

Сверточный зоопарк нейросетей

В.И. Слюсар

Центральный научно-дослідний інститут озброєння та
військової техніки Збройних Сил України, м.Київ, Україна

В работе предложена совокупность новых тензорно-матричных операций для обработки данных в нейронных сетях.

Введение. Необходимость миграции искусственного интеллекта (ИИ) в устройства пользователей требует перехода к более мощным вычислительным системам, разработкой которых сегодня занимаются многие ведущие фирмы мира. Речь идет, например, об использовании тензорных вычислительных устройств, более мощной процессорной техники и графических ускорителей, а также иных подходах. Вместе с тем, другим сдерживающим фактором является устаревшая, традиционная математическая концепция, когда для обработки данных используются обычные матричные операции. Поэтому продвижение в этой области будет существенным, если получится внедрить в библиотеки, применяемые для программирования нейронных сетей (НС) и их обучения, новые подходы, в том числе тензорно-матричный аппарат, предлагаемый ниже.

Изложение основного материала

Одним из компонентов указанного тензорно-матричного аппарата является торцевое произведение матриц, которое сводится к произведению каждого элемента левой матрицы на всю строку правой, соответствующую по номеру этому элементу. Альтернативой данной операции является столбцовое

произведение Хатри-Рао, при котором элемент левой матрицы умножается на соответствующий по номеру столбец правой. В [1 - 4] представлены свойства этих матричных операций, их комбинаций между собой, с кронекеровским произведением, и особенно важной является возможность перехода к адамарову произведению, что упрощает вычислительные операции. Среди известных свойств интерес представляет, в частности, эквивалентность данных операций произведению Кронекера в отношении векторов, а также теорема, связывающая торцевое и столбцовое произведения.

Операция торцевого произведения, как известно, используется в рамках тензорного скетча [5], где она позволяет заменить свертку от преобразования Фурье матрично-векторного произведения эквивалентным произведением Адамара соответствующих Фурье-сверток матриц. Это существенно снижает количество операций в том случае, если НС трансформирует исходное пространство больших данных в пространство меньшей размерности, что, как правило, наблюдается в сверточных сетях.

Следующим обобщением указанных матричных операций являются их блочные варианты, когда соответствующие процедуры выполняются поблочно. Например, каждая пара блоков 2-х матриц с одинаковым разбиением на блоки умножается по принципу торцевого произведения. Аналогичная операция возможна также для блочного столбцового произведения Хатри-Рао. В отношении блочных матриц выполняется то же свойство

транспонирования произведений и более сложное фундаментальное свойство перемножения блоков матриц с учетом их блочно-столбцовой и блочно-транспонированной структуры.

Необходимость введения блочных операций торцевого и столбцевого произведений Хатри-Рао обусловлена тем, что при дифференцировании например, столбцевого произведения матриц, переход к блочному варианту позволяет факторизовать операции дифференцирования и нахождения градиентов, очень важные при описании процедур обучения НС. Как известно, метод обратного спуска реализуется на основе вычисления градиента, при этом указанные блочные операции демонстрируют возможность упрощения такой процедуры и ее автоматизации. На этой же основе может быть сформирована нижняя граница Крамера-Рао [4], позволяющая определить потенциальную эффективность НС в условиях влияния шумов. В общем случае, с данной границей нужно сравнивать возможности ИИ в задачах оценивания параметров сигналов. Хотя в ряде случаев оказывается, что НС дают лучшую оценку точности, чем нижняя граница Крамера-Рао, этот вопрос нуждается в детальнейшем исследовании. Объясняются подобные факты тем, что НС учитывает большее количество факторов, чем та модель, в рамках которой описывается анализируемый процесс.

В случае НС в развитие известного тензорно-матричного аппарата автором было предложено проникающее торцевое произведение матриц. Суть его сводится к тому, что для 2-х матриц с

разбиением на блоки, произведение левой матрицы на правую заключается в адамаровом произведении левой матрицы, каждый блок которой как бы проникает внутрь правой и поблочно перемножается с блоками правой матрицы по принципу произведения Адамара. В этом случае, как показано в [1 - 4], выполняется перестановочное свойство, как в отношении матриц, так и в отношении произведений векторов, применяемых в НС.

Чтобы наглядно представить суть проникающего произведения, можно условиться, что матрица A , например, является входным изображением, а блок-матрица B - это набор весовых коэффициентов различных нейронов, например, первого слоя. Соответственно, в этом случае каждый элемент или пиксель изображения умножается на свой нейронный коэффициент.

Для того, чтобы формализовать сверточные НС, соответствующий результат проникающего произведения необходимо умножить на вектор единиц справа, либо на вектор-строку слева, либо на блочной вектор единиц справа, по процедуре блочного традиционного умножения. Чтобы получить скаляр, необходимо такое умножение осуществить слева и справа. При таком подходе описание нейрона сводится к функции активации с аргументом в виде проникающего произведения, умноженного на вектор единиц, что эквивалентно процедуре свертки. Такая запись позволяет компактно представить сложную архитектуру нейронных сетей, состоящую из нескольких слоев.

Следует отметить, что в библиотеке TensorFlow

уже реализован упрощенный вариант проникающего произведения, который состоит в том, что вектор умножается на матрицу целиком, или же матрица умножается на вектор. В этом случае каждый элемент вектора по методу произведения Адамара перемножается на соответствующий столбец матрицы. Однако, такая библиотека не работает в случае перемножения матриц и блочных матриц тем более, поэтому предложенный аппарат является достаточно полезным, чтобы ускорить реализацию вычислений при обучении нейросетей.

