Simulating Human Performance on the Traveling Salesman Problem

Bradley J. Best (bjbest@cmu.edu), Herbert A. Simon
(Herb_Simon@v.gp.cs.cmu.edu)
Carnegie Mellon University
Department of Psychology
Pittsburgh, PA 15213, USA

Abstract

This paper reports two empirical studies of human performance on the traveling salesman problem (TSP) and several computational simulations of human performance. Human performance was studied using the computer as a collection device to track mouse movements and individual moves within the problem. Data were collected for problem sets used by MacGregor and Ormerod (1996) and for problems constructed using random distributions. The introduction of a computer as a data collection device did not influence the quality of solutions to the problem. Several simulations were constructed to test the use of global and local processing in solving the TSP. The best fitting simulations and the data suggest a solution method involving a rapid global approximate solution followed by local exact solution.

Introduction

The traveling salesman problem (TSP) consists of finding the shortest complete tour through a series of points (cities), starting and ending with the same point. The Euclidean TSP deals with a set of points in the plane. The TSP is classified as nondeterministic polynomial (NP), or hard, by complexity theory. As such, the time that is required by the most efficient known algorithm for solving the problem is an exponential function of the number of points in the problem. In addition, the TSP is also known to be "complete", meaning that each NP-complete problem is reducible to every other NP-complete problem.

In a previous study of human performance on the traveling salesman problem, MacGregor and Ormerod (1996) found that human solvers produced consistently good solutions to the problem in a rapid and efficient manner. They tested the hypothesis that it is the number of interior points in a TSP that indicates difficulty for human solvers instead of the number of overall points. They found evidence supporting this hypothesis and also concluded that the solution method used by human solvers was likely to be perceptually based due to the rapidity of solution and lack of noticeable individual differences.

Ormerod and Chronicle (1999) noted that goodness of figure judgments and the optimality of solutions are linearly related. They found evidence that global perceptual processing might explain rapid TSP solution and cautiously suggested that perception of the convex hull of a TSP might be the process on which human solvers base their solutions.

Graham, Joshi, and Pizlo (1999) reported that the time required by human solvers to solve TSPs is a linear function of the problem size. They found no relationship between human performance and the convex hull or the minimum spanning tree. They also constructed a computational model and tested it along with several representative heuristics for solving TSPs. The heuristics did not provide good fits to their human data but the model they proposed, based on a pyramidal algorithm, did provide good fits to normalized tour length performance and proportion of optimal tours. This algorithm uses successive refinement to generate solutions. Initially, a rough solution is determined using intensity modes representing clusters of points in the problem (starting with no more than three nodes encompassing the entire problem). Later iterations break intensity modes into submodes and insert these into the current tour. Eventually, all of the intensity modes correspond to actual points in the problem and the tour solution is complete.

It is possible that human solvers are using a method very different than those proposed above. There is currently disagreement about the role that perception of the convex hull may play, the amount of parallelism (i.e., global perceptual processing) and seriality involved, and methods of measuring a simulation of human TSP performance.

Experiments

We designed two experiments to collect human solution data by computer including mouse moves and clicks. The first experiment replicates the problem set used by MacGregor and Ormerod (1996). The second experiment uses problems with uniform random distributions of points. In this report mouse movements are examined at the level of selection of individual moves (clicks), while in future reports they will be examined in more detail.

The TSPs were presented one at a time on a computer screen. Each problem started with a dialog box centered on the screen asking if the subject was ready to start. By clicking the “OK” button on the dialog, they centered the mouse. Subjects then used the mouse to select nodes until completing a tour. They could not backtrack or undo moves.

Experiment 1

The first experiment uses the problem set from MacGregor and Ormerod (1996). These problems are based on a set of points constructed by randomly perturbing the vertices of a regular polygon, and then adding interior points corresponding to a smaller interior polygon (also perturbed). Optimal solutions of these problems have a fairly regular shape and suggest a circular solution method with occasional dips to pick up interior points. We collected these data to determine if the use of a computer as a data collection device changed the nature of TSP solutions from the paper and pencil version of the task.

Experiment 1 Results

Subjects from the current experiment 1 completed the 14 problems used in experiments 1 and 2 from MacGregor and Ormerod (1996). The results for both the previous and current experiments are presented in Table 1 for comparison:

Table 1: MacGregor et. al. (1996) experiments 1 & 2 and current experiment % above optimal

	Optimal length	MacGregor Exp. 1	MacGregor Exp. 2	Computer Solvers
Avg.	585.3	3.1	6.3	5.6

On 8 of the 14 problems, the current subjects produced better solutions than the MacGregor and Ormerod subjects though their mean performance was slightly worse (5.6% vs. 4.7% above optimal). Given the small difference and the overlapping distributions, we concluded that the introduction of a computer and the inability to backtrack did not greatly impact the quality of subject solutions (positively or negatively).

