2014

Data Mining Based Hybridization of Meta-RaPS

Fatemah Al-Duoli
Old Dominion University

Ghaith Rabadi
Old Dominion University

Original Publication Citation

Al-Duoli, F., & Rabadi, G. (2014). Data mining based hybridization of meta-RaPS. *Procedia Computer Science, 36,* 301-307. doi:10.1016/j.procs.2014.09.097

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 36 (2014) 301 – 307

Procedia Computer Science

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014- Philadelphia, PA

Data Mining Based Hybridization of Meta-RaPS

Fatemah Al-Duoli[a], Ghaith Rabadi[a]*

[a]*Department of Engineering Management and Systems Engineering, Old Dominion University, Norfolk, Virginia 23529, USA*

Abstract

Though metaheuristics have been frequently employed to improve the performance of data mining algorithms, the opposite is not true. This paper discusses the process of employing a data mining algorithm to improve the performance of a metaheuristic algorithm. The targeted algorithms to be hybridized are the Meta-heuristic for Randomized Priority Search (Meta-RaPS) and an algorithm used to create an Inductive Decision Tree. This hybridization focuses on using a decision tree to perform on-line tuning of the parameters in Meta-RaPS. The process makes use of the information collected during the iterative construction and improvement phases Meta-RaPS performs. The data mining algorithm is used to find a favorable parameter using the knowledge gained from previous Meta-RaPS iterations. This knowledge is then used in future Meta-RaPS iterations. The proposed concept is applied to benchmark instances of the Vehicle Routing Problem.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: Vehicle Routing Problem; Metaheuristics; Meta-RaPS; Data mining; Supervised Learning

1. Introduction

Hybrid metaheuristics represent a class of algorithms that combine metaheuristics with each other or with other applicable algorithms. The resulting algorithms aim at taking advantage of the strengths of the hybridized algorithms while managing to keep the level of complexity down. For various real-life combinatorial optimization problems,

* Corresponding author. Tel.: +0-757-683-4918; fax: +0-757-683-5640.
E-mail address: grabadi@odu.edu.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2014.09.097

hybrid metaheuristic algorithms were able to find better results. The formulization of the field of hybrid metaheuristics started in the early 2000s [1]. In addition to combining metaheuristics with other metaheuristics, hybrid metaheuristics include integration with operational research algorithms (e.g. exact mathematical programming algorithms) and with machine learning and data mining techniques. Collaborations between metaheuristic algorithms and data mining algorithms are usually done with the objective of improving the data mining algorithms. In this paper, the objective is to demonstrate the integration of a metaheuristic algorithm called Meta-RaPS with an Inductive Decision Tree (IDT), a classification data mining algorithm. The following sections describe the main elements behind the proposed concept, which are Meta-RaPS, IDT algorithm, and the hybrid Meta-RaPS. The last two sections discuss the combinatorial optimization problem solved by the proposed algorithm and the computational experiments conducted to demonstrate the effectiveness of the hybrid Meta-RaPS.

2. Meta-RaPS

Meta-RaPS is a metaheuristic algorithm initially introduced as a modified form of the COMSOAL (Computer Method of Sequencing Operations for Assembly Lines) heuristic [2]. Moraga et al. [3] formally defined Meta-RaPS as "generic, high-level search procedures that introduce randomness to a construction heuristic as a device to avoid getting trapped at a local optimal solution." Meta-RaPS uses four parameters while solving a problem: priority percentage (p%), restriction percentage (r%), improvement percentage (i%), and number of iterations (I). These parameters are used during the construction and improvement phases.

The construction phase aims to incrementally build a single solution by identifying feasible components or moves, which make it a non-population-based metaheuristic. During this phase, Meta-RaPS employs a greedy heuristic to identify a set of feasible next moves. The set of feasible next moves are prioritized. The move with the best priority is selected p% of the time. During the remaining times, (100 - p%), the next feasible move is selected randomly from a restricted list and added to the solution being built. The restricted list contains a set of next moves with priority values that are within the best r% of the best priority move. The construction process is repeated until a solution is complete. Following the construction phase, the execution of the improvement phase is evaluated. The constructed solution may be improved if its value is within i% of the best constructed solution found so far. This is done to avoid wasting improvement time on bad solutions. The improvement phase typically involves using a local search heuristic. Both construction and improvement phases are repeated until the termination criteria are met, which in this case is reaching the maximum number of iterations "I". No knowledge is transferred from one iteration to the next making Meta-RaPS a multi-start metaheuristic. The best solution produced from all iterations is reported as the output of Meta-RaPS.

