performance Computing and Soft Computing. His areas of research interest have been Parallel and Distributed Computing, Mobile Computing, CAD, Soft Computing and Computational Neuroscience.