Добрый день! Меня зовут Динь Нгок Туан, студент группы R34372. Я представляю мою тему: Исследование алгоритмов оптимального управления, основанных на обучении с подкреплением.

Структура моей презентации включает в себя следующие части:

**Liệt kê ra**

**Актуальность**

Актуальность включается в том, что Оптимальные регуляторы обычно разрабатываются с использованием полных знаний о динамике системы. В данной части работы был исследован алгоритм обучения с интегральным подкреплением (IRL) для построения оптимального адаптивного регулятора.

**Постановка задачи**

Мы рассмотрим нелинейную динамическую систему с непрерывным временем. Функция ценности определена как выражение 2 и 3. Уравнение 3 имеет точную форму уравнения Беллмана с дискретным временем.

Целью является построение оптимального адаптивного регулятора с использованием обучения с подкреплением.

**IRL с использованием итерации по стратегии**

Далее был представлен алгоритм итерации по стратегии. Сначала нам нужно выбрать некоторую допустимую стратегию управления Начиная с j = 0, повторяйте по j до сходимости.

Для оптимизации управляющего сигнала (стратегия) нам необходимо минимизировать значение функции ценности.

У нас две этапе: оценки стратегии и улучшение стратегии. Для применения этого алгоритм мы использовали Функции аппроксимации значения.

С помощью Функции аппроксимации значения мы переписали форму (4) в такой вид нейтронных сети как показана на компоненте «критик». Параметр веса W неизвестен, Вектор активации можно вычислить данными, измеренными вдоль траекторий системы.

Структура реализации алгоритма итерации по стратегии представлена ​​на этих двух рисунках.

Критик работает в медленном масштабе и изучает ценность использования текущей стратегии. Когда Критик сходится к решению для параметров W, стратегия управления Атера обновляется, чтобы получить улучшенное значение.

**Итерации по стратегии для линейного квадратичного регулятора**

В этом случае параметр W выбирается по формуле (11), вектор регрессии phi определяется по формуле (12) с помощью кнокеровского произведения.

**Моделирование**

Моделирование для перевернутого маятника и тележки успешно выполнено. Подобрана формула для параметра W и вектора регрессии. Решение задачи алгебраического уравнения Риккати и Результаты адаптивного регулятора соответствуют друг друга.

Мы видим, что значения матриц K и P сошлись к оптимальному решению, как и в случае LQR, не зная, какая матрица A.

**Заключение**

В будущее время я планировал использовать этот алгоритм для нелинейной системы и исследовать алгоритм итерации по критерию, который может преодолеть недостаток алгоритм итерации по стратегии.