Empirical Study of Correlations in the Fitness Landscapes of Combinatorial Optimization Problems

Longfei Zhang
University of Electronic Science and Technology of China
Chengdu, Sichuan Province
southwind_zlf@outlook.com

Ke Li*
University of Exeter
Exeter, Devon
k.li@exeter.ac.uk

Shi Gu
University of Electronic Science and Technology of China
Chengdu, Sichuan Province
gus@uestc.edu.cn

ABSTRACT

One of the most common problem-solving heuristics is by analogy. For a given problem, a solver can be viewed as a strategic walk on its fitness landscape. Thus if a solver works for one problem, it is anticipated that it will be effective for problems within the same category whose fitness landscapes essentially share *structural similarity* with each other. However, due to the black-box nature, it is far from trivial to infer such similarity in real-world scenarios. To bridge this gap, this paper proposes two alternative approaches to empirically investigate the potential existence of structural similarity among different fitness landscapes. Specifically, we pick up three classic combinatorial optimization problems to constitute the benchmark set. We apply a local optima network construction routine to build a coarse-grained model to represent the fitness landscapes of different problems at various dimensions. Thereafter, we apply a graph embedding method, to empirically investigate the potential existence of correlations with respect to different local optima networks. From our empirical results, we are exciting to find some evidence of the existence of similarity not only for a given problem with various dimensions but also across different problems.

CCS CONCEPTS

• **Computing methodologies → Modeling methodologies**; **Discrete space search**;

KEYWORDS

Fitness landscape analysis, local optima networks, complex networks, graph embedding

ACM Reference Format:
Longfei Zhang, Ke Li, and Shi Gu. 2021. Empirical Study of Correlations in the Fitness Landscapes of Combinatorial Optimization Problems. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3459528

*Li is the corresponding author and he set the research agenda. Zhang implemented the experiments. Li and Gu co-supervised the research.

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07…$15.00
https://doi.org/10.1145/3449726.3459528

1 INTRODUCTION

The original idea of fitness landscape dates back to 1932 when Wright pioneered this concept in evolutionary biology [9]. In recent two decades and beyond, there have been many efforts devoted to developing computational models of fitness landscape [5], among which local optima networks (LONs) [3] have become the most popular one in the meta-heuristic community [4]. LONs are rooted from the study of energy landscapes in chemical physics [6] and its basic idea is to model the fitness landscape as a graph or network. In particular, the vertices of a LON are local optima (either minima or maxima depending on the problem definition) and the edges indicate certain search dynamics of a meta-heuristic algorithm, i.e., the transition from one local optimum to another. Since LONs are able to capture various characteristics of the underlying landscape (e.g., the number of local optima, their distribution and connectivity pattern), they are powerful tools for both fitness landscape analysis and visualization [2]. Metrics developed for analyzing and understanding graphs and/or complex networks are in principle useful for enabling the quantitative understanding of the structural characteristics of LONs. In addition, such metrics can then be used to composite quantitative features that are applicable for a wider range of applications related to automatic algorithm selection and/or configuration [7].

In practice, one of the most common problem-solving approaches is by analogy. Its basic assumption is that if a black-box optimization problems (BBOPs) solver is effective for one problem, we can expect its effectiveness for solving other problems belonging to the same category whose fitness landscape essentially share structural similarity with each other. In other words, the strategic walk on the fitness landscape induced by the solver may be extended to fitness landscapes of similar problems that are anticipated to share certain patterns or sub-structures but vary in different size or volume. Thus the inference of such similarity would not only deepen our understanding of the solver's behavior but also facilitates the design the BBOP solvers with respect to its fitness landscape potentially. However, it is in practice far from trivial to infer such similarity when encountering BBOPs. To the best of our knowledge, there is no dedicated research to investigate the correlations of the existing fitness landscape analysis literature. To bridge this gap, this paper proposes to use the technique from graph theory to empirically investigate the potential existence of correlation among different fitness landscapes. Specifically, we pick up three classic combinatorial optimization problems to constitute the benchmark set. We apply a LON construction routine to build a coarse-grained model to represent the fitness landscapes of different problems at various dimensions. Thereafter, we apply a graph embedding method, to empirically investigate the potential existence of structural similarity among different LONs. From our empirical results, we find some evidence that supports the existence of correlations not only for a given problem with various dimensions but also across different problems.

2 EXPERIMENTAL SETUP

This paper picks up three classic combinatorial optimization problems, including the number partitioning problem (NPP), traveling sales problem (TSP) and $0-1$ knapsack problem (KP), to composite the benchmark set. They are all NP-hard problems and have been widely used in the fitness landscape analysis literature.

Given the intrinsic complexity of fitness landscapes (e.g., a large number of local optima and their complex connectivity), LONs can be too complex to directly evaluate any structural information and compare the similarity. Instead of taking the whole graph as an integral entity, it might be more plausible to learn a feature representation (or spectrum) that captures certain topological sub-structure characteristics. In particular, this feature representation is invariant under graph isomorphism. In this paper, we consider applying a state-of-the-art graph embedding approach, FGSD method in particular [8], to serve this purpose. The FGSD method provides a simple yet powerful graph feature representation based on the multiset of node pairwise distances, providing that it exhibits certain uniqueness, stability, sparsity properties and is computationally fast with a $O(n^2)$ complexity where n is the number of vertices in a graph. In a nutshell, the FGSD method consists of two major steps. It first calculates the Moore-Penrose spectrum of the normalized Laplacian to capture certain inherent atomic sub-structures of the underlying graph. Then the histograms of the spectrum features are calculated to constitute a fixed length feature vector representations.

