

# JavaPermutationTools: A Java Library of Permutation Distance Metrics

Vincent A. Cicirello<sup>1</sup>

<sup>1</sup> Computer Science, School of Business, Stockton University, Galloway, NJ 08205

DOI: [10.21105/joss.00887](https://doi.org/10.21105/joss.00887)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

**Submitted:** 08 August 2018

**Published:** 01 September 2018

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

## Summary

Permutations are used to represent a wide variety of ordered data. For example, a permutation may represent an individual's preferences (or ranking) of a collection of products such as books or music. Or perhaps a permutation may represent a route for delivering a set of packages. Permutations can also represent one-to-one mappings between sets (e.g., instructors to courses at a fixed time). There are applications where measuring the distance between a pair of permutations is necessary. For example, a recommender system may want to compare the similarity of two individuals' preferences for music, perhaps to make song recommendations. Depending upon the application, the permutation features most important to distance calculation may be one or more of the absolute positions of the elements (e.g., for one-to-one mappings), the adjacency of elements (e.g., the earlier routing example), or the general precedence of pairs of elements (e.g., music preference example). Thus, it should be no surprise that there exists a large variety of permutation distance metrics in the research literature.

The motivation and origin of this library of permutation distance metrics is our research on evolutionary computation, and other metaheuristic approaches, for solving permutation optimization problems. A permutation optimization problem is a problem where solutions are represented by permutations of the elements of some set, and where the objective is to maximize or minimize some function. For example, a solution to a traveling salesperson problem (TSP) is a permutation of the set of cities, and the objective is to find the permutation that corresponds to the tour of the cities of minimal cost.

There are at least two applications of permutation distance that come from evolutionary computation. The first application is to maintain population diversity. The various forms of evolutionary computation, such as genetic algorithms, genetic programming, etc, solve problems through simulated evolutionary processes (Mitchell 1998). They maintain a population of solutions to the problem at hand, and this population evolves over many generations using operators such as mutation, crossover, etc. Just as it is in natural evolution, a diverse gene pool is important to evolutionary computation. In later generations, if variation among the members of the population declines, then the evolutionary search can stagnate. Approaches to population management (Sevaux and Sörensen 2005), such as in scatter search (Campos, Laguna, and Martí 2005; Martí, Laguna, and Campos 2005), attempt to maintain diversity among the population of solutions. They require a means of measuring distance between population members.

The second application of permutation distance for evolutionary computation is that of search landscape analysis. A fitness (or search) landscape (Mitchell 1998) is the space of possible solutions to an optimization problem spatially organized on a landscape with similar solutions as neighbors, and where elevation corresponds to fitness (or solution quality). Peaks (for a maximization problem) and valleys (for a minimization problem)

correspond to locally optimal solutions. The optimization problem is to find an optimal point on that landscape. Search landscape analysis is the term covering the theoretical and practical techniques for studying what characteristics of a problem make it hard, how different search operators affect fitness landscape topology, among others. There is a wide variety of work on fitness landscape analysis, including for permutation landscapes (Cicirello 2016, 2014; Hernando, Mendiburu, and Lozano 2016; Tayarani-N and Prugel-Bennett 2014; Cicirello and Cernera 2013; Sörensen 2007; Schiavinotto and Stützle 2007). Some fitness landscape analysis techniques use fitness distance correlation (FDC) (Jones and Forrest 1995), the Pearson correlation coefficient of the fitness of a solution vs its distance to the nearest optimal solution. The search landscape calculus (Cicirello 2016) is a fitness landscape analysis technique that examines local rates of change of fitness. These as well as others require distance metrics for the type of structure you are optimizing.

During the course of our research on fitness landscape analysis for permutation optimization problems (Cicirello 2016, 2014; Cicirello and Cernera 2013), we developed a Java library of permutation distance metrics. Most of the distance metrics in the literature are described mathematically with no source code available. Thus, our library offers convenient access to efficient implementations of a variety of distance metrics that share a common programmatic interface. The following table summarizes the distances available in this initial release of the library, along with their runtime ( $n$  is permutation length), whether they satisfy the requirements of a metric, and one or more key citations.

Distance	Runtime	Metric?	Citations
acyclic edge distance	$O(n)$	pseudo	(Ronald 1997, 1995)
cyclic edge distance	$O(n)$	pseudo	(Ronald 1997, 1995)
cyclic r-type distance	$O(n)$	pseudo	(Cicirello 2016)
deviation distance	$O(n)$	yes	(Cicirello 2016; Campos, Laguna, and Martí 2005)
deviation distance normalized	$O(n)$	yes	(Ronald 1998; Sörensen 2007)
edit distance	$O(n^2)$	yes	(Wagner and Fischer 1974; Sörensen 2007)
exact match distance	$O(n)$	yes	(Ronald 1998)
interchange distance	$O(n)$	yes	(Cicirello and Cernera 2013)
Kendall tau distance	$O(n \lg n)$	yes	(Kendall 1938; Meilă and Bao 2010; Fagin, Kumar, and Sivakumar 2003)
Lee distance	$O(n)$	yes	(Lee 1958)
r-type distance	$O(n)$	yes	(Campos, Laguna, and Martí 2005; Martí, Laguna, and Campos 2005)
reinsertion distance	$O(n \lg n)$	yes	(Cicirello 2016; Cicirello and Cernera 2013)
reversal distance	Init: $O(n!n^3)$ Compute: $O(n^2)$	yes	(Cicirello 2016; Caprara 1997)

Distance	Runtime	Metric?	Citations
squared deviation distance	$O(n)$	yes	(Sevaux and Sörensen 2005)

The library also provides distance metrics on sequences (Strings and arrays of various types), where unlike a permutation, a sequence may contain multiple copies of the same element. The following table lists the metrics on sequences that are provided ( $n \leq m$  are the lengths of the compared sequences).

