

TABLE OF CONTENTS

	Page
Acknowledgements	iii
Abstract (Thai)	iv
Abstract (English)	vi
List of Tables	xi
List of Figures	xii
Chapter 1 Introduction	1
1.1 Background and Motivation	1
1.2 Literature Review	3
1.3 Purposes of the Study	12
1.4 Research Scope and Method	12
1.5 Education/Application Advantages	13
1.6 Research Methodologies	13
1.7 Organization of Dissertation	14
 Chapter 2 Principles and Theories of the Study	 15
2.1 Fundamental of Clustering Algorithm	15
2.2 Top- <i>N</i> Recommendation Systems	20
2.2.1 Frequency-Based Method	21
2.2.2 Item-Based Top- <i>N</i> Recommendation Algorithm	21
2.2.3 User-Based Top- <i>N</i> Recommendation Algorithm	23

2.3 Genetic Algorithm (GA)	24
2.4 Memetic Algorithm (MA)	27
Chapter 3 Research Designs and Methods	30
3.1 Model of the Visual Clustering Method Based on the GA (VCM-GA)	30
3.1.1 Initial Population	31
3.1.2 Evaluation	31
3.1.3 Reproduction	34
3.1.4 Crossover Operation	34
3.1.5 Mutation	36
3.2 Model of the Visual Clustering Method (VCM) Based on the MA	36
3.3 Model of the Visual Clustering Method (VCM) Based on the K-means	38
3.4 Model of the Top- N Recommendation System Using VCM-GAs, VCM-MAs, and VCM-KM	38
3.5 Evaluation Measure for the Top- N Recommendation System	41
Chapter 4 Results and Discussion	42
4.1 Comparison of Top- N Recommendation Systems	42
4.2 Synthetic and Real-World Data Sets and Parameter Setting	43
4.2.1 Synthetic Data Sets	43
4.2.2 Real-World Data Sets	44

4.3 Results of Clustering on Real-World Data Sets Using VCM-GAs, VCM-MAs, and VCM-KM	49
4.4 Results of Clustering on Real-World Data Sets Using VCM-GAs and VCM-MAs	56
4.5 Experimental Results of Top-5 Recommendation Systems on Real-World Data Sets	72
4.5.1 Top-5 Recommendation Systems on KDD Data Set	73
4.5.2 Top-5 Recommendation Systems on TTS Data Set	81
4.5.3 Top-5 Recommendation Systems on ECR Data Set	83
4.5.4 Top-5 Recommendation Systems on RCM Data Set	89
4.5.5 Top-5 Recommendation Systems on MovieLens Data Set	95
4.6 Discussion	111
Chapter 5 Conclusion	114
References	117
Curriculum Vitae	120

LIST OF TABLES

Table	Page
2.1 A contingency table for binary variables	17
2.2 Example of the initial population	25
2.3 Example of the roulette wheel weighting	26
3.1 Example of three chromosomes in population	34
4.1 Sparsity level of the real-world data sets	46
4.2 Real-world data sets before clustering	68
4.3 Real-world data sets using VCM-KM after clustering	69
4.4 Real-world data sets using VCM-GA1 after clustering	69
4.5 Real-world data sets using VCM-GA2 after clustering	69
4.6 Real-world data sets using VCM-MA1 after clustering	69
4.7 Real-world data sets using VCM-MA2 after clustering	69
4.8 Dunn's index on five real-world data sets	71
4.9 Example of top- N recommendation on the KDD data set	73
4.10 Effectiveness of the parameter setting of the global search	74
4.11 Recommendation performance comparison on KDD data set when one and two items are selected into the basket (evaluated on test sets of 10-fold cross validation)	77
4.12 Recommendation performance comparison on TTS data set when one item is selected into the basket (evaluated on test sets of 10-fold cross validation)	81

LIST OF FIGURES

Figure	Page
1.1 Example of the cold-start problem	2
1.2 Example of the data sparsity problem	2
2.1 Clustering of a set of objects based on the k -means algorithm	19
2.2 Example of the sorted histogram used top- N frequency-based recommendation system	21
2.3 Concept of the item-based top- N recommendation system	22
2.4 Item-based collaborative filtering process for the top- N recommendation	22
2.5 Concept of the user-based top- N recommendation system	23
2.6 Process of the genetic algorithm	24
2.7 A simple crossover schematic	26
2.8 Process of the MA-based algorithm	28
3.1 (a) Users-items table, (b) Binary image	31
3.2 Example of initial population	31
3.3 (a), (b), and (c) Example of the different clusters in binary image	34
3.4 The hill climbing local search algorithm	38
3.5 Model of the proposed top- N recommendation systems	39
4.1 Number of items purchased	47
4.2 (a) Original 3-cluster binary image, (b) Row-column-interchanged binary image, (c)-(g) Result of Fig 4.2(b) using the VCM-GA1, VCM-GA2, VCM-MA1, VCM-MA2, and VCM-KM	50
4.3 Gray color values of three clusters	51

