

Отчет

Используемая среда:

cart-pole

Реализованная кастомная память:

**Persistent memory tokens
(далее token)**

Проведенные эксперименты:

**Рассмотрено влияние
кол-ва эпох, длин
контекстного окна и
размера батча на
обучение и процент
успешных эпизодов**

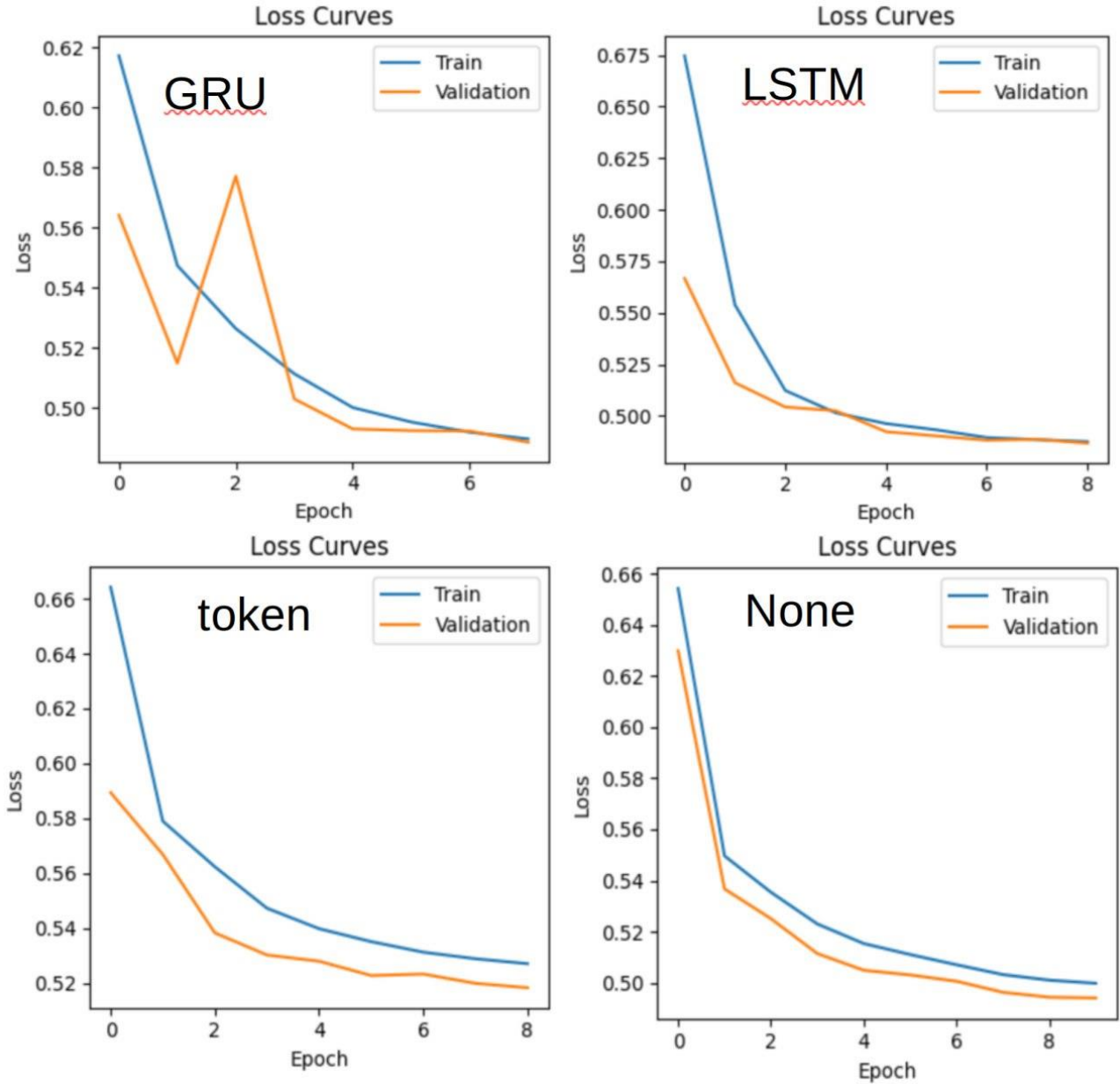
Описание выбранного метода памяти + ответы на вопросы из задания