Since its introduction, Meta-RaPS has been successfully used to solve several combinatorial problems including the Resource Constrained Project Scheduling Problem [4], the Vehicle Routing Problem [5], the Traveling Salesman Problem [6], the 0-1 Multidimensional Knapsack Problem [7], the Unrelated Parallel Machine Scheduling Problem with Setup Times [8], the Early/Tardy Single Machine Scheduling Problem [9], the Spatial Scheduling Problem with Release Times [10], and the Aerial Refueling Scheduling Problem with Due Data to Deadline Windows and Release Time [11].

3. Inductive Decision Tree (IDT)

Meta-RaPS is hybridized with a supervised data mining algorithm, which performs classification. Classification is a learning function that maps or classifies discrete input data into one of several predefined classes [12]. Examples of classification algorithms include Inductive Decision Trees (IDTs), neural networks, and support vector machines. The IDT generates classification rules without priori knowledge. This is achieved after completing a training phase, during which Training Examples (TE) are used as input. TE for supervised data mining algorithms consist of labeled data, which can be viewed as paired data: a vector of system attributes and the desired label. A desired label can be described as the desired outcome of learning or a class [13].

Hybridizing Meta-RaPS with an IDT involves making use of decision tree construction algorithms. The algorithm used here is the ID3 algorithm introduced by Quinlan [14]. ID3 constructs an IDT recursively by dividing the labeled TE based on the values of an attribute with the most information gain. The algorithm uses the concept of information entropy, which measures the ambiguity found when using an attribute to classify instances [15]. Thus, the best attribute is an attribute with the least entropy, which results in the most information gain. The attribute is selected to be the root node of the tree.

4. Hybrid Meta-RaPS

The goal of this effort is to identify an approach to combine the IDT data mining algorithm with Meta-RaPS. Talbi [16] introduced a framework to organize the various hybridization approaches between metaheuristics and data mining algorithms. This framework describes the time (online vs. offline), the objective (efficiency vs. effectiveness), and the objective behind the hybridization (e.g. solution encoding, initialization, parameter tuning, etc.). A survey mapping the aforementioned data mining hybridization categories to existing publications can be found in [17]. In this survey, the majority of the metaheuristics hybridized with data mining algorithms are population-based metaheuristics. This is an expected pattern as diverse population-based metaheuristics can take advantage of the knowledge-based guidance data mining algorithms provides. Very few non-population-based metaheuristics were surveyed; however none of the listed algorithms involved Meta-RaPS. Additionally, data mining algorithms were utilized to improve the parameter tuning in Particle Swarm Optimization (PSO) metaheuristic [18]. Regression models were used to mine the data generated by particles during typical PSO iterations and influence the parameter values of future PSO iterations.

In this paper, the proposed concept utilizes IDT to perform online Meta-RaPS parameters tuning based on the latest Meta-RaPS performance. The hybrid Meta-RaPS (Figure 1) operates in the same fashion as the original Meta-RaPS by using no prior knowledge and performing as a multi-start metaheuristic.

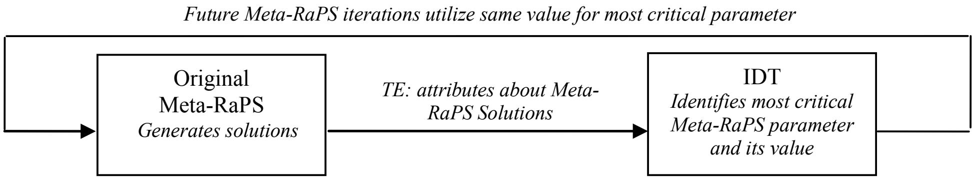

Figure 1 Hybrid Meta-RaPS Framework

After each iteration, a set of attributes about the solutions produced by Meta-RaPS is gathered. The attributes selected are: p% value, r% value, i% value, and the category (or label) of the solution. Unlike the original Meta-RaPS, the values of p%, r%, and i% are randomly selected within predefined high-level and a low-level ranges. With this mechanism, the values of parameters are not known in advance. This variability represents an opportunity to capture solution-specific information that is usually lost in the next Meta-RaPS iteration. The last attribute captured represent a label given to the solution produced. A solution can be labeled as a "good" solution if its performance is better than the previous group of solutions. The performance of the previous group of solutions is accounted for using a moving average of a predefined window size (avg%). A solution with an output that is worse than the previous solutions is labeled "bad". After executing TE% iterations, ID3 is executed to build an IDT using the training examples collected so far.

