

# A novel feature-based approach to characterize algorithm performance for the traveling salesperson problem

Olaf Mersmann · Bernd Bischl · Heike Trautmann ·  
Markus Wagner · Jakob Bossek · Frank Neumann

Published online: 28 March 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** Meta-heuristics are frequently used to tackle NP-hard combinatorial optimization problems. With this paper we contribute to the understanding of the success of 2-opt based local search algorithms for solving the traveling salesperson problem (TSP). Although 2-opt is widely used in practice, it is hard to understand its success from a theoretical perspective. We take a statistical approach and examine the features of TSP instances that make the problem either hard or easy to solve. As a measure of problem difficulty for 2-opt we use the approximation ratio that it achieves on a given instance. Our investigations point out important features that make TSP instances hard or easy to be approximated by 2-opt.

**Keywords** TSP · 2-opt · Classification · Feature selection · MARS

**Mathematics Subject Classifications (2010)** 90B06

---

The conference version of this article appeared in the proceedings of the *Learning and Intelligent Optimization Conference (LION)* 2012 [25].

O. Mersmann · B. Bischl · H. Trautmann (✉) · J. Bossek  
Statistics Faculty, TU Dortmund University, 44221 Dortmund, Germany  
e-mail: trautmann@statistik.tu-dortmund.de

O. Mersmann  
e-mail: olafm@statistik.tu-dortmund.de

B. Bischl  
e-mail: bischl@statistik.tu-dortmund.de

J. Bossek  
e-mail: jakob.bossek@tu-dortmund.de

M. Wagner · F. Neumann  
School of Computer Science, The University of Adelaide, Adelaide, SA 5005, Australia  
e-mail: markus.wagner@adelaide.edu.au

F. Neumann  
e-mail: frank.neumann@adelaide.edu.au

## 1 Introduction

For many NP-hard combinatorial optimization problems, meta-heuristic algorithms such as local search [15], simulated annealing [22], evolutionary algorithms [11], and ant colony optimization [9] have produced good results. Despite the numerous applications of meta-heuristics to hard combinatorial optimization problems, it is hard to understand the success of these algorithms from a theoretical point of view.

Strict mathematical investigations, as in the field of the runtime analysis of meta-heuristics, allow one to prove when and why such algorithms are able to solve certain types of problems. This field of research has gained increasing interest during the last decade and there are results for a wide range of meta-heuristic approaches such as simulated annealing [39], evolutionary algorithms [10], and ant colony optimization [28]. We refer the reader to the textbook of Neumann and Witt [29] for a comprehensive presentation of this research area.

The rigorous treatment of these algorithms in a strict mathematical sense is with no doubt desirable, but comes at the expense that one is usually only able to analyze simplified algorithms on only restricted classes of problems. With this paper, we follow a different approach. Our aim is to gain new theoretical insights into the behavior of meta-heuristics by investigating statistical properties of hard and easy instances of a given problem for a given algorithm. This relates to previous work in continuous domains for which the extraction of problem properties that might influence algorithm performance is an important and current focus of research, denoted as exploratory landscape analysis [3, 26]. For our investigations on combinatorial meta-heuristics, we choose one of the most famous NP-hard combinatorial optimization problems, namely the traveling salesperson problem (TSP). Given a set of  $N$  cities and positive distances  $d_{ij}$  to travel from city  $i$  to city  $j$ ,  $1 \leq i, j \leq N$  and  $i \neq j$ , the task is to compute a tour of minimal traveled distance that visits each city exactly once and returns to the origin.

In the general case (also known as the asymmetric TSP), the distances between two cities might even be different, depending on the direction taken. Many subclasses of the TSP can be defined depending on the constraints that the distances between cities have to satisfy. For example, the distances only have to satisfy the triangle inequality in the Metric TSP. The perhaps simplest NP-hard subclass of TSP is the Euclidean TSP where the cities are points in the Euclidean plane and the distances are the Euclidean distances between them. We will focus on the Euclidean TSP. It is well known that there is a polynomial time approximation scheme (PTAS) for this problem [2]. However, this algorithm is very slow, even for modest instance sizes.

A great number of heuristic approaches has been proposed for the TSP. Often local search methods are preferred in practice. The most successful algorithms rely on the well-known 2-opt swap operator, which removes two edges from a current tour and connects the resulting two parts by two other edges such that a different tour is obtained [16]. Despite the success of these algorithms for a wide range of TSP instances, it is still hard to understand 2-opt from a theoretical point of view. In the following we will denote the complete 2-opt heuristic by “2-opt algorithm” or simply “2-opt” whereas a related swap of two edges is denoted as “2-opt swap”.

In the past, theoretical studies regarding 2-opt have investigated the approximation behavior as well as the time to reach a local optimum. Chandra et al. [7] have studied the worst-case approximation ratio that 2-opt achieves for different classes of TSP instances. Furthermore, they investigated the time that a local search algorithm

based on 2-opt needs to reach a locally optimal solution. Englert et al. [12] have shown that there are even instances for the Euclidean TSP where a deterministic local search algorithm based on 2-opt would take exponential time to find a locally optimal solution. Furthermore, they have derived polynomial bounds on the expected number of steps until 2-opt reaches a local optimum for random Euclidean instances and proved that such a local optimum gives a good approximation for the Euclidean TSP. These results also transfer to simple ant colony optimization algorithms as shown in [20]. A parameterized analysis of evolutionary algorithms for the Euclidean TSP using a mutation operator based on 2-opt has been recently carried out in [37]. These results show that evolutionary algorithms are provably successful if the number of cities that lie in the interior of the convex hull of the given set of  $N$  cities is small.

Most previously mentioned investigations have in common that they either investigate the worst local optimum and compare it to a global optimal solution or investigate the worst case time that such an algorithm needs to reach a locally optimal solution. Although these studies provide interesting insights into the structure of TSP instances they do not give much insights into what is actually going on in the application of 2-opt based algorithms. In almost all cases the results obtained by 2-opt are much better than the actual worst-case guarantees given in these papers. This motivates the studies carried out in this paper, which aim to get further insights into the search behavior of 2-opt and to characterize hard and easy TSP instances for 2-opt.

In general, meta-learning is a subfield of machine learning, where learning algorithms are applied to meta-data about experiments. In this article, we take a statistical meta-learning approach to gain new insights into which properties of a TSP instance make it difficult or easy to solve for 2-opt. A general overview about how to measure hardness of instances for combinatorial optimization problems is given in [35]. By analyzing different features of TSP instances and their correlation we point out how they influence the search behavior of local search algorithms based on 2-opt. To generate hard or easy instances for the TSP we use an evolutionary algorithm approach similar to the one of [34]. However, instead of defining hardness by the number of 2-opt steps to reach a local optimum, we define hardness by the approximation ratio that such an algorithm achieves for a given TSP instance compared to the global optimal solution. This is motivated by classical algorithmic studies for the TSP problem in the field of approximation algorithms.

In a preliminary conference version [25] of this article, we have presented our initial approach to predict TSP problem hardness. Having generated instances that lead to a bad or good approximation ratio, the features of these instances are analyzed and classification rules are derived, which predict the type of an instance (easy, hard) based on its feature levels. In addition, instances of moderate difficulty in between the two extreme classes are generated by transforming hard instances into easy instances based on convex combinations of both. We call this procedure “morphing”.

In comparison to [25], we improve and extend our analyses as follows:

1. We will consider several modifications of our initially used EA. First, instances are forced to cover the whole extent of the underlying plane. Second, the rounding procedure is slightly altered and we investigate two different rounding strategies, differing in the sequence of rounding and mutation.

2. An improved point matching strategy compared to [25] ensures that points move as little as possible during the transformation. Systematic changes of the feature levels along this “path” are identified and used for a feature based prediction of the difficulty of a TSP instance for 2-opt-based local search algorithms.
3. Furthermore, we extend our studies to different instance sizes, whereas the analysis in [25] was restricted to a single instance size.
4. We further validate our feature set on TSPLIB instances.

The structure of the rest of this article is as follows. In Section 2, we give an overview about different TSP solvers, features to characterize TSP instances and indicators that reflect the difficulty of an instance for a given solver. Section 3 introduces an evolutionary algorithm for evolving TSP instances that are hard or easy to approximate and carries out a feature based analysis of the hardness of TSP instances. Finally, we finish with concluding remarks and an outlook on further research perspectives in Section 4.

## 2 Local search and the traveling salesperson problem

As mentioned above, local search algorithms are frequently used to tackle the TSP. They iteratively improve the current solution by searching for a better one in its predefined neighborhood. The algorithm stops when there is no better solution in the given neighborhood, or if a certain number of iterations has been reached.

Historically, 2-opt [8] was one of the first successful algorithms to solve larger TSP instances. It is a local search algorithm whose neighborhood is defined by the removal of two edges from the current tour. The resulting two parts of the tour are then reconnected by two other edges to obtain another complete tour. A few years later, this idea was extended to 3-opt [23] where three connections in a tour are first deleted, and then the best possible reconnection of the network is taken as a new solution. Lin and Kernighan [24] extended the underlying idea to more complex neighborhoods by making the number of performed swaps by 2-opt and 3-opt adaptive. Nowadays, variants of these seminal algorithms represent the state-of-the-art in heuristic TSP optimizers.

Among others, memetic algorithms and subpath ejection chain procedures have shown to be competitive alternatives to algorithms based on 2-opt and 3-opt swaps, with hybrid approaches still being investigated today. In the bio-inspired memetic algorithms for the TSP problem (see [27] for an overview) information about subtours is combined to form new tours via so-called “crossover operators”. Additionally, tours are modified via “mutation operators”, to introduce new subtours. The general idea behind the subpath ejection chain procedures is that in a first step a dislocation is created that requires further change. In subsequent steps, the task is to restore the system. It has been shown that the neighborhoods investigated by the ejection chain procedures form supersets of those generated by the Lin-Kernighan heuristic [14].

Contrary to the above-mentioned iterative and heuristic algorithms, Concorde [1] is an exact algorithm that has been successfully applied to TSP instances with up to 85 900 vertices. It follows a branch-and-cut scheme [30], embedding the cutting-plane algorithm within a branch-and-bound search. The branching steps create a search tree, with the original problem as the root node. By traversing the tree it is possible

**Algorithm 1** 2-Opt algorithm**Input:** Graph  $G$ **Output:** Hamiltonian cycle  $\pi$ Choose a  $\pi$ **while** true **do**

best improvement := 0

**for**  $i$  from 0 to number of cities **do**  $x_i := i^{th}$  node in  $\pi$         **for**  $j$  from  $i$  to number of cities **do**             $x_j := j^{th}$  node in  $\pi$             improvement :=  $distance(\{x_i, x_{i+1}\}) + distance(\{x_j, x_{j+1}\})$   
                                 $- distance(\{x_i, x_j\}) - distance(\{x_{i+1}, x_{j+1}\})$             **if** improvement > best improvement **then**                 $\pi := \pi'$                 best improvement :=  $c(\pi')$                 best current :=  $i$                 best other :=  $j$             **end if**        **end for**    **end for**    **if** best improvement > 0 **then**

comment : swap edges and reverse the cities in between

current := best current + 1; other := best other

**while** current  $\leq$  other **do**

increment current; decrement other

            tmp :=  $\pi[current]$              $\pi[current] := \pi[other]$              $\pi[other] := tmp$         **end while**    **else**

break while-loop

**end if****end while****return**  $\pi$ 

to establish that the leafs correspond to a set of subproblems that include every tour for our TSP.

## 2.1 Characterization of TSP instances

In general, the theoretical assessment of problem difficulty of a TSP instance prior to optimization is usually hard if not impossible. Thus, research has focussed on deriving and extracting problem properties, which characterize and relate to the hardness of TSP instances (e.g. [17, 21, 33, 34]). We refer to these properties as features in the following and subsequently provide an overview. Note that features which are based on knowledge of the optimal tour [18, 36] cannot be used to characterize an instance a priori to optimization and are therefore not relevant in the context of this paper.

A natural and considered feature is the number of cities  $N$  of the given TSP instance [17, 21, 33, 34]. In the following, the subset of features introduced in [17, 21, 33, 34] which we incorporated into our study will be detailed. Almost all mentioned features are included. However, due to the lack of details given in [21], some of the discussed features had to be omitted. Furthermore, we present several

new features we feel additionally relevant, i.e. features based on the minimum spanning tree (MST) and the angle between neighboring cities.

The features can be classified into eight groups which are detailed in the following. In total, 47 features are considered.

