Лицей «Физико-техническая школа»   
Санкт-Петербургского Академического университета

Отчет о практике

*Исследование взаимодействия алгоритмов обучения нейронных сетей без обратного переноса весов с обучением с подкреплением*

Работу выполнили:

Ученики: Лавров Дмитрий (11 Б класс),

Суворов Вячеслав (11 Б класс).

Научный руководитель:

Шпильман Алексей Александрович.

Место прохождения практики:

НИУ ВШЭ.

*Санкт-Петербург 2020*

Аннотация

В процессе научной работы мы изучили weight transport problem, реализовали различные алгоритмы обучения искусственных нейронных сетей и различные модели обучения с подкрепления. Было проведено тестирование алгоритмов на датасете MNIST и в применении к RL моделям. На основании проведенных экспериментов был сделан вывод о эффективности использования биологически правдоподобных алгоритмов в различных RL моделях.

Введение

Backpropagation (backprop или BP) - метод вычисления градиента, который используется при обновлении весов в искусственных нейронных сетях. Backpropagation является эффективным подходом для обучения глубоких нейронных сетей. Однако биологический нейрон не может получить информацию о своих аксонах, но при таком обучении активно используется информация о связях между нейронами. Из-за этого мозг не может использовать этот алгоритм обучения. Backpropagation имеет три биологически неправдоподобных требования:

1. Вес обратной связи должен быть таким же, как и вес прямой связи
2. forward и backward passes требуют различных вычислений
3. Информация о градиентах ошибок должна храниться отдельно от информации об активациях

В нашей работе мы используем биологически правдоподобные алгоритмы, далее мы будем их называть downstream weights алгоритмы (DW). Они не используют weight transport, но достаточно схожи с backpropagation.

Постановка задачи

Нами решалась задача написания биологически правдоподобных алгоритмов, в основе которых лежат знаниях человека о нервной системе, которая является быстрым и эффективным средством для анализа и обработки информации. Для анализа результатов работы мы провели ряд экспериментов, оценивая работу алгоритмов на различных наборах данных на различных RL моделях. После проведения всех экспериментов нашей задача была сделать вывод о эффективности алгоритмов.

Методика выполнения работы.

Здесь представлены биологически правдоподобные алгоритмы feedback alignment, weight mirrors и Kolen-Pollak. Но для начала рассмотрим самый примитивный алгоритм, не имеющий weight transport - Feedback alignment.

Feedback alignment (FA). В данном алгоритме мы создаем случайную матрицу B, которую мы не изменяем. Теперь, при обратном проходе мы используем не матрицу весов, а матрицу B (случайную). Данный алгоритм является биологически правдоподобным, но он плохо работает с очень глубокими и сверточными сетями из-за случайности весов B.

Kolen-Pollack algorithm (KP) - модифицированная версия FA. Обратное распространение происходит так же, он также обновляет матрицу обратного распространения.

Weight mirrors (WM). Цель алгоритма - построить матрицу так, чтобы она стала пропорциональна транспонированной матрице весов. Матрица обратного распространения стремится к матрице прямого распространения. Происходит поочередное обновление двух матриц - это и есть отличие этих двух алгоритмов

Обучение с подкреплением (англ. reinforcement learning или RL) - один из способов машинного обучения, в ходе которого испытуемая система (агент) обучается, взаимодействуя с некоторой средой. Приведем примеры RL моделей. Есть 2 основных группы. Value based methods (Q-learning, Deep Q-learning) и Policy based methods (A2C).

Q-learning - метод обучения с подкреплением. Алгоритм на основе полученных из среды данных формирует функцию полезности Q. На основе этой функции он может не случайно выбирать полезные действия, учитывая накопленный со средой опыт. Одно из преимуществ Q-обучения — то, что оно в состоянии сравнить ожидаемую полезность доступных действий, не формируя модели окружающей среды.

