**Robo Navigation through Interactive Delayed Targeted Contrastive Learning**

*Tenzin Singhay Bhotia (tbhotia3) Sakya Kanuparthy (skanuparthy6) Muskan Bawa (mbawa6)*

**Motivation**

The ability of autonomous robots to adeptly navigate [3] complex, dynamic indoor environments is crucial yet challenging, particularly when faced with unpredictable elements like moving obstacles and changing conditions. While integrating human feedback into navigation systems has shown promise, immediate training on such feedback can limit the robot's adaptability and learning efficiency. This project seeks to enhance the navigation capabilities of autonomous robots by innovating within an established interactive learning framework: TAMER [1,2] through a novel approach that incorporates **delayed training**, **contrastive learning** [4, 5]**,** and **targeted learning**.

**Project Description**

We will base our navigation in a 2D simulated environment from a feasibility standpoint, and might transition to 3D if time permits. We propose a novel system where an autonomous robot, designed for indoor navigation, integrates human feedback not in real-time, but after accumulating a diverse set of experiences. This approach allows for the utilization of contrastive learning on the collected batch of data, enhancing the robot's ability to distinguish between effective and less effective navigation strategies. Additionally, the system will employ an adaptive strategy for integrating new data with existing datasets, guided by a probability distribution that prioritizes learning from more challenging or frequently failed scenarios. A sample algorithm for updated TAMER.

1. Initialize: Policy, feedback dataset D\_feedback, contrastive dataset D\_contrast, D\_seen.

2. Contrastive Pairs: Gather human feedback, form contrastive state-action pairs.

3. Weighted Data Mix: for each x in D\_feedback : weight = prob(x | difficulty(D\_seen(x)))

4. Delay Train: Apply contrastive learning on D\_feedback+D\_contrast -> update policy.

4. Update D\_seen: D\_seen+=D\_feedback

5. Loop: Repeat, refine policy.

**Data**

The primary data will consist of state-action pairs from the robot's navigation attempts, augmented with human feedback on the actions' effectiveness. Sensor data, including visual and spatial information, will provide a rich context for each state-action pair.

**Plan to Collect Data**

Data will be collected in a 2D simulated environment as a grid world. We may explore further real-world indoor settings if time permits, with dynamic obstacles and variable lighting. The robot will execute navigation tasks within this simulation, with human operators providing targeted feedback on its actions. This controlled setting allows for efficient, comprehensive data collection across a broad spectrum of navigation challenges.

**Expected Outcome**

We anticipate a navigation system that can learn more effectively and adaptively from human feedback. By integrating delayed contrastive learning, the system is expected to achieve a deeper understanding of navigation decisions and their consequences, leading to significant improvements in handling difficult or unfamiliar navigation scenarios. The adaptive data integration approach is designed to sharpen the focus on rectifying navigation deficiencies, thereby enhancing the robot's overall performance and reducing the need for human intervention.

**Identification of Benchmark**

Performance benchmarks will be established based on the robot's success in navigating through predefined routes within the simulation, encountering a variety of obstacles and challenges. Metrics will include the completion rate, navigation efficiency, and the frequency of required human corrections. We will compare the base algorithm of TAMER, and will also include the effect of acquisition functions. Since our algorithm utilizes three different techniques in sync we will conduct ablation studies [6] to measure their relevance.

**Timeline**

Week 1-2: Development of the simulation environment, and baselines.

Week 3-4: Incorporation of delayed training and contrastive learning modules into TAMER, Week 4-5: Initial data collection and feedback, testing adaptive data integration strategy.

Week 5-6: Extensive testing, data collection, and iterative refinement across all frameworks.

Week 6-7: Comprehensive performance evaluation against benchmarks -> project dissemination.

**Heilmeier's Catechism**

**What are we trying to do?** We aim to improve indoor navigation for autonomous robots by enhancing established interactive learning frameworks (TAMER) with delayed contrastive learning and automatic strategic data integration.

**How is it done today, and what are the limits?** Current systems utilize real-time human feedback for immediate training, which can limit adaptability and efficiency, especially in dynamic environments.

**What's new in our approach?** The integration of delayed contrastive learning and adaptive data prioritization represents a significant innovation within the TAMER framework, promising substantial improvements in navigation capabilities.