*Distance Features:* These are based on summary statistics of the edge cost distribution. Here, the lowest, highest, mean and median edge costs are considered. The proportion of edges with distances shorter than the mean distance, the fraction of distinct distances, i.e. different distance levels, and the standard deviation of the distance matrix are included as well. The expected tour length for a random tour, given by the sum of all edge costs multiplied by  $2/(N - 1)$ , completes the list of suitable distance features.

*Mode Features:* Additional features [17] are the number of modes of the edge cost distribution and related features such as the frequency and quantity of the modes and the mean of the modal values. Enhancing the latter approach given in [17] we include a feature for computing the number of modes of the edge cost distribution [26].

*Cluster Features:* GDBSCAN [32] as recommended in [34] is used for clustering where reachability distances of 0.01, 0.05 and 0.1 are chosen. Derived features are the number of clusters and the mean distances to the cluster centroids.

*Nearest Neighbor Distance Features:* Uniformity of an instance is reflected by the minimum, maximum, mean, median, standard deviation and the coefficient of variation of the normalized nearest-neighbor distances (nnd) of each node [33, 34].

*Centroid Features:* The coordinates of the instance centroid together with the minimum, mean and maximum distance of the nodes from the centroid.

*MST Features:* Statistics which characterize the depth and the distances of the minimum spanning tree (MST). The minimum, mean, median, maximum and the standard deviation of the depth and distance values of the MST as well as the sum of the distances on the MST, which we normalize by dividing it by the sum of all pairwise distances.

*Angle Features:* This feature subset comprises statistics regarding the angles between a node and its two nearest neighbor nodes, i.e. the minimum, mean, median, maximum and the respective standard deviation.

*Convex Hull Features:* The area of the convex hull of the instance reflects the “spread” of the instance in the plane. Additionally, the fraction of nodes which define the convex hull is computed.

R [31] source code for the feature computation can be found online.<sup>1</sup> Note that the features have to be normalized appropriately in order to allow for a fair comparison of features across instances of different sizes  $N$ . Ideally, all instances should be normalized to the domain  $[0, 1]^2$  to get rid of scaling issues. However, the latter will not be an issue for our experiments as we explicitly generate instances which fill the  $[0, 1]^2$  plane.

In order to assess the difficulty of a given TSP instance, we will use the approximation ratio that an algorithm achieves for this instance as the optimization accuracy.

<sup>1</sup>[http://www.statistik.tu-dortmund.de/compstat\\_supplementary\\_material.html](http://www.statistik.tu-dortmund.de/compstat_supplementary_material.html)

The approximation ratio is given by the relative error of the tour length resulting from 2-opt compared to the optimal tour length and is a classical measure in the field of approximation algorithms [38]. Based on the approximation ratio that the 2-opt algorithm achieves,<sup>2</sup> we will classify TSP instances either as easy or hard. Afterwards, we will analyze the features of hard and easy instances.

### 3 Analysis of TSP problem difficulty

In this section, we analyze easy and hard TSP instances. We start by describing the evolutionary algorithm that we used to generate these instances. Later on, we characterize them using different features which we calculated and analyzed to determine which features make a TSP instance difficult or easy to solve for 2-opt.

#### 3.1 EA-based generation of easy and hard TSP instances

Our aim is to identify the features that are crucial for predicting the hardness of instances for the 2-opt algorithm. For this a representative set of instances is required which contains instances of varying degrees of difficulty. It turned out that the construction of such a set is a nontrivial task. The generation of instances in a random manner did not provide a sufficient spread with respect to the instance hardness. The same is true for moderately sized instances contained in the TSPLIB, i.e. lower than 1000 nodes, for which, in addition, the number of instances is not high enough to provide an adequate basis for our analysis. Higher instance sizes were excluded due to the large computational effort required for their analysis, especially the computation of the optimal tours.

Therefore, two sets of instances are constructed in the  $[0, 1]^2$ -plane, which focus on reflecting the extreme levels of difficulty. An evolutionary algorithm (EA) is used for this purpose (see Algorithms 2–5 for a description), which can be parameterized such that its aim is to evolve instances that are either as easy or as hard as possible for a given instance size. The approach is conceptually similar to [34] but focusses on approximation quality rather than on the number of swaps as in our view this indicator more adequately reflects problem hardness. In addition, the EA concept consists of a different mutation strategy. Initial studies showed that a second mutation strategy was necessary. “Local mutation” was achieved by adding a small normal perturbation to the location (normalMutation). “Global mutation” was performed by replacing each coordinate of the city with a new uniform random value (uniformMutation). This later step was performed with a very low probability. The two sequential mutation strategies together enable small local as well as global structural changes of the offspring resulting from the crossover operation. All parameters are given at the end of this section.

In contrast to our previous work in [25], a rescaling of the generated instances ensures the complete coverage of  $[0, 1]^2$  in that the minimum and maximum coordinates are placed on the boundary of the instance space (see Fig. 1). Therefore the

---

<sup>2</sup>It is important to note that the ratio will be approximated by an empirical average over different runs of the algorithms with different starting point.

**Algorithm 2** EA for evolving problem easy and hard TSP instances

---

```

function EA(popSize, instSize, numGenerations, time_limit, cells, repetitions, type,
rnd, mutationParameters)
  poolSize  $\leftarrow \lfloor \text{popSize}/2 \rfloor$ 
  for i = 1 to popSize do
    population[i]  $\leftarrow$  RESCALE(RANDOMINSTANCE(instSize))
    population[i]  $\leftarrow$  ROUND(population[i], cells)
    if rnd then
      population[i]  $\leftarrow$  NORMALMUTATION(population[i])
      population[i]  $\leftarrow$  CUTTOBOUNDARY(population[i])
    end if
  end for
  for generation = 1 to numGenerations do
    for k = 1 to popSize do
      fitness[k]  $\leftarrow$  COMPUTEFITNESS(population[k], repetitions)
    end for
    matingPool  $\leftarrow$  CREATEMATINGPOOL(poolSize, population, fitness)
    nextPopulation[1]  $\leftarrow$  population[BESTOF(fitness)] ▷ 1-elitism
    for k = 2 to popSize do
      parent1  $\leftarrow$  RANDELEMENT(matingPool)
      parent2  $\leftarrow$  RANDELEMENT(matingPool)
      offspring  $\leftarrow$  UNIFORMMUTATION(UNIFORMCROSSOVER(parent1, parent2))
      if !rnd then
        offspring  $\leftarrow$  NORMALMUTATION(offspring)
      end if
      offspring  $\leftarrow$  RESCALE(offspring)
      offspring  $\leftarrow$  ROUND(offspring, cells)
      if rnd then
        offspring  $\leftarrow$  NORMALMUTATION(offspring)
        offspring  $\leftarrow$  CUTTOBOUNDARY(offspring)
      end if
    end for
    population  $\leftarrow$  nextPopulation
    if over time_limit then
      return population
    end if
  end for
end function

```

---

**Algorithm 3** Mating pool creation

---

```

function CREATEMATINGPOOL(poolSize, population, fitness)
  for i = 1 to poolSize do
    matingPool[i]
       $\leftarrow$  BETTEROF(RANDELEMENT(population), RANDELEMENT(population))
    end for
  return matingPool
end function

```

---

area covered will not vary as much as in our previous work and instances become comparable in this regard.

In addition, two different rounding schemes are investigated which differ in the sequence of the rounding and normal mutation step. In the first case rounding



**Algorithm 4** Rescale instance

---

```

function RESCALE(instance)
  mins  $\leftarrow$  COLUMN_MINS(instance)
  maxs  $\leftarrow$  COLUMN_MAXS(instance)
   $\delta \leftarrow \text{maxs} - \text{mins}$ 
  scaled  $\leftarrow \emptyset$ 
  for city  $\in$  instance do
    scaled  $\leftarrow \text{scaled} \cup \{(\text{city} - \text{mins})/\delta\}$ 
  end for
  return scaled
end function

```

---

**Algorithm 5** Round instance

---

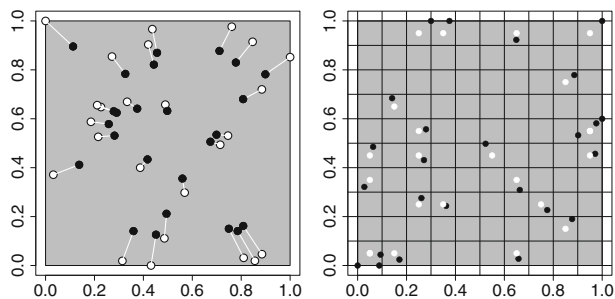
```

function ROUND(instance, cells)
  gridRnd  $\leftarrow$  CREATEGRID(resolution = cells)
  instRnd  $\leftarrow$  FLOOR(instance * cells) / cells
  for i = 1 to instSize do
    instRnd[i,]  $\leftarrow$  SetToCellCenter(instRnd, gridRnd)
  end for
  return instRnd
end function

```

---

**Fig. 1** Examples. *Left*: Rescaling of an instance of size 25. The original instance is reflected by black dots. *Right*: Rounding of an instance of size 25 to grid cell centers. The rounded instance is visualized by white dots



is applied after both mutation steps are complete (*!rnd*, denoted as *nrnd* in the following). After rescaling of the generated instance the points are rounded to force the cities to lie on a predefined grid. This is advantageous for some features which incorporate the proportion of distinct distances. Secondly, we consider normal mutation after the sequence of uniform mutation, rescaling and rounding (*rnd*). This strategy results in instances which resemble a grid structure but also include slight perturbations of the latter as it for instance occurs in circuit board problems. The rounding scheme conceptually differs from rounding to a predefined number of digits as previously considered in that in the current approach the points are rounded to the center of the grid cell they are placed in (see Fig. 1). By this means the probability that cities are located outside the boundary after normal mutation of the rounded points is very low. In these cases points are cut to the boundary of the plane.

The fitness function to be optimized is chosen as the approximation quality of 2-opt, estimated by the arithmetic mean of the tour lengths of a fixed number of 2-opt runs, on a given instance divided by the optimal tour length which is calculated using

Concorde [1]. In general other summary statistics instead of the arithmetic mean could be used as well such as the maximum, minimum or median approximation quality achieved. Note that randomness is only induced by varying the initial tour whereas the 2-opt algorithm is deterministic in always choosing the edge replacement resulting in the highest reduction of the current tour length. Depending on the type of instance that is desired, the `BETTEROF` and `BESTOF` operators are either chosen to minimize or maximize the approximation quality.

We use a 1-elitism strategy such that only the individual with the current best fitness value survives and will be contained in the next population. The rest of the population is obtained by choosing two parents from the mating pool, applying uniform crossover, uniform and normal mutation, rescaling and rounding in the appropriate order and adding the offspring to the population. This procedure is repeated until the population size is reached.

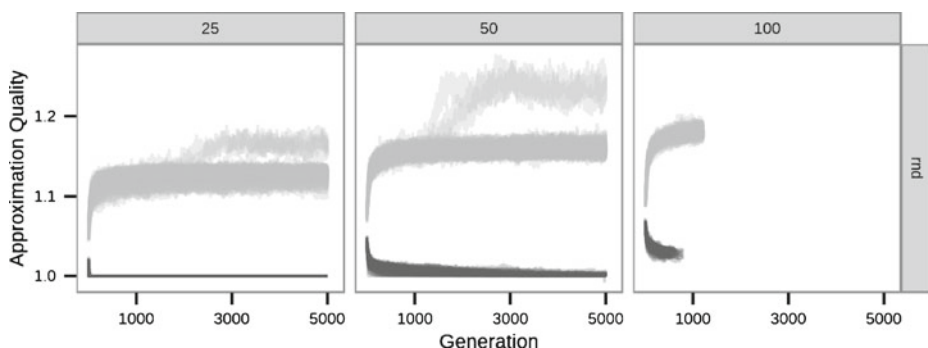
In the experiments, 100 instances each for the two instance classes (easy, hard) with fixed instance sizes of 25, 50 and 100 are generated. The remaining parameters are set as follows: *popSize* = 30, *generations* = 5000, *time\_limit* = 24h, *uniformMutationRate* = 0.001, *normalMutationRate* = 0.01, *cells* = 100, and the standard deviation of the normal distribution used in the *normalMutation* step equals *normalMutationSd* = 0.025. The parameter levels were chosen based on initial experiments. The number of 2-opt repetitions for calculating the approximation quality is set to 500. Again this was a trade-off between evaluation speed and the noise level of the fitness function.

