

**A Novel Framework for Web Log Mining using Transductive  
SVM Classifier and Ontology based Associative  
Classification**

Submitted by

**S. Chitra**

Supervisor

**Dr. B. Kalpana**

A Thesis Submitted to

Avinashilingam University for Women, Coimbatore – 641 043

In Partial Fulfilment of the Requirement for the Award of Degree of

**Doctor of Philosophy in Computer Science**

**November 2014**

## CERTIFICATE

This is to certify that the thesis entitled “A Novel Framework for Web Log Mining using Transductive SVM Classifier and Ontology based Associative Classification”, submitted to Avinashilingam University for Women, Coimbatore – 641 043, for the award of the degree of **Doctor of Philosophy in Computer Science** is a record of original research work done by **S. Chitra**, during the period of her study in the Department of Computer Science, Avinashilingam University for Women, Coimbatore - 641 043, under my supervision and guidance and the thesis has not formed the basis of any other Degree/ Diploma/ Associateship/ fellowship or similar title to any candidate of any university.

*As. Manojini*

Signature of the

Head of the Department

*K. L. Parvathi*

Signature of the Guide

**Dr. G. PADMAVATHI**  
M.Sc., M.Phil., Ph.D.  
Professor and Head  
Department of Computer Science  
Avinashilingam Institute for Home Science  
and Higher Education for Women  
Coimbatore - 641 043

*Parvathi 28/10/14*

Signature of the Dean


**Dr. A. PARVATHI**  
Dean, Faculty of Science  
Professor & Head  
Department of Mathematics  
Avinashilingam Institute for Home Science  
and Higher Education for Women  
Coimbatore - 641 043

## DECLARATION

I hereby declare that the matter embodied in this thesis is the result of investigations done by me in the department of Computer Science, Avinashilingam University for Women, Coimbatore - 641043, under the supervision and the guidance of **Dr. B. Kalpana**, Professor, Department of Computer Science, Avinashilingam University for Women, Coimbatore -641 043 and this has not formed the basis of any other Degree/ Diploma/ Associateship/ fellowship or similar title to any candidate of any university.



Signature of the Guide



Signature of the Candidate

## ACKNOWLEDGEMENT

I owe my reverential gratitude to Late **Ayya Dr. T. S. Avinashilingam** Avl., Founder, President and First Chancellor of Avinashilingam University for Women, Coimbatore for providing the temple of learning and also for his heavenly blessings. I owe my sincere and humble gratitude to late **Amma Dr. Rajammal P. Devadas** Avl., M.A., M.Sc., Ph.D. (Ohio State) D.Sc. (Madras), Hon. D.H.L. (Oregon State), Hon. D.Sc. University of Ulster (Northern Ireland), Former Chancellor, Avinashilingam University for Women, Coimbatore.

I record my sincere thanks to Late Anna **Mr. T. K. Shanmuganandam**, B.A., B.L., Former Chancellor, Avinashilingam University for Women, Coimbatore.

I record my sincere thanks to **Dr. T. K. S. Meenakshi Sundaram**, Chancellor, Avinashilingam University for Women, Coimbatore for providing the necessary infrastructure for the conduct of the research.

I express my sincere and heartfelt thanks to **Dr. Sheela Ramachandaran**, M.Sc., PG. Dip., Ph.D., Vice Chancellor, Avinashilingam University for Women, Coimbatore for providing necessary facilities and resources for the successful completion of this research work.

I express my heartfelt thanks to **Dr. Saroja Prabhakaran**, M.A., Dip. Ed, Ph.D., former Vice Chancellor, Avinashilingam University for Women, Coimbatore for motivating me to carry out the research, the constant encouragement, care and advice throughout the research.

I extend my gratitude to **Dr. Gowri Ramakrishnan**, M.Sc., M.Phil., Ph.D, Former Registrar and **Dr. (Mrs) A. Venmathi**, M.Sc., Dip.Ed., M.Phil., Ph.D., Registrar (In charge), Professor and Head, Department of Family Resource Management, Avinashilingam University for Women, Coimbatore, for providing the facilities to carry out the research.