**Who cares?** This advancement will be of interest to robotics researchers, AI practitioners, and industries that rely on autonomous navigation solutions.

**If successful, what difference will it make?** The project promises to significantly enhance the adaptability, efficiency, and overall performance of robotic navigation systems in complex indoor environments.

**What are the risks and payoffs?** The primary risk lies in the potential complexity and resource demands of the proposed system. The payoff includes groundbreaking improvements in autonomous robot navigation.

**How much will it cost?** Utilizing existing resources and simulations minimizes costs, with additional expenses potentially arising from software or hardware upgrades.

**How long will it take?** Designed to fit within a 8 week timeline, this project allows for thorough development, testing, and evaluation, aligning with academic semesters for feasible completion.

**What are the mid-term and final “exams” to check for success?** Mid-Term -> completed environment development/selection and updated TAMER algorithm. Final-Term -> complete results, with comparisons and benchmarks.

**References**

[1]: Knox, W. Bradley, and Peter Stone. "Tamer: Training an agent manually via evaluative reinforcement." 2008 7th IEEE international conference on development and learning. IEEE, 2008.

[2]: Knox, W. Bradley, and Peter Stone. "Interactively shaping agents via human reinforcement: The TAMER framework." Proceedings of the fifth international conference on Knowledge capture. 2009.

[3]: Wigness, Maggie, John G. Rogers, and Luis E. Navarro-Serment. "Robot navigation from human demonstration: Learning control behaviors." 2018 IEEE international conference on robotics and automation (ICRA). IEEE, 2018.

[4]: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

[5]: Eysenbach, Benjamin, et al. "Contrastive learning as goal-conditioned reinforcement learning." Advances in Neural Information Processing Systems 35 (2022): 35603-35620.

[6]: Meyes, Richard, et al. "Ablation studies in artificial neural networks." arXiv preprint arXiv:1901.08644 (2019).

Todo

1. Piazza post asking to reuse tamer pset for our project (will modify significantly ofcourse)
2. 29th - 4:40-5:30 pm -> paper discussion
   1. Allot tasks (slides, paper sections)
3. 30th -> section distribution
   1. Muskan -> 1st and 2nd (Intro + RW)
   2. Sakya -> 3rd and 4th (Transformers + StructFormer)
   3. Tenzin -> 5th and 6th (Data Generation + Experiments)
4. 31st (sunday) ->
   1. Questions - group
      1. Drawbacks
      2. Strengths
      3. Improvement suggestions
      4. Confusions??
   2. Create own slides - individual
   3. Create own script - individual
   4. Create 2 questions each person - individual
      1. Weakness of structformer -> collisions (refer to diffusionformer).
   5. and meet again (eve/night)
   6. 6-7: 1 hr
5. 1st ->
   1. Submit final slides in canvas
   2. Send attendance quiz
   3. Run through 2-3 times,
   4. polish script,
   5. decide (interactive discussion questions etc.)
6. 2nd April -> Project -> Setup environment and repo

Slide Rubrics (<https://docs.google.com/presentation/d/1AK-sj-_cTSZzGyS-tUN-W2Zn-OJnx1cqkEdSU1Nz_ZA/edit#slide=id.g2c7ccbaf5aa_0_27> )

1. Include lots of interactive visuals
2. Less text
3. Crisp to the point slides (main ideas)

Content Rubrics

1. Gripping content (no sleeping audience)
2. Simplify complex concepts/equations
   1. First principles
   2. Eg.
      1. Explain point cloud first,
      2. Explain transformers (self-attention, encode-decoder, “autoregressive”)
      3. Then structformers.
3. Pose Rhetorical question and answer yourself
4. Include all images in papers at least and video clips.
5. Think 1-2 steps further why any equation/concept/decision is like it is in the paper

Project…

1. Contrastive Learning
   1. Find resources -> blogish/code/library/video
   2. Types of algorithm to implement and test
2. 4/11 -> preliminary results with at least
   1. Complete implementation by 4 / 6th
   2. Coach implemented
   3. One contrastive Learning
   4. Skeletal structure set
   5. Preliminary Results (vs baseline)
3. 4/30th -> complete full project implementation
   1. Code
   2. Results (proper either defeat baseline/die trying (3-4 approaches)
   3. Report
4. Last 2 days for presentation prep