### 3.2 Characterization of the generated instances

Table 1 gives an overview of the instance generation process, i.e. the mean approximation qualities and average number of generations the EA managed to execute within the time limit. For all instance sizes, a sufficiently high performance discrepancy between the two evolved sets of hard and easy instances is generated while the absolute performance difference increases along with the instance size

**Table 1** Overview of generated instances: Mean approximation quality and mean number of EA generations within the time limit

Size	Class	Type	Mean approximation quality	Mean # of generations
25	easy	nrnd	1.00	5000.00
25	easy	rnd	1.00	5000.00
25	hard	nrnd	1.13	5000.00
25	hard	rnd	1.13	5000.00
50	easy	nrnd	1.00	3991.42
50	easy	rnd	1.00	3295.77
50	hard	nrnd	1.16	5000.00
50	hard	rnd	1.16	5000.00
100	easy	nrnd	1.03	454.93
100	easy	rnd	1.03	453.74
100	hard	nrnd	1.18	1194.59
100	hard	rnd	1.18	1204.87

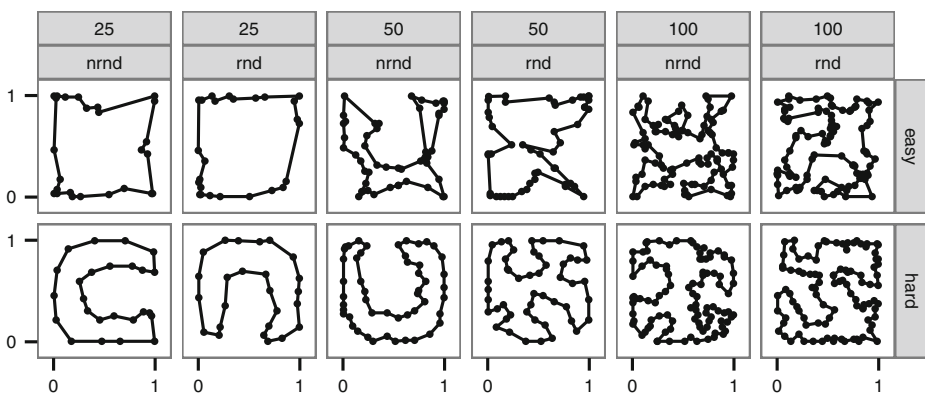


**Fig. 2** EA fitness in the course of the generations for all executed runs

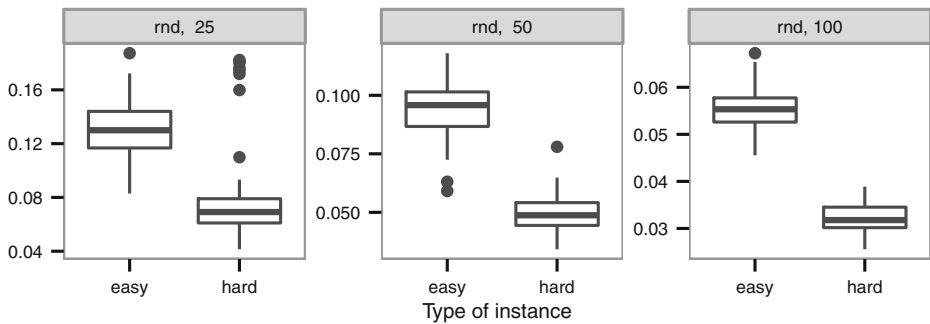
(Fig. 2). For instance sizes 25 and 50, the EA even evolves instances which 2-opt nearly manages to solve to optimality on average.

The average number of generations executed by the EA within the computing budget reflects the rising computational complexity when enlarging the instance size. While for the smallest instance size the maximum number of generations was reached, this amount decreases substantially for the higher instance sizes. Interestingly, the EA requires much higher computation times for generating the easy instances than it is the case for the hard ones. While the number of swaps carried out by 2-opt slightly increases in this situation (see discussion below), particularly problem hardness seems to increase for Concorde as the algorithm takes up much higher computation times than for the instances which are hard to approximate for 2-opt. No differences can be detected concerning the sequence of rounding and mutation.

In Fig. 3 exemplary EA instances of both classes are shown for the different instance sizes and rounding schemes together with the corresponding optimal



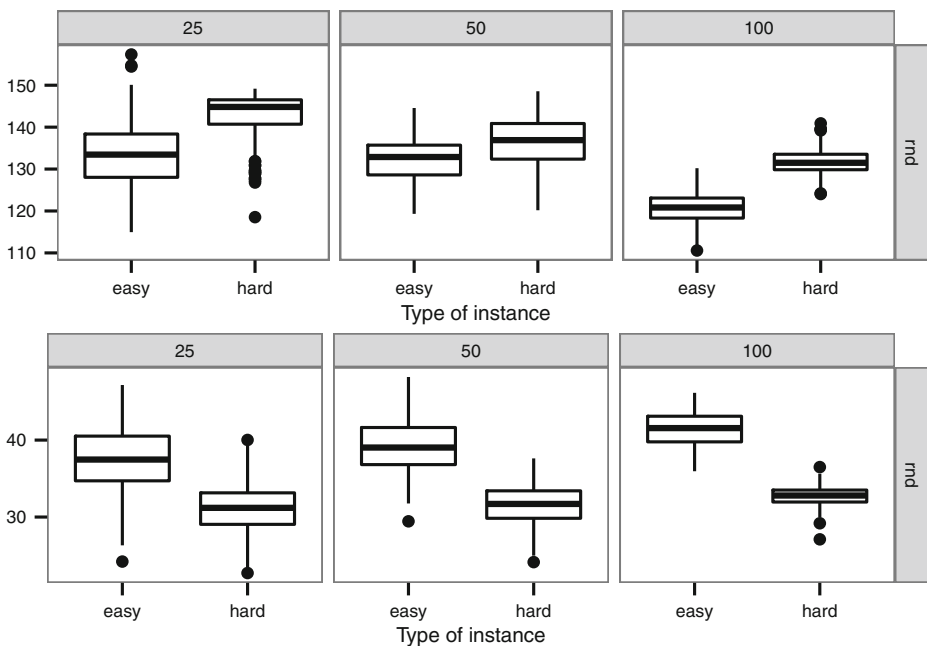
**Fig. 3** Examples of the evolved instances of both types (easy, hard) including the optimal tours computed by Concorde for different instance sizes. Note that there is no discernable difference between the *rnd* and *nrnd* type instances



**Fig. 4** Boxplots of the standard deviations of the tour length legs of the optimal tour, both for the evolved easy and hard instances

tours computed by Concorde. The main visual observations can be summarized as follows:

- The distances of the cities on the optimal tour appear to be more uniform for the hard instances than it is the case for the easy ones. This is supported by Fig. 4 that shows boxplots of the standard deviations of the edge weights on the optimal tour. There we see that respective standard deviations of the easy instances are roughly twice as high than for the hard instances for instance sizes of 100 which increases to a factor of three for the smallest instance size. Related to this context

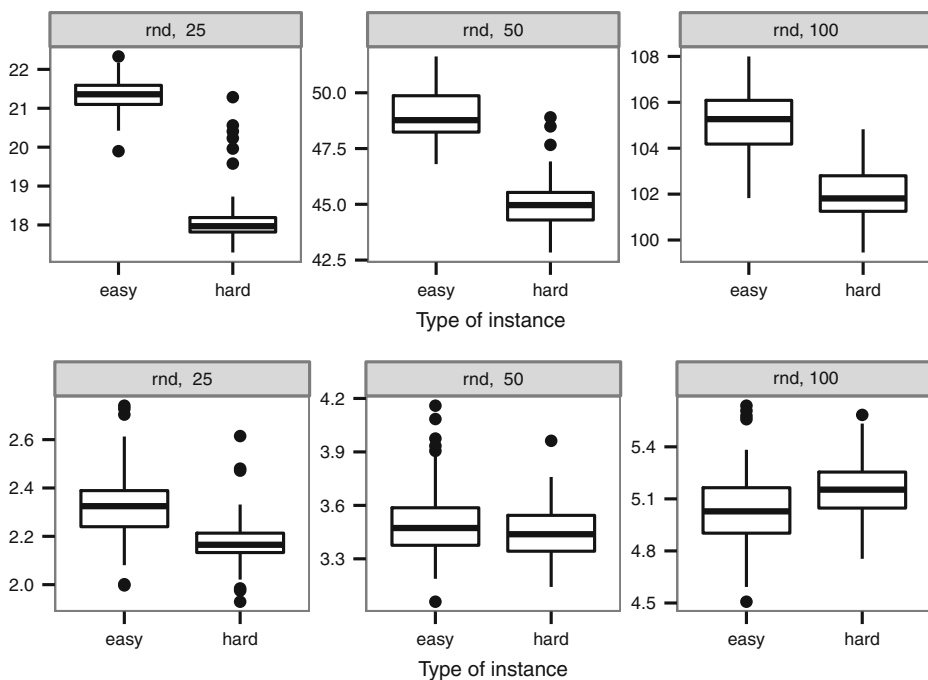


**Fig. 5** Boxplots of the mean (*top*) and standard deviations (*bottom*) of the angle between adjacent cities on the optimal tour

it is observable that the easy instances tend to consist of many small clusters of cities whereas this is not the case for the hard instances up to the same extent.

- Visually, the fraction of highly pointed angles within the class of easy instances exceeds the respective proportion within the class of hard instances. Figure 5 shows mean angles between neighboring points on the optimal tour and the corresponding standard deviations. The mean angles are significantly smaller for the easy instances than within the class of hard ones while the opposite is true for the respective standard deviations.
- The instance shapes for the smallest instance size structurally differ from the respective ones regarding the higher instance sizes. This is especially the case for the easy instances which exhibit an almost circular structure. Consequentially, the area within the convex hull enclosed by the points is much higher for high instance sizes than for smaller ones.
- U-shaped instances are prevalent within the class of generated hard instances while the respective frequency increases with decreasing instance size.
- No significant structural differences between the considered rounding schemes can be observed. Because of this we will focus on the *rnd* results for the rest of the paper. The results for the *nrnd* do not differ. Both sets of instances are available in the supplementary material.

Additionally, by analyzing the mean and standard deviation of the number of swaps executed by 2-opt (see Fig. 6) the choice of choosing the approximation quality



**Fig. 6** Mean number of swaps (*top*) executed by 2-opt and standard deviations (*bottom*) for different instance sizes

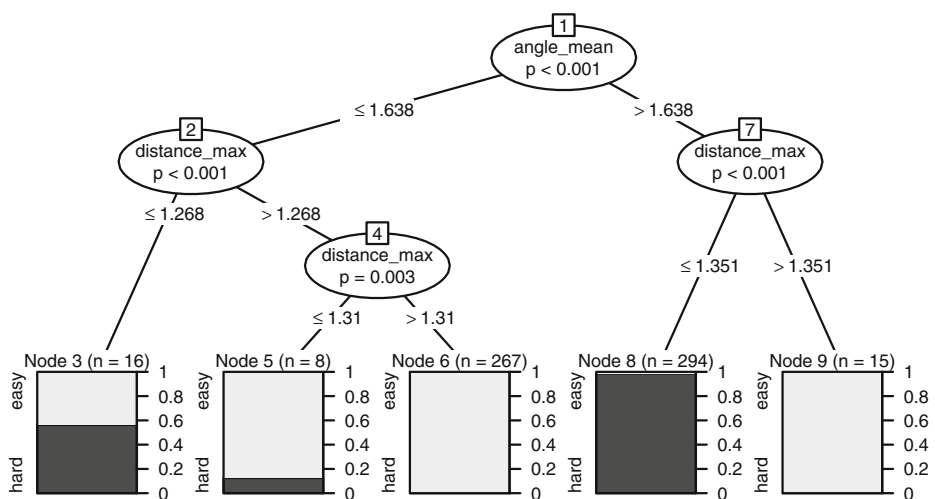
instead of the number of swaps as suggested in [34] as a meaningful performance indicator for 2-opt can be justified. It becomes obvious that there is no positive correlation between the number swaps and problem hardness measured by approximation quality. On the contrary, the opposite trend can be observed. However, it is questionable if this significant difference is really a relevant difference as the absolute deviations in the number of swaps are very small. Therefore, at least for 2-opt, the number of swaps is not an adequate indicator for algorithm performance.

