Humanlike Problem Solving in the Context of the Traveling Salesperson Problem

Alexandra Kirsch
Technische Universität München, Institut für Informatik (I9)
Boltzmannstr. 3
D-85748 Garching
kirsch@in.tum.de

Abstract

Computationally hard problems, like the Traveling Salesperson Problem, can be solved remarkably well by humans. Results obtained by computers are usually closer to the optimum, but require high computational effort and often differ from the human solutions. This paper introduces Greedy Expert Search (GES) that strives to show the same flexibility and efficiency of human solutions, while producing results of similarly high quality. The Traveling Salesperson Problem serves as an example problem to illustrate and evaluate the approach.

Motivation and Problem Statement

Most of us have had some fun with automotive navigation systems. In one instance on a drive of several hours duration the satnav constantly informed us that it was replanning the route because of congestion. However, we couldn't find any change in the planned trajectory nor was there a traffic jam. We conjectured that the route was replanned at a distance of several hundred kilometers, which we would reach in the course of some hours — and thus was of no interest at all.

This is a typical example of misunderstandings between machines and users that are due to different handling of everyday problem solving tasks. Because most such problems are NP complete, humans and machines face the same challenge of trading off solution quality with computational efficiency. The computational approach usually strives for optimality of the solution as long as the computations can be performed in reasonable time. The problems are usually assumed to be static, changes in the environment are simply considered as a new problem to compute from scratch. In contrast, human problem solving is extremely efficient and well-adapted to changing situations, while the solutions are most of the time still close to the theoretical optimum.

Newell and Simon [1972] have derived the general framework of search algorithms from the observation of human behavior. Even though pure decision-making as presented in this paper is not a cognitive system in its own right, it is undoubtedly an important ingredient of cognition. In addition, at the end of the paper, I describe some extensions to make the algorithm more flexible through learning, which is another important ingredient of cognition.

I propose Greedy Expert Search (GES) as an algorithm to solve problems in a more human-like way. An explanation for the efficiency of human problem solving could be that we don't "compute" complete, global solutions for a problem, but proceed in a stepwise manner. Without a complete solution in mind (but possibly some sketch of the further strategy), the reaction to changes in the environment is a lot more natural than with complete global solutions.

Based on this hypothesis GES strives to *reduce computational complexity* and allows for *flexibility*, while maintaining a *solution quality that is acceptable for everyday tasks*. By acceptable I mean that it is not too far from a theoretical optimum and that it resembles human solutions. Achieving the latter requirement is also important for human-machine interaction.

To illustrate and evaluate the approach I use the Traveling Salesperson Problem (TSP) as an example. A planar TSP (the only type used in this paper) gives a set of points in 2D space, which can be interpreted as cities on a map and the task is to connect these points in the shortest possible way, thus allowing a traveling salesperson to visit all the cities with minimum costs. The TSP is a well-known problem for computer scientists. As an NP complete problem it has received attention in the context of search and optimization algorithms. On the other hand, the TSP has also been examined by cognitive psychologists because it underlies a large set of real-world problems like planning a vacation trip or a shopping tour.

The work from psychology has been used as valuable input for the present work in two respects: 1) it provides theories about strategies and knowledge that seem to be used by humans to solve TSPs, and 2) it provides data on human solutions of TSPs, so that I can evaluate my approach with respect to similarity to human solutions.

This paper makes the following contributions:

- it introduces Greedy Expert Search (GES) as a general paradigm of an efficient search algorithm for everyday problems;
- it evaluates the GES algorithm for the Traveling Salesperson Problem, taking into account solution quality and similarity to human solutions.

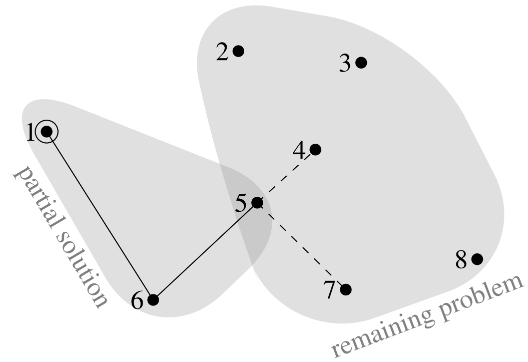


Figure 1: Illustration of GES solving a TSP. Point 1 is the start point, path 1-6-5 has already been decided on.

This work is explicitly not intended to develop a model of human thinking, nor is its purpose to develop exceptionally good solutions of TSPs (available algorithms are highly optimized and outperform humans on theoretical problem instances by far). The purpose of this work is an algorithm for a wide class of problems that occur in everyday situations and to produce solutions that fulfill human expectations. Because of the ample literature in psychology on the TSP, this is an appropriate test case, but it is not the only interesting domain (bin packing would be another example of an NP complete problem that occurs in everyday situations).

