The Minotaur and the Labyrinth

Sam Epstein

April 27, 2024

Imagine the following scenario. A hero is trapped in a labyrinth, which consists of long corridors connecting to small rooms. The intent of the the hero is to reach the goal room, which has a ladder in its center reaching the outside. The downside is the hero is blindfolded. The upside is there is a Minotaur present to guide the hero.

At every room, the Minotaur tells the hero the number of corridors n leading out (including the one which the hero just came from). The hero states a number between 1 and n and the Minotaur takes the hero to corresponding door. However the hero faces another obstacle, in that the Minotaur is trying to trick him. This means the mapping the Minotaur uses is a function of all the hero's past actions. Thus if a hero returns to the same room, he may be facing a different mapping than before. This process continues for a very large number of turns. The question is how much information is needed by the hero to find the exit? Using **Kolmogorov Game Derandomization**, we get the following surprising good news for the hero. Let c be the number of corridors.

The hero can find the exit using $\log c + \epsilon$ *bits.*

The error term ϵ is logarithmic and also is dependent on the mutual information of the entire construct with the halting sequence, which is negligible except in exotic cases.

The basic idea is as follows. Take a random hero who chooses a corridor with uniform probability. Then the hero is performing a random walk on the graph[1] (of the labyrinth). Assuming the number of turns is greater than the graph's mixing time, the probability the hero is at exit at the end is not less than $1/2c$. Then the following theorem can be applied. $\mathbf{K}(x)$ is the prefix Kolmogorov complexity of x. $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$ is the amount of information the halting sequence $\mathcal{H}$ has about x.

Theorem. *If probabilistic agent* $\mathbf{p}$ *wins against environment* $\mathbf{q}$ *with at least probability* p*, then there is a deterministic agent of complexity* $\mathbf{K}(\mathbf{p}) - \log p + \mathbf{I}(\langle p, \mathbf{p}, \mathbf{q}\rangle; \mathcal{H})$ *that wins against* $\mathbf{q}$.

An open problem is to find a simpler proof to this theorem than the current one.

[1] The graph is assumed to not be bipartite.