### 3.3 Classification of instance hardness

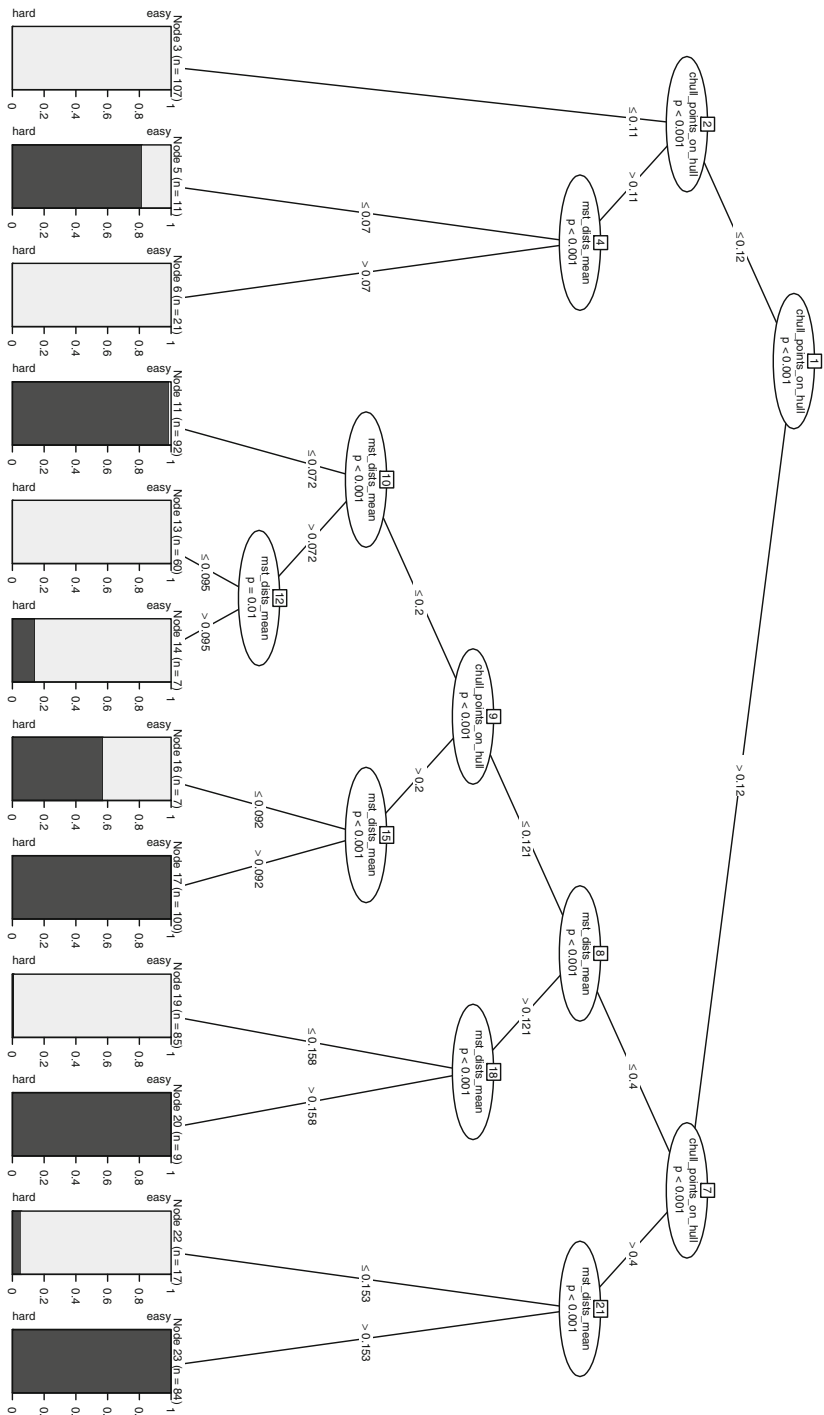
A decision tree [6] is used to differentiate between the two instance classes. Independent from the instance sizes and rounding schemes an almost perfect classification of instances into the two classes based on only two features is possible. Figure 9 visualizes the values of two exemplary feature combinations which can be used for this purpose. It becomes obvious that the classification task is almost trivial as the instance classes could be separated in a quite satisfactory manner with one feature already.

The corresponding classification rules are presented in Figs. 7 and 8 for *rnd* together with the ten-fold cross-validated classification accuracies.

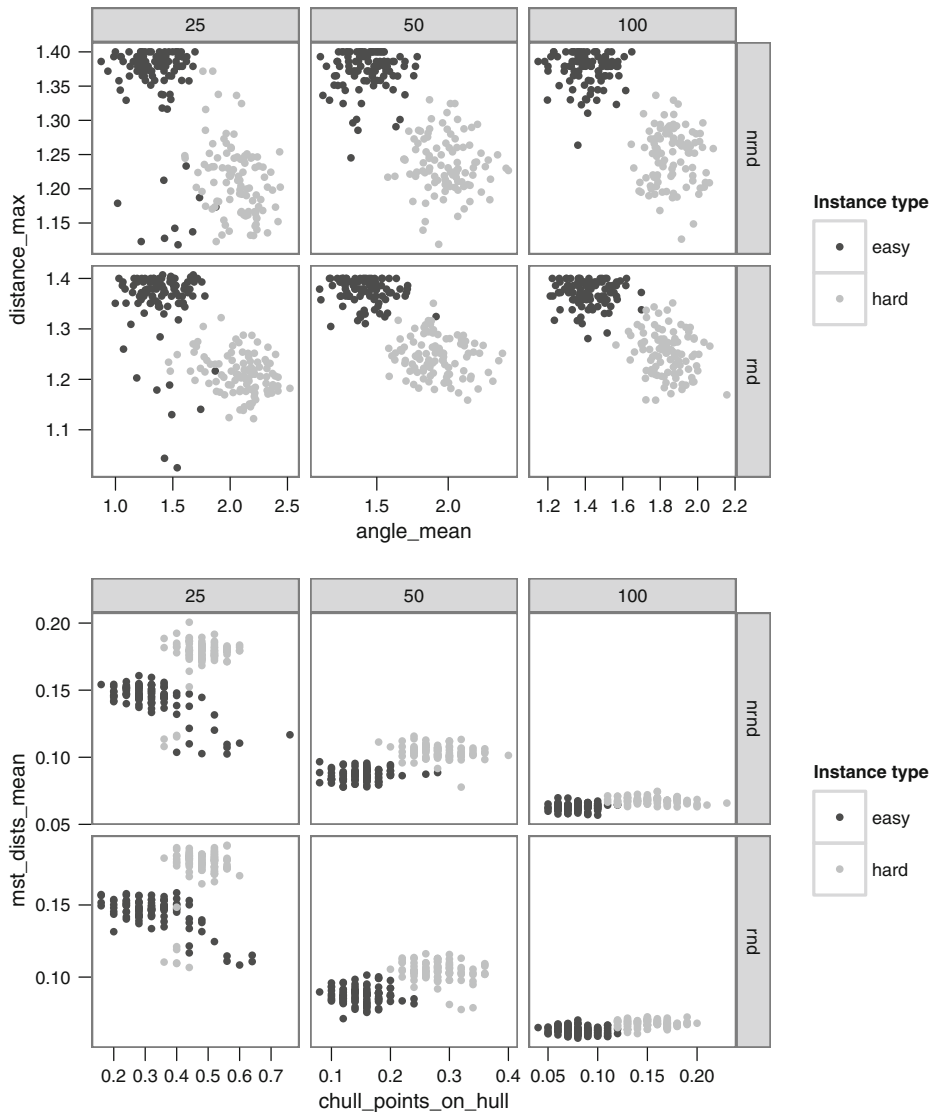
The first rule is perfectly in line with the exploratory observations of Section 3.2. The mean angles between the cities on the optimal tour were found to be significantly higher for the hard instances than the for the easy ones (see Fig. 5). This corresponds to the classification rule which is based on the feature “angle\_2nn\_mean” measuring the mean angle between the nodes and their two nearest neighbors. Secondly, the easy instances exhibit a more uniform distribution of the tour length legs on the optimal tour (see Fig. 4). This observation coincides with the characteristics of the feature *dist\_max*. The higher the maximum distance between two cities the lower is the probability of a low standard deviation of the tour length legs.



**Fig. 7** Classification rules for *rnd* for the first feature combination given in Fig. 9. Mean classification accuracy equals 0.968



**Fig. 8** Classification rules for *rmd* for the second feature combination given in Fig. 9. The cross-validated accuracy of this rule is 0.975



**Fig. 9** Scatt1erplot of exemplary feature combinations which allow an accurate separation of easy and hard instances

The second rule comprises two of the new features introduced in this paper. A lower fraction of points on the convex hull of all points together with smaller mean distance of the minimum spanning tree indicates a low instance hardness. In general the rules are more accurate for the higher instances sizes than for instance size 25.

Summarizing, an accurate feature-based separation of the easy and hard instances can be successfully achieved, even with various combinations of two features. Arising from this, we will investigate in the next sections how instances of moderate difficulty in between the evolved easy and hard instances can be generated and if an explicit



prediction of the expected approximation quality of 2-opt on all instances is possible based on the available features. This prediction problem in our view is much more interesting as well as challenging than the classification task analyzed within this section and will thus be investigated in more detail.

### 3.4 Morphing hard into easy instances

We are now in the position to separate easy and hard instances with the classification rules presented in Section 3.3. In this section, instances in between, i.e. of moderate difficulty, are considered as well. Starting from the result in [12] that a hard TSP instance can be transformed into an easy one by slight variation of the node locations, we studied the “transition” of hard to easy instances by morphing a single hard instance into an easy instance by a convex combination of the points of both instances, which generates an instance in between the original ones (Alg. 6).

---

#### Algorithm 6 Morphing

---

```

function MORPH(hardInstance, easyInstance,  $\alpha$ , cells, rnd)
  easyInstance  $\leftarrow$  pointMatching(hardInstance, easyInstance)
  instMorph  $\leftarrow$   $\alpha \cdot \text{hardInstance} + (1 - \alpha) \cdot \text{easyInstance}$ 
  instMorph  $\leftarrow$  RESCALE(instMorph)
  instMorph  $\leftarrow$  ROUND(instMorph, cells)
  if rnd then
    instMorph  $\leftarrow$  NORMALMUTATION(instMorph)
    instMorph  $\leftarrow$  CUTTOBOUNDARY(instMorph)
  end if
  return instMorph
end function

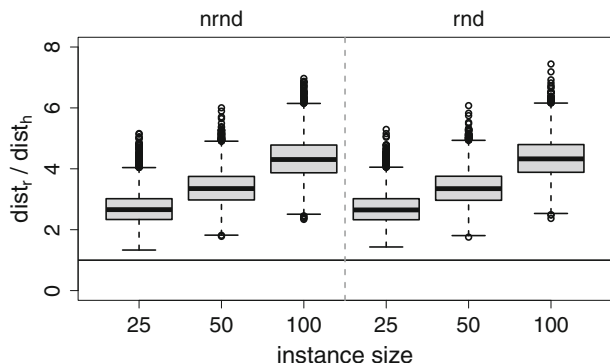
```

---

The point matching between the input instances is improved w.r.t. [25] in that a greedy approach is used which successively selects the point pair for which the pairwise Euclidean distance is minimal. Depending on the type of rounding scheme used in the EA that generated the instances, a normal mutation step and successive cutting to the boundary might be required after rescaling and rounding to ensure that the newly constructed instance is of the desired instance type.

