

THE UNIVERSITY OF ADELAIDE

DOCTORAL THESIS

**Diversity Optimization and Parameterized
Analysis of Heuristic Search Methods for
Combinatorial Optimization Problems**

Author:
Wanru Gao

Principle Supervisor:
Prof. Frank Neumann

Co-Supervisor:
Dr. Markus Wagner

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy*

in the

Optimisation and Logistics
School of Computer Science

December 8, 2016

Declaration of Authorship

I, Wanru Gao, declare that this thesis titled, “Diversity Optimization and Parameterized Analysis of Heuristic Search Methods for Combinatorial Optimization Problems” and the work presented in it are my own.

- I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.
- I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.
- I also give permission for the digital version of my thesis to be made available on the web, via the University’s digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed:

Date:

“Courage is not the absence of fear, but rather the assessment that something else is more important than fear.”

Franklin D. Roosevelt

THE UNIVERSITY OF ADELAIDE

Abstract

Faculty of Engineering, Computer & Mathematical Science

School of Computer Science

Doctor of Philosophy

Diversity Optimization and Parameterized Analysis of Heuristic Search Methods for Combinatorial Optimization Problems

by Wanru Gao

Heuristic search algorithms belong to the most successful approaches for many combinatorial optimization problems which have wide real world applications in various areas. The heuristic algorithms usually provide solutions with acceptable quality in reasonable time-frame which is different from exact algorithms. Fixed-parameter approach provides a way for understanding how and why heuristic methods perform well for prominent combinatorial optimization problems. In this thesis, there are two main topics discussed.

Firstly, we integrate the well-known branching approach for the classical combinatorial optimization problem, namely minimum vertex cover problem, to a local search algorithm and compare its performance with the core component of the state-of-the-art algorithm. After that, we investigate how well-performing local search algorithms for small or medium size instances can be scaled up to solve massive input instances. A parallel kernelization technique is proposed which is motivated by the assumption that huge graphs are composed of several easy to solve components while the overall problem is hard to solve.

Using evolutionary algorithms to generate a diverse set of solutions where all of them meet certain quality criteria has gained increasing interests in recent years. As the second section, we put forward an evolutionary algorithm which allows us to maximize the diversity over a set of solutions with good quality and then focus on the theoretical analysis of the algorithm to provide understanding of how evolutionary algorithms maximize the diversity of a population and guarantee the quality of all solutions at the same time. Then the idea is extended to evolving hard/easy optimization problem instances with diverse feature values. The feature-based analysis of heuristic search algorithms plays an important role in understanding the behaviour of the algorithm and our results show good classification of the problem instances in terms of hardness based on different combinations of feature values.

Acknowledgements

I would like to express my special appreciation and thanks to many people who have encouraged, inspired and helped me in the PhD study. Without their generous help, it will be impossible for this thesis to be finished. I would like to thank

- First of all, to my supervisor Frank Neumann. It has been an honour to be one of his PhD students. He has provided me great support, guidance and inspiration in doing research. I appreciate all his contributions of time, ideas and funding to make my PhD study inspiring and productive.
- To my co-supervisor Markus Wagner for his kind advices and encouragement in both researching and teaching.
- To all of the co-authors of the papers, Mojgan Pourhassan, Samadhi Nallaperuma, Benjamin Doerr, Tobias Friedrich and Timo Kötzing. I learned a lot in collaborations with others.
- To all the researchers and practitioners whose work is referred in my study for their great efforts in previous research. None of the studies is able to be conducted without the foundations set by other researchers.
- To all reviewers of the papers and the examiners of this thesis for their valuable comments.
- To the past and present group members in our research group that I have had the pleasure to work with, Bradley Alexander, Mingyu Guo, Sergey Polyakovskiy, Samadhi Nallaperuma, Mojgan Pourhassan, Shayan Poursoltan, Junhua Wu, Feng Shi, Aneta Neumann and other fellow students in the HDR office. It has been a great experience to work in the research group of Optimisation and Logistics in School of Computer Science. Sharing ideas in both research and daily life helps me progress a lot.
- To the School of Computer Science, Genetic and Evolutionary Computation Conference (GECCO) and the International Conference on Parallel Problem Solving from Nature (PPSN) for supporting my conference travel and research visit.
- To everyone in the group of Algorithm Engineering in Hasso Plattner Institute, Potsdam, Germany for sharing research ideas and providing feedback on my research during my stay with them.
- To all of the teachers who have taught or directed me in different areas for encouraging me to keep studying.
- Lastly, to my beloved family, especially my mom Jing and my dad Mingli, my relatives and friends for supporting me all the time in both life and research.

