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.

Signature of the Head of the Department

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

Signature of the Guide

28/4/14

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.

Kalpana B.
Signature of the Guide

25/11/14
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.

WELCARE
HEALTHY LIVING

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 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, Singanallur, Coimbatore - 641005.
Ph : 0422 - 2597704 / Toll Free : 18004251855,
E-mail : info@welcareindia.com / Website : www.welcareindia.com

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

S.No.	Abbreviation	Description
1.	AG	Acyclic Graph
2.	AIS	*Artificial Immune System* (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.	FMCSP	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	*International Business Machines* Corporation

S.No.	Abbreviation	Description
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