# Combinatorial optimization methods for the $(\alpha, \beta)$-k Feature Set Problem

**Amir Salehipour**


---

Faculty of Engineering and Built Environment
School of Electrical Engineering and Computing
University of Newcastle, NSW, Australia

# Combinatorial optimization methods for the $(\alpha, \beta)$-k Feature Set Problem

Amir Salehipour

March 2019

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

# Statement of Originality

The thesis contains no material which has been accepted for the award of any other degree or diploma 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. I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

Amir Salehipour 28 March 2019

# Dedication

To my parents, my wife and my son, and my sisters.

# Acknowledgment

My first acknowledgment must go to my lovely wife, Dr Leila Moslemi Naeni. I want to thank her for all the support that she gave me during these long four years of my research.
Secondly, I would like to thank my supervisors, Prof. Pablo Moscato and Prof. Regina Berretta for providing me with this opportunity to pursue my PhD studies at the University of Newcastle, Australia. I would like to thank all the past and present members of the CIBM team (Center for Bioinformatics, Biomarker Discovery and Information-Based Medicine, the University of Newcastle) for their friendly support, chats, and helpful ideas. Starting with ex-member Dr Carlos Riveros, an amazing supportive person, and a true friend, who helped me many times with programming, and modeling. Then, Dr Renato Vimieiro, Dr Mateus Rocha de Paula and Dr Ahmed Shamsul Arefin for their friendship. Finally, thanks to all my colleagues and friends that supported me thorough this journey: Ademir, Amer, Claudio, Francia, Heloisa, Inna, Lukas, Luke, Marta, Mohammad, Nader, Nisha and Shannon.
I am very thankful to the University of Newcastle for my scholarships and for the opportunity to be part of an incredible academic environment. My Doctorate was fully funded by the University of Newcastle's Postgraduate Research Scholarships; those financial supports are very much appreciated.
I would like to thank my mother and my father, for the love and support that they have given me from the other side of the world. You gave me the strength and motivation to continue and finish my PhD.

Amir Salehipour
March 2019

# Contents

# List of Figures

# List of Tables

# List of Algorithms

# Abstract

This PhD research thesis proposes novel and efficient combinatorial optimization-based solution methods for the $(\alpha, \beta)$-k Feature Set Problem. The $(\alpha, \beta)$-k Feature Set Problem is a combinatorial optimization-based feature selection approach proposed in 2004, and has several applications in computational biology and Bioinformatics. The $(\alpha, \beta)$-k Feature Set Problem aims to select a minimum cost set of features such that similarities between entities of the same class and differences between entities of different classes are maximized.

The developed solution methods of this research include heuristic and exact methods. While this research focuses on utilizing exact methods, we also developed mathematical properties, and heuristics and problem-driven local searches and applied them in certain stages of the exact methods in order to guide exact solvers and deliver high quality solutions. The motivation behind this stems from computational difficulty of exact solvers in providing good quality solutions for the $(\alpha, \beta)$-k Feature Set Problem. Our proposed heuristics deliver very good quality solutions including optimal, and that in a reasonable amount of time.

The major contributions of the presented research include: 1) investigating and exploring mathematical properties and characteristics of the $(\alpha, \beta)$-k Feature Set Problem for the first time, and utilizing those in order to design and develop algorithms and methods for solving large instances of the $(\alpha, \beta)$-k Feature Set Problem; 2) extending the basic modeling, algorithms and solution methods to the weighted variant of the $(\alpha, \beta)$-k Feature Set Problem (where features have a cost); and, 3) developing algorithms and solution methods that are capable of solving large instances of the $(\alpha, \beta)$-k Feature Set Problem in a reasonable amount of time (prior to this research, many of those instances pose a computational challenge for the exact solvers).

To this end, we showed the usefulness of the developed algorithms and methods by applying them on three sets of 346 instances, including real-world, weighted, and randomly generated instances, and obtaining high quality solutions in a short time. To the best of our knowledge, the developed algorithms of this research have obtained the best results for the $(\alpha, \beta)$-k Feature Set Problem. In particular, they outperform state-of-the-art algorithms and exact solvers, and have a very competitive performance over large instances because they always deliver feasible solutions, and obtain new best solutions for a majority of large instances in a reasonable amount of time.

# Awards, Publications, and Outcomes

Part of the material presented in this research thesis has been already presented, and published in peer-reviewed conferences. The list of publications and presentations is provided below. It is worth mentioning that during my PhD studies I won one major award of the University of Newcastle for conducting an outstanding original research on the $(\alpha, \beta)$-k Feature Set Problem.

**Awards**

1. Winner of the "2015 Research Poster Prize Competition", awarded by Faculty of Engineering and Built Environment, University of Newcastle, 2015.

2. Two PGRSS travel grants (approximately, $4,500), awarded by the Faculty of Engineering and Built Environment, University of Newcastle, 2016.

**Conference proceedings**

1. "Tight lower bounds and a hybrid heuristic for a problem of selecting features", orally presented at EURO 2016 International Conference (peer-reviewed). Poznan, 3 – 6 July 2016.

2. "An optimization approach towards selecting features in biological datasets", orally presented at 24th National Conference of the Australian Society for Operations Research (peer-reviewed). Canberra, 16 – 18 November 2016.

**Working papers**

1. "Efficient solution methods for the Min k $(\alpha, \beta)$-k Feature Set Problem" (75% complete; will be submitted to an international journal in next few months).

2. "A heuristic algorithm for the $(\alpha, \beta)$-k Feature Set Problem" (80% complete; will be submitted to a top tier journal in the next few weeks).