

Contents

List of Figures	7
List of Tables	8
1 INTRODUCTION	1
1.1 Importance of feature selection	2
1.2 Existing approaches	2
1.3 Motivation	3
1.4 Problem Formulation	3
1.5 Contribution	3
1.6 Organization of the report	4
2 BACKGROUND	5
2.1 Subset generation	7
2.2 Subset evaluation	7
2.3 Stopping criteria	7
2.4 Result validation	8
2.5 Ensemble method	8
3 RELATED WORKS	9
3.1 Filter Method	9
3.1.1 Relief	9
3.1.2 Correlation-Based Feature Selection (CFS)	10
3.1.3 Fast Correlation-Based Filter (FCBF)	11
3.2 Wrapper Method	12

3.2.1	Sequential forward selection (SFS)	12
3.2.2	Sequential backward elimination (SBE)	13
3.3	Hybrid Method	15
3.4	Discussion	15
4	PROPOSED METHOD	16
4.1	Search method for filter phase	17
4.2	Attribute evaluation functions : E_i	18
4.3	Majority function-1: f_1	18
4.4	Search method used for wrapper phase	18
4.5	Classifiers	19
4.6	Majority function-2: f_2	19
4.7	Complexity Analysis	19
5	EXPERIMENTAL RESULTS	21
5.1	Environment used	21
5.2	Datasets used	21
5.2.1	UCI datasets	21
5.2.2	Intrusion datasets	22
5.2.3	Real life datasets	22
5.3	Results	22
5.3.1	On Monk datasets	22
5.3.2	On NSL-KDD dataset	25
5.3.3	On TU-IDS datasets	31
5.4	Discussion	31
6	CONCLUSION AND FUTURE WORK	33

List of Figures

2.1	Steps in feature selection	5
3.1	Filter method	10
3.2	Wrapper method	12
4.1	Multi-level hybrid model	17
5.1	Accuracy curve for Monk 1 dataset.	24
5.2	Accuracy curve for Monk 3 dataset.	25
5.3	Feature feature correlation (a)	28
5.4	Feature feature correlation (b)	28
5.5	Feature feature correlation (c)	29
5.6	Feature feature correlation (d)	29
5.7	Feature feature correlation (e)	30

List of Tables

2.1	Evaluation measures : A general comparison	6
3.1	Feature selection approaches : A general comparison	14
5.1	Result of MLH on Monk 1 dataset	23
5.2	Result of MLH on Monk 3 dataset	23
5.3	Validation using Monk datasets	23
5.4	Comparison of MLH with other algorithms	24
5.5	Result of MLH on NSL-KDD	26
5.6	Classification accuracy on NSL-KDD (2 class)	27
5.7	Classification accuracy of MLH on NSL-KDD (using protocol specific 2 class framework)	27
5.8	Classification accuracy of MLH on TU-IDS datasets	31
5.9	Classification accuracy of MLH on TU-IDS datasets	31