

УДК 004.93

*Бондаренко В. О., аспірант,  
Олейник А. А., канд. техн. наук, доц.,  
Субботин С. А., д-р техн. наук, проф.  
Национальный университет «Запорожская политехника»*

## **НЕПРЕРЫВНО КОНСТРУКТИВНЫЕ ГЛУБОКИЕ НЕЙРОННЫЕ СЕТИ**

Традиционно алгоритмы глубокого обучения обновляют вес сети, тогда как архитектура сети выбирается вручную, используя метод проб и ошибок. В данном исследовании предложено два новых подхода, которые автоматически обновляют структуру сети, а также изучают ее вес. Новизна данного подхода заключается в параметризации, где глубина или дополнительная сложность постоянно инкапсулируются в пространстве параметров, которые добавляют дополнительную сложность.

Глубокое обучение включает в себя несколько уровней нелинейной обработки информации. Это позволяет изучать архитектуры, которые реализуют функции в виде повторяющихся композиций более простых функций, тем самым изучая уровни абстракции с лучшим обобщением и возможностью представления.

Хотя глубокое обучение полезно, содержание множества количества слоев может быть проблематичным: во-первых, когда больше слоев, весов, пространство и вычислительная сложность выше; во-вторых, когда имеется больше свободных параметров, существует более высокий риск переобучения; в-третьих, если сеть глубокая, существует проблема исчезновения (взрыва) градиентов, когда ошибка распространяется по многим слоям, и один относительно простой механизм заключается во введении стробирующих механизмов.

Начиная с 1990-х годов, было много подходов для оптимизации архитектуры сети, начиная от ранних инкрементальных методов добавления скрытых модулей один за другим или начиная с большой сети и сокращая ее, до более сложных современных подходов, таких как эволюционные алгоритмы или обучение с подкреплением, а также методы стимулирующего стиля.

Цель исследования аналогично - изучить сетевую архитектуру на основе данных. Основное отличие состоит в том, что вместо поиска в дискретном пространстве всех архитектур, параметризовали модели таким образом, что само понятие сложности или глубины само по себе непрерывно, делая модель от начала до конца дифференцируемой, и позволяя спуску градиента искать по архитектуре в дополнение к их

па. Предлагается два метода построения и изучения структуры глубокой нейронной сети, где сложность сети на уровне скрытого блока или слоя кодируется непрерывными параметрами. Эти параметры корректируются вместе с весами сети во время градиентного спуска, что подразумевает мягкое изменение структуры сети вместе с весами сети.

В первом методе в туннельных сетях, основанных на сетях магистралей, связанный с каждым скрытым блоком, является непрерывным параметром, если этот параметр не активен, блок просто копирует свои входные данные в свой выходной в обход нелинейности, эффективно увеличивающего глубину сети.

Во втором методе подающих надежды персептроны, основанные на подающих надежды деревьях, есть параметр, связанный с каждым слоем, указывающий, необходима ли дальнейшая нелинейная обработка. Первоначально мы начинаем с одного слоя, а во время обучения с градиентным спуском, когда это необходимо, этот параметр может стать активным, что вызывает создание еще одного полного слоя, эффективно увеличивая глубину сети.

Эксперименты на синтетических двуспиральных данных показывают, как туннельные сети и начинающие персептроны могут адаптироваться к различным размерам для различных сложностей задач, используя один и тот же набор гиперпараметров, адаптируя количество единиц для туннельных сетей и количество слоев для начинающих персептронов. Что касается реальных задач распознавания цифр и изображений, наблюдаем, что туннельные сети достигают лучшей производительности, обеспечивая лучшую регуляризованную модель и используя меньшее количество параметров, по сравнению с сетями магистралей. Точно так же начинающие персептроны показали сопоставимые или лучшие показатели. По сравнению с туннельными сетями, начинающие персептроны, кажется, растут больше и сокращают меньше. Устанавливая скорость обучения по убыванию, наблюдается, что разные слои растут с разной скоростью и используются по-разному. В сочетании с регуляризацией это позволяет туннельным сетям сохранять некоторые неиспользуемые верхние уровни линейными, что дает возможность фактически удалить их из сети в конце.

В будущем будет интересно увидеть различные области применения конструктивных нейронных сетей. Одним из потенциальных подходов является включение сверточных слоев в конструктивную архитектуру, как это обычно используется в приложениях компьютерного зрения. Другим возможным направлением является применение методов условий последовательного прогнозирования, аналогичных рекуррентным сетям магистралей.