

# An Empirical Analysis of Web Navigation Prediction Techniques

Honey Jindal, Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India

Neetu Sardana, Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India

## ABSTRACT

With the advancement of Information Technology, web is growing rapidly and it has become necessary part of our daily lives. It is mandate to study the navigation behavior of the user to improve the quality of web site design for personalization and further recommendation. Analysis of web navigation behavior heavily relies on navigational models. This paper is an effort to give insights of current state-of-the-art techniques used for web navigation prediction. These navigation models are broadly classified into three categories: sequential mining, classification and clustering. Analytical analysis is performed on all the categories used in web navigation prediction. Further empirical analysis is performed on popular techniques of each category Markov Model (sequential mining), Support vector machine (classification) and K-means (clustering) on the common platform to measure the effectiveness of these techniques.

## KEYWORDS

Markov, Navigation, Prediction, Techniques, Web

## 1. INTRODUCTION

With the rapid growth of information technology, the usage of web has become the necessary part in our daily lives. There are immeasurable users across the world who interact daily with the web. Every user wants to have a system which presents their desired information within a single click. Therefore, web mining is required to organize the data in efficient ways.

Web navigation is an integral part of web. Web navigation refers to navigating web pages present on web which is organized as hypertext or hypermedia. The user can navigate the website using text links, breadcrumbs, navigation bars, sitemap, dropdown menu, flyout menu and named anchor. One of the major task in web navigation (Thi Thanh el al., 2014) is to predict the next web page for the user(s) by observing his navigational browsing patterns. This problem is referred as Web Navigation Prediction (WNP). WNP becomes one of the most prominent research areas due to their extensive usage in various applications for personalization (Qingyan Yang et al., 2010), pre-fetching and caching, e-commerce and website designing (Ruili Geng et al., 2015).

DOI: 10.4018/jcit.2017010101

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Several techniques have been proposed by the researchers for WNP in the past. This paper has done in depth analysis of various approaches present in literature for Web Navigation Prediction. The main categories identified for WNP are sequential mining, classification and clustering. Merits and demerits of all the techniques belonging to each category are analyzed. Further empirical analysis is performed on techniques of sequential mining, classification and clustering. The techniques compared are KMMM, SVM and K-Means.

The organization of the paper is as follows: In Section 2, we present the related work and contributions of the paper. Section 3, provides brief insights of WNP techniques. Section 4, presents the analytical inferences drawn from the literature. Section 5, describes the experimental results. Finally, Section 6 presents the conclusion and future trends are described in Section 7.

## **2. RELATED WORK**

Predicting the web navigation behaviour of the user has attracted the attention of many researchers and practitioners. There are many papers which present the broad survey on web usage mining. Surveys (Srivastava et al., 2000; Hu Chen, 2003; Pani & Saroj, 2011; Singh, Brijendra, 2010) present an overview of web usage mining algorithms and majorly focused on the pattern extraction techniques. In these papers, the phases required in pattern construction and the general data mining techniques are discussed.(Srivastava et al., 2000; Prasad, 2015; Facca et al., 2005) presented the survey in the context of various dimensions of web usage mining like data sources, navigation techniques (like association rule mining, sequential mining and clustering), and software tools used for analysis, applications and security.

(Renata Iváncsy et al., 2006) focused on frequent pattern mining techniques. The paper provides the brief summary of how frequent patterns can be obtained from the web logs. (Baraskar et al., 2015; Neve, 2013; Kapil, 2014) surveyed the sequential mining techniques and does not provide the insights of other WNP techniques. (Kumar et al., 2014) done a theoretical underpinning of user navigation using Markov based models. Observations are highlighted based on prior facts. Models are not implemented in a common platform to provide better insights. Most of these papers surveyed specifically on sequential mining only.

In prior studies, hybrid approaches were introduced to overcome the shortcomings of baseline navigation prediction approaches. (Bhawana et al., 2015) investigated the effectiveness of two Markov based models. (Thwe et al., 2014; Vishwakarma et al., 2013) analyzed the application of Markov models on clusters. (Sampath et al., 2013a, 2013b, 2014) present the effect of association rule mining on Markov models. They observed that integration of association rule mining in Markov model improves the prediction accuracy.

In contrast to these studies the goal of our study is to perform in-depth analysis of Web Navigation Prediction techniques. Performances of WNP techniques are compared on a common platform to conclude the best navigation prediction model. Our work differs from the existing work in many ways.

The contributions of this paper are summarized as:

- Analytical study is performed on the various techniques used for web navigation prediction;
- Empirical analysis has been done on all the categories;
- We conduct experiments on each category of Web Navigation Prediction using CTI dataset.

## **3. WEB NAVIGATION PREDICTION TECHNIQUES**

WNP techniques aim at discovering the knowledge from the web that can help in personalization and recommendation to the web users. Techniques available for WNP can be classified into three categories:

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/an-empirical-analysis-of-web-navigation-prediction-techniques/178467](http://www.igi-global.com/article/an-empirical-analysis-of-web-navigation-prediction-techniques/178467)

## Related Content

---

### Microarray Data Mining

Li-Min Fu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1224-1230).

[www.irma-international.org/chapter/microarray-data-mining/10978](http://www.irma-international.org/chapter/microarray-data-mining/10978)

### Outlier Detection

Sharanjit Kaur (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1476-1482).

[www.irma-international.org/chapter/outlier-detection/11015](http://www.irma-international.org/chapter/outlier-detection/11015)

### Multilingual Text Mining

Peter A. Chew (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1380-1385).

[www.irma-international.org/chapter/multilingual-text-mining/11001](http://www.irma-international.org/chapter/multilingual-text-mining/11001)

### Variable Length Markov Chains for Web Usage Mining

José Borgesand Mark Levene (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 2031-2035).

[www.irma-international.org/chapter/variable-length-markov-chains-web/11098](http://www.irma-international.org/chapter/variable-length-markov-chains-web/11098)

### On Explanation-Oriented Data Mining

Yiyu Yao (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 842-848).

[www.irma-international.org/chapter/explanation-oriented-data-mining/10918](http://www.irma-international.org/chapter/explanation-oriented-data-mining/10918)