The root node of the tree produced by ID3 can either be p%, r%, or i% while the leaf nodes present a solution label: "good" or "bad." Though the tree structure will show all the parameters and link them to the leaf nodes (both "good" and "bad"), the relationship of interest is between the root node and a leaf node labeled "good." This relationship represents the parameter (root node) that is strongly associated with the "good" leaf node (Figure 2). The knowledge represented by this relationship is passed to future Meta-RaPS iteration by fixing the value of the root node parameter. During the next TE% Meta-RaPS iterations, additional solutions are produced by Meta-RaPS. These new solutions are added to the training set. At the end of the TE% iterations, the IDT tree is built again and a possible new root node is identified and a different possible new root node parameter is passed to the next Meta-RaPS iterations.

Though knowledge is passed after a tree is built, this knowledge may be ineffective in producing better solutions.

To avoid continuing to use inapplicable knowledge, a kick-out mechanism is incorporated to the algorithm. After fixing the value of the root node parameter identified by the tree for avg% iterations, the performance of Meta-RaPS solutions is evaluated. If a solution is improving relative to the performance of the moving average of the last avg% solutions, then continue to use the knowledge. Else, stop using the knowledge and revert back to randomly selecting values for Meta-RaPS' parameters.

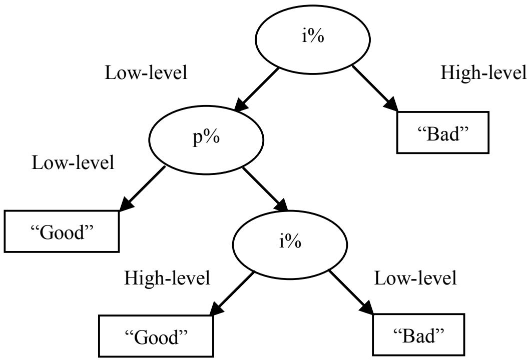

Figure 2 Proposed IDT using Meta-RaPS parameters as attributes

5. The Capacitated Vehicle Routing Problem (CVRP)

To demonstrate the effectiveness of the proposed concept, the hybrid Meta-RaPS is used to solve the CVRP. This problem describes the layout of a system with a central depot and various destinations or customers. The problem was initially introduced by Dantzig and Ramser [19] as the Truck Dispatching Problem. The problem consists of a central depot, where products are stored, and a fleet of trucks to deliver the products to geographically dispersed customers. The goal is to find optimized routes that minimize the total distance traveled while satisfying the vehicle capacity constraints and serving all customers. The input for the CVRP consists of the coordinates of each customer to be served. The coordinates are used to calculate the distances between the depot and each customer. Therefore, the default number of routes in a CVRP is equal to the number of customers in the problem. Each route can be described as [0, *i*, 0], where the depot is usually located at vertex 0 and *i* represents customer *i*. The symmetrical CVRP is used to illustrate the proposed concept. The VRP is a combinatorial problem as it is considered a special version of the TSP if the traveling salesman is constrained by returning to the central depot after visiting each city. Similar to the TSP, the VRP is an NP-hard problem [20]. To solve the CVRP, the underlying heuristics to be used in Meta-RaPS are detailed in the following sub-sections:

5.1. Construction Phase

The construction phase uses a greedy heuristic to incrementally build a solution. The heuristic selected is the Clarke and Wright (C&W) savings algorithm [21]. This algorithm solves the CVRP by merging routes in an effort to maximize the savings by reducing the total distance traveled. The concept of savings is achieved in the following equation:

$$S_{ij} = d_{0i} + d_{0j} - d_{ij} \geq 0 \qquad \forall arc(i, j) \tag{1}$$

The distance between two destinations is referred to as d. Thus, the distance between customer i and the depot is d_{0i}, the distance between customers i and j is d_{ij}. S_{ij} represents the distance saved by directly connecting customers i and j. To use the C&W algorithm to solve a CVRP instance, the distances between each customer must be calculated. The distances can be saved in a distance matrix. The distances are then used to create a savings matrix, where equation (1) is used to calculate the savings if an arc were to connect two customers directly. The savings matrix is then sorted in a descending order, where the arc with the most savings is first. The outcome of sorting will be referred to as the sorted savings list.

Every arc in the sorted savings list represents a possibility to merge two CVRP routes by connecting two customers together. However, to merge two routes using an arc from the sorted savings list, the arc is examined against three C&W conditions. The first condition checks if the customers in the arc belong to two separate existing routes. The second condition verifies that the truck capacity is not exceeded if two routes are merged using the arc under consideration. The last condition checks if both customers in the arc are either first or last in their existing routes. After evaluating every arc in the sorted savings list, several of the initial routes are expected to be merged.