Approach

The motivation of Greedy Expert Search is to show similar *flexibility* and *efficiency* as humans. For the Traveling Salesperson Problem it has been measured that humans can solve such problems in near-linear "computation" time [Graham, Joshi, and Pizlo, 2000]. However, the TSP is an NP complete problem and with growing problem size the number of possible paths grows exponentially. The basic idea of GES (and greedy algorithms in general) for reducing effort is to treat each step of the path construction procedure as an independent problem. Thus each additional point only affects running time polynomially.

Building solutions step by step is not only efficient, but also leads naturally to high flexibility. For example, a standard AI planner generates plans from the beginning to the end and if the world changes, a new plan has to be generated or the original plan has to be changed. The philosophy of GES is not to generate a complete plan, but to decide the next action on the spot.

So far GES looks like a purely reactive approach and indeed it is a greedy algorithm as its name indicates. However, it makes extensive use of different forms and sources of knowledge to compensate for the drawbacks of the greedy search. So for a planning agent, a hierarchical structure of goals would be used as knowledge to guide the search. This knowledge is integrated into GES in the form of so-called *experts*. *Direction experts* filter the possible operators in each step, while *evaluation experts* assess the suggested operations to decide which one to execute next.

Figure 1 illustrates the General Expert Search algorithm using a Traveling Salesperson Problem as an example. A state for a TSP is composed of a partial solution that cannot be changed anymore and the remaining problem. In Figure 1 the path up to point 5 is already decided. To determine the next step of the solution, a set of direction experts is asked for good candidate points. Although all remaining points could potentially be added to the path, the direction experts in this example only consider points 4 and 7 worthy for further consideration. Then the candidate points are evaluated by different independent evaluation experts. One expert could favor point 4 because of its closeness to point 5, another might prefer point 7 as it follows the outer contour of the problem. Depending on the confidence of each evaluation expert and a combination function, one of the points is chosen and added to the partial solution. The process is repeated until there are no unconnected points in the remaining problem left.

Algorithm

I use the standard definition of search problems as described by Newell and Simon [1972], in which a search problem is defined by an initial state, a goal test and a set of operators with a successor function, which returns the expected next state when an operator is applied in a specific state. In addition a problem definition usually contains a cost function, which is replaced in GES by a set of experts.

Figure 2 shows the GES algorithm. In each search step GES determines promising operators, called *horizon*, by consulting the direction experts. This serves mainly to reduce the computational complexity, but can also serve as a filter to keep an agent from considering completely abstruse operators. Using the successor function, the expected resulting next states are evaluated by the evaluation experts. The operator that is rated most highly by the evaluation experts is executed, which results in the next state.

Beside the problem, this algorithm has as input a set of direction experts, a set of evaluation experts and combination functions for the horizons (i.e. the combination of the single results of the direction experts) and the state evaluation. The combination functions will not be considered closely in this paper. For the combination of horizons I use the union of the single horizons, and the evaluation combination function is a weighted sum of the results of the evaluation experts, each returning a value indicating the quality of the proposed state and a confidence (which corresponds to the weight).

Parametrization for the TSP

The Traveling Salesperson Problem is a well-known problem from the literature [Golden et al., 1980]. Given a set of nodes the task is to find the shortest path that includes all nodes at least once and returns to the initial node. In this paper the start node of the tour is given.

A prerequisite for GES to be efficient is the use of efficient experts. For the TSP I used simple experts, avoiding global computations involving the whole problem as much as possible. When global computations were unavoidable, they were mostly performed as an offline step before the search starts (this is done in the function *initialize* in the algorithm in Figure 2).

```
function GES (problem, direction-experts, evaluation-experts,combine-horizon, combine-evaluations)

  function GES-get-action(s)
    if problem.goal-test(s)
      then return null
      else
        ops ← combine-horizon(map(λe: e.get-operators(s,problem.operators), e ← direction-experts))
        ops-eval ← map(λo: cons (o, combine-evaluations(map(λe: e.evaluate(o,s), e ← evaluation-experts))),
                      o ← ops)
        a ← max(ops-eval :key second)
        return a

  foreach e in direction-experts e.initialize()
  foreach e in evaluation-experts e.initialize()
  s ← problem.initial-state
  loop
    a ← GES-get-action(s)
    if a=null
      then return
      else s ← execute(a)
```

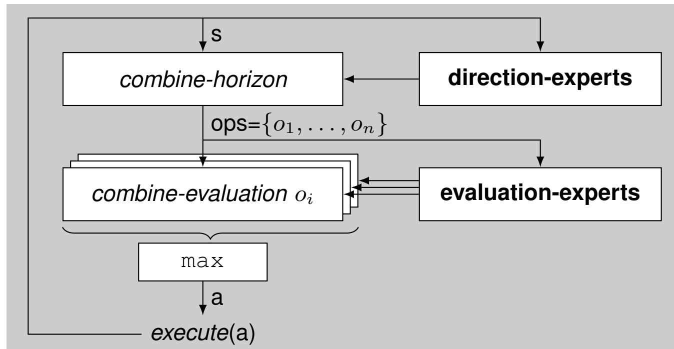


Figure 2: General GES algorithm. Keywords are shown in typewriter text, functions are in italics and complex data structures (e.g. objects) are typed in bold. The gray box illustrates the approach.

