ГЛУБОКОЕ МНОГОЗАДАЧНОЕ МЕТАОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ

А. М. Герасимчик

Белорусский государственный университет, г. Минск;
hherasimchyk@gmail.com;
науч. рук. – В. В. Краснопрошин, д-р. техн. наук, проф.

Задача глубокого обучения с подкреплением является важной задачей в сфере робототехники. В работе был предложен оригинальный подход, в котором для повышения эффективности и скорости обучения нейронной сети была использована многозадачная функция потерь с её стохастической аппроксимацией с одновременным возмущением. С помощью вычислительных экспериментов по обучению агента в среде ML1 фреймворка MetaWorld проведено сравнение предложенного способа с известным алгоритмом метаобучения MAML. Проведены эксперименты, результаты которых показывают, что представленные алгоритмы эффективнее как по качеству решения задачи, так и по скорости обучения, по сравнению с исходным алгоритмом.

Ключевые слова: искусственный интеллект, нейронная сеть, метаобучение, многозадачное обучение, глубокое обучение, глубокое обучение с подкреплением, обучение с подкреплением.

ВВЕДЕНИЕ

Искусственные нейронные сети являются одним из основных и распространенных инструментов при создании интеллектуальных систем. Также, они активно используются для анализа данных в таких сферах, как робототехника, компьютерное зрение, обработка естественного языка и др. Процесс обучения нейронных сетей есть один из самых трудоемких этапов.

Несмотря на значительный научный прогресс в глубоком обучении с подкреплением, существующие алгоритмы не обладают достаточной эффективностью для решения задач в реальном мире. К тому же, такие алгоритмы требуют значительного количества времени на обучение, что усложняет процесс разработки. Для решения данных проблем в последнее время особенно актуально использование алгоритмов метаобучения или «обучения учиться». В работе предлагается подход к метаобучению с подкреплением с использованием многозадачного оптимизатора весов. Экспериментально показывается, что предлагаемый подход является более эффективным, чем известный алгоритм MAML.

МНОГОЗАДАЧНОЕ ОБУЧЕНИЕ

Во время метаобучения с подкреплением традиционно вклад каждой задачи учитывается равным. Однако, такое обучение неэффективно, т. к. для решения каждой новой задачи нужно обучать нейронную сеть заново, что влечет большие затраты по времени. Для решения данной проблемы предлагается использование подхода многозадачного обучения.

Многозадачное обучение представляет такой подход, при котором модель обучается предсказывать несколько задач одновременно [1]. Было предложено использовать жесткое разделение параметров для всех скрытых слоев сверточной сети [3]. Таким образом, для всех задач используется одна нейронная сеть, а наличие нескольких задач отражается только на функции потерь. Для этой цели в [3] была предложена следующая многозадачная функция потерь:

$$\mathcal{L}_{\xi_t}^{MT}\left(\omega_t, \{Q_{t_i}\}_{i=1}^{M}\right) = \sum_{i=1}^{M} \frac{1}{\left(\omega_t^{(i)}\right)^2} \mathcal{L}_{\theta,t_i}(Q_{t_i}) + \sum_{i=1}^{M} log\left(\omega_t^{(i)}\right)^2 \quad (1)$$

где веса ω_t являются гиперпараметрами.

Как показано в [4], производительность модели чрезвычайно чувствительна к выбору многозадачных весов ω_t. В данной работе вопрос рассматривается с точки зрения оптимизации и предлагается использовать в качестве многозадачного оптимизатора весов подход стохастической аппроксимации с одновременным возмущением из-за их устойчивости к неопределенностям и успешного применения в различных задачах машинного обучения [5].

Для оптимизации нестационарной оптимизационной задачи из [3], предлагается алгоритм “SPSA-Delta”, где стохастический оптимизатор многозадачных весов строит следующие оценки многозадачных весов на итерации t:

$$\begin{cases} L_t^{\pm} = L_t(\widehat{\omega}_{t-1} \pm \beta_t \Delta_t) \\ \widehat{\omega}_t = \widehat{\omega}_{t-1} - \alpha_t \Delta_t \dfrac{L_t^{+} - L_t^{-}}{2\beta_t} \end{cases} \quad (2)$$

где L_t – наблюдения обучающего эпизода t, $\widehat{\omega}_t$ – оценка ω_t, Δ_t – вектор, состоящий из независимых случайных величин с

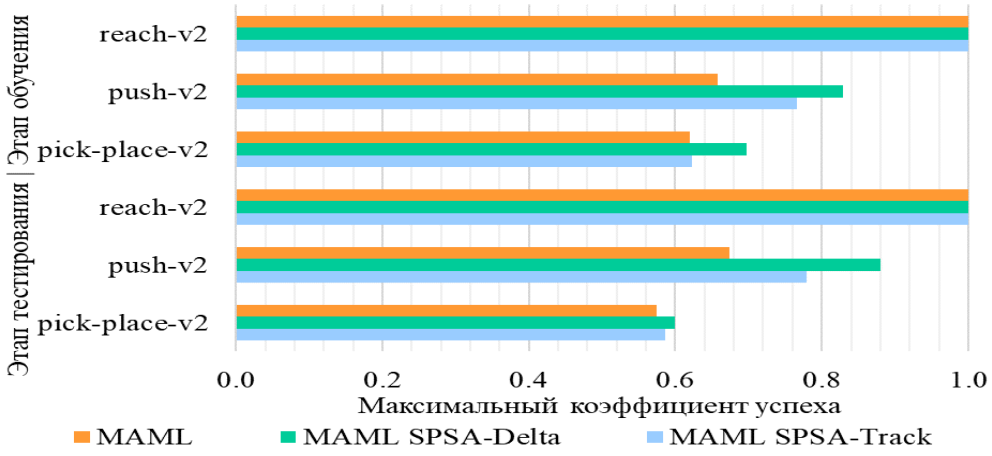
распределением Бернулли, $\widehat{\omega}_0$ — вектор с начальными значениями, $\{\alpha_t\}$ и $\{\beta_t\}$ — последовательности положительных чисел.