I express my heartfelt thanks to **Dr. G.P.Jeyanthi**, M.Sc., M.Phil., Ph.D, Controller of Examinations, Avinashilingam University for Women, for her valuable help in carrying out the research work.

I express my sincere thanks to **Dr. A. Parvathi**, M.Sc., M.Phil., Ph.D., Former Dean, Faculty of Science and **Dr. Vasugi Raaja**, M.Sc., Dip.Ed, M.Phil., Ph.D., Dean Faculty of Home Science, Avinashilingam University for Women, Coimbatore, for her spontaneous and timely help and amenities provided for the successful completion of this research work.

I express my sincere and heartfelt thanks to **Dr. G. Padmavathi**, M.Sc., M.Phil., Ph.D., Professor and Head, Department of Computer Science, Avinashilingam University for Women, Coimbatore. Her encouragement, motivation, and support helped me to complete my research work.

I feel extremely privileged and fortunate to have worked under the able supervision and professional guidance of my supervisor, **Dr. B. Kalpana**, M.Sc., M.Phil., Ph.D., Professor, Department of Computer Science, Avinashilingam University for Women, Coimbatore. Her constant encouragement, motivation, valuable advice, and support helped me to explore and gain deep knowledge in the field of research. I am greatly indebted to her for the invaluable suggestions, constant support, motivation and encouragement not only for this research but also during the entire duration of this course.

She helped to define my research goals and showed the way to achieve them. Her resourceful guidance and systematic approach urged me to put my best possible efforts in completing my work and documentation in time. Her sympathetic friendly nature, patience, timely counselling, willing to help at anytime, anywhere in any situation during the entire period of the study has moulded this research into a reality.

I record my gratitude to all the **faculty members** and the **non teaching staff** of Department of Computer Science, and members of Computer Centre, Avinashilingam University for Women, Coimbatore for their encouragement and support. I also record my thanks to my friends, colleagues for their support, encouragement and cooperation rendered for the completion of this research.

I thank **Dr. M. Karunamoorthi**, Associate Professor of English (Retd.), Government Arts College, Coimbatore, and **Dr. Kalyani Suresh** of Amrita Viswa Vidyapeedam, Coimbatore, for spending their valuable time in grammatical correction of this thesis.

I record my sincere thanks to **Mrs. Sasirekha Kathirvelu** for providing the welcare dataset for experimentation.

I thank my mother **Mrs. B. Dhanabagyam**, my father late **Mr. S. P. Sundaram**, my husband **Dr. V. Sivakumar**, my sister **Dr. S. Shanthi** for their blessings, patience, wishes, prayers, encouragement and support extended to me throughout this research.

Last but not least, I thank my daughter **S. Shwetha**, my son **S. Shreehari**, my niece **D. Anurekha** and my nephew **D. Aravindan** for their cooperation.

Above all, I raise my humble heart in adoration to **God Almighty** who has showered his blessings and provided me physical and mental strength to execute this research successfully.

## TO WHOSOEVER IT MAY CONCERN

We permit **S. Chitra, 09PH79, Ph. D.**, research scholar(part-time), Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore – 641 043, to use our web log data of our web site [www.welcareindia.com](http://www.welcareindia.com) for a period of January 2013 to December 2013. The web log data obtained should be used for research purpose only and not for any other commercial use.

**Dated: 23.12.2013**

**For S & T Welcare Equipments (P) Ltd**



**Authorized Signatory**



### **S&T Welcare Equipments (P) Ltd.**

"S&T Arcade", No.20, Vasanth Nagar, Singanailur, Coimbatore - 641005.  
Ph : 0422 - 2597704 / Toll Free : 18004251855,  
E-mail : [info@welcareindia.com](mailto:info@welcareindia.com) / Website : [www.welcareindia.com](http://www.welcareindia.com)