Each expert provides in addition to its result a confidence, depending on the specific problem. In this work, the confidence of each expert is provided manually. Ideally the confidence should be computable from the problem structure and could also be learned.

In the following I describe the direction and evaluation experts used for solving TSPs and their motivation from psychological findings about human TSP solving.

Direction Experts The task of the direction experts is to identify points that are worth considering in the next step of the solution. They can follow a global or local strategy: the NEIGHBORHOOD direction expert only has a local view, whereas the CONVEX-HULL and PINWHEEL experts follow a more global strategy.

The NEIGHBORHOOD direction expert is inspired by the well-known nearest neighbor heuristic for solving TSPs: it simply returns the n unvisited points that are closest to the last point of the current solution.

The two other direction experts are inspired by the statements of participants in various studies, for example by Tenbrink and Wiener [2009], in which the participants described their strategies as following a circle-like path. Based on this impression of human solution strategies and on the finding that an optimal tour follows points on the convex hull in order (which is equivalent to the fact that an optimal tour has no crossing paths) [Golden et al., 1980], algorithms using the convex hull have been suggested. The CONVEX-HULL expert used here first computes the convex hull offline and during the search suggests points that are in the area between the current point on the convex hull and the next one. In the situation in Figure 1 the CONVEX-HULL would not suggest point 4, because it is too far from the line between point 6 and 7 (the currently visited portion of the convex hull), nor would it suggest point 8, for instance, as this would skip point 7, which is the next one on the convex hull.

Also striving for a circular solution, the PINWHEEL direction expert follows an even simpler strategy: it computes the center of mass of the given TSP points and then suggests the points as they lie on a projected circle around this center. In the situation of Figure 1 it might suggest points 7 and 8. With a combination of a pinwheel and nearest neighbor strategy Best and Simon [2000] could explain human solution behavior to a large extent.

Evaluation Experts The task of the evaluation experts is to provide a value between 0 and 1 as a measure of the goodness of a proposed operator. The experts evaluate different aspects of a potential next state: 1) the quality of the partial solution, 2) the complexity of the remaining problem and 3) overall solution that is taking shape.

Nearest Neighbor Evaluation Expert The evaluation expert POINT-DISTANCE returns a measure of how far the possible new point is away from the last point of the current trajectory. Using only POINT-DISTANCE together with the NEIGHBORHOOD direction expert results in the well-known nearest-neighbor heuristic algorithm for TSPs.

Convex Hull Evaluation Experts Some evaluation experts are derived from algorithms using the convex hull as a global guideline. These experts don't mimic any of the algorithms from the literature directly, because those usually require some kind of global evaluation (for example for the cheapest insertion heuristic, the closeness of a point to each arc in the convex hull is calculated [Golden et al., 1980]).

The INDENTATION expert is based on the assumption that humans prefer solutions with few indentations [MacGregor and Ormerod, 1996]. I loosen this criterion to prefer small indentations to large indentations. This is similar to the largest angle insertion heuristic [Golden et al., 1980; MacGregor, Ormerod, and Chronicle, 2000], but without any global comparison to other hypothetical insertion operations. In Figure 1 when considering the point after point 6, point 5 would receive a low rating, whereas point 7 would be considered good.

A similar idea underlies the CHEAPEST-INSERTION expert. It is inspired by the cheapest insertion strategy [Golden et al., 1980; MacGregor, Ormerod, and Chronicle, 2000] but doesn't perform any global comparison. It favors the insertion of a point if the sum of the distances from the point to be considered to the current convex hull points (e.g. in Figure 1 the distances 6-5 and 5-7) is only slightly larger than the distance of the convex hull points (points 6 and 7 in the example).

A third convex hull heuristic comes from the assumption that problems with more points inside the convex hull are more difficult to solve for humans [MacGregor and Ormerod, 1996; Vickers et al., 2006]. The INNER-POINTS expert favors remaining problems with few inner points, leading to an early inclusion of inner points.

Other Evaluation Experts Following the assumption that humans choose partial solutions in a way that facilitates an easy completion of the remaining problem (also underlying the INNER-POINTS expert) I use the spatial expansion of the problem as a rough estimation of problem complexity, using the diagonal of the bounding box of the problem. This is not an ideal measure since it is not invariant to rotations, but it has the advantage of simplicity. The expert PROBLEM-DIAMETER relates the reduction of the problem diameter to the diameter of the current sub-problem, making a larger reduction preferable to a smaller one.