CONTENTS

Declaration of Authorship	i
Abstract	iii
Acknowledgements	iv
Contents	v
List of Figures	ix
List of Tables	xi
List of Abbreviations	xii
1 Introduction	1
2 Combinatorial Optimization and Heuristic Search	5
2.1 Introduction	5
2.2 Local Search	6
2.2.1 Metaheuristics for Local Search	6
2.3 Evolutionary Algorithms	8
2.3.1 General Issues of Evolutionary Algorithms	8
2.3.1.1 Representation	9
2.3.1.2 Variation Operator	9
2.3.1.3 Selection Mechanism	10
2.3.1.4 Other Issues	11
2.3.2 Main Types of Evolutionary Algorithms	11
2.3.2.1 Genetic Algorithms	11
2.3.2.2 Evolution Strategies	11
2.3.2.3 Evolutionary Programming	12
2.3.2.4 Genetic Programming	12

2.3.3	Evolutionary Multi-objective Optimization	12
2.4	Combinatorial Optimization	13
2.4.1	Travelling Salesman Problem	14
2.4.1.1	Exact Algorithms for TSP	15
2.4.1.2	Heuristics and Approximation Algorithms for TSP	16
2.4.2	Minimum Vertex Cover Problem	18
2.5	Diversity in Evolutionary Algorithm	19
2.5.1	Diversity in Objective Space	19
2.5.2	Diversity in Decision Space	20
2.5.3	Diversity Measurement	21
2.6	Conclusion	22
3	Algorithm Analysis and Problem Complexity in Evolutionary Algorithms	23
3.1	Introduction	23
3.2	Classical Computational Complexity Analysis	24
3.2.1	Fitness-based Partitions	24
3.2.2	Deviation Inequalities	25
3.2.3	Drift Analysis	27
3.3	Parameterized Analysis	28
3.3.1	Some Basic Definitions	28
3.3.2	Bounded Search Tree	29
3.3.3	Kernelization	30
3.4	Feature-based Analysis	30
3.4.1	Features Selection for Characterizing Problem Instances	31
3.4.2	Feature-based Analysis for Problem Hardness	32
3.5	Conclusion	33
4	Heuristic Algorithms for Minimum Vertex Cover Problem	34
4.1	Introduction	34
4.2	Background	35
4.3	Initialization Strategies	36
4.3.1	Different Initialization Approaches	36
4.3.2	Theoretical Analysis	38
4.3.3	Experimental Results	39
4.4	Two Local Search Algorithms for MVC	43
4.4.1	Edge-based Local Search Algorithm	43
4.4.2	Vertex-based Local Search Algorithm	44
4.4.3	Experimental Analysis	44
4.5	Conclusion	45
5	Scaling up Local Search for MVC in Massive Graphs	48
5.1	Introduction	48
5.2	Substructures in Massive Graphs	50
5.3	Parallel Kernelization for MVC	51
5.4	Experimental Results on DIMACS and BHOSLIB Benchmarks	52
5.4.1	Results for Original DIMACS and BHOSLIB Benchmarks	53
5.4.2	Results for Combined DIMACS and BHOSLIB Benchmarks	54
5.4.3	Results for Combination of Existing Hard Instances	55
5.5	Experimental Results on Real World Graphs	56
5.6	Remarks and Conclusions	59