When using the C&W heuristic in the construction phase of Meta-RaPS, the initial steps (forming the distance matrix, forming the savings matrix, and creating the sorted savings list) are performed. The sorted savings list represents priorities. The construction phase will loop over every arc in the sorted savings list. If a random value ≤ *p%*, then use the arc with the highest priority, which is the first arc in the list. Else, create a restricted list using the *r%* and randomly select an arc. The selected arc is then evaluated against C&W conditions. If the arc passes all three conditions, then two routes are merged. This process is repeated until all the arcs have been evaluated. The total distance of the final routes is then calculated.

5.2. Improvement Phase

If the total calculated distance of the constructed solution is within *i%* of the best constructed solution found so far, then apply a local search heuristic. The local search heuristic chosen here is pair-wise exchange [22]. The heuristic takes every route in the constructed solution, exchanges each adjacent pairs of nodes, and recalculates the distance. If the calculated distance is shorter, then improvement is attained and the improved solution is reported.

6. Computational Study

The study consists of using both the original Meta-RaPS and the proposed hybrid Meta-RaPS to solve the 14 Christofides, Mingozzi, and Toth [23] (CMT) CVRP input instances. The input problems vary in size from 50 to 199 customers. Half of the input instances have truck capacity as the only constrain while the other half represents the same problems, but with total route time as an additional constraint. Solving the problems using the original Meta-RaPS establishes pre-hybridization baseline. To tune the parameters used in the original Meta-RaPS, a two-level factorial design was followed. Table 1 shows the two-levels selected for each Meta-RaPS parameter.

Table 1 Two-level Factorial Design Values for Meta-RaPS Parameters

Levels	Original Meta-RaPS Parameters				Hybrid Meta-RaPS Parameters	
	p%	r%	i%	I	TE%	Avg%
Low	40	10	20	500	5	4
High	80	70	60	1000	10	8

To identify the appropriate design factor levels, five randomly selected problems were run for 20 times with every combination of design factors. The combination of parameters that led to best solution quality (least deviation from optimal value) is 80% for p%, 10% for r%, 60% for i%, and 1000 iterations for I.

These selected values are also used in the hybrid Meta-RaPS to identify high and low levels for the parameters. For example, a high-level p% value is a randomly generated value that is greater than 80. The inverse is true for a low-level p% value. Additionally, hybrid Meta-RaPS introduces two new parameters (TE% an Avg%). The same two-level factorial design process was followed to identify their values. This led to selecting 5 for TE% and 8 for Avg%.

All input instances were run 20 times. The average deviation from the known optimal values associated with each input problem is reported in Table 2. The last column represents the difference or relative performance, equation 2, that hybrid Meta-RaPS reached compared to the original Meta-RaPS. This shows that the hybrid Meta-RaPS outperformed the original Meta-RaPS in all but 4 instances. The improvement ranges from 1.8 to 96.95%.

$$Diff\,\% = \frac{Hybrid - Original}{Hybrid} * 100 \qquad (2)$$

Table 2 Results of Original and Hybrid Meta-RaPS

Input Problem	Original Meta-RaPS	Hybrid Meta-RaPS	Diff%
vrpnc1	5.580746	3.714091	50.26
vrpnc2	4.127416	3.462221	19.21
vrpnc3	3.371006	4.11224	-21.99
vrpnc4	4.082892	4.286668	-4.99
vrpnc5	5.847576	5.741699	1.844
vrpnc6	3.135661	4.591166	-46.42
vrpnc7	4.971092	4.28175	16.10
vrpnc8	7.598772	7.309541	3.957
vrpnc9	7.247888	6.536675	10.88
vrpnc10	8.547171	7.178763	19.06
vrpnc11	1.18162	0.599948	96.95
vrpnc12	0.921488	0.503471	83.03
vrpnc13	1.329579	1.46228	-9.98
vrpnc14	0.255836	0.216922	17.94

7. Conclusion

The main objectives of this effort are to continue advancing the performance of Meta-RaPS and to demonstrate a data mining hybridization approach. The process of parameter tuning has a big impact on the performance of metaheuristic algorithms. The proposed concept helps in identifying values by making use of the latest attained attributes of the solutions produced by Meta-RaPS. The hybridization approach is a simple, problem-independent way to integrate data mining with metaheuristics. This approach can also be viewed as metaheuristic-independent. Currently, Meta-RaPS' parameters represent the attributes used to build the IDT. The same approach can be employed with a different metaheuristic algorithm, where the associated parameters are used to build the IDT.

References

1. Talbi E. G., "A Taxonomy of Hybrid Metaheuristics", Journal of Heuristics, 8(5), pp. 541–564, 2002.

2. Arcus, A. L., COMSOAL: a computer method of sequencing operations for assembly lines. Int Jr of Production Research, 4, 259–277, 1966.