- ✓ Память в форме Persistent Memory Tokens обрабатывает сжатое представление исторического контекста эпизодов. Эти токены не зависят от текущего входного контекста и представляют собой глобальные обучаемые векторы, которые взаимодействуют с последовательностью через механизмы внимания (self-attention).
- ✓ Persistent memory tokens добавляются к входной последовательности перед подачей в Decision Transformer. В архитектуре Transformer они участвуют в self-attention слое так же, как и обычные токены (state, action, return-to-go), но сами не зависят от конкретной последовательности, а являются обучаемыми параметрами модели.
- ✓ Обучение стандартное – как исходное
- ✓ Архитектура – nn.Transformer

Результаты обучения на дефолтных параметрах

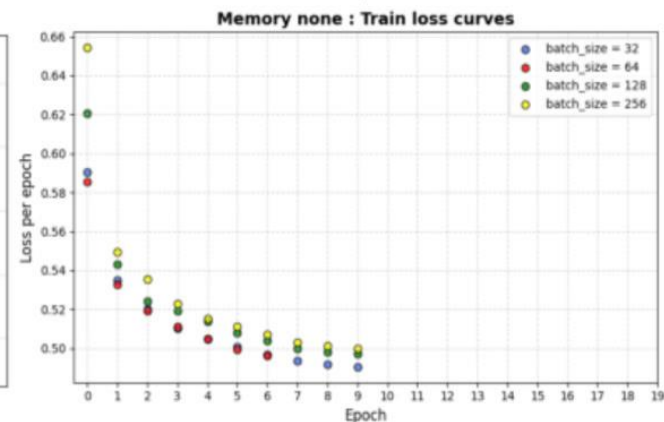
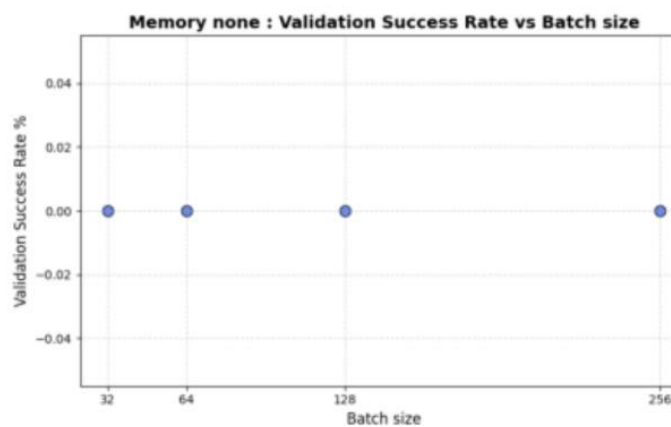
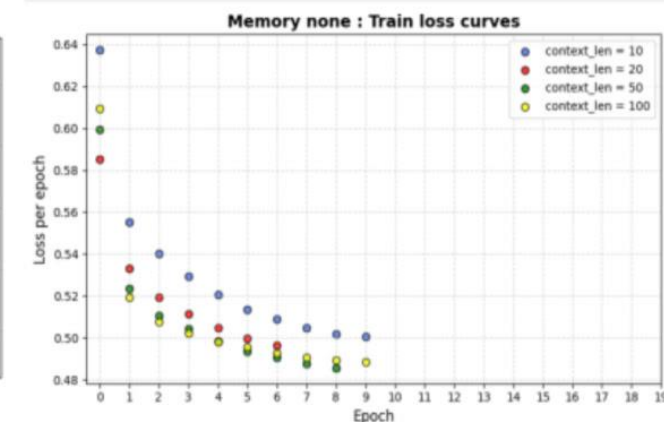
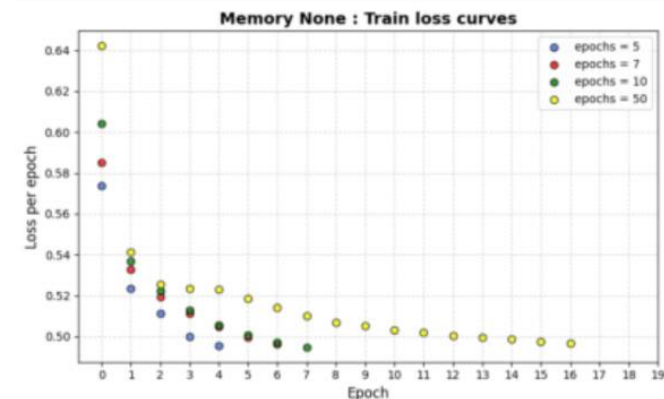
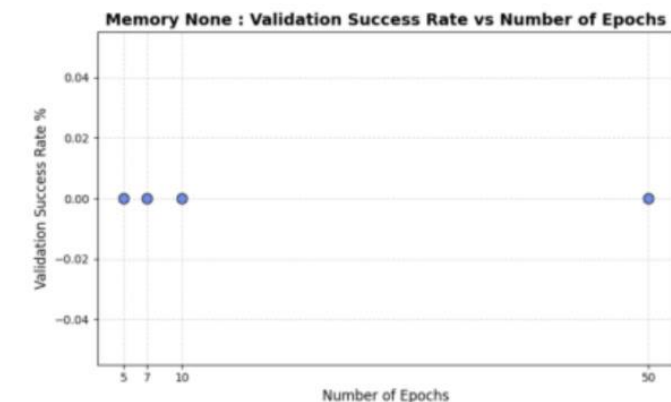
| Тип памяти | Success rate % | Mean Steps |
|------------|----------------|------------|
| None | 0% | 32 |
| GRU | 100% | 495 |
| LSTM | 100% | 487 |
| token | 0% | 12 |

