DeepGoogol: Learning Optimal Stopping Theory

Final Project Proposal ◇ Computational Cognitive Modeling

Isaac Haberman, Erica Detemmerman, Erica Dominic

I. Optimal Stopping Theory

In mathematics, optimal stopping is a theory concerned with choosing an action that maximizes expected reward or minimizes expected cost. The canonical example of an optimal stopping problem is the Game of Googol. In the Game of Googol, a sequence of *n* distinct integers between 1 and googol (10^{100}) is generated, with all valid sequences being equally likely. This sequence is presented to a player element by element. The player chooses when to stop iterating through the sequence, and wins if they stop on the largest number in the sequence. Importantly, the player cannot revert to a previously seen element. Say $n = 100$. If the player chooses when to stop randomly, the probability of winning is 1%, as the largest element has equal probability of appearing anywhere in the sequence.

To maximize the probability of winning the game, some simple calculations [1] prove that the optimal strategy is to observe *1/e* elements of the sequence and choose the next element that is larger than the previous maximum. Using this strategy, the player's probability of winning is $1/e \approx 0.37$. Not only is this a vast improvement on the 1% success rate, but this 37% success rate holds for all *n*.

This project will explore how cognitive models and humans learn (or fail to learn) the optimal strategy for the Game of Googol. To do so, we will build and test a few computational agents and present the game to a small group of human participants. Time permitting, we will also vary parameters of the game, such as the sequence length and the range of integer values.

II. Cognitive Modeling

To test and compare cognitive models for the Game of Googol, we will build a Python testing environment and bots for easy comparison and modeling. We propose testing the following algorithms:

1. Monte Carlo Methods
    a. First visit Monte Carlo
2. Temporal Difference Methods
    a. SARSA
    b. Q-learning
3. Deep Reinforcement Learning
    a. Deep Q-learning

The work of Becker, et al. [2] suggests that an agent can effectively learn the optimal stopping rule for the Game of Googol using Monte Carlo simulations. In Homework 2 we found that temporal difference methods can master similar games, and are computationally faster. Finally, since each state of a sequence has an explicit value and two actions, we believe that the game lends itself to a deep reinforcement learning agent. Using an RNN or one of its variants, we should be able to train an agent to effectively learn the optimal strategy.

III. Human Learning

In addition to testing several cognitive models, we will collect data from a small group of human participants. Each participant will be instructed on how to play the Game of Googol, but no information on optimal stopping theory. Each participant will play several games, for each game the player's choice and the game outcome will be recorded. With that information we will attempt to measure human learning (i.e. success rate) over time. Time permitting, we will vary certain factors of the game; for instance, the sequence length or the range of integers. Coster and Avabeck [3] showed that people undervalue trying another attempt, especially when the sequence length increases. This is caused by the pressure of time or a strong desire for the reward to be finally obtained [4]. Because the experimental situation and being observed will likely increase people's sense of time pressure, to counteract this added time pressure we will tell our participants at the beginning that the task is not timed.

IV. Sources

[1] Ferguson, Thomas S. "Who Solved the Secretary Problem?" *Statistical Science*, vol. 4, no. 3, 1989, pp. 282-296. https://www.math.upenn.edu/~ted/210F10/References/Secretary.pdf
[2] Becker, Sebastian et al. "Deep Optimal stopping." arXiv.org. 29 Jan 2019. https://arxiv.org/abs/1804.05394
[3] Costa, Vincent D., Averbeck, Bruno B. "Frontal–Parietal and Limbic-Striatal Activity Underlies Information Sampling in the Best Choice Problem", *Cerebral Cortex*, vol. 25, no. 4, April 2015, pp. 972–982. https://doi.org/10.1093/cercor/bht286
[4] Christian, "Optimal Stopping, How to Find the Perfect Apartment, Partner, and Parking Spot." https://medium.com/galleys/optimal-stopping-45c54da6d8d0

Outline

I. Optimal Stopping & Mathematical Theory
- **Source:** Optimal Stopping https://en.wikipedia.org/wiki/Optimal_stopping
- **Source**: Secretary Problem https://en.wikipedia.org/wiki/Secretary_problem
- **Source**: A Solution to the Game of Googol https://projecteuclid.org/download/pdf_1/euclid.aop/1176988613
- **Source**: Who Solved the Secretary Problem? https://www.math.upenn.edu/~ted/210F10/References/Secretary.pdf

II. Machine Learning
- Monte Carlo
- Reinforcement Learning
- **Source**: Deep Optimal Stopping https://arxiv.org/pdf/1804.05394.pdf
- **Source**: Q-Learning Algorithms for Optimal Stopping Based on Least Squares http://www.mit.edu/~dimitrib/lspe-optstop-ecc.pdf

III. Human Learning
- **Source**: Frontal–Parietal and Limbic-Striatal Activity Underlies Information Sampling in the Best Choice Problem (neuropsychology) https://academic.oup.com/cercor/article/25/4/972/336412
- **Source**: Optimal Stopping, How to Find the Perfect Apartment, Partner, and Parking Spot (funny human-centric article about the practical applications of optimal stopping) https://medium.com/galleys/optimal-stopping-45c54da6d8d0

Submission instructions: due Monday, April 1 (one half page written). Submit via email to instructors-ccm-spring2019@nyuccl.org with the file name lastname1-lastname2-lastname3-ccm-proposal.pdf