

Слияние знаний в импульсных нейронных сетях

Д.И. Антонов

Ульяновский государственный технический университет
Ульяновск, Россия
d.antonov@ulireran.ru

С.В. Сухов

УФИРЭ им. В.А. Котельникова РАН
Ульяновск, Россия
ssukhov@ulireran.ru

Аннотация—Показана методика организации слияния знаний (knowledge fusion) импульсной искусственной нейронной сети (ИмНС, Spiking Neural Network) без доступа к первоначальным исходным данным. Разработан метод ранжирования выходных нейронов ИмНС по степени влияния на процесс классификации на основе знания о распределении весов сети. Эксперименты проводились на свободно доступных наборах данных в среде SpyukeTorch.

Ключевые слова— импульсные нейронные сети, спайковые нейронные сети, слияние знаний, ансамбль сетей, прунинг сетей.

1. ВВЕДЕНИЕ

За последнее десятилетие разработчики моделей глубокого обучения (deep learning) добились больших успехов в различных направлениях, в частности в области машинного зрения (computer vision, pattern recognition). Более широкое использование потенциала уже обученных моделей можно реализовать объединением полезной информации нескольких моделей. Один из таких эффективных методов относится к методам ансамблей нейронных сетей – бэггинг (bagging). При бэггинге несколько сетей параллельно обрабатывают тестовые данные и по итогам проводят голосование за консолидированный результат. Но хорошая точность такого способа объединения знаний достигается при условии, что все элементарные классификаторы обучаются и работают параллельно (независимо друг от друга), что может быть довольно громоздким решением.

Работа с громоздкими нейронными сетями стала возможной благодаря широкому распространению графических процессоров (GPU), производящих большое количество параллельных вычислений. Но большую сеть, вмещающуюся на сервере, не разместить в небольших цифровых устройствах. Обходят эту проблему, уменьшая размер сети методом обрезки или прунинга (pruning) за счёт сокращения части её параметров.

Попытка прямого дообучения на новых данных обученной ранее сети сталкивается с явлением катастрофического забывания (catastrophic forgetting) предыдущих знаний [1]. В данной работе мы описываем способ избежать катастрофического забывания используя так называемое слияние знаний (knowledge fusion) [2]. Слияние знаний в нашем случае использует идеи прунинга и даёт возможность приближаться к точности бэггинга, избегая его тяжеловесности.

2. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим две импульсные нейронные сети, обученные каждая на своем наборе данных. Будем полагать, что первые слои в обеих сетях одинаковы и выполняют функцию выделения признаков. Последний слой сети выполняет функцию классификатора. Наша цель – определить наименее важные нейроны в выходном слое с целью обрезки соответствующих связей.

Обрезанные связи могут быть заменены связями из другой сети, в результате чего образуется сеть, способная уверенно классифицировать объекты из двух наборов, но обладающая таким же набором параметров, как и исходная сеть. В работе представлено решение поставленной задачи для варианта, который не требует доступа к первоначальному набору обучающих данных.

3. ЭКСПЕРИМЕНТЫ

Для построения ИмНС мы использовали пакет SpyukeTorch, кодирующий информацию временем прихода импульсов. Использованная архитектура ИмНС была взята из работ [3, 4]. Обучение нейронных сетей производилось на выборках из наборов изображений рукописных цифр MNIST [5] и букв EMNIST [6], содержащих по 10 классов каждый и состоящие из 24000 тренировочных и 4000 тестовых изображений.

Архитектура ИмНС состоит из пакета для предобработки поступающих изображений (пакет фильтров разности гауссиан, Difference of Gaussians или DoG-фильтров) и собственно сети (Рис.1). Сеть состоит из последовательно расположенных трех частей, содержащих свёрточный слой (convolution layer) и слой подвыборки или пуллинговый слой (pooling layer). Тренировка первых 2-х свёрточных слоёв производится «без учителя» на основе принципа пластичности временной зависимости импульса (spike-timing-dependent plasticity или STDP), а тренировка 3-го слоя - на базе принципа пластичности временной зависимости импульса с подкреплением (reward-modulated STDP или R-STDP).

Способ слияния знаний, реализованный нами, в существенной мере следовал из двух экспериментов по определению степени важности нейронов выходного слоя для классификации изображений.

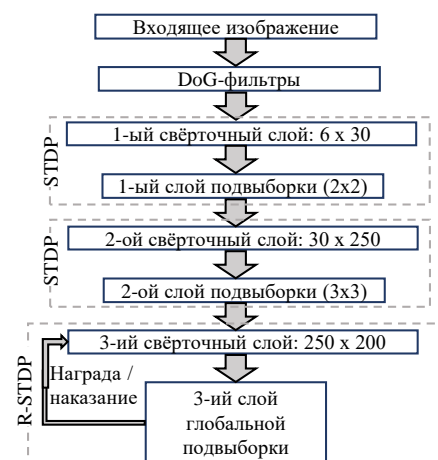


Рис.1. Архитектура нейронной сети

Первый эксперимент заключался в подсчёте количества использований сетью каждого нейрона

выходного слоя в процессе классификации. Было выяснено, что нейроны сильно неравнозначны по степени активности. За время распознавания упомянутого выше тестового набора данных из 4000 изображений частота использования каждого из 200 нейронов выходного слоя колебалась примерно от 0 до 120 раз (Рис.2).