6	Diversity Maximization for Single-objective Problems in Decision Space	60
6.1	Introduction	60
6.2	Background	61
6.2.1	Decision Space Diversity Measurement	61
6.2.2	Population-based Evolutionary Algorithm with Diversity Maximization	62
6.2.3	Classical Runtime Analysis Method for Evolutionary Algorithms	63
6.3	Diversity Maximization for OneMax Problem	64
6.3.1	OneMax Problem	64
6.3.2	Analysis of Large Threshold	66
6.3.3	Analysis of Smaller Threshold	69
6.4	Diversity Maximization LeadingOnes Problem	76
6.4.1	LeadingOnes Problem	76
6.4.2	Runtime Analysis for $(\mu + 1)$ -EA _D on LeadingOnes	77
6.5	Analysis of Vertex Cover Problem	80
6.5.1	Analysis of Complete Bipartite Graphs	81
6.5.1.1	$\epsilon < 1/3$	82
6.5.1.2	$\epsilon = 1/3$	83
6.5.1.3	$1/3 \leq \epsilon < 1/2$	85
6.5.2	Analysis of Paths	88
6.5.2.1	Paths with Even Number of Nodes	89
6.5.2.2	Paths with Odd Number of Nodes	92
6.6	Conclusion	93
7	Diversity Maximization for Multi-objective Problems in Decision Space	94
7.1	Introduction	94
7.2	Preliminaries	95
7.2.1	OneMinMax Problem	95
7.2.2	Hypervolumn-based Algorithm with Diversity Maximization	96
7.3	Search Space Diversity Optimization	98
7.3.1	Diversity Maximization for Multi-objective Problem	98
7.3.2	Experimental Analysis for Diversity Maximization for OneMinMax	99
7.3.3	Runtime Analysis for Diversity Maximization for OneMinMax	100
7.4	Conclusion	103
8	Feature-Based Diversity Optimization for Problem Instance Classification	104
8.1	Introduction	104
8.2	Preliminaries	105
8.2.1	Traveling Salesman Problem	105
8.2.2	Features of TSP Instance	106
8.2.3	Evolutionary Algorithm for Evolving Instance with Diversity Optimization	107
8.2.4	Experiments Setup	109
8.3	Range of Feature Values	109
8.4	Instance Classification Based on Multiple Features	113
8.4.1	Diversity Maximization over Single Feature Value	113
8.4.2	Diversity Maximization over Multiple Feature Values	116
8.5	Instance Classification Using Support Vector Machine	117
8.5.1	SVM with Linear Kernel	117
8.5.2	Nonlinear Classification with RBF Kernel	118
8.6	Diversity optimization for Instance Hardness	120
8.7	Conclusion	121

9 Conclusions	123
Bibliography	126

LIST OF FIGURES

2.1	The general procedure of an evolutionary algorithm expressed as a flow chart.	8
2.2	An example for a case that two individuals that are next to each other in objective space but are mapped to two points that are far away in decision space.	20
4.1	The histograms show the frequency that each algorithm gets the initial vertex cover of certain size for two sample instances from <i>DIMACS</i> benchmark set.	40
4.2	The histograms show the frequency that each algorithm gets the initial vertex cover of certain size for sample instances that randomly generated or from real world graph.	40
4.3	The box plots show the distribution of the results from 101 independent runs of seven different initialization approach on different minimum vertex cover instances.	42
4.4	The graphs shows the improvement of two local search algorithms in three example instances over iterations.	46
6.1	The $\mu \times n$ matrix represents the individuals in a population. In the example, it is a matrix for a population with 4 individuals which are all 8 bits in length. The 7th column is all-1-bit column and the 3rd column is 0-bit column as defined.	69
6.2	The example of global optimum for maximizing population diversity in One-Max where population size $\mu \leq \frac{n}{n-v}$.	71
6.3	The $\mu \times n$ matrix represents the solutions of MVC in a population. In the example, each row represent an individual. In each row, the left ϵn bits and right $(1 - \epsilon)n$ represent the existence of nodes from set V_1 and set V_2 in the individual respectively.	81
6.4	When $\epsilon = 1/3$, the population of MVC with maximized diversity should include cover sets with different types.	84
6.5	In the complete bipartite graph with large ϵ , it is possible at some stage, the last individual of one type is replaced by an individual of the other type.	87