It has been observed in all studies that the solutions of humans seldom contain intersecting lines. This is not surprising as it has been shown that optimal solutions of TSPs cannot contain intersections [Golden et al., 1980]. There is an ongoing debate whether people avoid intersections deliberately or if the results are intersection-free because they are usually near the optimal solution [Vickers et al., 2003]. I have included an evaluation expert AVOID-INTERSECTIONS that rates a point to be inserted with 0 if it adds an intersection and with 1 if it does not. In most instances that this expert makes no difference, because intersections are usually caused by earlier decisions and at the moment when the intersection is produced there are only few or no alternatives left that would not lead to an intersection. This expert might be more effective if a lookahead mechanism for evaluation experts were added to GES.

A more indirect way of avoiding intersections is to avoid lines that "split" the problem, i.e. to avoid situations in which large parts of the remaining problem lie on both sides of the partial solution. The expert AVOID-SPLITTING calculates the number of points of the remaining problem lying on each side of the line that would potentially be added. In Figure 1 connecting point 4 to the current solution would result in splitting the problem (point 7 would lie on one side of the line from point 5 to 4, point 2 would be on the other side).

The FOLLOW-LINES expert is a simple implementation of the observation by Vickers et al. [2006] that the aesthetic form of a trajectory is somehow equivalent to good TSP solutions. I only consider straight lines as elements of aesthetic forms for reasons of simplicity, although humans surely use a more sophisticated pattern recognition. This expert would prefer point 4 as the next candidate in Figure 1 as it almost perfectly continues the line from point 6 to 5. In contrast, point 7 would be ranked low by this expert as it introduces a sharp bend.

Table 1: Used combinations of experts. The numbers in the evaluation expert fields indicate the confidences.

	direction			evaluation experts							
	NEIGHBORHOOD	CONVEX-HULL	PINWHEEL	POINT-DISTANCE	INDENTATION	CHEAPEST-INSERTION	INNER-POINTS	PROBLEM-DIAMETER	AVOID-INTERSECTIONS	FOLLOW-LINES	AVOID-SPLITTING
NN	✓			1							
CH-1		✓			0.3	1	1				
CH-2	✓	✓		1	1	1				1	
PW-1	✓		✓	1				0.2	1	1	1
PW-2	✓		✓	1				0.2	1	1	
PW-3	✓		✓	1				0.2			0.5

Evaluation

The claims of GES are flexibility, efficiency and acceptable solution quality. The first two points have been explained as design principles in the last section. But the most efficient algorithm is useless if the computed results are of low quality. Therefore we will now have a look at the solutions produced for TSPs.

Experimental Configurations

Table 1 shows the combinations of experts used for the evaluation. The configuration NN is a classical nearest neighbor approach. The CH-1 and CH-2 variants make use of the convex hull direction expert, for CH-1 all evaluation experts use the convex hull, whereas CH-2 is a mixed variant. PW-1, PW-2 and PW-3 use the pinwheel direction expert and evaluation experts that are independent of the convex hull.

The variation CH-2 was specifically designed as an efficient solution: all the experts work with constant runtime in each online computation step. Overall the computation is still quadratic because of the precomputation, but this combination shows that the linear function for the solution times of humans that Graham, Joshi, and Pizlo [2000] have observed might be explainable with GES using appropriate experts.

The expert combinations and confidence values shown in Table 1 have been determined informally on different data sets. Depending on the data set, the best combinations of experts and confidences varies significantly. The choice shown here are configurations that have worked well (on average)

Table 2: Comparison of GES solution quality to measurements of MacGregor and Ormerod [1996]. The problem identifiers are composed of the number of points in the TSP and the number of inner points (i.e. points that are not on the convex hull). The best result per problem is typeset in bold, the worst in italics.

								GES			participants	
Problem	NN	CH-1	CH-2	PW-1	PW-2	PW-3	min	mean	max	min	mean	max
10-1	**1.27**	2.47	2.47	**1.27**	*37.92*	**1.27**	1.27	7.78	37.92	0.00	0.80	9.69
10-2	**0.00**	16.77	16.77	21.89	*23.94*	3.88	0.00	13.88	23.94	0.00	3.51	23.12
10-3	**2.92**	6.12	3.89	10.22	*10.45*	4.55	2.92	6.36	10.45	0.00	2.76	15.76
10-4	**0.00**	*20.38*	4.88	4.88	14.88	7.75	0.00	8.79	20.38	0.00	3.77	11.71
10-5	8.98	*16.37*	14.14	12.25	**0.00**	8.98	0.00	10.12	16.37	0.00	3.05	14.40
10-6	11.52	17.36	**7.29**	17.07	13.44	*18.26*	7.29	14.16	18.26	0.00	3.84	16.08
20-4	**5.02**	7.16	7.24	*8.42*	*8.42*	**5.02**	5.02	6.88	8.42	0.52	2.91	15.94
20-6	9.29	**4.63**	9.90	7.74	19.60	*20.01*	4.63	11.86	20.01	0.00	6.03	20.26
20-8	4.07	**3.59**	9.23	9.23	9.23	*18.55*	3.59	8.98	18.55	1.70	5.11	14.46
20-10	*24.74*	11.64	22.56	13.39	19.57	**7.76**	7.76	16.61	24.74	1.45	3.29	6.61
20-12	**2.30**	8.97	23.42	14.65	*37.26*	9.23	2.30	15.97	37.26	0.62	4.56	20.36
20-14	**16.61**	19.84	21.35	*33.03*	31.03	19.31	16.61	23.53	33.03	1.84	9.61	25.18
20-16	17.34	24.71	20.59	*26.08*	18.92	**10.32**	10.32	19.66	26.08	1.00	8.48	23.70