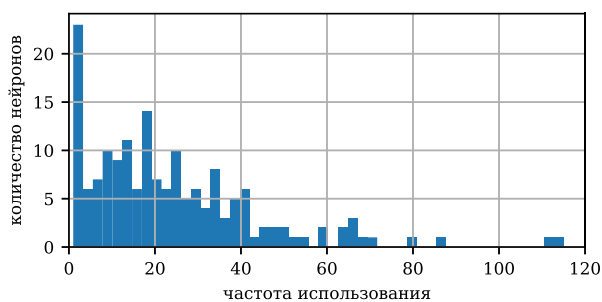


Рис.2 Частота использования нейронов выходного слоя при тестировании набора в 4000 изображений

Второй эксперимент заключался в поиске зависимости активности нейрона от параметров сети. Наиболее высокой оказалась корреляция между степенью активности нейрона и суммой наибольших 10% весов, связанных с каждым нейроном.

Процедура определения степени важности нейронов была использована для создания метода слияния знаний. Предварительно было получено две сети, имеющие идентичную архитектуру, описанную выше. Первая из них была получена тренировкой на выборке из набора изображений цифр MNIST в течение 50 эпох с точностью классификации (accuracy) около 94%. Точность классификации A определялась по следующей формуле:

$$A = D_{\text{correct}} / D_{\text{total}},$$

где D_{correct} – количество правильно классифицированных данных, D_{total} – общее количество данных. Вторая сеть была натренирована на изображениях рукописных букв набора EMNIST. Процесс обучения отличался от обучения первой сети. Первые два слоя второй сети были идентичны первым двум слоям первой сети, уже натренированным на изображениях цифр, и повторно они не тренировались на изображениях букв. Данные первые два слоя выполнили роль универсального выделителя признаков. Третий слой сети был натренирован в течение 50 эпох на наборе из изображений рукописных букв. В итоге натренированная вторая сеть научилась распознавать буквы с точностью около 90%, но её точность классификации цифр была около 8%.

На базе двух сетей методом слияния знаний была создана третья сеть, имеющая такую же архитектуру, как и первые две, и максимально получающая знания от первой и второй сетей. Первый шаг процедуры слияния знаний состоял в прунинге 3-их слоёв первой и второй сети. Третий свёрточный слой данного типа сети содержит веса, относящиеся к 200 нейронам, по 20 нейронов на каждый из 10 классов. В каждом классе нейроны были проранжированы нами по параметру частичной суммы весов. Расчёт параметра частичной суммы весов состоял в выделении из весов, относящихся к каждому нейрону, максимальных 10% весов и суммировании их значений. После ранжирования

нейронов, из каждого класса первой сети было выделено по 10 нейронов с максимальным параметром частичной суммы весов. Такое же выделение нейронов была сделано во второй сети.

Второй шаг процедуры состоял в непосредственном слиянии знаний в создаваемой сети. Первые два слоя без изменений были перенесены из исходных сетей в итоговую сеть. А третий слой был скомпонован из нейронов 3-их слоёв первой и второй сетей таким образом, что каждый класс получил по 10 предварительно выделенных нейронов из первой сети и по 10 нейронов из второй.

В результате проведённого слияния знаний была получена сеть, которая смогла распознавать как цифры со средней точностью 87%, так и буквы со средней точностью 75%. Точность бэггинга на этом же наборе составила 91% на изображениях цифр и 78% на изображениях букв. Таким образом, метод слияния знаний позволил получить сеть вдвое меньшего размера с падением точности классификации на 3-4%.

4. ЗАКЛЮЧЕНИЕ

Полученная методика слияния знаний ИмНС достаточно проста и результативна, но имеет, как минимум, два ограничения:

- тренировочные наборы должны быть достаточно похожими и первые слои сетей должны быть достаточно универсальными, чтобы эти слои смогли бы выделить из наборов общие признаки,
- количество «сливаемых» знаний ограничено сверху количеством выходных нейронов.

Авторы видят дальнейшее развитие своей работы в разработке методики слияния знаний для более общего вида сетей.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 20-07-00974 и № 18-47-732006.

ЛИТЕРАТУРА

- [1] McCloskey, M. Overcoming catastrophic forgetting in neural networks / M. McCloskey, N.J. Cohen // Psychology of Learning and Motivation – Advances in Research and Theory. – 1989. – Vol. 24(C). – P. 109-165.
- [2] Akhlaghi, M.I. Knowledge Fusion in Feedforward Artificial Neural Networks / M.I. Akhlaghi, S.V. Sukhov // Neural Processing Letters. – 2018. – Vol. 48(1). – P. 257-272.
- [3] Mozafari, M. SpykeTorch: Efficient simulation of convolutional spiking neural networks with at most one spike per neuron / M. Mozafari, M. Ganjtabesh, A. Nowzari-Dalini, T. Masquelier // Frontiers in Neuroscience. – 2019. – Vol. 13. – P. 1-12.
- [4] Mozafari, M. Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks / M. Mozafari, M. Ganjtabesh, A. Nowzari-Dalini, S.J. Thorpe, T. Masquelier // Pattern Recognition. – 2019. – Vol. 94. – P. 87-95.
- [5] The MNIST database of handwritten digits Web Site [Electronic resource]. — Access mode: <http://yann.lecun.com/exdb/mnist/>.
- [6] The EMNIST Dataset [Electronic resource]. — Access mode: <https://www.nist.gov/products-and-services/emnist-dataset>.