# CONTENTS

---

<b>CHAPTER No.</b>	<b>TITLE</b>	<b>PAGE No.</b>
	<b>LIST OF FIGURES</b>	
	<b>LIST OF TABLES</b>	
	<b>LIST OF ABBREVIATIONS</b>	
	<b>ABSTRACT</b>	
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	Overview of the Research	1
1.2	Overview of Web Mining	4
1.2.1	Categorization of Web Mining Approaches	5
1.3	Web Usage Mining	7
1.3.1	Application Areas	9
1.3.2	Requirements of Web Usage Mining	11
1.4	Web Usage Data	11
1.4.1	Data Sources	12
1.4.2	Types of Web Data	14
1.5	General Web Log Mining System	16
1.5.1	Raw Server Logs	17
1.5.2	Preprocessing	19
1.5.3	Pattern Discovery	19
1.5.4	Pattern Analysis	23
1.6	Motivation and Objectives	25
1.7	Chapter Formulation	28
1.8	Chapter Summary	29
<b>2</b>	<b>REVIEW OF LITERATURE</b>	<b>30</b>
2.1	Web Usage Mining Techniques	32
2.1.1	Web Personalization	35
2.1.2	Site Improvement	37
2.1.3	Business Intelligence	38

---

---

<b>CHAPTER No.</b>	<b>TITLE</b>	<b>PAGE No.</b>
2.1.4	Site Modification	39
2.1.5	Usage Characteristics	39
2.2	Log File Analysis for Web Usage Mining	41
2.3	Web Page Recommendation	43
2.4	Web Page Prediction Techniques	45
2.4.1	Preprocessing	46
2.4.2	Pattern Discovery and Analysis Techniques	52
2.5	Conclusion	70
<b>3</b>	<b>METHODOLOGY</b>	<b>71</b>
3.1	Web Log Data	72
3.2	Research Framework	75
3.2.1	Preprocessing Algorithms	78
3.2.2	Potential User Identification	81
3.2.3	Web Log Associative Classification	83
3.3	Chapter Summary	84
<b>4</b>	<b>PREPROCESSING ALGORITHMS</b>	<b>85</b>
4.1	Cleaning Web Log Data	86
4.2	User Identification	88
4.3	Session Identification	89
4.3.1	Calculation of Browsing Time	91
4.3.2	Discretization	91
4.3.3	Acyclic Graph Construction	92
4.3.4	Pattern Extraction	102
4.4	Chapter Summary	104
<b>5</b>	<b>POTENTIAL USER IDENTIFICATION</b>	<b>105</b>
5.1	Categorization of Web Users	105
5.2	Classification	107
5.2.1	Attribute Identification	108
5.2.2	Building the Classifier	111
5.3	Chapter Summary	122

---

---

<b>CHAPTER No.</b>	<b>TITLE</b>	<b>PAGE No.</b>
<b>6</b>	<b>WEB LOG ASSOCIATIVE CLASSIFICATION</b>	<b>123</b>
6.1	Semantic Web Mining	127
6.2	Related Techniques	130
6.2.1	Fuzzy Associative Rule Mining	130
6.2.2	Periodic Pattern Mining	131
6.2.3	Lattice	131
6.3	OSIPSO Algorithm	134
6.4	Chapter Summary	146
<b>8</b>	<b>SUMMARY AND CONCLUSION</b>	<b>169</b>
8.1	Future Research Directions	174
	<b>BIBLIOGRAPHY</b>	176
	<b>PUBLICATIONS RELATED TO RESEARCH WORK</b>	209

---

## LIST OF TABLES

<b>TABLE No.</b>	<b>TITLE</b>	<b>PAGE No.</b>
3.1	Important Terms In Web Log Data	73
3.2	Web Log File Format	74
4.1	List Of Graph Relations	94
5.1	Characteristics Exhibited By Potential And Non-Potential Users	106
5.2	Attribute Identification	109
5.3	Discretized Values Of The Attributes	109
5.4	Training Data	110
6.1	Semantically Optimized Web Usage Log	136

