

Contents

1	Introduction	5
1.1	Feature selection techniques	6
1.2	Motivation	7
1.3	Objective	8
1.4	Contribution	8
1.5	Organisation of dissertation	9
2	Related Work	10
2.1	Existing feature selection methods	10
2.2	Discussion	11
3	An effective feature-feature correlation measure : ENMRS	13
3.1	NMRS	13
3.2	ENMRS	14
3.3	Validation	14
3.4	Discussion	17
4	An effective feature-class correlation measure : crRelevance	18
4.1	Proposed crRelevance	18
4.1.1	Algorithm and complexity analysis	20
4.2	Validation of crRelevance	21
4.3	Discussion	24
5	An effective feature selection technique : FPEC	25
5.1	Proposed FPEC	25
5.1.1	Algorithm of proposed method	26
5.2	Validation of FPEC	27
5.2.1	Determination of co-efficient values	27

5.2.2	Experimental results	28
5.3	Discussion	28
6	Conclusion and Future work	31
6.1	Future work	32

List of Figures

1.1	Feature Selection Techniques	6
3.1	Normal Distribution	16
3.2	Lognormal Distribution	17
3.3	Beta Distribution	17
5.1	Size of optimal feature subset and Accuracy of three classifier w.r.t coefficient n_1 and n_2 on Subset of Yeast cell cycle dataset	28
5.2	Size of optimal feature subset and Accuracy of three classifier w.r.t coefficient n_1 and n_2 on Wine dataset	30

List of Tables

2.1	Comparison of feature selection methods	12
3.1	p values for cluster 1	15
3.2	p values for cluster 2	16
4.1	An example dataset	19
4.2	Sorted order of F_1 of example dataset	19
4.3	crRelevance scores for Iris dataset	22
4.4	Accuracy of 3 classifiers on optimal feature set based on cr- Relevance score	23
4.5	crRelevance scores for Wine dataset	23
4.6	crRelevance scores for Glass dataset	23
5.1	Symbolic Representation	27
5.2	Dataset description	27
5.3	size of feature subset and accuracy of different classifier w.r.t coefficient n_1 and n_2	29
5.4	Size of optimal feature subset and Accuracy of three classifier for $n_1 = 0.3$ and $n_2 = 0.7$ on three datasets	29