

## Abstract

The main purpose of web personalization systems is providing personalized suggestions about the items on the web that users will be interested in. The most popular way to know a user's interests will be using explicit voting system for the web page. However voting might change user's normal behavior and be expensive to give the user an incentive to vote. So, without asking users anything, by looking at his/her behavior a personalized system can predict user's interest level after their behavior pattern is modeled.

There are a few systems using implicit rating and those systems use only a small number of implicit indicators, which might not include more predictive indicators. It is very hard to detect most of the user's behavior on the web browser level unless Microsoft makes IE open source code. Using the data collected by "kixbrowser" that was built for this paper's experiments, we could build predictive linear, non-linear, neural network and mouse-click-only models (the most common predictor) for each user.

Through the experiment, and prediction in this study, we learn that we can model only using a user's implicit interest indicators and predict the user interest level according to his behavior in the web page.

# Table of contents

## Abstract

<b>1.Introduction</b>	1
<b>1.1 Problem statement</b>	2
<b>1.2 Solution approach</b>	2
<b>1.3 Organization of the thesis</b>	3
<b>2. RELATED RESEARCH</b>	5
<b>2.1 Introduction</b>	5
<b>2.2 Implicit systems</b>	6
<b>2.2.1 The Tapestry text filtering system</b>	7
<b>2.2.2 GroupLens</b>	8
<b>2.2.3 InfoScope</b>	8
<b>2.2.4 URN</b>	9
<b>2.2.5 Curious Browser</b>	9
<b>2.2.6 Powerize server 1.0</b>	10
<b>2.2.7 Jeremy Goecks' Agent</b>	12
<b>2.3 Comparison between different systems based on features</b>	13
<b>2.4 Discussion</b>	16

<b>3. Monitoring user's behavior</b>	18
<b>3.1 Architecture</b>	18
<b>3.2 Implicit interest indicators from user's behavior</b>	19
<b>3.2.1 Duration</b>	19
<b>3.2.2 Distance of mouse movement</b>	20
<b>3.2.3 Distance of Scrollbar movement</b>	21
<b>3.2.4 Number of mouse clicks</b>	21
<b>3.2.5 The number of highlighting text or sentence</b>	22
<b>3.2.6 Frequency</b>	22
<b>3.2.7 Recency</b>	22
<b>3.2.8 The number of rolls over the hyperlinks</b>	23
<b>3.2.9 The page after typing characters</b>	23
<b>3.2.10 Button/Ctrl key</b>	24
<b>3.2.11 File size</b>	26
<b>3.3 User's Explicit Interest</b>	26
<b>3.4 Overall Methodology to generate data</b>	28
<b>3.4.1 Building a web browser (kixbrowser)</b>	28
<b>3.4.2 Connecting recording functions to all implicit indicators             related to user's interest level and recorded into the raw             log file</b>	29 30
<b>3.4.3 Getting the data by doing experiments</b>	30
<b>3.4.4 Building a parser (extractor) for the raw log data to the             log data</b>	
<b>3.5 Extracting Data</b>	31

3.5.1 Get each user's raw-log file	32
3.5.2 Extracting raw-log file to ascii log file for analysis	34
3.5.3 Importing this ASCII log file into a statistical package and Excel spread sheet	35
<b>4. Evaluation of implicit interest indicators</b>	<b>36</b>
4.1. Collecting data from users	36
4.2. Correlations between each candidate interest indicator and explicit rating	37
4.3 Regression Analysis on individual implicit interest indicators	48
4.4 Regression analysis on multiple implicit interest indicators	55
4.4.1 Linear predictor system	56
4.4.2 Non-linear predictor system	57
4.4.3 Neural Network Analysis	60
4.5 Mouse clicks only regression system	64
4.6 Comparative analysis of regression techniques	65
4.7 Discussion	68
<b>5 Summary of results &amp; future work</b>	<b>70</b>
5.1 Summary of contribution	70
5.2 Strength and weakness	71
5.2.1 Advantages of this system	71
5.2.2 Disadvantages of this system or something not covered in the experiments	72
5.3 Future work	73
<b>REFERENCE</b>	<b>74</b>

<b>Appendix A</b>	78
<b>Appendix B</b>	91