3. Moraga R.J., DePuy G.W., Whitehouse G.E., "Metaheuristics: A Solution Methodology for Optimization Problems. In: A.B. Badiru (ed.), Handbook of Industrial and Systems Engineering", CRC Press, FL, 2006.

4. DePuy G.W., Whitehouse G.E., "A simple and effective heuristic for the multiple resource allocation problem, International Journal of Production Research", Vol. 32, 424-31, 2001.

5. Moraga R. J., "Meta-RaPS: An Effective Solution Approach for Combinatorial Problems", Ph.D. thesis, University of Central Florida, Orlando, FL, 2002.

6. DePuy G.W., Moraga R.J., Whitehouse G.E. "Meta-RaPS: a simple and effective approach for solving the traveling salesman problem", Transportation Research Part E, Vol. 41, p. 115–130, 2005.

7. Moraga R.J., DePuy G.W., Whitehouse G.E., "Meta-RaPS approach for the 0–1 multidimensional knapsack problem, Computers & Industrial Engineering, Vol. 48, No. 2, 83-96, 2005.

8. Rabadi G., Moraga R.J., Al-Salem A., "Heuristics for the unrelated parallel machine scheduling problem with setup times", Journal of Intelligent Manufacturing, Vol. 17, 85-97, 2006.

9. Hepdogan S., Moraga R.J., DePuy G.W., Whitehouse G.E., "A Meta-RaPS For The Early/Tardy Single Machine Scheduling Problem", International Journal of Production Research, Vol. 47, No. 7, 1717-1732, 2009.

10. Garcia C., Rabadi G., "A Meta-RaPS algorithm for spatial scheduling with release times". Int. J. Planning & Scheduling, Vol. 1, Nos. 1/2, 19-31, 2011.

11. Kaplan S., Rabadi G., "A Simulated Annealing and Meta-RaPS Algorithms for the Aerial Refueling Scheduling Problem with Due Date-to-Deadline Windows and Release Time", Engineering Optimization, V. 45, No. 1, P.67-87, 2013.

12. Fayyad U. M., Piatetsky-Shapiro G., Smyth P., "From data mining to knowledge discovery: an overview, Advances in knowledge discovery and data mining", American Association for Artificial Intelligence, Menlo Park, CA, 1996.

13. Park S.C., Paman N., Shaw M.J., "Adaptive scheduling in dynamic flexible manufacturing systems: a dynamic rule selection approach", IEEE Trans. Robot. Autom, v 13, p 486–502, 1997.

14. Quinlan J. R., "Learning efficient classification procedures and their application to chess end games," in Machine Learning: An ArtiJcial Intelligence Approach, R. S . Michalski, J. G. Carbonell, and T. M. Mitchell, Eds. Palo Alto, CA: Tioga Publishing, pp. 463-482, 1983.

15. Shannon, C. E., "A mathematical theory of communication". Bell System Technical Journal, 27, 379-423, 1948.

16. Talbi E. G., "Metaheuristics, From Design To Implementation", University of Lille, John Wiley & Sons, Inc., New Jersey, 2009.

17. Jourdan L., Dhaenens C., Talbi E.G., "Using data mining techniques to help metaheuristics: A short survey". In Hybrid Metaheuristics (HM'2006), Gran Canaria, Spain, LNCS Vol. 4030, pp. 57–69, 2006.

18. Lessmann S., Caserta M., Arango I. M., "Tuning metaheuristics: A data mining based approach for particle swarm optimization". Expert Syst. Appl. 38(10) pp. 12826-12838, 2011.

19. Dantzig, G.B., Ramser, J.H., "The Truck Dispatching Problem". Management Science 6 (1), pp. 80–91, 1959.

20. Toth P., Vigo D.,"TheVehicle Routing Problem", Monographs on Discrete Mathematics and Applications, SIAM, Philadelphia, 2002.

21. Clarke, G., and Wright, J., "Scheduling of Vehicles from a Central Depot to a Number of Delivery Points," Operations Research, Vol. 12,No. 4, pp. 568-581, 1964.

22. Leopold C., "Arranging program statements for locality on the basis of neighbourhood preferences". Int J of Approximate Reasoning 19, pp. 73-90, 1998.

23. N. Christofides, A. Mingozzi, and P. Toth, "The vehicle routing problem," Combinatorial Optimization,
N. Christofides, A. Mingozzi, P. Toth, and C. Sandi (Editors), Wiley, Chichester, pp. 315–338, 1979.