over all data sets and were not specifically optimized for the problems discussed in this paper.

Comparison of Tour Lengths

The standard approach for comparing TSP solutions is the length of the trajectory, usually measured in percentage above the optimal tour length (PAO). I use a data set of MacGregor and Ormerod [1996] and compare the results obtained with GES to those of the participants in their study.

A small difficulty in the comparison is that the participants of MacGregor and Ormerod [1996] could choose the starting point of the trajectory, whereas in our case it was predefined. The human results might have been a little worse if the start point was given, but this should not have a strong effect.

Table 2 provides the PAO of each parametrization of GES as well as the minimum, maximum and average values of these solutions. For a comparison the respective values of humans from MacGregor and Ormerod [1996] are shown[1]. The best solution obtained with GES usually lies between the minimum and mean value of the participants, showing that comparable results to humans can be achieved, although not yet at the level of the best human solutions.

The values of each parametrization show that there are large differences in the quality of the solution found by each combination of experts. This explains why the mean value of the generated solutions are much larger than the average of the human solutions: some parametrizations are inappropriate for some problems. With a reliable mechanism to set the expert confidences for a given problem, this large fluctuation could be reduced.

A surprising finding is that the maximum values of GES are not much higher, in some cases even lower, than the worst human solutions. Since the expert combinations were not filtered for each problem, I had expected some solutions to be a lot worse than what humans would produce. The numbers, however, show that the GES algorithm with appropriate experts produces stable behavior, even if the expert combinations are poorly chosen.

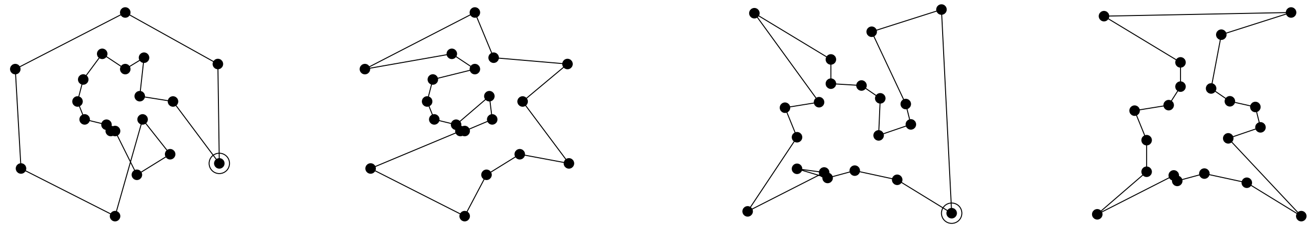
Visual Comparison

In their paper entitled "Some tours are more equal than others" Tak, Plaisier, and van Rooij [2008] argue that the tour length is often not a reliable measure for the likeness of a computed solution to a human solution. Since one of the goals of GES is to produce solutions that are understandable for humans, let's have a closer look at the forms of the solutions.

The last two lines of Table 2 show rather disappointing results for GES. Although these two problems also seem to be particularly demanding for humans, the best GES results are still far from the mean value of human solutions. The best computational solution I found was with an often-cited convex hull algorithm using a cheapest insertion strategy (CHCI) [Golden et al., 1980], which is a simple, yet efficient algorithm that disqualifies in its standard form as a human model because of its global approach (in contrast humans seem to construct solutions sequentially [MacGregor, Ormerod, and Chronicle, 2000]) and because it often outperforms humans. In the problem 20-14 it finds a solution with 13.61% above optimum and for 20-16 with 1.46% above optimum.

Figure 3 visualizes the best solution found with GES and the CHCI solution. Except for a small "loop" in each of the GES solutions, the CHCI solutions don't seem to be superior at first glance, although when calculating the path lengths a significant difference is visible. I would expect that the solutions found with GES would be accepted by people as a reasonable solution (except for the loops).

[1] The values given in MacGregor and Ormerod [1996] are contradictory: the PAO of the minimum and mean values do not always correspond to the absolute tour lengths (in particular for problems 20-4 and 20-6). I have used the PAO as calculated from the absolute tour length taken from tables 1 and 4 of MacGregor and Ormerod [1996].



