

LIST OF FIGURES

Figure No.	Title	Page No.
1.1	Examples of Negative Online Reviews	6
1.2	Examples of Positive Online Reviews	7
1.3	Examples of Ham Online Reviews	9
1.4	Examples of Spam Online Reviews	11
1.5	Example of Review With Human Friendly Clues to Detect Spam	14
1.6	Reviews from Individual Spammer	16
1.7	Reviews from Group Spammers	16
1.8	Unsupervised and Supervised Classification	19
1.9	Steps in Review Spam Detection System	20
1.10	Need For OSRD Systems	21
2.1	Clustering Process	39
3.1	Proposed Research Methodology	62
3.2	Interaction of Algorithms and Research Phases	63
4.1	BoW Feature Example	78
4.2	N-Grams as Features	83
4.3	Review Length Example	88
4.4	Sentiment Classes of Reviews – Examples	89
4.5	Amazon Sales Rank Example	94
4.6	Feature Selection Algorithm	97
4.7	Steps in Proposed Hybrid Feature Selection Algorithm	99
4.8	Steps Involved in MRMR-IG	103
4.9	Markov Blanket Filter	105
4.10	The MRMR-IG Algorithm	106

Figure No.	Title	Page No.
4.11	Steps in MRMR-MI	109
4.12	Combining Features	110
4.13	Feature Integration Algorithm	111
4.14	ACO-Based Feature Selection	113
4.15	GA-Based Feature Fusion and Selection	115
4.16	Enhanced ACO-Based Feature Selection Algorithm	118
5.1	Support Vector Machine Hyperplane	121
5.2	The EoC System	123
5.3	Enhanced SVM Classifier	126
5.4	Optimization Procedure to Remove Irrelevant SVs	128
5.5	Hyper Planes and Margins	129
5.6	Mapping Feature Points into Feature Space Using Kernel Function	129
5.7	Target Class Assignment While Using Mahalanobis Distance	131
5.8	Steps in the Design of the Proposed Enhanced EoC System	136
6.1	KNN Classification Algorithm	141
6.2	K-Means Algorithm	144
6.3	EM Algorithm	149
6.4	Steps in Type 1 Hybrid Systems	151
6.5	Steps in Type 2 Hybrid Systems	152
6.6	Steps in Type 3 Hybrid Systems	153
7.1	Confusion Matrix	156
7.2	Precision (%) of Feature Selection Algorithms	160
7.3	Recall (%) of Feature Selection Algorithms	160
7.4	F Measure(%) of Feature Selection Algorithms	161
7.5	Accuracy (%) of Feature Selection Algorithms	161

Figure No.	Title	Page No.
7.6	Speed (Seconds) of Feature Selection Algorithms	165
7.7	Effect of Feature Selection Algorithms on Ham/Spam Detection System in Terms of Speed (Seconds)	165
7.8	Precision (%) of OSRD Systems	167
7.9	Recall (%) of OSRD Systems	168
7.10	F Measure(%) of OSRD Systems	168
7.11	Accuracy (%) of OSRD Systems	169
7.12	Speed (Seconds) of OSRD Systems	169
7.13	Precision (%) of Hybrid Systems	172
7.14	Recall (%) of Hybrid Systems	173
7.15	F-Measure (%) of Hybrid Systems	173
7.16	Accuracy (%) of Hybrid Systems	174
7.17	Speed (Seconds) of Hybrid Systems	174

LIST OF TABLES

Table No.	Title	Page No.
3.1	Features Extracted	64
2.2	Proposed Hybrid Systems	69
4.1	Review Centric Features	74
4.2	POS Tagset (Partial List)	80
4.3	POS Tagging - Example	81
4.4	Similarity Score Using N-Grams	83
4.5	Reviewer Centric Features	87
6.1	List of Proposed Hybrid Systems	139
6.2	Details of the Hybrid Ensemble Systems	140
7.1	Coding Scheme Used – Feature Selection	159
7.2	Parameter Settings of ACO and GA	159
7.3	Efficiency Gain (%) Of FS_MGA Algorithm	164
7.4	Coding Scheme Used – Enhanced Classification Systems	167
7.5	Coding Scheme – Hybrid Systems	172