4.4 Result of the clustering on three clusters in the view point of the gray color values	52
4.5 Results of the clustering in the view point of the row and column positions	53
4.6 (a) Original 5-cluster binary image, (b) Row-column-interchanged binary image, (c)-(g) Result of Fig 4.6(b) using the VCM-GA1, VCM-GA2, VCM-MA1, VCM-MA2, and VCM-KM	54
4.7 (a) Original 7-cluster binary image, (b) Row-column-interchanged binary image, (c)-(g) Result of Fig 4.7(b) using the VCM-GA1, VCM-GA2, VCM-MA1, VCM-MA2, and VCM-KM	55
4.8 KDD binary image	56
4.9 TTS binary image	56
4.10 RCM binary image	57
4.11 ECR binary image	57
4.12 MovieLens binary image	57
4.13 Result of Fig 4.8 on KDD image using the VCM-KM	58
4.14 Result of Fig 4.9 TTS image using the VCM-KM	58
4.15 Result of Fig 4.10 on RCM binary using the VCM-KM	59
4.16 Result of Fig 4.11 on ECR image using the VCM-KM	59
4.17 Result of Fig 4.12 on MovieLens image using the VCM-KM	60
4.18 Result of Fig 4.8 on KDD image using the VCM-GA1	60
4.19 Result of Fig 4.9 on TTS image using the VCM-GA1	60
4.20 Result of Fig 4.10 on RCM image using the VCM-GA1	61
4.21 Result of Fig 4.11 on ECR image using the VCM-GA1	61
4.22 Result of Fig 4.12 on MovieLens image using the VCM-GA1	61

4.23 Result of Fig 4.8 on KDD image using the VCM-GA2	62
4.24 Result of Fig 4.9 on TTS image using the VCM-GA2	62
4.25 Result of Fig 4.10 on RCM image using the VCM-GA2	62
4.26 Result of Fig 4.11 on ECR image using the VCM-GA2	63
4.27 Result of Fig 4.12 on MovieLens image using the VCM-GA2	63
4.28 Result of Fig 4.8 on KDD image using the VCM-MA1	64
4.29 Result of Fig 4.9 on TTS image using the VCM-MA1	64
4.30 Result of Fig 4.10 on RCM image using the VCM-MA1	64
4.31 Result of Fig 4.11 on ECR image using the VCM-MA1	65
4.32 Result of Fig 4.12 on MovieLens image using the VCM-MA1	65
4.33 Result of Fig 4.8 on KDD image using the VCM-MA2	66
4.34 Result of Fig 4.9 on TTS image using the VCM-MA2	66
4.35 Result of Fig 4.10 on RCM image using the VCM-MA2	66
4.36 Result of Fig 4.11 on ECR image using the VCM-MA2	67
4.37 Result of Fig 4.12 MovieLens image using the VCM-MA2	67
4.38 The k -cluster level sensitivity study in the VCM-GA2	75
4.39 The k -cluster level sensitivity study in the VCM-MA2	76
4.40 Average precision sensitivity of neighborhood size using user-based recommendation system on ECR data set	84
4.41 Average recall sensitivity of neighborhood size using user-based recommendation system on ECR data set	84
4.42 Average F1 sensitivity of neighborhood size using user-based recommendation system on ECR data set	85

4.43	Average precision sensitivity of neighborhood size using item-based recommendation system on ECR data set	85
4.44	Average recall sensitivity of neighborhood size using item-based recommendation system on ECR data set	86
4.45	Average F1 sensitivity of neighborhood size using item-based recommendation system on ECR data set	86
4.46	Average precision on ECR data set using the proposed methods and the common methods	87
4.47	Average recall on ECR data set using the proposed methods and the common methods	88
4.48	Average F1 on ECR data set using the proposed methods and the common methods	89
4.49	Average precision on RCM data set using the user-based methods	90
4.50	Average recall on RCM data set using the user-based methods	90
4.51	Average F1 on RCM data set using the user-based methods	91
4.52	Average precision on RCM data set using the item-based methods	91
4.53	Average recall on RCM data set using the item-based methods	92
4.54	Average F1 on RCM data set using the item-based methods	92
4.55	Average precision on RCM using the proposed methods and common methods	93
4.56	Average recall on RCM data set using the proposed methods and common methods	94
4.57	Average F1 on RCM data set using the proposed methods and common methods	95

4.58 Average precision sensitivity of neighborhood size using user-based recommendation system on MovieLens data set	97
4.59 Average recall sensitivity of neighborhood size using user-based recommendation system on MovieLens data set	98
4.60 Average F1 sensitivity of neighborhood size using user-based recommendation system on MovieLens data set	99
4.61 Average precision sensitivity of neighborhood size using item-based recommendation system on MovieLens data set	101
4.62 Average recall sensitivity of neighborhood size using item-based recommendation system on MovieLens data set	102
4.63 Average F1 sensitivity of neighborhood size using item-based recommendation system on MovieLens data set	104
4.64 Average precision on MovieLens data set using the proposed methods and the common methods	106
4.65 Average recall on MovieLens data set using the proposed methods and the common methods	108
4.66 Average F1 on MovieLens data set using the proposed methods and the common methods	110