В многозадачном метаобучении можно наблюдать нестабильное изменение параметров ω_t, которое влияет на обучение параметров модели. По этим причинам в качестве многозадачного оптимизатора весов предлагается использовать другой подход стохастической аппроксимации [6], назовем его “SPSA-Track”:

$$\begin{cases} L_{2t} = L_{2t}(\widehat{\omega}_{2t-2} + \beta_n \Delta_t) \\ L_{2t-1} = L_{2t-1}(\widehat{\omega}_{2t-2} - \beta_t \Delta_t) \\ \widehat{\omega}_{2t-1} = \widehat{\omega}_{2t-2} \\ \widehat{\omega}_{2t} = \widehat{\omega}_{2t-1} - \alpha_t \Delta_t \dfrac{L_{2t} - L_{2t-1}}{2\beta_t} \end{cases} \quad (3)$$

ОПИСАНИЕ И РЕЗУЛЬТАТЫ ПРОВЕДЕННЫХ ЭКСПЕРИМЕНТОВ

Для проведения экспериментов в области метаобучения с подкреплением в качестве базового алгоритма метаобучения для цикла внешней оптимизации был выбран алгоритм MAML, который является одним из самых цитируемых алгоритмов в данной области. Для цикла внутренней оптимизации использовался простой и популярный алгоритм REINFORCE.



Максимальный коэффициент успеха алгоритмов в тестовой среде ML1

В качестве прикладной задачи для решения и оценки эффективности разработанных алгоритмов рассмотрим тест ML1 в среде MetaWorld [7]. Алгоритмы оцениваются по трем задачам из ML1: reach-v2, push-v2, pick-place-v2, где варьируется или позиция, которую нужно достигнуть, или целевая позиция объекта. Позиции цели не указаны в состояниях

мира, что вынуждает алгоритмы метаобучения с подкреплением адаптироваться к цели методом проб и ошибок.

На рисунке представлен максимальный коэффициент успеха, усредненный по 5 запускам, в тестовой среде ML1 MetaWorld. Исходя из полученных результатов, видно, что все 3 алгоритма отлично справляются с решением задачи reach-v2 как на этапе обучения, так и на этапе тестирования.

На более сложных задачах push-v2 и pick-place-v2 алгоритм MAML SPSA-Delta является более эффективным среди всех рассмотренных. Улучшение относительно базового алгоритма составило 17% на этапе обучения и 21% на этапе тестирования на задаче push-v2, 8% и 3% на задаче pick-place-v2 соответственно. Однако, на задаче pick-place-v2 способность метода MAML SPSA-Delta к обобщению не сильно выше алгоритма MAML SPSA-Track (60% у MAML SPSA-Delta и 59% у MAML SPSA-Track).

Максимальный коэффициент успеха, усредненный по всем задачам тестовой среды ML1

Алгоритмы	Этап обучения	Этап тестирования
MAML	76%	75%
MAML SPSA-Delta	84%	83%
MAML SPSA-Track	80%	79%

В таблице представлено сравнение среднего максимального коэффициента успеха, достигнутого каждым алгоритмом в тестовой среде ML1.

ЗАКЛЮЧЕНИЕ

Исходя из результатов экспериментов в тестовой среде ML1 MetaWorld тремя различными методами глубокого метаобучения с подкреплением: исходным алгоритмом MAML, MAML SPSA-Delta, MAML SPSA-Track, предложенный метод MAML SPSA-Track показывает улучшение эффективности в среднем на 4%, а MAML SPSA-Delta на 8% соответственно. Более того, последний затрачивает в среднем в 2 раза меньше времени на обучение на задачах push-v2 и pick-place-v2. Согласно полученным результатам, видно, что использование многозадачной функции потерь и её стохастической аппроксимации с одновременным возмущением позволяет значительно улучшить эффективность алгоритмов глубокого обучения с подкреплением.

Библиографические ссылки

1. Zhang Y., Yang Q. An overview of multi-task learning / Y. Zhang, Yang Q. // National Science Review. – 2018. – P. 30–43

2. Ruder S. An overview of multi-task learning in deep neural networks / S. Ruder // In 2017 European Control Conference (ECC). – 2017. – P. 301-312
3. Boiarov A., Granichin O., Granichina O. Simultaneous perturbation stochastic approximation for few-shot learning / A. Boiarov, O. Granichin, and O. Granichina // In 2020 European Control Conference (ECC). – 2020. – P. 350–355
4. Kendall A., Gal Y., Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics / A. Kendall, Y. Gal, R. Cipolla // In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2018. – P. 7482–7491
5. Granichin O., Volkovich Z. Randomized Algorithms in Automatic Control and Data Mining / O. Granichin, Z. Volkovich // Intelligent Systems Reference Library. Springer Nature. – 2015. – P. 136–159
6. Granichin O., Amelina N. Simultaneous perturbation stochastic approximation for tracking under unknown but bounded disturbances / O. Granichin, N. Amelina IEEE Transactions on Automatic Control. – 2014. – P. 1653–1658
7. Yu T., Quillen D., He Z. Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement / T. Yu, Quillen D., Z. He // CoRL 2019. – 2019. – P. 203-231.