Несмотря на то, что по графикам лосса, видим сходимость в каждом типе памяти начина с 6 эпохи, реально обучаются только типы памяти на основе рекуррентных сетей – GRU, LSTM



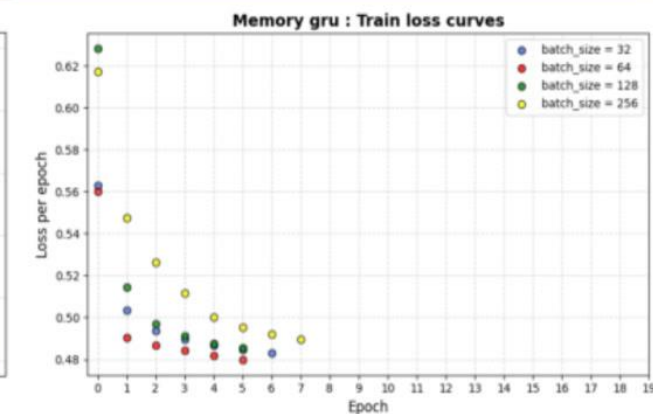
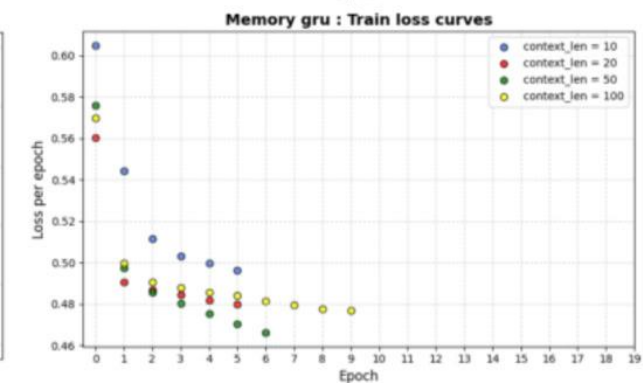
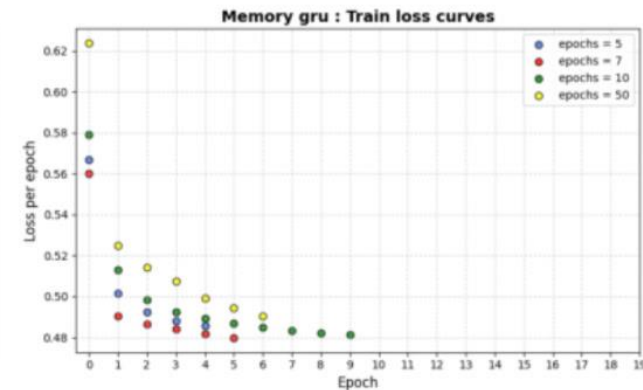
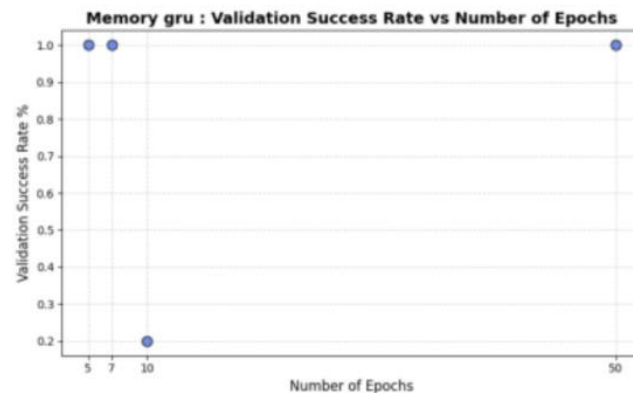
Эксперименты с памятью None