Figure 10 shows the positive effect of the greedy heuristic point matching strategy in contrast to a random point matching as utilized in [25]. A simulation was con-

**Fig. 10** Effect of heuristic vs. random point matching strategy. Boxplots of the sums of all interpoint distances of the random approach ( $\text{dist}_r$ ) relative to the heuristic ones ( $\text{dist}_h$ ) are given for the two different rounding concepts and varying instance sizes

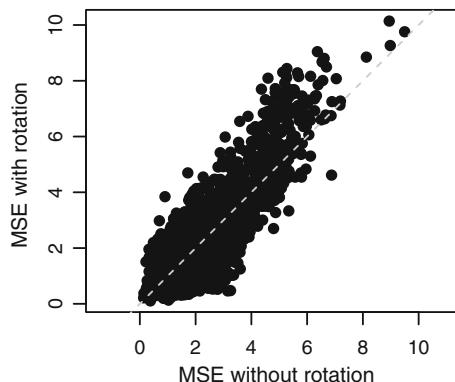


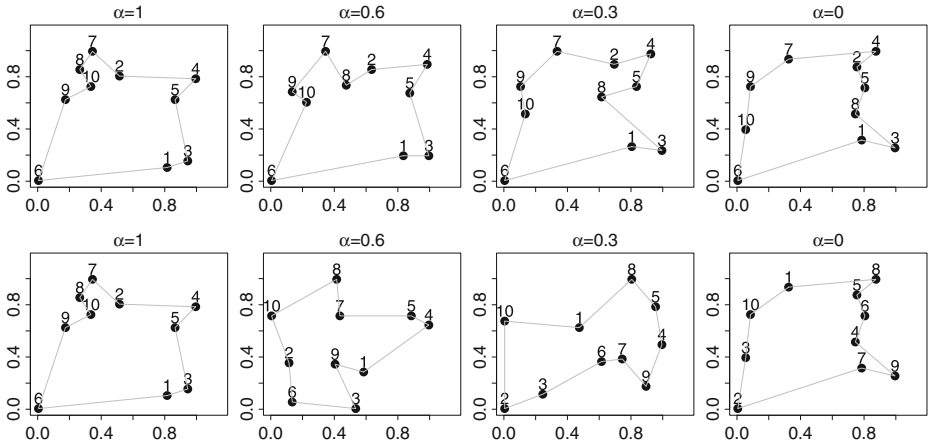
ducted in the following way: Two random instances are generated in the  $[0, 1]^2$ -plane, rescaled, rounded and (possibly) normally mutated afterwards to reflect the EA rounding scheme which applies rounding before normal mutation (*rnd*). Afterwards, the sum of interpoint distance after random point matching ( $dist_r$ ) and greedy heuristic point matching ( $dist_h$ ) are calculated and divided by each other ( $dist_r / dist_h$ ). Results are shown for instance sizes  $\{25, 50, 100\}$  and both rounding schemes. It becomes obvious that the interpoint distances resulting from the greedy approach are much smaller than the respective ones of the random strategy. The latter distance sums are on average roughly twice as high for the instance size 25 and increases linearly to a factor of four for instance size 100. The effects are visually identical for both rounding concepts. For the instance sizes we consider here, no randomly picked matching was ever better than the matching returned by the greedy algorithm.

One might argue that the greedy point matching should be improved further. One obvious idea is to rotate the instances such that they align “better”. This would amount to a normalization w.r.t. rotation of all instances, a desirable feature because the length of the optimal tour as well as the approximation quality are invariant under rotation because both only depend on the edge weights which are invariant under rotation. To normalize the instances we calculated the covariance matrix of the points of a TSP instance and rotated the points around their centroid such that the main axis, i.e. the eigenvector corresponding to the largest eigenvalue is parallel to the X axis. This results in two possible rotational angles. We opted to rotate by the angle which lead to the point configuration which had more more points to the right of and above the centroid. Performing this normalization followed by a greedy point matching did not increase the quality of the matching (the total distance travelled by each point during the morphing does not improve on average). This is illustrated in Fig. 11. We therefore opted to not normalize instances. In future studies it might be a logical next step to try and integrate this form of normalization into the EA generating the instances.

Morphing examples are shown in Fig. 12. Based on the initial instances ( $\alpha = 1$  and  $\alpha = 0$ ), instances emerge from each other with decreasing  $\alpha$ . Clearly, the advantageous effect of the improved point matching becomes visible as the transitions of the morphed instances are much smoother than in the former case of random point matching. Furthermore, in case of random point matching instances tend to concentrate on the center part of the  $[0, 1]^2$  - plane.

**Fig. 11** Comparison of morphing quality of rotationally normalized and unnormalized instances. Normalization does not lead to a significant improvement in the quality of the point matching and subsequent morphing



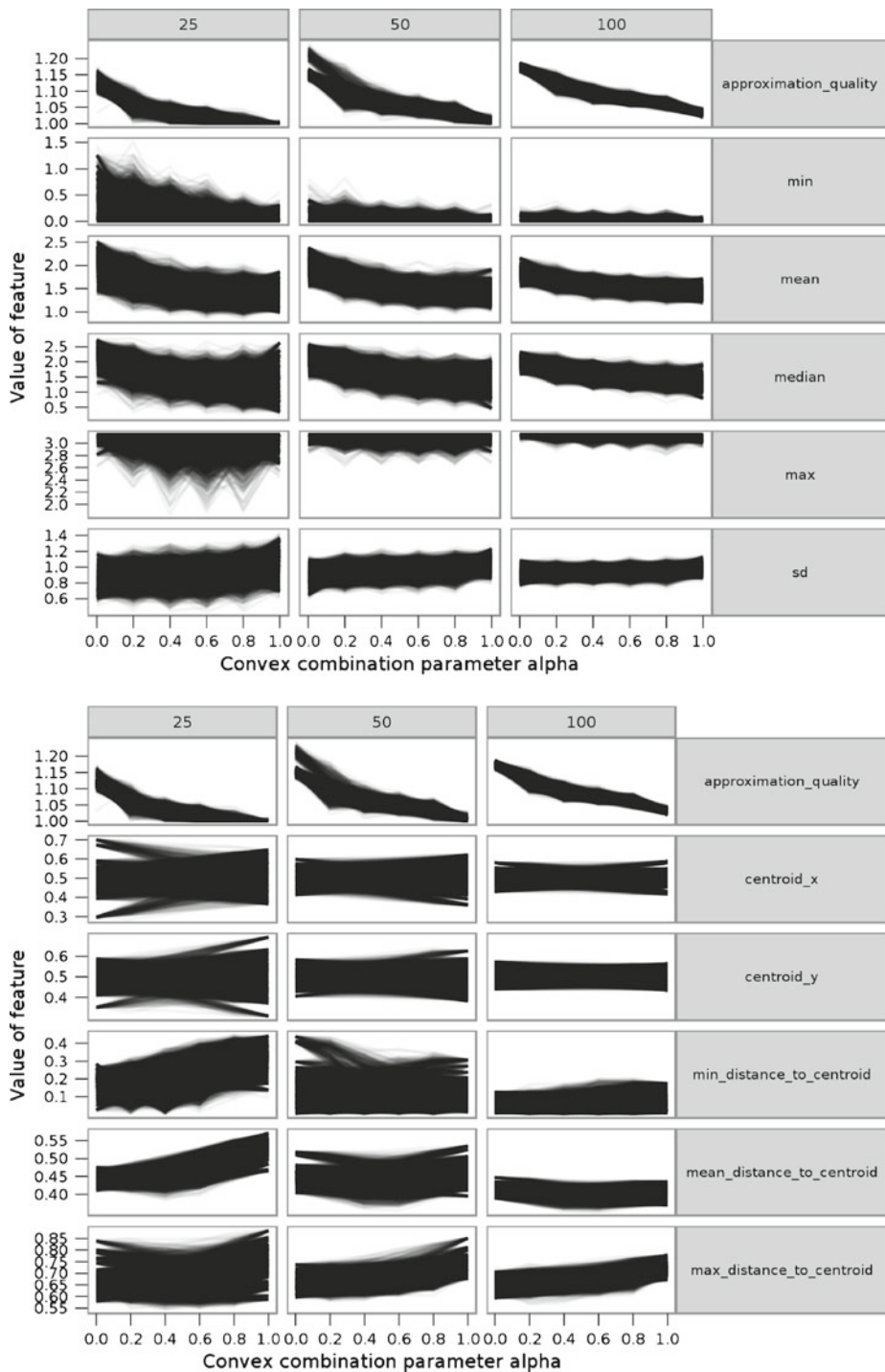


**Fig. 12** Example: Morphing of one instance into a different instance for different  $\alpha$ -levels of the convex combination (*nrnd*) with heuristic greedy (*above*) and random point matching (*below*). Optimal tours are visualized in grey

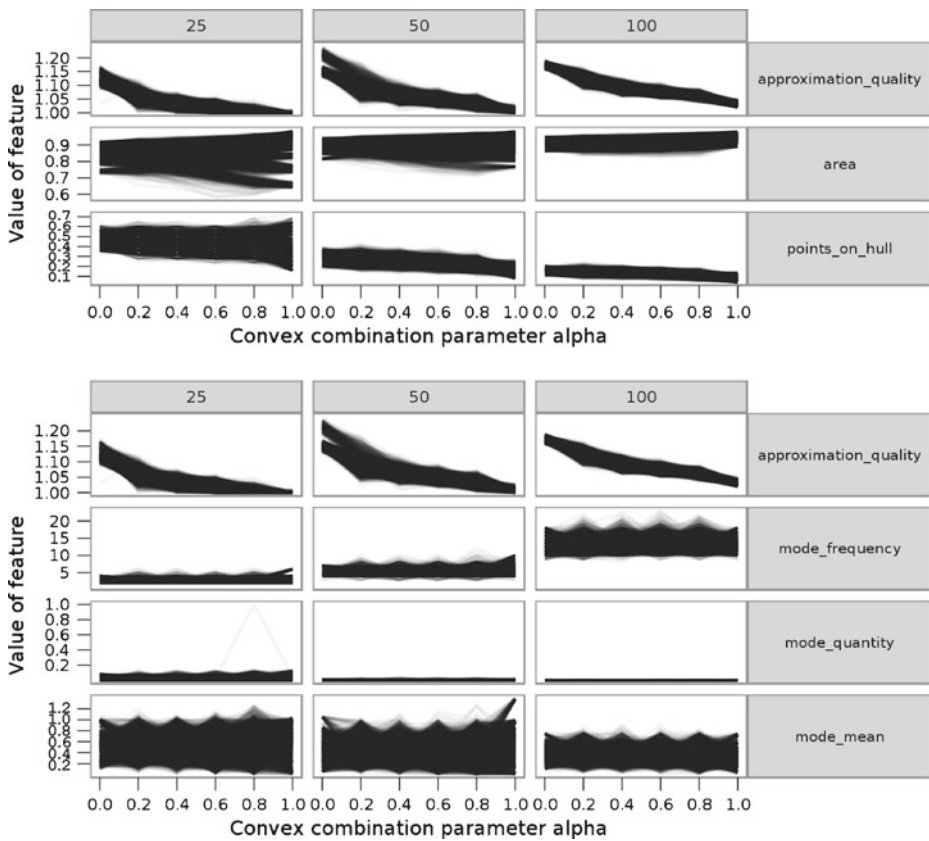
The morphing strategy is applied to all possible combinations of single hard and easy instances of the two evolved instance sets using 6 levels of  $\alpha$ , i.e.  $\alpha \in \{0, 0.2, 0.4, \dots, 1\}$ . Each generated instance is characterized by the levels of the features discussed in Section 2.1. Thus, the changes of the feature levels with increasing  $\alpha$  can be studied which is of interest as it should lead to an understanding of the influence of the different features on the approximation quality.

Figures 13, 14, 15, 16 and 17 show the approximation quality for the instances of all morphing sequences for the various  $\alpha$  levels in the top subfigure. Starting from a hard instance on the left side of each individual plot ( $\alpha = 0$ ) the findings of [12] are confirmed. The approximation quality of 2-opt quickly increases with slight increases of  $\alpha$ . Additionally, the feature levels of the generated instances are visualized arranged by feature groups. We concentrate on the subset of data which is based on rounding before the normal mutation step as the presented observations coincide for both rounding schemes.

Obviously, many features do not show any systematic relationship with the approximation quality for all considered instance sizes, e.g. most features related to the centroid, the clustering as well as the modes of the edge cost distribution. Interestingly, some features exhibit different tendencies for the smallest and highest instance size, e.g. the features reflecting the mean and minimum distance to the centroid in Fig. 13. This is due to the different structural shapes of small and large instance classes, more specifically the almost circular structure of the easy instances for an instance size of 25 (see Fig. 3). The same reasoning holds for the features associated with the mean and median distances and the standard deviation of the distance matrix. Exceptional behavior of the feature levels occurs for the features related to the MST depth (mean, median, max and sd) for which a systematic nonlinear decrease can be observed only for the instance size 100. In most cases the change of the feature in relation to  $\alpha$  is similar for all instance sizes, what does change is the variance of the feature.