(a) CH-1 (79.5) (b) CHCI (75.1) (c) PW-3 (65.5) (d) CHCI (60.2)

Figure 3: 20-node problem with 14 (left) and 16 (right) interior points from [MacGregor and Ormerod, 1996]. The numbers is parentheses are the tour lengths.

Next I compare the solutions found with GES to those found in the study of Tak, Plaisier, and van Rooij [2008] for one specific problem.[2]

But first of all, some words of caution are necessary. The data sets of Tak, Plaisier, and van Rooij [2008] have a very special design with lots of equidistant points. This leads in many cases to the algorithmic results being more or less random, depending on the order in which the points of the problems are represented internally. Therefore the following qualitative comparisons must be taken with a grain of salt. It is well possible that a slight change in the implementation could have large effects on the results. Also, the participants could choose the starting point of the tour like in the experiment of MacGregor and Ormerod [1996].

Figure 4 presents the problem CIRCLE and the most popular solution among the participants (Figure 4(a)). Surprisingly, the solution shown in Figure 4(b) is not listed by Tak et al. However, in a similar problem called SQUARE 11% of the participants came up with a similar circle-like solution and when asking a friend of mine to complete this problem, he also produced this solution with the comment "I'll just go in rounds".

The solution in Figure 4(b) was found by two configurations of GES: CH-2 and PW-1. In general, PW-1 typically produces solutions with very clear lines. Often this algorithm "forgets" to collect all relevant points in its outer tour and then has to take a second inner tour to include the remaining inner points. This general behavior might be due to the strategy of following a circle-like tour together with the urge to follow straight lines, resulting in clear lines but avoiding necessary deviations from these lines. Although both the PINWHEEL direction expert and the FOLLOW-LINES evaluation expert are also present in variant PW-2, it generated a different solution (Figure 4(c)). This might be due to the very tight decision at the start whether to move right or downward. As PW-2 doesn't use the AVOID-SPLITTING expert, both points are regarded as equally well.

The solution in Figure 4(d) produced by variant CH-1 is very long compared to the other solutions and no participant in Tak et al.'s study came up with a solution like this. Still a person might attribute some "strategy" to this tour: first going in a circle and then collecting the remaining points in a zigzag pattern. Zigzag patterns were indeed mentioned (although rarely) as human strategies in the study of Tenbrink and Wiener [2009]. This behavior also occurred with CH-1 in other problems, but interestingly not in the ones of MacGregor and Ormerod [1996], which have been specifically designed as challenging regarding inner points of a problem.

Discussion and Future Work

First I will discuss how far GES fulfills its goals of flexibility, efficiency and solution quality, specifically in the context of the Traveling Salesperson Problem. This is followed by a brief discussion whether GES could also qualify as a model for human problem-solving.

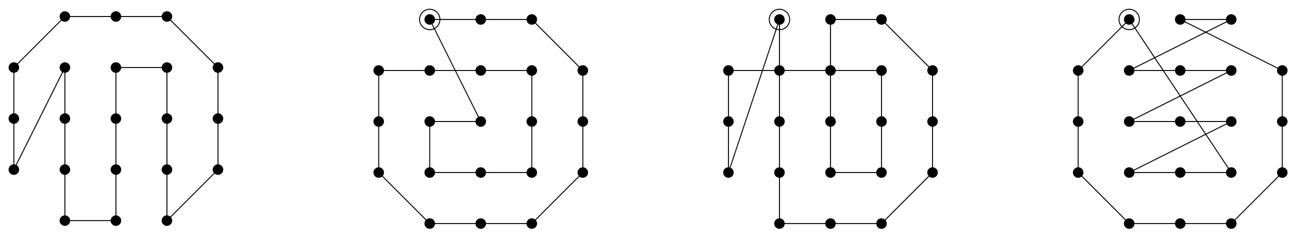
Flexibility and Efficiency

The flexibility to react to changes in the world is inherent in GES: since only the solution up to the current state has been computed, the next step is always chosen in the situation. The question for GES is rather how to avoid purely reactive behavior. In the TSP this question did not arise since the world didn't change. But the CONVEX HULL and PINWHEEL direction experts point into the direction of how to include global strategies into the local search process. In contrast to hierarchical search methods [Mock, 2002; Holte et al., 1996] I believe that the problem solving algorithm doesn't have to be hierarchical, but the knowledge used in the search should have some hierarchical structure.

Another form of flexibility is the adaptation to different problem instances or situations. The concept of experts in GES is the basis for such adaptiveness, but to exploit it fully, the confidences of the experts would have to be determined automatically. The most flexible way to do this would be machine learning. The main challenge for learning the confidences is the representation of problem types. For the TSP there are some possible features that might play a role to classify problems, for example the number of points that are not on the convex hull, the distance of such inner points from the "border" of the problem, the ostensibility of clusters or regions. But even for people it is difficult to classify which problems are easy to solve [MacGregor, Ormerod, and Chronicle, 2000; Vickers et al., 2003].

