A Heuristic Solution to the Closest String Problem Using Wave Function Collapse Techniques

Xu, Shirley (School: The Bishop's School)

The Closest String problem (CSP) is an NP-Complete problem which seeks to find the geometrical center of a set of input strings: given k strings of length L and a non-negative integer d, construct a solution string t, if it exists, such that the Hamming distance between t and each input string is no larger than d. This project proposes WFC-CSP, a novel heuristic algorithm inspired by Wave Function Collapse (WFC) techniques to solve CSP. Experimental results show that the WFC-CSP algorithm is highly reliable and efficient. The single iteration complexity of WFC-CSP is tractable with respect to number of strings k, string length L, and the alphabet size. Furthermore, the target maximum Hamming distance d does not affect the algorithm's complexity within an iteration. As more iterations are allowed, WFC-CSP's success rate of finding solution strings that satisfy the maximum Hamming distance requirement increases. In comparison to the Fixed-Parameter algorithm (FP-CSP) Gramm et al. proposed in "Fixed-parameter algorithms for closest string and related problems," the WFC-CSP algorithm is significantly faster when d is larger or equal to 16, while FP-CSP's runtimes become unviable. When compared to an Ant Colony Optimization-based metaheuristic approach to CSP (Ant-CSP) that Faro et al. proposed in "Ant-CSP: An ant colony optimization algorithm for the closest string problem," WFC-CSP offers a consistently higher success rate in finding solution strings, and, in many cases, also has a faster run time. The Closest String Problem has wide applications in the fields of computational biology and coding theory.

Awards Won:
Second Award of $2,000
American Mathematical Society: Third Award of $500
Mu Alpha Theta, National High School and Two-Year College Mathematics Honor Society: First Award of $ 1,500