DQN (Deep Q Network) - модифицированный Q-learning. В этом алгоритме для построения Q-функции используется нейронную сеть.

DDQN (Deep Double Q Network) - модифицированный DQN. Он решает проблему переоценки награды в алгоритме DQN с помощью изменения алгоритма обучения.Это первый алгоритм с двумя классами для искусственных нейронных сетей.

A2C (Actor-Critic Methods) - метод обучения с подкреплением. Данный метод использует две нейронных сети. Actor, который измеряет, насколько хорошо предпринятые действия (он действует). Critic который оценивает действие нашего агента.

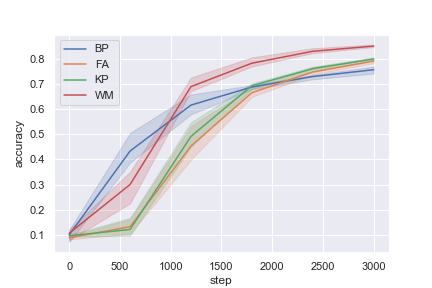
Для реализации и сравнения алгоритмов обучения с подкреплением была использована библиотека gym. Эта библиотека предоставляет пользователю среду для обучения агента.

Нашей задачей была реализация алгоритмов обучения нейронных сетей. Они были реализованы с использованием библиотеки PyTorch. DW и backpropagation мы обучали на датасете MNIST (рукописные цифры), а также на разных группах RL моделей. После реализации и обучения алгоритмов мы сделали выводы целесообразности использования биологически правдоподобных алгоритмов в RL по сравнению с backprop. Датасет MNIST мы выбрали так как это хорошо изученный датасет. На котором можно эффективно обучать модели различной архитектуры

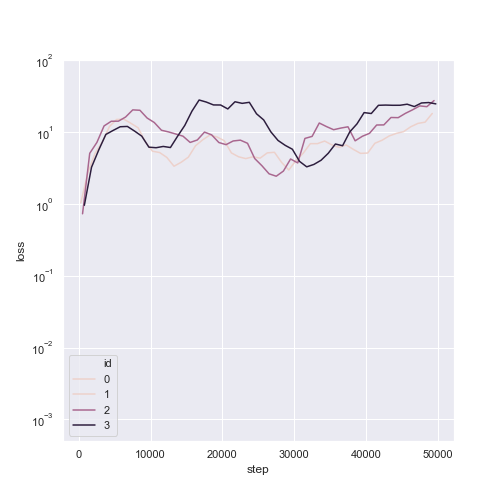
Результаты

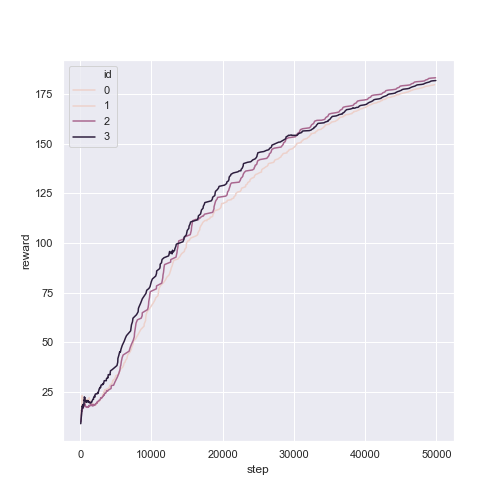
Для отображения графиков была использована библиотека matplotlib с frontend seaboard. Данные были сглажены экспоненциально взвешенным скользящим средним.