The computational efficiency for finding a solution was also a basic design principle. By shifting from a global view to local decisions, GES bears the potential of linear runtime

[2] For the comparison I used the publicly available data at `tsp.wtak.nl`. The tour lengths presented here have been converted with the factor 0.567 as indicated in [Tak, Plaisier, and van Rooij, 2008].



(a) 25% of participants (29.4) (b) CH-2, PW-1 (29.3) (c) PW-2 (32.5) (d) CH-1 (37.2)

Figure 4: Comparison of solutions using the data set CIRCLE from Tak, Plaisier, and van Rooij [2008].

for a complete problem, given that the experts return their decision at each decision step in constant time. In the TSP as presented in this paper none of the direction experts has a constant runtime, because in theory all remaining points could be candidates and even for finding the nearest next point, all remaining points have to be regarded (this step can be accelerated by a precomputation, but then the offline cost is quadratic). Also evaluation experts that take into account the overall form need linear computation time at each decision point. For good TSP solutions some global knowledge seems to be indispensable. These problems might be removed with parallel computations.

Solution Quality

The experiments with the TSP have shown that GES can produce solutions that are in the range of human solutions. Still, in some cases the results are not comprehensible from a human point of view (for example the completely unnecessary loop in Figure 3(a)). This is due to the lack of appropriate experts to judge the overall form of a figure. An interesting approach would be a library of common shapes such as circles, rectangles, zigzag lines, etc. to match those shapes with the developing TSP solution and to use them as direction and evaluation experts. This idea moves very much into the direction of image understanding and it must be checked if there are efficient algorithms readily available for such a pattern matching.

Another piece of knowledge that is missing in the experts used here is hierarchical or regional knowledge about the problem. Several studies in psychology [Wiener, Ehbauer, and Mallot, 2009; Tenbrink and Seifert, 2011; Kong and Schunn, 2007] have shown that people use information about clusters or regions (like countries or points marked in a certain color). Using such information would also be beneficial with GES.

In its present form, GES makes the rather strong assumption that a solution to a problem can be found in a reactive way without the need for backtracking. This mechanism seems to be sufficient for people in simple everyday activities like tooth brushing or preparing coffee as stated by Schwartz et al.: "Indeed, for the types of simple, routine tasks we study, it is appropriate to say that the action plan is elaborated *through* action, rather than prior to action" [Schwartz et al., 1991].

For more unusual tasks, it is obvious that humans do hypothesize and reconsider possible solutions, as shown in the experiments of Hayes-Roth and Hayes-Roth [1979]. For example, in the TSP to avoid line crossings, it would be helpful to sketch future solutions and validate if they contain crossings. Therefore, the next development of GES will contain the possibility to create future solutions and either use them as input for experts or as a backtracking mechanism. To some extent this contradicts the basic scheme of GES to achieve linear solution times by constructing the solution stepwise. But if this mechanism is used economically, its additional power should trade off the slightly higher computational effort.

A Model for Human Problem Solving

Even though GES is intended as an efficient algorithm that shows humanlike flexibility and solution quality, it might also serve as a basis for a psychological model to explain human problem solving strategies.

In the case of the TSP GES directly accounts for the observation of Tenbrink and Wiener [2009] that there are several mechanisms at work, which are used differently depending on the problem at hand. Table 2 shows how different combinations of experts are beneficial for different problems. Also the findings of Graham, Joshi, and Pizlo [2000] that people solve TSPs in near-linear runtime can be explained by GES.

Finally, and maybe most importantly, GES is not specifically designed for TSPs, but — if instantiated with appropriate knowledge — could also account for human problem-solving abilities in other domains and thus serve as a general model.

Related Work

This work builds on work of different areas including search, planning, behavior-based decision-making and cognitive architectures. Space only permits to mention some specific pieces of work that had a particular influence on the work in this paper.

The latest headline project of IBM, the Watson program that could beat expert human players in the game of Jeopardy, uses a similar approach as GES with experts providing candidates for answers and evaluating such candidates in parallel [Ferrucci et al., 2010]. Watson additionally uses machine learning to determine the weights of experts. The scope of the Watson project is a lot larger than the work presented here and the focus was more on outperforming humans than to imitate them.

For the specific domain of the Traveling Salesperson Problem, several theories of how humans solve such problems have been proposed. One promising hypothesis is the use of the convex hull as an outline for the solution and then to add inner points following a global optimum. This theory has been mostly dismissed by the observation that humans seem not to create a global plan, but rather to follow a kind of circle-like path [MacGregor, Ormerod, and Chronicle, 2000; Vickers et al., 2003]. Some researchers have tried to combine the convex hull with a step-wise connection of points [MacGregor, Ormerod, and Chronicle, 2000] similarly to using the convex-hull direction expert in GES, while others argue that human performance can just as well be explained without the convex hull theory following a more reactive strategy [Vickers et al., 2003].

