CS 461 - ARTIFICIAL INTELLIGENCE

Initial Term Project Proposal

Fall 2023-2024

Project Name

InVista: Inverse Reinforcement Learning for Vision-Language Navigation

Members

- Zeynep Hanife Akgül - 22003356
- Arshia Bakhshayesh - 22001468
- Huzaifa Huzaifa - 22301342
- Emre Karataş - 22001641
- Faaiz Khan - 22001476

1. Project

1.1 Description

Vision-Language Navigation (VLN) is the task of training AI systems to navigate and interact with an environment based on visual inputs as well as natural language instructions.

This project's primary objective is to address the challenge of training an autonomous agent to navigate indoor environments using a Room-to Room (R2R) dataset, locate specific objects, and process natural language to understand coarse instructions while using Inverse Reinforcement Learning (IRL). IRL, analogous to RL, is a framework that aims to find out the reward function of an agent by observing its behavior in a given environment.

The project deliverables include a trained agent capable of interpreting expert behavior, understanding complex scenarios, processing natural language instructions, and making informed decisions in real time, ultimately achieving successful indoor navigation and object localization.

The significance of this project lies in its innovative approach to integrating IRL techniques, allowing the agent to actively seek human guidance when faced with uncertain or challenging situations, demonstrating the potential of IRL in creating adaptive and intelligent autonomous systems for complex tasks.

The technologies that will be used in this project (tentative; may be changed in the future) are the Matterport3D indoor R2R dataset [1] and Python to perform IRL, amongst other things. This project will build on already existing codebases and established frameworks to support the IRL and VLN tasks - i.e. Natural Language Processing (NLP), etc; ensuring a solid foundation for the future.

1.2 Goals

The goals of this project are to:

- successfully initialize the simulation environment and establish a basic natural language processing module for understanding simple sub-tasks,
- develop and implement the Inverse Reinforcement Learning (IRL) framework, integrating expert demonstrations and simulated behaviors into the agent's training pipeline,
- compare the use and effects of IRL versus other technologies like Imitation Learning and Behavioral Cloning,
- enhance the natural language processing module to improve the agent's understanding of nuanced instructions, enabling it to handle a broader range of coarse language inputs,
- evaluate the trained agent on VLN evaluation metrics, and
- provide a trajectory visualization for some example language-based navigation episodes completed by the trained agent.

2. Literature

The focus of this project is a paper that introduces Vision-based Navigation with Language-based Assistance (VNLA), which involves guiding an agent with visual perception via language to locate objects within photorealistic indoor environments. A comprehensive framework called Imitation Learning with Indirect Intervention (I3L) is developed to model the language-based guide. In this particular framework, the agent is required to locate a specific object inside the environment. The agent spawns at a random location inside the environment. It then tries to find out the desired object. If it fails to do so or is lost, it signals its instructor for additional guidance. The instructor then responds with a sub-goal instruction, which the agent can prepend to the original instruction to make its way out to the object [2].

3. References

[1] Matterport3D: Learning from RGB-D data in indoor environments, https://niessner.github.io/Matterport/ (accessed Oct. 23, 2023).

[2] K. Nguyen, D. Dey, C. Brockett, and B. Dolan, “Vision-based navigation with language-based assistance via imitation learning with indirect intervention,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. doi:10.1109/cvpr.2019.01281