Experiment 2

In the second experiment 9 subjects worked on a problem set generated with uniform random x and y coordinates. These subjects are categorized as good, fair, or bad based on their performance. Good subjects are defined as those subjects that complete the problems with less than a 3% average deviation from the optimal path length across the problem set. Fair subjects are defined as those subjects that have a greater than 3% deviation from the optimal path length but less than 10%. Any subjects with greater than 10% deviation from the optimal path length are classified as poor (1 subject), and are not analyzed further here.

The first three subjects completed the problems by selecting a starting point, and continuing until completing a tour. The remaining 6 subjects were assigned a starting point to try to reduce solution variance to make comparison more straightforward.

Experiment 2 Results: Speed and accuracy

The good subjects and fair subjects had overlapping latency distributions. The good subjects' average solution took 23.8 seconds while the fair subjects' average solution took 22.8 seconds. There is no apparent speed accuracy tradeoff between the subject groups. Table 2 gives the average solution time by subject for each problem:

Table 2: Latency and % above optimal across subject groups					
Good subjects			Fair subjects		
Subject	Average latency	% above optimal	Subject	Average latency	% above optimal
s1	31.3	1.08	s5	20.6	4.58
s2	15.3	0.59	s8	27.5	3.58
s3	16.1	2.19	s9	20.2	4.04
s4	14.0	0.71			
s7	42.2	1.16			
Overall	23.8	1.15	Overall	22.8	4.06

In the first experimental condition subjects were free to choose their own starting point for the problem, while in the second condition, each subject was assigned a starting point. The main difference between the conditions is in the latency for the first two moves. For the "choose start" condition, subjects took 2306ms to make the first move, with the other moves latencies taking from 797ms to 997ms. In the "assigned start" condition, subjects took only 1892ms to make their first move, but then took 1554ms for the second move while the remaining moves took between 964ms and 705ms on average.

The latency pattern for the first two moves indicates that subjects may be planning their global solution strategy before making the first move in the choose start solution, while in the assigned start condition this planning time is distributed across the first two moves.

Simulations: Subject solution methods compared to simulations

Subjects appear to solve these TSPs by rotating around a central point and picking up points in the local region as they go. This suggests a global plan for direction of rotation and the general shape of the figure and local decisions that pick a path through the points of immediate concern. We developed several algorithms using global and local aspects of this strategy to simulate human performance on the problem.

Results are shown below for good subjects and these algorithms with the random problems. Algorithm performance is shown by: (a) comparing overall path length to the good subjects' path length and to the optimal path length and (b) determining how many of the subjects' individual moves can be guessed by the algorithm.

Algorithm descriptions:

Nearest neighbor: The nearest node is selected next unless it is also the starting node and other nodes are still open (tour completion constraint). This is a purely local strategy that produces results unlike tours completed by human solvers (MacGregor & Ormerod, 1996).

Pinwheel: A center of mass for the points is calculated first. The radial angle from the center of mass to the current point is calculated. The next point with the smallest amount of rotation around the center of mass is selected in the rotational direction of the simulation (clockwise or counterclockwise).

Local Search: Local search combines the global plan of rotation around a center of mass with the local plan of looking ahead at the next five points to select the shortest path to the final (fifth) point. On the first move, the algorithm only considers points with a positive rotation around the center of mass from the starting point (clockwise simulations). On following moves the algorithm additionally considers points that result in a small rotation counter to the current direction of rotation. For large problems (20 node) the allowable counter-rotation is 20 degrees and for small problems (10 node) it is 30 degrees.

Simulation Results

Simulation performance was evaluated in two ways. First, each algorithm predicted each subject's moves on each problem using the current subject's problem state. The measure used to evaluate performance in this case is the percentage of total subject moves that the algorithm correctly predicts (Tables 3 & 4). The second way the simulations were evaluated was by calculating the average overall path length for the algorithm run on the problem (averaged across all possible starting points), and then comparing this path length to the optimal path length, the subject performance, and the other algorithms (Table 5).

Predicting the next move

The nearest neighbor algorithm explains a large part of subject's choices while completing this task. Overall, 80% of all subject choices can be explained using this local processing decision. Broken down by subject, there are variations in the ability of this algorithm to predict subject moves, but the range is fairly narrow (76% to 84%).

The pinwheel algorithm predicts 78% of subject's choices. The pinwheel algorithm does equally well in predicting each subject (ranging from 77% to 78%). It does vary substantially by individual problem, however. For problem 2, it predicts 98% of subject moves, while for other problems the prediction rate is lower (ranging to a low of 57%).

The search algorithm, which combines aspects of global planning (rotation around a center) and local search for nearness, predicts fully 93% of individual subject moves. Broken down by subject, the algorithm's predictions range from 96% to 91%. Broken down by problem, algorithm performance reaches 100% on some problems and never goes below 82%, indicating that the algorithm is performing well on most problems.

