Do Vision-Language Models Solve the Traveling Salesperson Problem Like Humans?

Richard Doan
School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, United States

Hugh Weston
School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, United States

Jacob Grinberg
School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, United States

Chao Tang
School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, United States

Sashank Varma
School of Interactive Computing, Georgia Institute of Technology, Atlanta, Georgia, United States

Abstract

Humans produce near-optimal (i.e., within ~10% of minimum-length) tours for small Traveling Salesperson Problem (TSP) instances of up to 120 points despite the problem's NP-hardness. The current study asks whether vision-language models (VLMs) employ human-like strategies when solving TSP instances. Using the stimuli and human data of Marupudi et al. (2022, *Cognitive Science Conference*), we found that Gemini 3 Pro, GPT-5.2, and Claude Opus 4.5 generally show the same performance as humans, including a linear increase in deviation from optimal tour length with increasing problem size. The VLMs also behave consistently with heuristics and strategies that have been documented for humans: convex hull, crossing avoidance, and divide-and-conquer via visual clustering. Finally, they show a linear increase in total tokens consumed with increasing problem size, paralleling the linear increase in human solution times with increasing problem size. Directions for future research include extending to larger problem instances and to adversarial examples.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*