**Fig. 13** Angle (*top*, based on 2 nearest neighbors) and Centroid Features (*bottom*): Approximation quality and feature values for different  $\alpha$  levels of all conducted morphing experiments

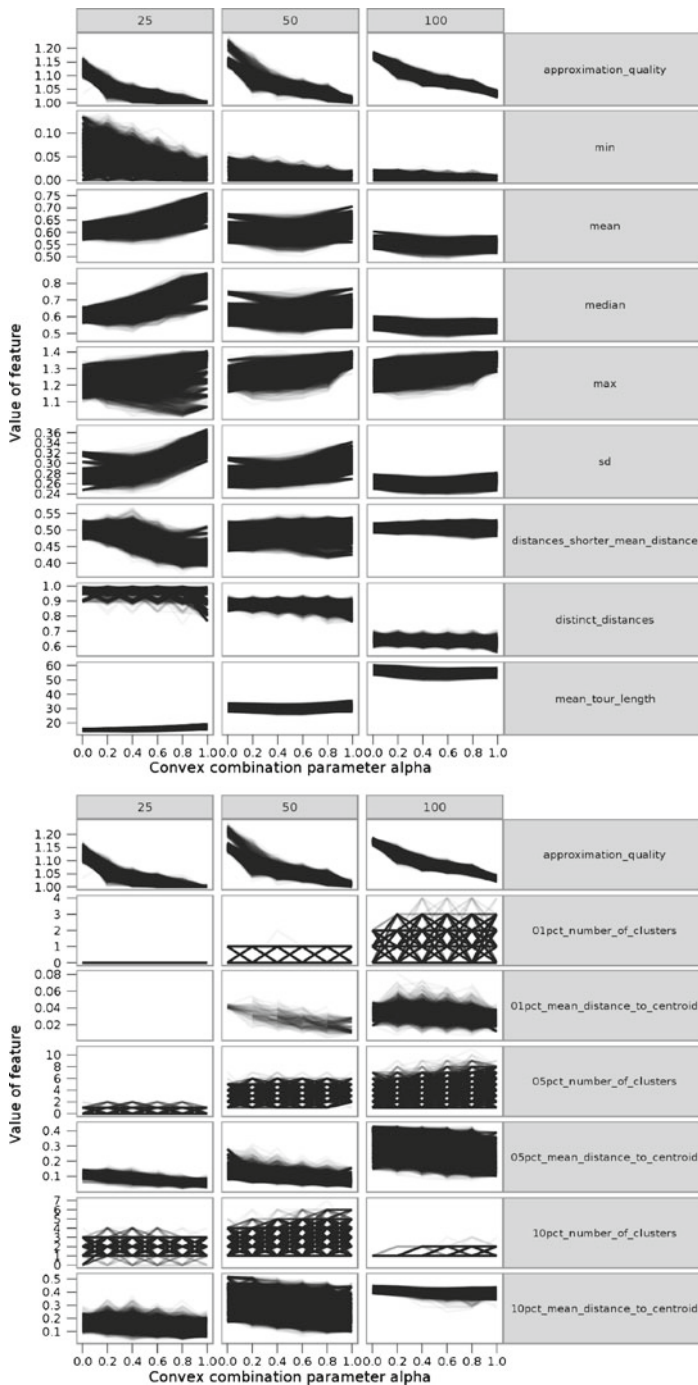


**Fig. 14** Convex Hull (*top*) and Mode (*bottom*) features: Approximation quality and feature values for different  $\alpha$  levels of all conducted morphing experiments

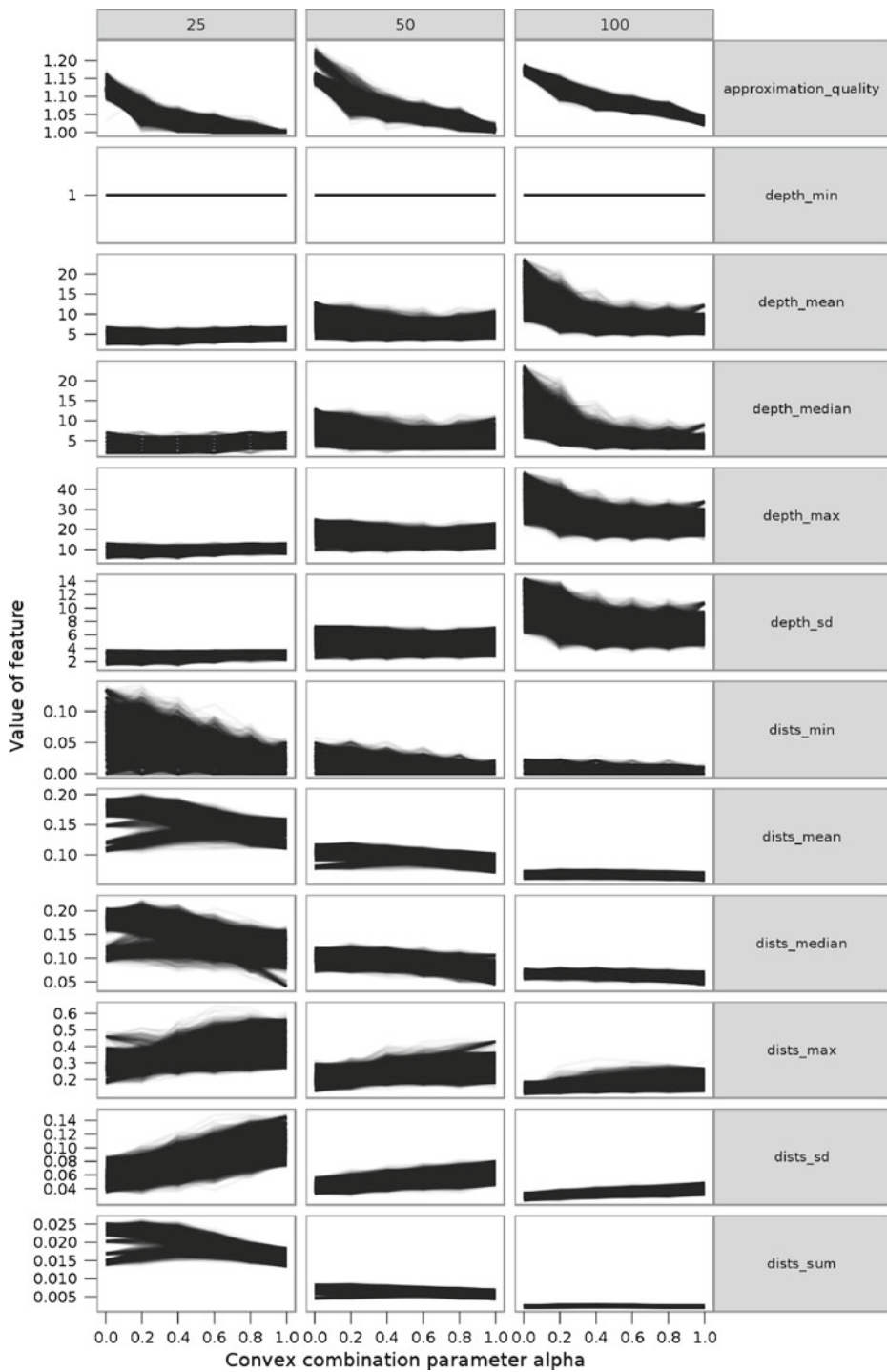
However, systematic nonlinear relationships with the approximation quality can be detected for the mean and median distances on the MST as well as the standard deviation (Fig. 16), the maximum distance between the cities and the respective standard deviation (sd) (Fig. 15) and the coefficient of variation of nearest neighbor distances (Fig. 17). Additional promising features in Fig. 13 are the fraction of points on the convex hull, the area of the convex hull, and the mean angle between adjacent cities as well as the maximum distance to the centroid in Fig. 14. Naturally, the features included in the two exemplary classification rules above form a subset of the mentioned relevant features.

### 3.5 Feature-based prediction of TSP problem hardness

In order to get a more accurate picture of the relationship between the approximation quality and the features of the full data set including all morphed instances a Multivariate Adaptive Regression Splines (MARS) [13] model is constructed in order to directly predict the expected approximation quality of 2-opt on a given

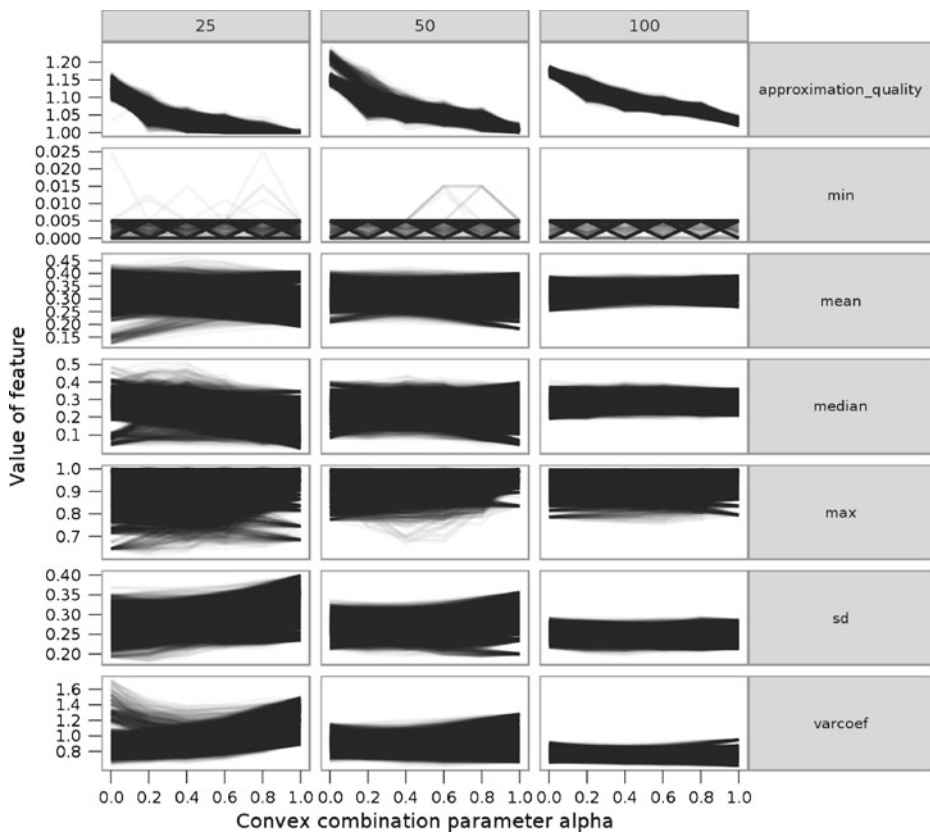


**Fig. 15** Distance (*top*) and Cluster (*bottom*) features: Approximation quality and feature values for different  $\alpha$  levels of all conducted morphing experiments. The annotations “01pct”, “05pct” and “10pct” identify different levels of the reachability distance as a parameter of GDBSCAN [32] used for clustering



**Fig. 16** MST features: Approximation quality and feature values for different  $\alpha$  levels of all conducted morphing experiments





**Fig. 17** Nearest neighbor distance features: Approximation quality and feature values for different  $\alpha$  levels of all conducted morphing experiments

instance based on the candidate features. As before in the classification experiment, the regression model is again validated by a ten-fold cross-validation procedure.

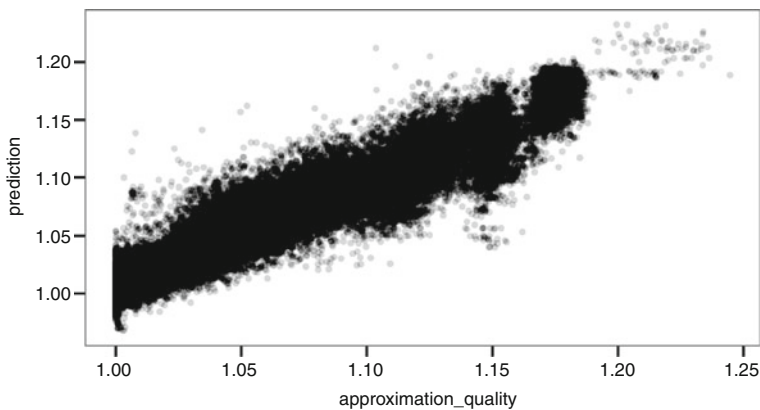
We used MARS with second degree interaction effects to model the relationship between the approximation quality and the calculated instance features. Other modeling approaches, such as  $k$ -nearest neighbors and linear models, were also considered but some initial experiments on a subset of the data showed that MARS provided competitive results and scaled well to the full dataset. The final model is shown in Table 2. We achieved a root mean squared error (RMSE) of 0.0170 for *rnd* and 0.0165 for *nrnd*. This compares favorably to a simple model that always predicts the mean (RMSE for *rnd* equals 0.0512 and for *nrnd* 0.0516) which we outperform by a factor of 3. In other words, given the features of a TSP instance, we expect to predict, on average, the approximation quality of a 2-opt solution to within  $\pm 1.6\%$  of the true approximation ratio. In the following, we concentrate our analysis on *rnd* as the results almost coincide. We also present a scatterplot of the true approximation quality values vs. their predictions (from the cross-validation) in Fig. 18 to show how predictive accuracy varies across the whole range of problem hardness.