Several theories propose a hierarchical approach of first determining a coarse route and then connecting the problem points along the coarse solution [Best and Simon, 2000; Kong and Schunn, 2007; Graham, Joshi, and Pizlo, 2000]. This can be seen as a generalization of the convex hull approach with different global strategies. In GES such mechanisms could be included by experts that identify point clusters.

Tenbrink and Wiener [2009] summarize these efforts after their own experiments very much to the point: "TSP-related strategies have hitherto often been treated as real alternatives that are mutually exclusive. Our results indicate that they may be better represented as a repertory of strategies and subprocesses that are available to humans when solving TSP tasks. The relative weight of each particular suprocess or strategy may differ substantially between individuals and subtasks."[Tenbrink and Wiener, 2009] This describes exactly the idea behind the use of different experts in GES.

Conclusion

I have introduced GES as a search algorithm that emphasizes the use of different knowledge sources in problem solving. The Traveling Salesperson Problem, which has received attention both from psychology and computer science, has served as an example to demonstrate the benefits of GES: flexibility, efficiency and acceptable quality of the search, which includes similarity to human solutions.

The approach heavily relies on problem-relevant knowledge and it has to be seen for other problems than the TSP how that knowledge can be represented and acquired.

Acknowledgments

With the support of the Technische Universität München — Institute for Advanced Study, funded by the German Excellence Initiative.

References

Best, B. J., and Simon, H. A. 2000. Simulating human performance on the traveling salesman problem. In *Proceedings of the Third International Conference on Cognitive Modeling*.

Ferrucci, D.; Brown, E.; Chu-Carroll, J.; Fan, J.; Gondek, D.; Kalyanpur, A. A.; Lally, A.; Murdock, J. W.; Nyberg, E.; Prager, J.; Schlaefer, N.; and Welty, C. 2010. Building watson: An overview of the deepqa project. *AI Magazine*.

Golden, B.; Bodin, L.; Doyle, T.; and Stewart, W. 1980. Approximate travelling salesman algorithms. *Operations Research* 28:694–711.

Graham, S. M.; Joshi, A.; and Pizlo, Z. 2000. The traveling salesman problem: A hierarchical model. *Memory & Cognition* 28(7):1191–1204.

Hayes-Roth, B., and Hayes-Roth, F. 1979. A cognitive model of planning. *Cognitive Science* 3(4):275–310.

Holte, R.; Perez, M. B.; Zimmer, R. M.; and Macdonald, A. J. 1996. Hierarchical a*: Searching abstraction hierarchies efficiently. In *Proceedings of the National Conference on Artificial Intelligence*.

Kong, X., and Schunn, C. D. 2007. Global vs. local information processing in visual/spatial problem solving: The case of traveling salesman problem. *Cognitive Systems Research* 8:192–207.

MacGregor, J., and Ormerod, T. 1996. Human performance on the traveling salesman problem. *Perception & Psychophysics* 58(4):527–539.

MacGregor, J.; Ormerod, T.; and Chronicle, E. 2000. A model of human performance on the travelling salesperson problem. *Memory & Cognition* 7:1183–1190.

Mock, K. 2002. Hierarchical heuristic search techniques for empire-based games. In *International Conference on Artificial Intelligence*.

Newell, A., and Simon, H. 1972. *Human Problem Solving*. Upper Saddle River, New Jersey: Prentice Hall.

Schwartz, M. F.; Reed, E. S.; Montgomery, M.; Palmer, C.; and Mayer, M. H. 1991. The quantitative description of action disorganisation after brain damage: A case study. *Cognitive Neuropsychology* 8(5):381–414.

Tak, S.; Plaisier, M.; and van Rooij, I. 2008. Some tours are more equal than others: The convex-hull model revisited with lessons for testing models of the traveling salesperson problem. *The Journal of Problem Solving* 2(1).

Tenbrink, T., and Seifert, I. 2011. Conceptual layers and strategies in tour planning. *Cognitive Processing* 12:109–125.

Tenbrink, T., and Wiener, J. 2009. The verbalization of multiple strategies in a variant of the traveling salesperson problem. *Cognitive Processing* 10:143–161.

Vickers, D.; Lee, M.; Dry, M.; and Hughes, P. 2003. The roles of the convex hull and the number of potential intersections in performance on visually presented traveling salesperson problems. *Memory & Cognition* 31(7):1094–1104.

Vickers, D.; Lee, M. D.; Dry, M.; Hughes, P.; and McMahon, J. A. 2006. The aesthetic appeal of minimal structures: Judging the attractiveness of solutions to traveling salesperson problems. *Perception & Psychophysics* 68(1):32–42.

Wiener, J.; Ehbauer, N.; and Mallot, H. 2009. Planning paths to multiple targets: memory involvement and planning heuristics in spatial problem solving. *Psychological Research* 73:644–658.