

Об алгебраических свойствах операций, используемых при построении современных свёрточных нейронных сетей

Фадеев Егор Павлович¹★

fadeevegor@yandex.ru

Зубюк Андрей Владимирович¹

zubuk@compd2.phys.msu.ru

¹Москва, МГУ имени М.В.Ломоносова, физический факультет

Сегодня для решения задач компьютерного зрения, распознавания речи и др. широко используются искусственные нейронные сети (ИНС). Многие архитектуры ИНС включают блоки, состоящие из свёрточных слоёв, активаций и пулингов. Известно, что одни виды этих операции позволяют достигать больших показателей качества, чем другие (например, ИНС с max-пулингами, как правило, превосходят ИНС с усредняющими пулингами). Однако, обоснованного объяснения этому не дано.

В большинстве случаев свёртки, активации и пулинги рассматриваются как манипуляции над пикселями изображений (отсчётами, если речь идёт об иных данных): свёртки — как выделение локальных «признаков» (например, границ объектов, характеризующихся большими градиентами яркостей), пулинги — как выделение «наиболее интенсивных» признаков и отсечение «менее интенсивных» и т. д.

В настоящей работе предлагается рассматривать свёртки, активации и пулинги как операторы, действующие в евклидовых векторных пространствах, элементами (далее — *векторами*) которых являются изображения и др. многомерные массивы. Авторы полагают, что исследование алгебраических свойств этих операторов позволит лучше понять отличия различных видов используемых при построении ИНС операций (например, max-пулингов и усредняющих пулингов), а также определить, какие их свойства ключевым образом влияют на качество ИНС в целом.

С предложенной точки зрения свёрточный слой представляет собой линейный оператор, его сингулярными базисами являются базисы Фурье.

Активация — нелинейная функция одной переменной, применяется к каждой координате вектора в *каноническом* базисе в отдельности (каноническим будем называть ортонормированный базис, каждый вектор которого есть массив, состоящий из един-

ственной единицы и остальных нулей). Одной из наиболее часто используемых в современных ИНС функций активации является ReLU, определённая как $\text{ReLU}(x) = \max\{0, x\}$. Применение ReLU к каждой координате вектора в каноническом базисе есть не что иное, как *проецирование* на неотрицательный ортант канонического базиса — коническую оболочку, натянутую на векторы канонического базиса.

Пулинг — линейный или нелинейный оператор, понижающий размерность [2, 3]. Композиция активации ReLU и одного из наиболее часто применяемых пулингов — *max-пулинга* — также может быть выражена через оператор *проецирования* [1, 4]. Действительно, рассмотрим действие такой композиции на фрагмент f изображения, попавший в одно «окно пулинга» размером $n \times n$. Пусть e_{ij} , $i, j = 1, \dots, n$, — векторы, образующие канонический базис (для фрагмента изображения). Пусть A_{ReLU} — оператор по координатному применению активации ReLU, и P_{max} — оператор *max-пулинга*. Тогда $P_{\text{max}}A_{\text{ReLU}}f$ — норма проекции f на множество $\bigcup_{i,j=1}^n \{\alpha e_{ij}, \alpha \geq 0\}$, т. е. на объединение «рёбер» неотрицательного ортанта канонического базиса.

Как видно из вышеизложенного, блоки типа «свёртка — активация — пулинг» с алгебраической точки зрения представляют собой линейное преобразование с последующим проецированием на нелинейное множество, тесно связанное с базисом, отличным от сингулярных базисов линейного преобразования. В работе исследована зависимость качества ИНС в задаче классификации изображений из базы CIFAR10 от выбора сингулярных базисов линейного преобразования и базиса, определяющего множества, на которые осуществляется проецирование. Показано, что в ряде случаев замена рассмотренных выше базисов Фурье и канонического базиса на отличные от них базисы не снижает качество ИНС. Это позволяет выдвинуть гипотезу о том, что критическое влияние на качество ИНС оказывают отмеченные выше общие алгебраические свойства рассмотренных операций, а не их конкретный вид. Однако, эта гипотеза требует дальнейшего изучения.

Работа поддержана грантом РФФИ № 17-07-00832.

- [1] Зубюк А. В., Байков В. Г. Идемпотентная нейронная сеть как реализация морфологического метода узнавания объектов по изобра-

-
- жениям // Москва: Техническое зрение в системах управления - 2018, 2018 — p. 53–54.
- [2] *Passricha V., Aggarwal R.K.* A comparative analysis of pooling strategies for convolutional neural network based Hindi ASR // J Ambient Intell Human Comput, Berlin: Springer Berlin Heidelberg, 2019. — p. 5–25.
- [3] *Rippel O., Snoek J., Adams R.P.* Spectral Representations for Convolutional Neural Networks // In: Proceedings of the 28th international conference on neural information processing systems (NIPS'15), vol 2, 2015 — p. 2449–2457.
- [4] *Jin L., Li S., Hu B., Liu M.* A survey on projection neural networks and their applications // Applied Soft Computing, 2019. — vol. 76, p. 533–544.