In this paper, we consider Pearson correlation coefficient (PCC) [1] to help quantify the similarity of different dimensions LONs. PCC is used to measure the statistical association between two continuous variables. Its value ranges from −1 to 1. In particular, a larger PCC value indicates a stronger correlation. Given two random variables $\mathbf{x}^1$ and $\mathbf{x}^2$, their PCC is calculated as:

$$PCC(\mathbf{x}^1, \mathbf{x}^2) = \frac{\mathsf{cov}(\mathbf{x}^1, \mathbf{x}^2)}{\sigma(\mathbf{x}^1)\sigma(\mathbf{x}^2)}, \tag{1}$$

where $\mathsf{cov}(\cdot,\cdot)$ evaluates the covariance and $\sigma(\cdot)$ represents the standard deviation.

3 RESULTS

From this PCC heatmap matrix, we observe some clear divisions of blocks. Specifically, the LONs of NPP with $d \leq 20$ have been highly correlated with those of $0-1$ KP with $d \leq 35$; while the LONs of higher dimensional NPP (i.e., with $d \geq 25$) have shown significant correlations with the LONs of TSP across all dimensions. On the other hand, the correlation between TSP and $0-1$ KP is rather marginal until the dimension d goes up to 50. Nevertheless, this observation still agrees with those obtained from the statistical analysis of four network metrics, i.e., there do exist certain patterns or sub-structures sharing across the fitness landscapes of different combinatorial optimization problems. In particular, since the fitness landscapes of TSP always share certain similar patterns across

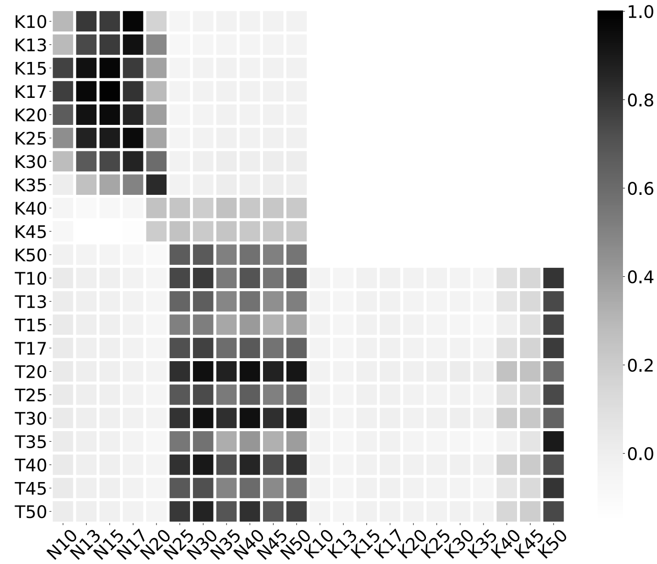

Figure 1: Heatmap of PCC values for cross comparison of NPP, TSP and $0-1$ KP at all dimensions.

different dimensions, their structural similarity with the fitness landscapes of other problems are also across all dimensions.

4 CONCLUSIONS

In this paper, we empirically investigate the potential existence of correlations among the fitness landscapes of different combinatorial optimization problems. We use a graph embedding method from the deep learning literature that compresses LONs into feature vectors with a fixed length and evaluate their inter-correlations. From our empirical results, we find some evidence that support the existence of correlations among different LONs.

REFERENCES

[1] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 1–4.
[2] Gabriela Ochoa and Katherine Malan. 2019. Recent advances in fitness landscape analysis. In *GECCO'19: Proc. of the 2019 Genetic and Evolutionary Computation Conference Companion*. ACM, 1077–1094.
[3] Gabriela Ochoa, Marco Tomassini, Sébastien Vérel, and Christian Darabos. 2008. A study of NK landscapes' basins and local optima networks. In *GECCO'08: Proc. of the 2008 Genetic and Evolutionary Computation Conference*. ACM, 555–562.
[4] Gabriela Ochoa, Sebastien Verel, Fabio Daolio, and Marco Tomassini. 2014. *Recent Advances in the Theory and Application of Fitness Landscapes*. Springer, Chapter Local Optima Networks: A New Model of Combinatorial Fitness Landscapes.
[5] Hendrik Richter. [n. d.]. Fitness Landscapes and Evolutionary Dynamics. In *Nostradamus: Proc. of the 2012 Nostradamus conference on Modern Methods of Prediction, Modeling and Analysis of Nonlinear Systems*, Ivan Zelinka, Otto E. Rössler, Václav Snásel, Ajith Abraham, and Emilio S. Corchado Rodríguez (Eds.).
[6] Frank H. Stillinger. 1995. A Topographic View of Supercooled Liquids and Glass Formation. *Science* 267, 5206 (1995), 1935–1939.
[7] German Treimun-Costa, Elizabeth Montero, Gabriela Ochoa, and Nicolás Rojas-Morales. 2020. Modelling parameter configuration spaces with local optima networks. In *GECCO'20: Proc. of the 2020 Genetic and Evolutionary Computation Conference*. ACM, 751–759.
[8] Saurabh Verma and Zhi-Li Zhang. 2017. Hunt For The Unique, Stable, Sparse And Fast Feature Learning On Graphs. In *NIPS'17: Proc. of the 2017 Annual Conference on Neural Information Processing Systems*. 88–98.
[9] Sewall Wright. 1932. The roles of mutations, inbreeding, crossbreeding and selection in evolution. In *Proc. of the 11th International Congress of Genetics*, Vol. 1. 356–366.