LIST OF ABBREVIATIONS

Abbreviation	Description
ACO	Ant Colony Optimization
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
BFPA	Binary Flower Pollination Algorithm
BIF	Best Individual Features
BoW	Bag-of-Word
BRR	Burst Review Ratio
CFS	Correlation-based Feature Selection
CFV	Candidate Feature Vector
ECL	Ensemble Classification
EM	Expectation Maximization
EM_ESVM	Hybrid Model Using EM in Step 1 and ESVM in Step 2
EoC	Ensemble of Classifier
ERD	Early Rating Deviation
ES	Ensemble SVM Classification System
ES_SO	Enhanced Ensemble SVM Classification System with Speed Optimizers
ES_SO+ED	Enhanced Ensemble SVM Classification System with Speed Optimizers and Euclidean Distance Measure
ES_SO+MD	Enhanced Ensemble SVM Classification System with Speed Optimizers and Mahalanobis Distance Measure
ESVM	Enhanced SVM
ETF	Early Time Frame
eWOM	Electronic Word of Mouth
FCR	Feature to Class Relationship
FFR	Feature to Feature Relationship

FN	False Negative
FP	False Positive
FS	Feature Selection
FS_ACO	Feature Selection Using ACO Algorithm
FS_ACO+GA	Feature Selection Using ACO + GA Algorithm
FS_MGA	Feature Selection Using FS_MIMG and ACO + GA Algorithm
FS_MIMG	Feature Selection using MRMR_IG and MRMR-MI Algorithm
FS_MRMR	Feature Selection using MRMR Algorithm
GA	Genetic Algorithms
GCS	Group Content Similarity
GD	Group Deviation
GRD	General Rating Deviation
GS	Group Size
HHS	Ham Hit Score
HMS	Ham Miss Score
iBPSO	Improved Binary Particle Swarm Optimization
ICS	Individual Content Similarity
ID3	Iterative Dichotomiser 3
IETF	Individual Early Time Frame
IG	Information Gain
IMC	Individual Member Coupling
IRD	Individual Rating Deviation
KDE	Novel Kernel Density Estimation
KM	K-Means
KM_ESVM	Hybrid Model Using KM in Step 1 and ESVM in Step 2
KNN	K-Nearest Neighbour
KNN_ESVM	Hybrid Model Using KNN in Step 1 and ESVM in Step 2
LASSO	Least Absolute Shrinkage and Selection Operator

LSTM	Long Short Term Memory
MCS	Member Content Similarity
MI	Mutual Information
ML	Machine Learning
MRMR	Maximum Relevant Minimum Redundant
MRMR-IG	Maximum Relevant Minimum Redundant-Information Gain
MRMR-MI	Maximum Relevant Minimum Redundant-Mutual Information
MS	Mean Shift
MS_ESVM	Hybrid Model Using MS in Step 1 and ESVM in Step 2
NB	Naive Bayes
NB_ESVM	Hybrid Model Using NB in Step 1 and ESVM in Step 2
NLP	Natural Language Processing
OSRD	Online Spam Detection System
PDF	Probability Distribution Function
POS	Part of Speech
PSO	Particle Swarm Optimization
PWCC	Product Word Composition Classifier
RDF	Radial Basis Function
RGS	Ratio of Group Size
ROC	Receiver Operating Characteristic
RVP	Ratio of Verified Purchase
SAGE	Sparse Additive Generative
SAMS	Sentiment Attribute Matching Stack
SBS	Sequential Backward Selection
SC	Support Count
SFS	Sequential Forward Selection
SHS	Spam Hit Score

SMS	Spam Miss Score
SRM	Structural Risk Minimization
SU	Symmetrical Uncertainty
SVM	Support Vector Machine
SVM_ESVM	Hybrid Model Using SVM in Step 1 and ESVM in Step 2
SVM_KM_ESVM	Hybrid Model Using SVM, KM in Step 1 and ESVM in Step 2
SWNN	Sentence Weight Neural Network
TF	Term Frequency
TG	Targeting Group
TN	True Negative
TP	True Positive
TW	Time Window
USC	Uncorrelated Shrunken Centroid
WWW	World Wide Web