**Table 2** Results of the MARS model for *rnd*

Spline	Coefficient
(Intercept)	1.133
$h(\text{mst\_dists\_sd}-0.0369208)$	-0.716
$h(0.0369208-\text{mst\_dists\_sd})$	1.707
$h(\text{angle\_2nn\_mean}-1.45864)$	0.100
$h(1.45864-\text{angle\_2nn\_mean})$	0.012
$h(\text{chull\_points\_on\_hull}-0.28)$	-0.012
$h(0.28-\text{chull\_points\_on\_hull})$	-0.147
$h(\text{mst\_depth\_max}-12)$	0.005
$h(12-\text{mst\_depth\_max})$	0.001
$h(\text{distance\_max}-1.26151)$	-0.199
$h(1.26151-\text{distance\_max})$	0.039
$h(\text{cluster\_10pct\_mean\_distance\_to\_centroid}-0.496846)$	-0.788
$h(0.496846-\text{cluster\_10pct\_mean\_distance\_to\_centroid})$	-0.003
$h(\text{angle\_2nn\_mean}-1.45864)*h(\text{distance\_mean}-0.606019)$	-0.795
$h(\text{angle\_2nn\_mean}-1.45864)*h(0.606019-\text{distance\_mean})$	-0.561
$h(\text{mst\_depth\_median}-12)$	-0.018
$h(12-\text{mst\_depth\_median})$	0.002
$h(\text{distance\_sd}-0.307304)$	-0.458
$h(0.307304-\text{distance\_sd})$	0.695
$h(\text{centroid\_max\_distance\_to\_centroid}-0.72554)*h(1.26151-\text{distance\_max})$	22.418
$h(0.72554-\text{centroid\_max\_distance\_to\_centroid})*h(1.26151-\text{distance\_max})$	-0.718
$h(0.28-\text{chull\_points\_on\_hull})*h(\text{mst\_dists\_mean}-0.0671864)$	1.737
$h(0.28-\text{chull\_points\_on\_hull})*h(0.0671864-\text{mst\_dists\_mean})$	-15.553
$h(\text{distance\_mean\_tour\_length}-56.3844)$	0.005
$h(56.3844-\text{distance\_mean\_tour\_length})$	-0.002
$h(\text{mst\_dists\_mean}-0.105423)$	-0.224
$h(0.105423-\text{mst\_dists\_mean})$	-1.681
$h(\text{centroid\_min\_distance\_to\_centroid}-0.314769)*h(\text{mst\_depth\_max}-12)$	0.089
$h(0.314769-\text{centroid\_min\_distance\_to\_centroid})*h(\text{mst\_depth\_max}-12)$	-0.003
$h(\text{angle\_2nn\_mean}-1.45864)*h(\text{angle\_2nn\_median}-2.16409)$	-0.006
$h(\text{angle\_2nn\_mean}-1.45864)*h(2.16409-\text{angle\_2nn\_median})$	-0.100
$h(\text{angle\_2nn\_mean}-1.81252)*h(\text{mst\_depth\_max}-12)$	-0.013
$h(1.81252-\text{angle\_2nn\_mean})*h(\text{mst\_depth\_max}-12)$	-0.003
$h(\text{mst\_dists\_sd}-0.0369208)*h(\text{mst\_dists\_sum}-0.0178643)$	170.696
$h(\text{distance\_sd}-0.274633)*h(12-\text{mst\_depth\_median})$	0.075
$h(\text{mst\_depth\_sd}-11.632)$	0.011
$h(11.632-\text{mst\_depth\_sd})$	0.025
$h(\text{mst\_depth\_mean}-18.8)$	0.011
$h(18.8-\text{mst\_depth\_mean})$	-0.017
$h(\text{angle\_2nn\_mean}-1.82054)*h(0.105423-\text{mst\_dists\_mean})$	6.850
$h(1.82054-\text{angle\_2nn\_mean})*h(0.105423-\text{mst\_dists\_mean})$	0.699
$h(\text{distance\_max}-1.26151)*h(\text{mst\_dists\_sd}-0.0495085)$	3.614
$h(\text{distance\_max}-1.26151)*h(0.0495085-\text{mst\_dists\_sd})$	5.079
$h(\text{mst\_depth\_max}-12)*h(\text{mst\_dists\_mean}-0.0829785)$	0.066
$h(\text{mst\_depth\_max}-12)*h(0.0829785-\text{mst\_dists\_mean})$	0.082

Because MARS models are highly non-linear, it is hard to visualize them. In Fig. 19 nine features which are frequently used in the splines of the model are visualized. The top of the figure shows a scatter plot of the feature against the approximation quality.



**Fig. 18** Scatterplot for true approximation quality values vs. predictions for *rnd*. Predictions are from ten-fold cross-validation

The color of each point represents the error the model makes when predicting this point and gives us some insight into where the model fits the data well and where it deviates significantly. The only real structure to be seen in the plots is a small cluster of red points with an approximation quality of about 1.15 that is visible in every panel of the plot. This shows that our model fits the data fairly well. The bottom part of the figure shows a variant of a partial dependency plot. Instead of averaging over all observations as in the partial dependency plot, we use a weighted average, where we give observations that are close to the feature value a higher weight - we call this the weighted partial dependency plot. That is, for a feature  $x$  with value  $x^*$ , we calculate

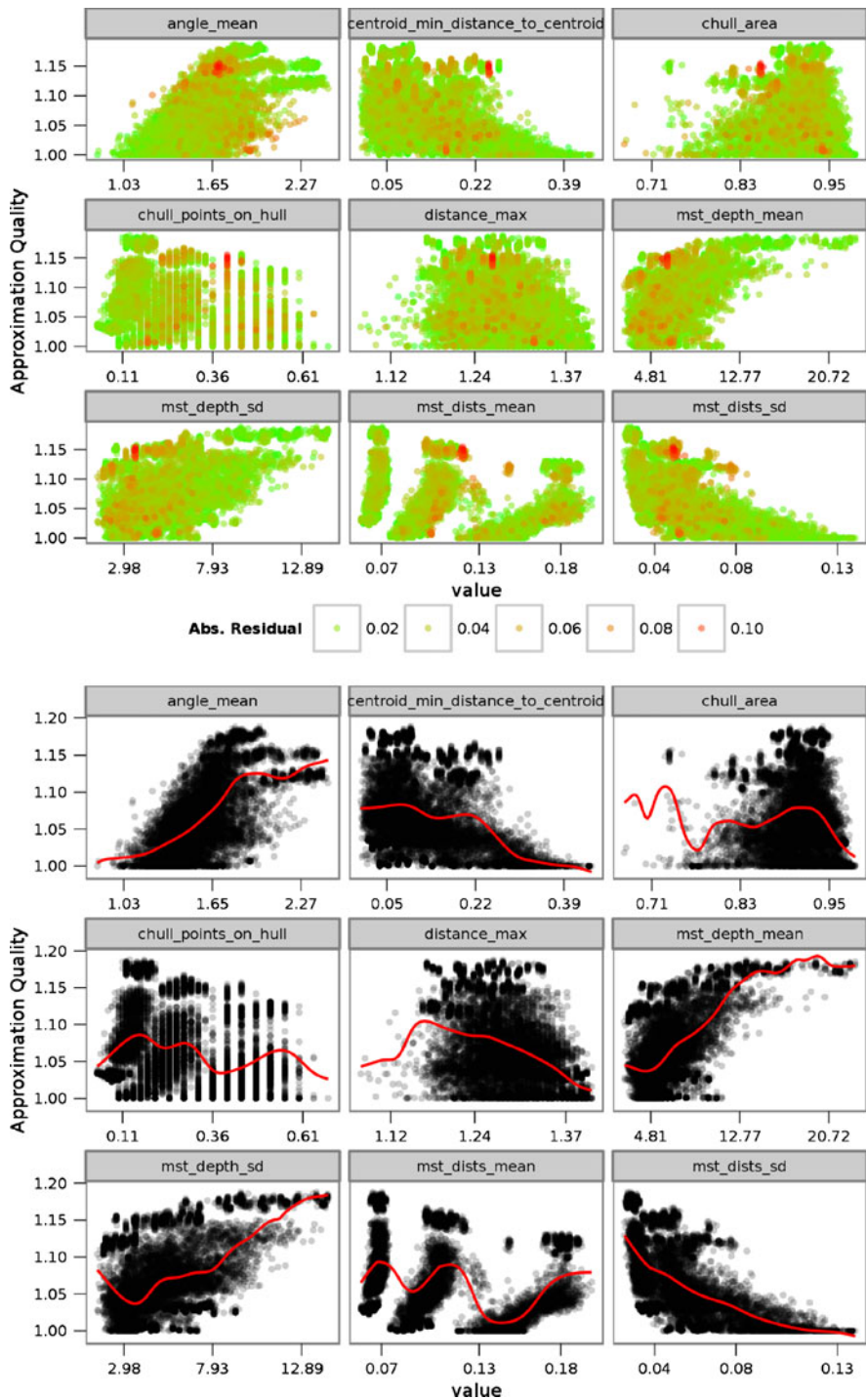
$$f(x^*) = \frac{1}{\sum_{i=1}^N w_i(x^*)} \sum_{i=1}^n w_i(x^*) * m((x^*, \mathbf{d}_i \setminus x))$$

where  $\mathbf{d}_i$  denotes the features of the  $i$ -th instance,  $w_i(x^*)$  the weight assigned to the  $i$ -th observation and  $m((x^*, \mathbf{d}_i \setminus x))$  the predicted approximation quality for the  $i$ -th feature vector if we set feature  $x$  to  $x^*$ . We chose to use a Gaussian weighting function

$$w_i(x^*) = \alpha \phi(x_i - x^*)$$

where  $\alpha$  is set to a fourth of the standard deviation of the feature and  $\phi$  denotes the density function of the standard normal distribution. We see that the average response of the model fits the point clouds quite well.

We further studied whether we could handle the regression problem with a substantially smaller feature set in order to simplify our model. For this purpose we performed a sequential forward search, which iteratively adds the best feature w.r.t. the RMSE. As regression model we again used MARS with second order interaction effects. Such a forward search is a simple feature selection wrapper as introduced in [19]. In this procedure we perform an outer resampling loop (here 10-fold cross-validation) to create training and test sets. For each training set we perform the feature selection by forward search. For each training set and feature selection run the outer training set is resampled again (here simple hold-out with 2/3



**Fig. 19** MARS model: Nine most frequently used features for *rnd*. *Top*: Residual scatter plot, *Bottom*: Weighted partial dependency plot

**Table 3** Results of the MARS model with feature selection by forward search

Feature list	RMSE
Empty model	0.05113
+ mst_dists_sd	0.03440
+ angle_2nn_mean	0.02515
+ mst_dists_mean	0.02240
+ mst_depth_median	0.02036

for training and 1/3 for testing) in a so-called inner loop. The RMSE of each feature set is measured and greedily optimized according to this inner resampling. We stop the selection when the performance in RMSE does not improve by at least  $\sqrt{5} \cdot 10^{-5}$ . The outer resampling ensures unbiased performance results and the whole procedure is sometimes called nested resampling [4].

The final results are 10 potentially different feature sets, but in our case we always end up with the four features displayed in Table 3, which are also always selected in the same order. We also display the (mean) RMSE for the feature sets during the search on the inner test sets (the numbers are averaged across all 10 feature selections). The unbiased RMSE in the outer cross-validation is 0.02037.

From the results we gain further insights into which features reduce the RMSE the most and that we can build an acceptable model with only four features. But it must still be noted that we perform substantially worse than selecting the full model. The reader should be aware of the fact that a MARS model already performs an internal feature selection which is somewhat similar to our approach, but faster. This last step was mainly undertaken to study in further detail how well we can predict the approximation quality with a model with a really low number of features.