Table 3: Percent guesses correct by subject			
Subject	Nearest	Pinwheel	Search
1	82	77	96
2	84	78	94
3	77	77	91
4	76	77	91
5	79	78	91
Average	80	78	93

Table 4: Percent guesses correct by problem														
Problem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Avg.
Subject % above	0.7	0.0	0.0	1.1	0.0	0.0	1.8	3.4	2.9	1.0	1.7	1.8	0.1	1.3	1.1
Nearest % guess	80	84	78	78	90	86	66	80	78	86	69	73	84	82	80
Pinwheel % guess	73	98	93	84	93	80	67	77	57	84	67	64	71	78	78
Search % guess	90	100	100	96	100	98	82	88	85	96	81	94	90	95	93

Overall path length

In Table 5, optimal length is given in abstract units used by the computer and '% above' shows what percentage the subjects or algorithm exceeded the optimal length by. The nearest neighbor and pinwheel algorithms in general perform an order of magnitude worse than human solvers of the problem (14.6% and 12.7% above optimal vs. 1.1% above optimal for the good subjects). The search algorithm produces results that average 4.2% above optimal. While not as good as the good human solvers, the algorithm does perform as well as the fair human solver of these problems (4.1% above optimal, see above).

Table 5: Comparison of overall path lengths generated by subjects and simulations									
Prob.	Optimal	Subject length	Subject % above	Nearest length	Nearest % above	Pinwheel length	Pinwheel % above	Search length	Search % above
1	1287	1296	0.7	1352	5.1	1418	10.1	1341	4.2
2	1026	1026	0.0	1137	10.8	1118	9.0	1046	1.9
3	906	906	0.0	968	6.9	938	3.6	919	1.4
4	954	964	1.1	1068	11.9	1045	9.6	954	0.0
5	984	984	0.0	1025	4.2	1080	9.8	984	0.0
6	933	933	0.0	1056	13.2	1059	13.5	953	2.1
7	1092	1110	1.8	1316	20.5	1297	18.8	1205	10.3
8	1560	1614	3.4	1880	20.5	1823	16.9	1765	13.1
9	1463	1506	2.9	1719	17.5	1651	12.8	1508	3.1
10	1353	1367	1.0	1621	19.8	1457	7.7	1373	1.4
11	1515	1541	1.7	1803	19.0	1758	16.0	1646	8.6
12	1521	1548	1.8	1805	18.7	1732	13.9	1562	2.7
13	1413	1415	0.1	1553	9.9	1701	20.4	1534	8.6
14	1286	1303	1.3	1622	26.1	1483	15.3	1309	1.8
Avg	1235	1251	1.1	1423	14.6	1397	12.7	1293	4.2

Discussion

MacGregor and Ormerod (1996) compared human TSP performance to heuristic algorithms including the Nearest Neighbor, Largest Interior Angle, and Convex Hull algorithms. They concluded that these heuristics did little to explain human performance, and that human solvers instead used a perceptual process to solve the problem.

The current results indicate that the Nearest Neighbor algorithm does explain a very large portion of human TSP solution methods. Although the overall solution obtained using a Nearest Neighbor algorithm does not produce solutions of the same quality as human solvers, a majority of the moves made by human solvers are to the closest point.

When the local processing constraints suggested by the Nearest Neighbor algorithm are combined with a global plan for the general shape of the solution, as in the search algorithm described above, the fit of the simulation to human performance is quite close. This fit is close at the level of overall path lengths as described in other research (e.g., Graham, Joshi, and Pizlo, 1999). Even more important, however, is the ability of the search algorithm to accurately predict individual subject moves.

The search algorithm presented here relies on a two-stage process, in which the first stage involves exploration of the entire problem (identifying the center of mass of all points in the problem) and the second stage involves locally working out the details. As a result, the second stage is completed in time proportional to the number of points. This provides an explanation for the time required for human performance, which is a roughly linear function of the number of points in the problem.

In addition to matching the human performance data collected, this two-stage process also receives support from empirical studies in other areas. Initial diversive exploration of the whole display followed by specific exploration of parts of the display was noted by Berlyne (1971). This pattern of visual search has also been recognized in eye-tracking studies of medical diagnosis of X-rays by radiologists (Nodine and Kundel, 1987) and initial perception of artwork (Locher and Nodine, 1987).

References

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

Graham, S. M., Joshi, A., and Pizlo, Z. (submitted). The traveling salesman problem: A perceptual model.

Locher, C. J., and Nodine, C. F. (1987). Symmetry catches the eye. In J. K. O'Regan and A. Levy-Schoen (eds.), *Eye movements: From physiology to cognition.* North-Holland: Elsevier Science Publishers.

MacGregor, J. N., and Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception and Psychophysics,* 58, 527-539.

Nodine, C. F., and Kundel, H. L. (1987). Perception and display in diagnostic imaging. *Radiographs*, 7, 1241-1250.

Ormerod, T. C., and Chronicle, E. P. (1999). Global perceptual processing in problem solving: The case of the traveling salesperson. Perception and Psychophysics, 61, 1227-1238.