Выводы: без наличия памяти, на данных перебираемых параметрах обычный self.attention не может поймать нужные зависимости



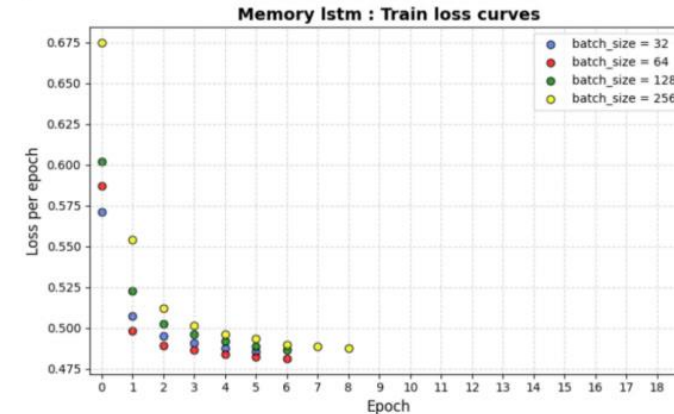
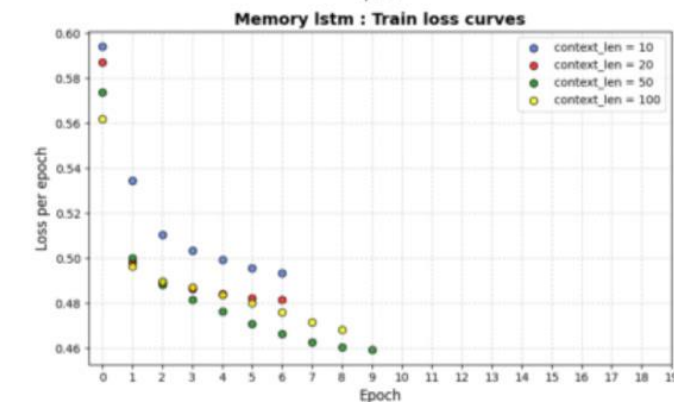
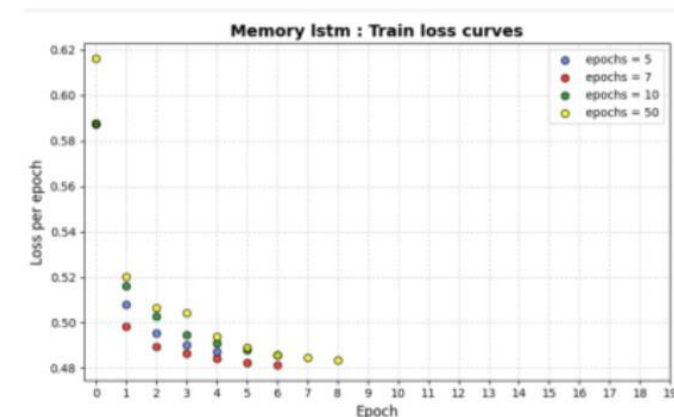
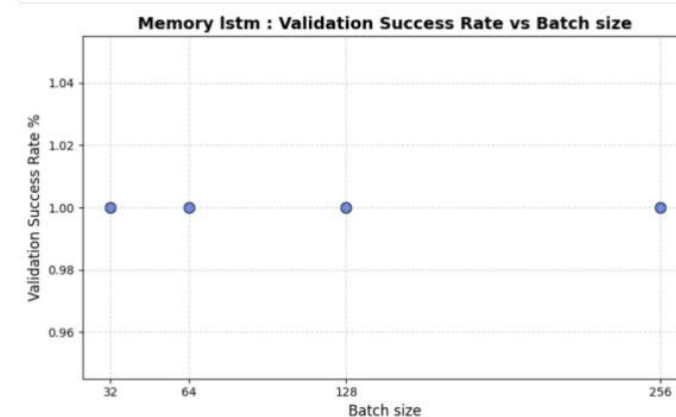
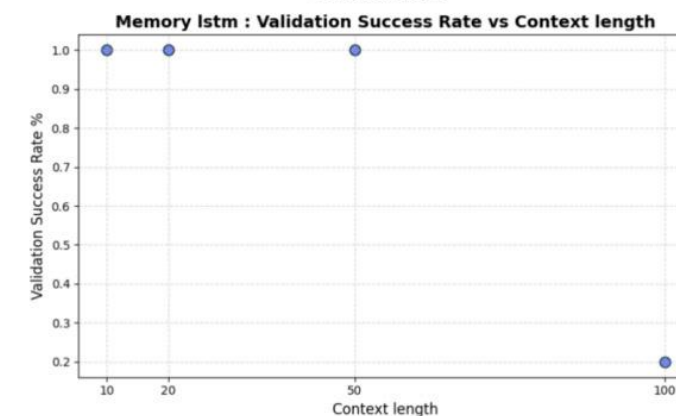
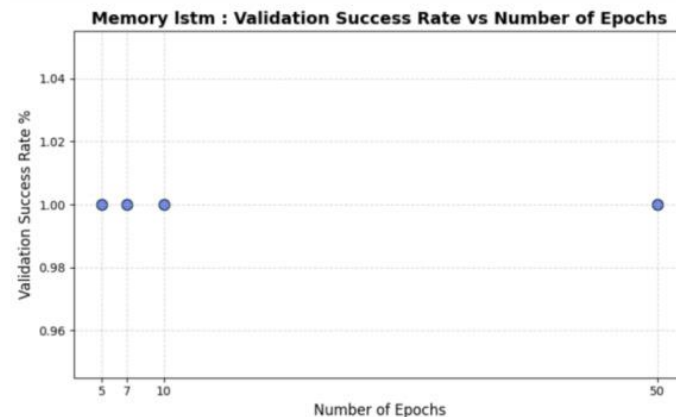
Эксперименты с памятью GRU

Выводы: чем больше длина контекста тем дольше сходится и тем хуже качество. Аналогичное поведение при увеличении размера батча



Эксперименты с памятью LSTM

Выводы: поведение очень
похоже на GRU но более
резкое



Эксперименты с памятью token

Выводы: видим, что положение очень похоже на None, что логично, так как по построение token – это добавлене дополнительных память-токенов в трансформер, однако видим, что при должной настройке размера батча, можно обучить частично модель в отличии от None памяти, однако все-еще данный тип сильно проигрывает рекуррентным типам памяти