6.6	Arrange all $(k + 1)$ optimal solutions for the MVC problem of a path with $n = 2k$ according to the ascending order of numbers of '01' pairs.	89
7.1	The 8×7 matrices represent some populations for which there is no 1-bit flip can improve the population diversity. The last rows report the numbers of 1-bits in corresponding columns.	100
8.1	The boxplots show the distribution of some feature values of a population consisting of 100 different hard or easy TSP instances of different number of cities without or with diversity mechanism.	110
8.2	Some examples of the evolved hard TSP instances of different number of cities are shown with an optimal tour computed by Concorde.	112
8.3	2D Plots of feature combinations which provide a separation between easy and hard instances. The blue dots and orange dots represent hard and easy instances respectively.	113
8.4	2D Plots of feature combinations which does not provide a clear separation between easy and hard instances. The blue dots and orange dots represent hard and easy instances respectively.	113
8.5	3D Plots of combining experiment results from maximizing the diversity over features <i>mst_dists_mean</i> , <i>nnds_mean</i> and <i>chull_area</i> , which provides a good separation of easy and hard instances.	115
8.6	3D Plots of combining experiment results from maximizing the diversity over features <i>mst_dists_mean</i> , <i>chull_area</i> and <i>centroid_mean_distance_to_centroid</i> , which provides a good separation of easy and hard instances.	115
8.7	3D Plots of combining experiment results from maximizing the diversity over features <i>mst_dists_mean</i> , <i>chull_area</i> and <i>mst_depth_mean</i> , which provides a good separation of easy and hard instances.	115
8.8	3D Plots of combining experiment results from maximizing the diversity over features <i>mst_dists_mean</i> , <i>nnds_mean</i> and <i>chull_area</i> with considering of weighting, which provides a good separation of easy and hard instances. Legend is the same as that in Figure 8.5.	116
8.9	3D Plots of experiment results from maximizing the diversity over approximation ratio in the feature space of feature combination <i>chull_area</i> , <i>mst_dist_mean</i> and <i>mst_depth_mean</i> . The color of the dot reflect the hardness of the problem.	121
8.10	3D Plots of experiment results from maximizing the diversity over approximation ratio in the feature space of feature combination <i>chull_area</i> , <i>angle_mean</i> and <i>cluster_10pct_mean_distance_to_centroid</i> . The color of the dot reflect the hardness of the problem.	121

LIST OF TABLES

4.1	Experimental results on instances comparing the statistics between Algorithm 4.1 and 4.2.	41
4.2	Performance comparison between Algorithm 4.5 and 4.6 on some sample instances. The average vertex cover size is listed after running each algorithm for certain number of iterations.	46
5.1	This table contains the results from NuMVC-PK and NuMVC on the BHOSLIB and DIMACS benchmark.	53
5.2	This table contains instances that have been tested on, which are generated by duplicating one existing hard instance in BHOSLIB benchmark.	54
5.3	This table contains instances that have been tested on, which are generated by combining several existing hard instance in BHOSLIB benchmark.	55
5.4	Experimental results of NuMVC-PK and single run of NuMVC on instances from some real world graphs about social networks.	57
5.5	Experimental results of NuMVC-PK and single run of NuMVC on instances from some real world graphs about collaboration networks.	57
5.6	Experimental results on instances from some real world graphs with different parameter settings. $\mu = 3$ in all cases.	58
5.7	Experimental results on instances from some duplicated benchmark graphs with different parameter settings. $\mu = 3$ in all cases.	59
8.1	This table lists the accuracy of SVM with RBF kernel separating the hard and easy instances in 21 different two-feature space.	119
8.2	This table lists the accuracy of SVM with RBF kernel separating the hard and easy instances in 35 different three-feature space.	119

LIST OF ABBREVIATIONS

OMM	One Min Max
RLS	Randomized Local Search
EA	Evolutionary Algorithm
GA	Genetic Algorithm
NP	Nondeterministic Polynomial
TSP	Traveling Salesman Problem
MVC	Minimum Vertex Cover
MOP	Multi-objective Optimization Problem
MOEA	Multi-Objective Evolutionary Algorithm
w.l.o.g.	without loss of generality
i.e.	id est (that is)
e.g.	exempli gratia (for example)

To my beloved parents...