Do Quality Indicators Prefer Particular Multi-Objective Search Algorithms in Search-Based Software Engineering? (Hot Off the Press track at GECCO 2021)

Shaukat Ali
Simula Research Laboratory
Fornebu, Norway
shaukat@simula.no

Paolo Arcaini
National Institute of Informatics
Tokyo, Japan
arcaini@nii.ac.jp

Tao Yue
Simula Research Laboratory and
Nanjing University of Aeronautics and Astronautics
Nanjing, China
taoyue@ieee.org

ABSTRACT

In Search-based Software Engineering (SBSE), researchers and practitioners (SBSE users) using multi-objective search algorithms (MOSAs) often select commonly used MOSAs to solve their search problems. Such a selection is usually not justified, and the main selection criterion is the MOSA popularity. On the other hand, SBSE users are usually aware of the desired qualities of solutions of their search problem, captured by Quality Indicators (QIs). Consequently, to guide SBSE users in selecting MOSAs for their specific SBSE problems, we study preference relationships between QIs and MOSAs with an empirical evaluation. Given a QI or a quality aspect (e.g., convergence), we suggest a MOSA that is highly likely to produce solutions representing the QI or the quality aspect. Based on our experiments' results, we provide insights and suggestions for SBSE users to choose a MOSA based on experimental evidence.

This is an extended abstract of the paper [2]: S. Ali, P. Arcaini, and T. Yue, "Do Quality Indicators Prefer Particular Multi-Objective Search Algorithms in Search-Based Software Engineering?", 12th International Symposium on Search-Based Software Engineering (SSBSE 2020).

CCS CONCEPTS

• **Software and its engineering → Search-based software engineering**.

KEYWORDS

Search-Based Software Engineering, Multi-Objective Search

ACM Reference Format:
Shaukat Ali, Paolo Arcaini, and Tao Yue. 2021. Do Quality Indicators Prefer Particular Multi-Objective Search Algorithms in Search-Based Software Engineering? (Hot Off the Press track at GECCO 2021). 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/10.1145/3449726.3462721

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

ACM ISBN 978-1-4503-8351-6/21/07.
https://doi.org/10.1145/3449726.3462721

1 EMPIRICAL STUDY

Motivation. Given a Quality Indicator (QI) or a quality aspect (e.g., uniformity), guiding Search-based Software Engineering (SBSE) researchers and practitioners, i.e., SBSE users, in choosing a MOSA that has a high probability of finding solutions representing the selected QI or the quality aspect.

Application Context. In practical applications, typically, SBSE users are aware of the quality of solutions they are interested in. They want to choose one MOSA without performing large-scale experiments to compare multiple existing MOSAs. Indeed, in such cases, users are usually limited by time budgets, implying that it is practically impossible to perform such extensive experiments.

Empirical Study. The design of our empirical evaluation is shown in Fig. 1.[1] We used real-world, industrial, and open-source SBSE

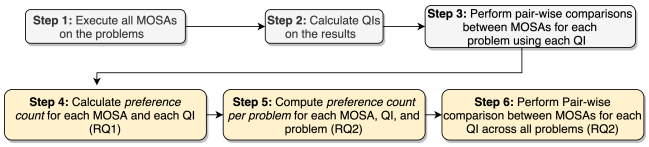

Figure 1: Design of the Experiment

problems, whose details are given in [1]. The data from [1] had 100 runs of commonly used MOSAs for solving SBSE problems (Step 1 in Fig. 1), i.e., NSGA-II, MoCell, SPEA2, PAES, SMPSO, and CellDE. The data also include computed QIs that are commonly used in SBSE (Step 2, Fig. 1), i.e., Generational Distance (GD), Euclidean Distance (ED), Epsilon (EP), Generalized Spread (GS), Pareto Front Size (PFS), Inverted Generational Distance (IGD), Hypervolume (HV), and Coverage (C). Step 3 in Fig. 1 consists in performing relevant statistical tests to compare MOSA pairs with each commonly used QI; for each QI, we computed which MOSA performed *significantly better* than another MOSA. Steps 4 and 5 in Fig. 1 were performed in this paper to answer the following two research questions (RQs):

- **RQ1**: How frequently a QI prefers a particular MOSA? We study the percentage of times that a QI prefers a particular MOSA to understand the overall preferences of a QI.
- **RQ2**: How frequently a QI prefers a particular MOSA across the different SBSE problems? We study the preferences of QIs across the problems when studying pairs of MOSAs.

[1] Data, scripts, and results at https://github.com/ERATOMMSD/QIsPreferences.