1. Для датасета MNIST

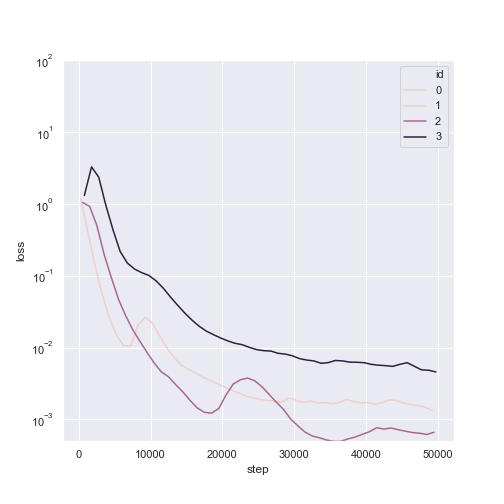


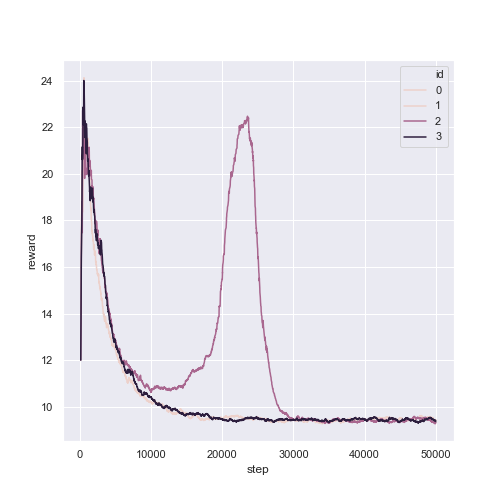
1. DQN
   1. Backpropagation



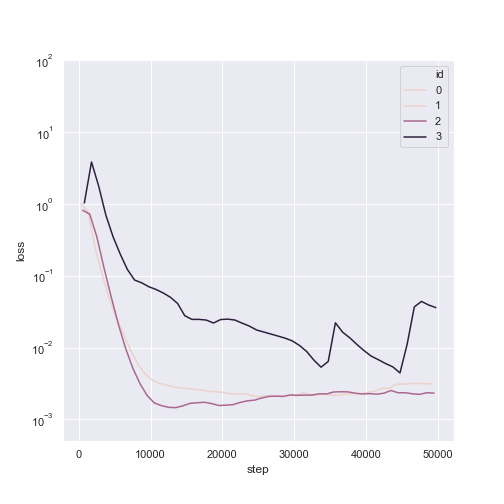


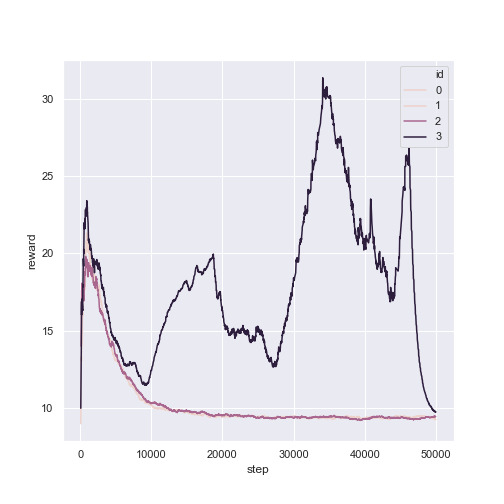
* 1. Feedback alignment



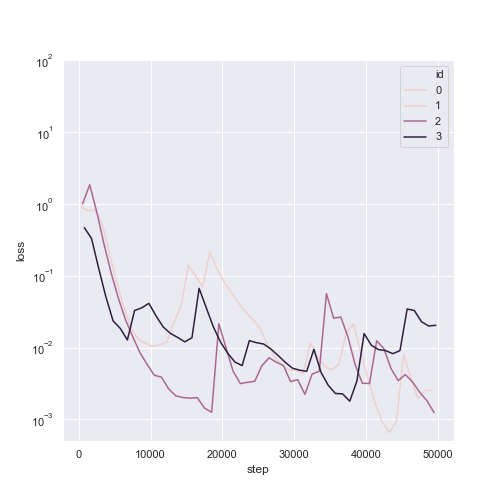


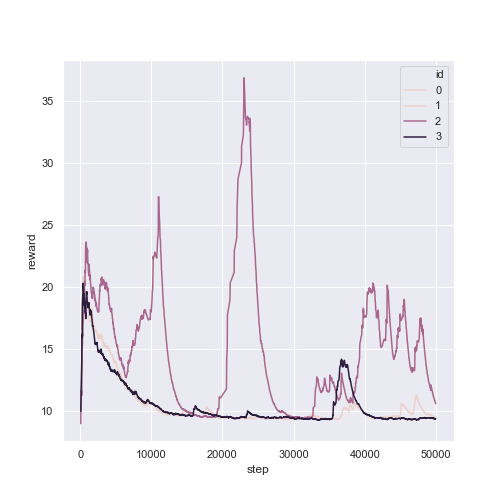
* 1. Kolen-Pollack



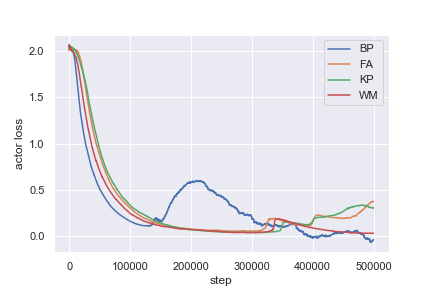


* 1. Weight Mirrors

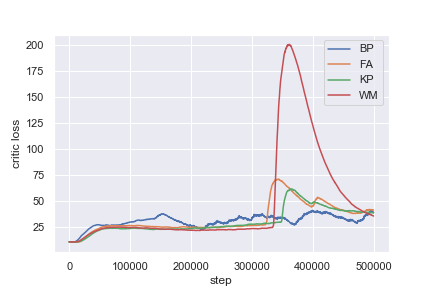




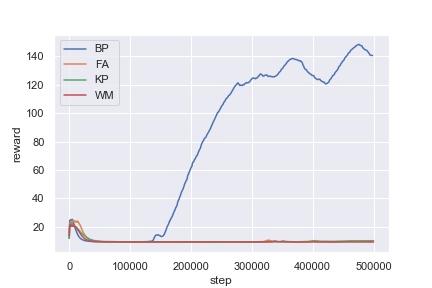
1. A2C
   1. Actor



* 1. Critic



* 1. Reward



Вывод

На датасете MNIST все алгоритмы обучались с незначительным отличием в скорости. Что говорит о работоспособности каждого из алгоритмов.

В DQN у агентов с нейронными сетями без обратного переноса весов loss стремился к , однако награда за эпизод не поднималась выше 50, что говорит о том, что агенты фактически не обучались. Средний максимум предсказания награды агентами оставался не больше 2 на протяжении всего обучения. Агент с нейронной сетью обучающейся через backpropagation обучился успешно.

В A2C награда за эпизод для агентов с нейронными сетями без обратного переноса весов также оставалась низкой.

Backprop хорошо подстраивается под изменение входных данных, это происходит в связи с тем, что у backprop моделей нет обучаемых параметров, отвечающих за обратное распространение ошибки. Мы предполагаем, что в RL моделях, в связи с особенностью алгоритма обучения, weight mirrors работает хуже backprop. Это происходит в связи с тем, что матрица обратного распространения подстраивается под входные данные.

Список литературы

1. Mohamed Akrout, Collin Wilson, Peter C. Humphreys, Timothy Lillicrap, Douglas Tweed. “Deep Learning without Weight Transport.”
2. Deep Mind. “Human Level Control Through Deep Reinforcement Learning.”
3. LeCun, Y. & Cortes, C. (2010), 'MNIST handwritten digit database'.

Ссылки:

1. freeCodeCamp. “An intro to Advantage Actor Critic methods” <https://www.freecodecamp.org/news/an-intro-to-advantage-actor-critic-methods-lets-play-sonic-the-hedgehog-86d6240171d/>

2. PyTorch documentation.

<https://pytorch.org/docs/stable/index.html>