## LIST OF FIGURES

<b>FIGURE No.</b>	<b>TITLE</b>	<b>PAGE No.</b>
1.1	General Framework of Web Mining	5
1.2	Categorization of Web Mining	6
1.3	Web Usage Mining Architecture	9
1.4	Application Areas	10
1.5	Potential Data Sources	13
1.6	Web Data Categorization	15
1.7	Phases of Web Usage Mining System	17
1.8	Format of Web Log File	17
1.9	Sample Web Log File	17
1.10	Data Mining Methods	24
3.1	General Architecture of Web System with Web Access Prediction	77
3.2	Tasks in Next Page Prediction System	78
3.3	Research Framework	79
4.1	Graph Example	92
4.2	DAG Example	96
4.3	MAG Example	100
4.4	Preprocessing Procedure	103
5.1	Classifier and Decision Boundaries	111
5.2	Basic Classification Model	112
5.3	Paradigms in Web Log Data Classification	115
5.4	Process of Classification	116
5.5	Support Vector Machine Hyperplane	118
5.6	The TSVM Algorithm	121

<b>FIGURE No.</b>	<b>TITLE</b>	<b>PAGE No.</b>
6.1	2-Dimensional Representation of Concept Relationship	133
6.2	Lattice Example	134
6.3	OSIPSO Procedure	135
6.4	Example of Web Usage Context	137
6.5	Lattice construction	138
6.6	Generation of personalized Ontology	140
6.7	OSIPSO Algorithm	142
6.8	Associative Access Pattern Rule Mining	145

## LIST OF ABBREVIATIONS

S.No.	Abbreviation	Description
1.	AG	Acyclic Graph
2.	AIS	<i>Artificial Immune System</i> (association rules)
3.	CF	Collaborative Filtering
4.	CLF	Common Log Format
5.	C-Logs	Concept Logs
6.	CRM	Customer Relationship Management
7.	CSB-mine	Conditional Sequence Base mining algorithm
8.	CSS	Cascading Style Sheets
9.	DAG	Directed Acyclic Graph
10.	DNS	Domain Name Server
11.	DT	Decision Tree
12.	ECLAT	EquivalenceCLAssTransformation
13.	E-Commerce	Electronic Commerce
14.	EIAA	European Interactive Advertising Association
15.	ELF	Extended Log Format
16.	EM	Expectation-Maximization
17.	FAP	Frequent Access Patterns
18.	FCA	Formal Concept Analysis
19.	FM	Feature-Matrices
20.	FMCSF	Fast Mining of Closed Sequential Patterns
21.	FP	Frequent-Pattern
22.	GIF	Graphics Interchange Format
23.	HDAG	Hierarchical Directed Acyclic Graph
24.	HTML	Hypertext Markup Language
25.	HTTP	Hypertext Transfer Protocol
26.	IBM	<i>International Business Machines Corporation</i>

<b>S.No.</b>	<b>Abbreviation</b>	<b>Description</b>
27.	IP	Internet Protocol
28.	ISP	Internet Service Providers
29.	JPEG	Joint Photographic Experts Group
30.	LAPIN_WEB	LAst Position INduction for WEB log
31.	LCM	Linear time Closed pattern Miner
32.	MAG	Mixed Ancestral Graph
33.	NLP	Natural Language Processing
34.	OLS	Ordinary Least Squares
35.	OSIPSO	Optimum Session Interval based Particle Swarm Optimization
36.	OWL	Web Ontology Language
37.	PAG	Partial Ancestral Graph
38.	PSO	Particle Swarm Optimization
39.	RDF	Resource Description Framework
40.	SCLM	Software Configuration and Library Manager
41.	SETM	SET-oriented Mining (association rules)
42.	SOM	Self Organizing Map
43.	SVM	Support Vector Machine
44.	SWARS	Sequential Web Access-based Recommender System
45.	TSVM	Transductive Support Vector Machine
46.	URL	Uniform Resource Locator
47.	W3C	World Wide Web Consortium
48.	WEEV	Web Ecology and Evolution Visualization
49.	WP	Web Personalization
50.	WUM	Web Usage Mining
51.	WWW	World Wide Web
52.	XML	Extended Markup Language