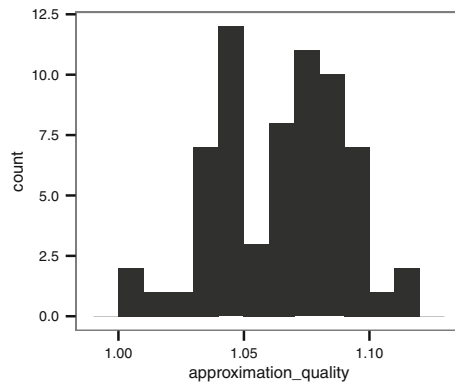
### 3.6 Feature validation on TSPLIB

In order to further validate our proposed feature set, we use it to predict the approximation quality of 2-opt on the 65 TSPLIB instances listed in Table 4 (mainly

**Table 4** Considered TSPLIB instances

a280.tsp	fl417.tsp	kroB100.tsp	pr299.tsp
att48.tsp	fri26.tsp	kroB150.tsp	pr439.tsp
bayg29.tsp	gil262.tsp	kroB200.tsp	pr76.tsp
bays29.tsp	gr120.tsp	kroC100.tsp	rat195.tsp
berlin52.tsp	gr137.tsp	kroD100.tsp	rat99.tsp
bier127.tsp	gr17.tsp	kroE100.tsp	rd100.tsp
brazil58.tsp	gr202.tsp	lin105.tsp	rd400.tsp
brg180.tsp	gr21.tsp	lin318.tsp	st70.tsp
burma14.tsp	gr229.tsp	linhp318.tsp	swiss42.tsp
ch130.tsp	gr24.tsp	pcb442.tsp	ts225.tsp
ch150.tsp	gr431.tsp	pr107.tsp	tsp225.tsp
d198.tsp	gr48.tsp	pr124.tsp	u159.tsp
d493.tsp	gr96.tsp	pr136.tsp	ulysses16.tsp
dantzig42.tsp	hk48.tsp	pr144.tsp	ulysses22.tsp
eil101.tsp	kroA100.tsp	pr152.tsp	
eil51.tsp	kroA150.tsp	pr226.tsp	
eil76.tsp	kroA200.tsp	pr264.tsp	

**Fig. 20** Histogram of 2-opt approximation quality on TSPLIB instances



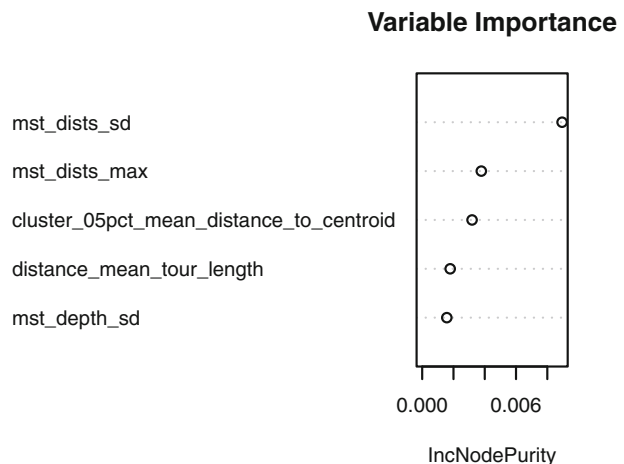
the ones with less than 500 cities where we could either directly access or infer 2D city coordinates).

All instances have been rescaled to  $[0, 1]^2$  and the approximation quality has been calculated exactly as before. Figure 20 shows a histogram of the approximation quality distribution.

We now use a random regression forest [5] to predict the approximation quality on these instances based on our features (a MARS model works, too, but less well; we did not use a random forest in the previous section because of the large number of observations and resulting memory and time problems). Because of the very low number of observations we validate this by performing leave-one-out cross-validation. This results in an RMSE of 0.0138, compared to an RMSE of 0.0249 if we always predicted the mean value, which is both on a comparable scale as the analysis in the previous section.

Finally, we fit the forest on all 65 instances, compute the variable importance (measured by the the mean increase in node purity w.r.t. the underlying regression trees, for details see [5]) and plot the five most influential features in Fig 21.

**Fig. 21** Variable importance of five most influential features of random regression forest



## 4 Summary and outlook

In this paper we investigated concepts to predict TSP problem hardness for 2-opt based local search strategies on the basis of experimental features that characterize the properties of a TSP instance. A crucial aspect was the generation of a representative instance set as a basis for the analysis. This turned out to be far from straightforward. Therefore, it was only possible to generate very hard and very easy instances using sophisticated (evolutionary) strategies. Summarizing, we managed to generate classes of easy and hard instances of different sizes for which we are able to predict the correct instance class based on the corresponding feature levels with only marginal errors. Several feature combinations, which are cheap to compute even for large instances, could be identified as key features for differentiating between hard and easy instances, and the results are supported by exploratory analysis of the evolved instances and the respective optimal tours. However, it should be noted that most probably not the whole space of possible hard instances is covered by using our evolutionary method, i.e. probably only a subset of possible characteristics or feature combinations that make a problem hard for 2-opt can be identified by the applied methodology.

Instances of moderate difficulty were constructed by morphing hard into easy instances where the effects of the transition on the corresponding feature levels could be studied. A MARS model was successfully applied to predict the approximation quality of 2-opt independent from the instance size based on the features of the generated instances with very high accuracy. We strongly believe that it should be straight forward to apply the same methodology to other algorithms and use these models to derive a strategy for the algorithm selection problem in the context of the TSP.

Moreover, we investigated two different rounding schemes within the evolutionary algorithm for instance generation which either result in instances exhibiting points on a regular grid or slightly perturbed points. However, the experimental results did not show any significant differences between the different concepts.

The analysis offers promising perspectives for further research, specifically a systematic comparison to other local and global search as well as hybrid solvers with respect to the influence of the feature levels of an instance on the performance of the respective algorithms. The investigation of much higher instances sizes would be very interesting as well. However, it has to be kept in mind that the computational effort intensely increases with increasing instance size as the optimum solution, e.g. computable via Concorde, is required to calculate the approximation quality of 2-opt.

Finally, it is open how representative the generated instances are for real-world TSP instances. We did try to directly predict the TSPLIB approximation quality with the models obtained from our synthetic instances, but this did not work out very well, possibly

Instead we successfully validated the usefulness of the feature set directly on TSPLIB. Therefore, it is very desirable to collect and create a much larger pool of small to medium sized, real-world, TSP instances for comparison experiments. In general, more work needs to be done here.

There is also the question of how well these models can extrapolate to much larger instance sizes. This would again be a desirable property in the context of algorithm selection for very large instances for which it is not feasible to calculate the global optimal tour.

In closing we would like to mention that all source code used in these experiments is available online (see Footnote 1) for anyone to use and extend. We will also publish an already extended R package under the name of *tspmeta* on the R package server CRAN, which we will further improve to enable feature-based analysis of TSP instances.

**Acknowledgements** This work was partly supported by the Collaborative Research Center SFB 823, the Graduate School of Energy Efficient Production and Logistics and the Research Training Group “Statistical Modelling” of the German Research Foundation.

## References

1. Applegate, D., Cook, W.J., Dash, S., Rohe, A.: Solution of a min-max vehicle routing problem. *INFORMS J. Comput.* **14**(2), 132–143 (2002)
2. Arora, S.: Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. *J. ACM* **45**(5), 753–782 (1998)
3. Bischl, B., Mersmann, O., Trautmann, H., Preuss, M.: Algorithm selection based on exploratory landscape analysis and cost-sensitive learning. In: *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation, GECCO '12*, pp. 313–320. ACM, New York, NY, USA (2012)
4. Bischl, B., Mersmann, O., Trautmann, H., Weihs, C.: Resampling methods in model validation. *Evol. Comput. J.* **20**(2), 249–275 (2012)
5. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
6. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: *Classification and Regression Trees*. Wadsworth, Belmont, CA (1984)
7. Chandra, B., Karloff, H.J., Tovey, C.A.: New results on the old k-Opt algorithm for the traveling salesman problem. *SIAM J. Comput.* **28**(6), 1998–2029 (1999)
8. Croes, G.A.: A method for solving traveling-salesman problems. *Oper. Res.* **6**(6), 791–812 (1958)
9. Dorigo, M., Stützle, T.: *Ant Colony Optimization*. MIT Press (2004)
10. Droste, S., Jansen, T., Wegener, I.: On the analysis of the (1+1) evolutionary algorithm. *Theor. Comput. Sci.* **276**, 51–81 (2002)
11. Eiben, A., Smith, J.: *Introduction to Evolutionary Computing*. Springer (2007)
12. Englert, M., Röglin, H., Vöcking, B.: Worst case and probabilistic analysis of the 2-opt algorithm for the tsp: extended abstract. In: Bansal, N., Pruhs, K., Stein, C. (eds.) *SODA*, pp. 1295–1304. SIAM (2007)
13. Friedman, J.H.: Multivariate adaptive regression splines. *Ann. Stat.* **19**(1), 1–67 (1991)
14. Glover, F.: Ejection chains, reference structures and alternating path methods for traveling salesman problems. *Discrete Appl. Math.* **65**(1–3), 223–253 (1996)
15. Hoos, H.H., Stützle, T.: *Stochastic Local Search: Foundations & Applications*. Elsevier/Morgan Kaufmann (2004)
16. Johnson, D.S., McGeoch, L.A.: The traveling salesman problem: A case study in local optimization. In: Aarts, E.H.L., Lenstra, J.K. (eds.) *Local Search in Combinatorial Optimization*. Wiley (1997)
17. Kanda, J., Carvalho, A., Hruschka, E., Soares, C.: Selection of algorithms to solve traveling salesman problems using meta-learning. *IJHIS* **8**(3), 117–128 (2011)
18. Kilby, P., Slaney, J., Walsh, T.: The backbone of the travelling salesperson. In: *Proc. of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05*, pp. 175–180. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2005)
19. Kohavi, R., John, G.H.: Wrappers for feature subset selection. *Artif. Intell.* **97**(1–2), 273–324 (1997)
20. Kötzing, T., Neumann, F., Röglin, H., Witt, C.: Theoretical properties of two ACO approaches for the traveling salesman problem. In: *Proc. of ANTS 2010, LNCS*, vol. 6234, pp. 324–335 (2010). Extended journal version appears in *Swarm Intelligence*
21. Kovárik, O., Málek, R.: Meta-learning and meta-optimization. Tech. rep., CTU Technical Report KJB2012010501 003, Prague (2012). [http://cig.felk.cvut.cz/research/publications/Meta-learning\\_and\\_meta-optimization.pdf](http://cig.felk.cvut.cz/research/publications/Meta-learning_and_meta-optimization.pdf)
22. van Laarhoven, P., Aarts, E.: *Simulated Annealing: Theory and Applications*. Springer (1997)



23. Lin, S.: Computer solutions of the travelling salesman problem. *Bell Syst. Tech. J.* **44**(10), 2245–2269 (1965)
24. Lin, S., Kernighan, B.: An effective heuristic algorithm for the traveling salesman problem. *Oper. Res.* **21**, 498–516 (1973)
25. Mersmann, O., Bischl, B., Bossek, J., Trautmann, H., Wagner, M., Neumann, F.: Local search and the traveling salesman problem: A feature-based characterization of problem hardness. In: Hamadi, Y., Schoenauer, M. (eds.) *Learning and Intelligent Optimization. Lecture Notes in Computer Science*, pp. 115–129. Springer Berlin Heidelberg (2012)
26. Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., Rudolph, G.: Exploratory landscape analysis. In: *Proc. of the 13th Annual Conference on Genetic and Evolutionary Computation, GECCO '11*, pp. 829–836. ACM, New York, NY, USA (2011)
27. Merz, P., Freisleben, B.: Memetic algorithms for the traveling salesman problem. *Complex Syst.* **13**(4), 297–345 (2001)
28. Neumann, F., Witt, C.: Runtime analysis of a simple ant colony optimization algorithm. *Algorithmica* **54**(2), 243–255 (2009)
29. Neumann, F., Witt, C.: *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Springer (2010)
30. Padberg, M., Rinaldi, G.: A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems. *SIAMR* **33**(1), 60–100 (1991)
31. R Development Core Team: *R: R Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2011). <http://www.R-project.org>. ISBN 3-900051-07-0
32. Sander, J., Ester, M., Kriegel, H., Xu, X.: Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining Knowl. Discov.* **2**(2), 169–194 (1998)
33. Smith-Miles, K., van Hemert, J.: Discovering the suitability of optimisation algorithms by learning from evolved instances. *Ann. Math. Artif. Intell.* **61**(2), 87–104 (2011)
34. Smith-Miles, K., van Hemert, J.I., Lim, X.Y.: Understanding tsp difficulty by learning from evolved instances. In: Blum, C., Battiti, R. (eds.) *LION*, vol. 6073, pp. 266–280. *Lecture Notes in Computer Science*. Springer (2010)
35. Smith-Miles, K., Lopes, L.: Measuring instance difficulty for combinatorial optimization problems. *Comput. OR* **39**(5), 875–889 (2012)
36. Stadler, P.F., Schnabl, W.: The landscape of the traveling salesman problem. *Phys. Lett.* **A161**, 337–344 (1992)
37. Sutton, A.M., Neumann, F.: A parameterized runtime analysis of evolutionary algorithms for the euclidean traveling salesperson problem. In: Hoffmann, J., Selman, B. (eds.) *AAAI. AAAI Press* (2012)
38. Vazirani, V.V.: *Approximation Algorithms*. Springer (2001)
39. Wegener, I.: Simulated annealing beats Metropolis in combinatorial optimization. In: *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP '05)*, vol. 3580, pp. 589–601. *Lecture Notes on Computer Science*. Springer (2005)