

Contents

1	Introduction	1
1.1	Motivation and Fundamentals	1
1.1.1	Globalization and Data Mining	1
1.1.2	Basic Data Mining Concepts	2
1.1.3	Dynamic Data Mining	4
1.1.4	Fuzzy Sets	4
1.1.5	Rough Sets	6
1.1.6	Soft-Computing Clustering Algorithms	7
1.1.7	Support Vector Clustering	8
1.2	Objectives	9
1.2.1	General Objective	9
1.2.2	Specific Objectives	9
1.3	Methodology	9
1.4	Expected Results	10
1.4.1	Final Products	10
1.4.2	Original Contributions	10
1.5	Papers Overview	10
1.5.1	A Rough-Fuzzy Approach for Support Vector Clustering	10
1.5.2	Dynamic Rough-Fuzzy Support Vector Clustering	11
1.5.3	Rough-Fuzzy Support Vector Domain Description for Outlier Detection	11
2	A Rough-Fuzzy Approach for Support Vector Clustering	12
2.1	Introduction	12
2.2	Literature Overview on Support Vector Clustering	13
2.2.1	General Introduction to Support Vector Clustering	13
2.2.2	Mathematical Description of Support Vector Clustering	14
2.2.3	Extensions of Support Vector Clustering	17
2.3	A Rough-Fuzzy Approach for Support Vector Clustering	18
2.3.1	Overview of the Proposed Method	18
2.3.2	Development of the Proposed Method	19
2.3.3	Comprehensible Application on Example Data Set	22
2.4	Experiments and Discussion	24
2.4.1	Description of Data Sets and Experimental Set-Up	24
2.4.2	Results and Discussion	25
2.5	Conclusions and Future Work	31

3	Dynamic Rough-Fuzzy Support Vector Clustering	33
3.1	Introduction	33
3.2	Literature Overview	35
3.2.1	Foundations of Dynamic Clustering	35
3.2.2	Related Dynamic Clustering Algorithms	36
3.2.3	Conceptual Comparative Analysis of Algorithms	38
3.3	Rough-Fuzzy Support Vector Clustering	39
3.3.1	Training Phase	40
3.3.2	Labeling Phase	41
3.3.3	Fuzzification Phase	42
3.4	Dynamic Rough-Fuzzy Support Vector Clustering	43
3.4.1	Addressed Structural Changes	44
3.4.2	Development of Dynamic RFSVC	46
3.4.3	Pseudo-code of Dynamic RFSVC and Conceptual Comparison	50
3.5	Experiments and Discussion	51
3.5.1	Description of the Data Sets and Set-Up	51
3.5.2	Results for Benchmark and Synthetic Data Sets	52
3.5.3	Application to NOAA Weather Data	56
3.6	Conclusions and Future Work	58
4	Rough-Fuzzy Support Vector Domain Description for Outlier Detection	60
4.1	Introduction	60
4.2	Literature Overview	61
4.2.1	Support Vector Domain Description	61
4.2.2	Outlier Detection Methods and SVDD	63
4.3	Rough-Fuzzy Support Vector Domain Description	64
4.3.1	Overview of the Proposed Method	64
4.3.2	Development of the Proposed Method	65
4.4	Computational Experiments and Discussion	66
4.4.1	Description of Data Sets and Experimental Set-Up	66
4.4.2	Experimental Results	67
4.5	Conclusions	69
	Bibliography	71

List of Tables

2.1	Soft Computing Variations of Support Vector Clustering	17
2.2	Recent Improvements for SVC	18
2.3	Membership Matrix for Bounded Support Vectors	24
2.4	Data Sets Characteristics	25
2.5	Algorithms' Parameters	26
2.6	Validity Measures for Artificial Data Sets	29
2.7	Validity Measures for Benchmark Data Sets	29
2.8	Validity Measures for Real World Data Sets	30
2.9	Sample of the Membership Matrix for XO Data Set	30
3.1	Related Approaches of Dynamic Clustering	39
3.2	Structural Changes Addressed by Dynamic Clustering Algorithms	39
3.3	Outputs of RFSVC Algorithm	43
3.4	Conceptual Comparison between D-RFSVC and other Algorithms	51
3.5	Data Sets Characteristics	52
3.6	Algorithm's Parameters for $t = 0$	52
3.7	S1-Gaussian: Membership Matrix for $t = 1$	54
3.8	S1-Gaussian: Membership Matrix for $t = 2$	54
3.9	Weather Data Set: Membership Matrix	58
4.1	Results for Two Circles Data Set	68
4.2	Results for XO Data Set	68
4.3	Results for Banknote Data Set	68
4.4	Results for Iris Data Set	69
4.5	Results for Abalone Data Set	69

List of Figures

2.1	General Idea of SVC	14
2.2	Data Points Outside the Sphere in the Higher-Dimensional Space	18
2.3	Support Vector and its zone of Influence	20
2.4	Motivation Data Set	22
2.5	Results of the Training Phase of RFSVC Algorithm for Motivation Data Set	23
2.6	Results of the Labeling Phase of RFSVC Algorithm for Motivation Data Set	23
2.7	RFSVC Results for 2D Data Sets	27
2.8	RFCM Results for 2D Data Sets	28
2.9	RPCM Results for 2D Data Sets	28
3.1	General Idea of RFSVC	42
3.2	Creation of Clusters by Accumulation	44
3.3	Creation of Clusters by Merging	45
3.4	Creation and Expansion of Clusters for S1-Gaussian data set	53
3.5	Contraction and Elimination of Clusters for Three Spheres data set	53
3.6	Change of Shape and Merging of Clusters for XO data set	54
3.7	Parallel Coordinates Plot for the XO Data Set	56
3.8	Parallel Coordinates Plot for XO Data Set: Tracking of Cluster 3.	57
3.9	Structural Changes in Weather Data.	57
4.1	Flow Chart of Proposed Method	65
4.2	Rough-Fuzzy Target Class: Blue points are examples of observations that could be outliers while cyan points are in the lower-approximation; orange points belong to the fuzzy boundary and red points are the support vectors.	67