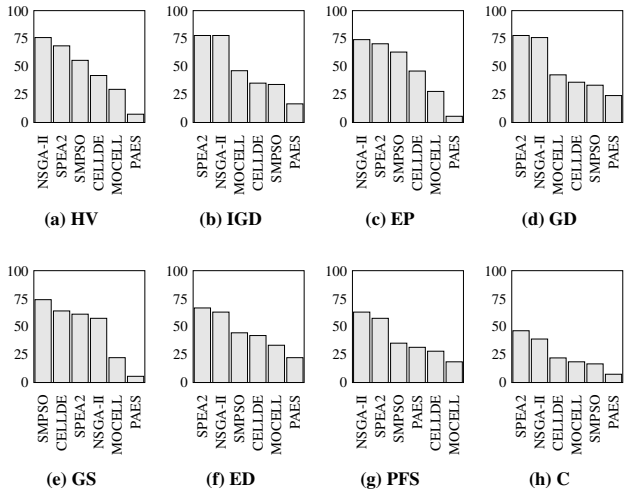

Figure 2: RQ1 – *Preference count*

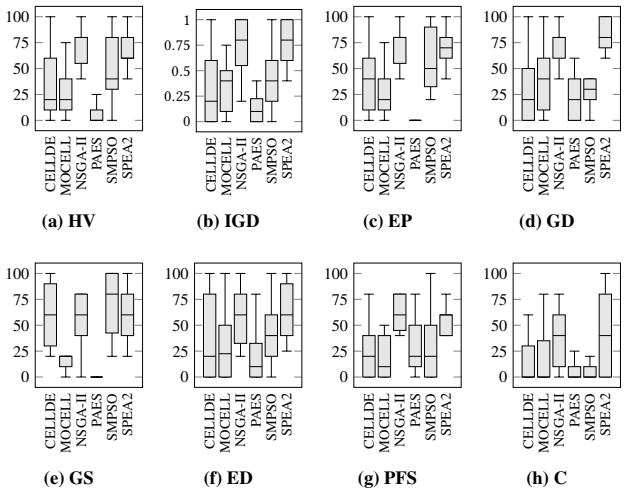

Figure 3: RQ2 – *Preference count per problem*

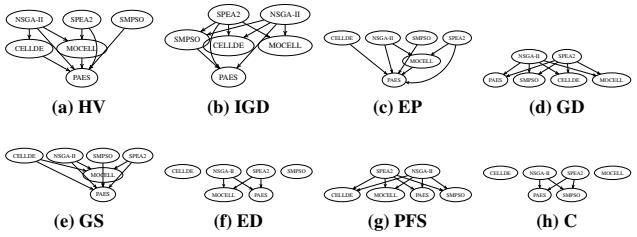

Figure 4: RQ2 – *Significant preference* between pairs of MOSAs

2 SUMMARY OF THE RESULTS

RQ1. We introduce the *preference count* $PC(A, Q)$ as the percentage of times Q prefers MOSA A when compared to another MOSA in any problem (*Step 4* in Fig. 1). Fig. 2 reports, for each QI, the preference counts for all the MOSAs (in the decreasing order).

We can observe that some QIs have strong preferences for some MOSAs. For example, GD prefers SPEA2 the most (77.78%). Some QIs also have low preferences for particular MOSAs as, e.g., GS for PAES (5.56%). There are also QIs that don't have particular preferences for specific MOSAs as, for example, C has the low preference counts for all the MOSAs. Finally, we can also notice that some MOSAs (i.e., NSGA-II and SPEA2) are preferred by most of the QIs, probably reflecting their inherent quality.

RQ2. We compute the *preference count per problem* $PC(A, Q, P)$ as the percentage of times that A is preferred by Q (when compared to another MOSA) for problem P (*Step 5* in Fig. 1). Fig. 3 reports, for each QI Q and each MOSA A, the distribution of the metric across the search problems. We observe that for MOSAs that have a high *preference count* (see Fig. 2), the *preference count per problem* varies significantly across the problems. Instead, for MOSAs that are usually not preferred, the variance is low. We can conclude that the problem characteristics influence the effectiveness of a MOSA A, and so a QI Q may prefer a MOSA A only on some problems.

We also checked whether a MOSA A is *significantly preferred* over another MOSA B by a QI QI (across problems). To do this, we applied the Mann-Whitney U test and the $\hat{A}_{12}$ statistics. Fig. 4 reports a representation of the significant preference relation for each QI. An arrow from MOSA A to MOSA B indicates a *significant preference* of A over B. Some MOSAs are always significantly preferred over others (e.g., NSGA-II and SPEA2 over PAES). As expected, the most preferred MOSAs (see Fig. 2) are also those significantly preferred. Some MOSAs are worst than some MOSAs but also better than some others (e.g., MOCELL in HV, EP, and GS): they are MOSAs that can produce good solutions, although not optimal ones.

3 CONCLUSIONS

In [2], we did an empirical evaluation for assessing the preferences of QIs for MOSAs. Although the study was for SBSE problems, we believe it is of interest for all domains in which MOSAs are used.

ACKNOWLEDGMENTS

This work is supported by the Co-evolver project (No. 286898/F20) funded by the Research Council of Norway, National Natural Science Foundation of China under Grant No. 61872182. Paolo Arcaini is supported by ERATO HASUO Metamathematics for Systems Design Project (No. JPMJER1603), JST.

REFERENCES

[1] Shaukat Ali, Paolo Arcaini, Dipesh Pradhan, Safdar Aqeel Safdar, and Tao Yue. 2020. Quality Indicators in Search-Based Software Engineering: An Empirical Evaluation. *ACM Trans. Softw. Eng. Methodol.* 29, 2, Article 10 (March 2020), 29 pages. https://doi.org/10.1145/3375636

[2] Shaukat Ali, Paolo Arcaini, and Tao Yue. 2020. Do Quality Indicators Prefer Particular Multi-Objective Search Algorithms in Search-Based Software Engineering? In *Search-Based Software Engineering - 12th International Symposium, SSBSE 2020, Bari, Italy, October 7-8, 2020, Proceedings*. Lecture Notes in Computer Science, Vol. 12420. Springer International Publishing, Cham, 25–41. https://doi.org/10.1007/978-3-030-59762-7_3