Distance	Runtime	Metric?	Citations
edit distance	$O(n * m)$	yes	(Wagner and Fischer 1974)
exact match distance	$O(n)$	yes	(Ronald 1998)
Kendall tau distance	$O(n \lg n)$	yes	(Kendall 1938)
longest common subsequence distance	$O(n * m)$	yes	(Wagner and Fischer 1974)

The source repository (<https://github.com/cicirello/JavaPermutationTools>) contains source code of the library, programs that provide example usage of key functionality, as well as programs that reproduce results from papers that have used the library. API documentation is hosted on the web (<https://jpt.cicirello.org/>).

## References

- Campos, Vicente, Manuel Laguna, and Rafael Martí. 2005. “Context-Independent Scatter and Tabu Search for Permutation Problems.” *INFORMS Journal on Computing* 17 (1):111–22. <https://doi.org/10.1287/ijoc.1030.0057>.
- Caprara, Alberto. 1997. “Sorting by Reversals Is Difficult.” In *Proceedings of the First Annual International Conference on Computational Molecular Biology*, 75–83. ACM.
- Cicirello, Vincent A. 2014. “On the Effects of Window-Limits on the Distance Profiles of Permutation Neighborhood Operators.” In *Proceedings of the International Conference on Bioinspired Information and Communications Technologies*, 28–35. <https://doi.org/10.4108/icst.bict.2014.257872>.
- . 2016. “The Permutation in a Haystack Problem and the Calculus of Search Landscapes.” *IEEE Transactions on Evolutionary Computation* 20 (3):434–46. <https://doi.org/10.1109/TEVC.2015.2477284>.
- Cicirello, Vincent A., and Robert Cernera. 2013. “Profiling the Distance Characteristics of Mutation Operators for Permutation-Based Genetic Algorithms.” In *Proceedings of the 26th International Conference of the Florida Artificial Intelligence Research Society*, 46–51. AAAI Press.
- Fagin, Ronald, Ravi Kumar, and D. Sivakumar. 2003. “Comparing Top K Lists.” *SIAM Journal on Discrete Mathematics* 17 (1):134–60.
- Hernando, Leticia, Alexander Mendiburu, and Jose A. Lozano. 2016. “A Tunable Generator of Instances of Permutation-Based Combinatorial Optimization Problems.” *IEEE Transactions on Evolutionary Computation* 20 (2):165–79. <https://doi.org/10.1109/TEVC.2015.2433680>.
- Jones, Terry, and Stephanie Forrest. 1995. “Fitness Distance Correlation as a Measure of Problem Difficulty for Genetic Algorithms.” In *Proceedings of the 6th International Conference on Genetic Algorithms*, 184–92. Morgan Kaufmann.

- Kendall, M. G. 1938. “A New Measure of Rank Correlation.” *Biometrika* 30 (1/2):81–93.
- Lee, C. 1958. “Some Properties of Nonbinary Error-Correcting Codes.” *IRE Transactions on Information Theory* 4 (2):77–82. <https://doi.org/10.1109/TIT.1958.1057446>.
- Martí, Rafael, Manuel Laguna, and Vicente Campos. 2005. “Scatter Search Vs. Genetic Algorithms: An Experimental Evaluation with Permutation Problems.” In *Metaheuristic Optimization via Memory and Evolution*, 263–82. Springer.
- Meilă, Marina, and Le Bao. 2010. “An Exponential Model for Infinite Rankings.” *Journal of Machine Learning Research* 11:3481–3518.
- Mitchell, Melanie. 1998. *An Introduction to Genetic Algorithms*. MIT Press.
- Ronald, Simon. 1995. “Finding Multiple Solutions with an Evolutionary Algorithm.” In *Proceedings of the Ieee Congress on Evolutionary Computation*, 641–46. IEEE Press.
- . 1997. “Distance Functions for Order-Based Encodings.” In *Proceedings of the Ieee Congress on Evolutionary Computation*, 49–54. IEEE Press.
- . 1998. “More Distance Functions for Order-Based Encodings.” In *Proceedings of the Ieee Congress on Evolutionary Computation*, 558–63. IEEE Press.
- Schiavinotto, Tommaso, and Thomas Stützle. 2007. “A Review of Metrics on Permutations for Search Landscape Analysis.” *Computers & Operations Research* 34 (10):3143–53. <https://doi.org/10.1016/j.cor.2005.11.022>.
- Sevaux, Marc, and Kenneth Sörensen. 2005. “Permutation Distance Measures for Memetic Algorithms with Population Management.” In *Proceedings of the Metaheuristics International Conference (Mic2005)*, 832–38.
- Sörensen, Kenneth. 2007. “Distance Measures Based on the Edit Distance for Permutation-Type Representations.” *Journal of Heuristics* 13 (1). Kluwer Academic Publishers:35–47. <https://doi.org/10.1007/s10732-006-9001-3>.
- Tayarani-N, Mohammad-H., and Adam Prugel-Bennett. 2014. “On the Landscape of Combinatorial Optimization Problems.” *IEEE Transactions on Evolutionary Computation* 18 (3):420–34. <https://doi.org/10.1109/TEVC.2013.2281502>.
- Wagner, Robert A., and Michael J. Fischer. 1974. “The String-to-String Correction Problem.” *Journal of the ACM* 21 (1):168–73.