Другим типом произведения, расширяющего возможности работы с НС, является проникающие кронекеровское произведение, суть которого состоит в том, что каждый блок левой матрицы умножается по принципу прямого произведения целиком на всю матрицу справа с помощью проникающего произведения. Данная процедура полезна при математической формализации обработки в нейросети видеопотока или множества кадров с различных видеокамер одновременно. При этом каждый кадр изображения взвешивается на свой вес соответственно во множестве нейронов, в том числе нескольких слоев. С помощью данных операций было предложено аналитическое описание последовательности выполнения обработки данных в различных слоях моделей LeNet5 [6] и AlexNet [7].

Рассмотренный тензорно-матричный аппарат позволяет наращивать арсенал методов, используемых при описании НС, поскольку проникающее произведение и его блочные версии дают исходные массивы взвешенных данных, либо

пикселей, с которыми далее можно выполнять произвольные операции. Традиционно это могут быть свертки, однако, в данном случае можно ввести, например, процедуру преобразования Фурье или вейвлетного преобразования. Это позволит выделить дополнительные признаки, особенно если ядро свертки больше 2-х, например, 4, 6 и т.д. Это направление предстоит еще в дальнейшем исследовать на эффективность. Аналогично указанная процедура может комбинироваться с операцией maxPooling. На данный момент известны два ее варианта: выбор максимальных значений, либо формирование средневзвешенной суммы по всему блоку. Предлагается предварительно перед этим выполнять то же преобразование Фурье либо вейвлет-преобразование и далее выбирать максимальный отклик. Это позволит в определенной степени усреднить шум в заданной области и как-то спрогнозировать спектральный состав фрагмента данных. Аналогичное обобщение возможно на случай 3-хмерных входных массивов, когда используется, например, RGB-матрица, либо многомерный тензор, объединяющий множество данных с видеокамер в различных спектральных диапазонах. В этом случае в НС применяются 3D либо многомерные свертки, причем 3D свертка уже используется в данное время, однако, комбинация ее с преобразованием Фурье, вейвлетами или другими функциями пока не применялась.

Данный подход может быть обобщен на различные варианты комбинаций проникающего произведения и суммирования всех элементов

матричного блока, выполнения макспулинга и т. д. Возможно, например, введение промежуточной операции, когда сумма ищется только по строкам блока, и далее макспулинг применяется построчно, в результате чего формируется векторный отклик, либо же аналогичная сумма формируется по столбцам. Возможно применить макспулинг после преобразования Фурье и т.п. По сути, речь идет “зоопарке” сверток для НС по аналогии с названием web-сайта “Зоопарк нейросетей” [8]. В принципе, каждую такую операцию можно обозначать своим математическим символом либо же вводить специальные наименования функций, словесно описывающие суть манипуляций с данными.

Что касается более сложных сетей, в которых имеют место не только сужающийся вид архитектуры, но и расширяющиеся ветви (например, GAN), то в этом случае рассмотренный матричный аппарат является недостаточным и может быть дополнен процедурой обобщенного проникающего произведения. Суть его заключается в том, что каждый блок левой матрицы умножается по принципу проникающего торцевого произведения на соответствующую строку блоков правой матрицы. Соответственно, синтез расширяющейся архитектуры НС ограничивается только строками блоков. В настоящий момент для перехода к расширяющимся структурам применяется только прямое произведение матриц, т.е. один элемент умножается на какую-то матрицу весов формата 2×2 , 3×3 , 4×4 и т.д., что позволяет постепенно увеличить размерность и перейти от малой размерности на

входе к увлечённой – на выходе НС. В предлагаемом подходе применяется более гибкий аппарат, в рамках которого можно делать расширение только по строкам или как альтернатива - по столбцам блоков, что обеспечивается транспонированным обобщенным проникающим произведением. В более общем случае, такая процедура может выполняться поблочно и применительно к блочному транспонированному обобщенному проникающему произведению. Такие тензорно-матричные процедуры позволяют наращивать структуры НС, применяя их ансамбли, кластеры или конгломераты различной иерархии, в том числе фрактальной структуры [9].

Аналогичное обобщение предлагается использовать и для проникающего кронекеровского произведения в блочном варианте реализации, когда эта процедура выполняется поблочно. Далее можно применять, как было указано, различные варианты сверток, 3-хмерных и многомерных макспулингов.

Выводы. В заключение, следует отметить, что предложенное многообразие сверточных операций в сочетании с развитым тензорно-матричным аппаратом предоставит разработчикам дополнительные возможности по освоению новых концепций НС и их практической реализации. Важным условием при этом является интеграция новых тензорно-матричных операций в библиотеки фреймворков проектирования и обучения НС.

Литература

1. Slyusar, V. I. (1998) End matrixs products in radar applications. *Radioelectronics and Communications Systems*, Vol. 41, no. 3.

2. Slyusar, V. (1999). A Family of Face Products of Matrices and its Properties. *Cybernetics and systems analysis*, 3(35), 379–384.
3. Slyusar, V. I. (2003) Generalized face-products of matrices in models of digital antenna arrays with nonidentical channels. *Radioelectronics and Comm. Systems*, Vol. 46; pp. 9 - 17.
4. Слюсар В.И. Информационная матрица Фишера для моделей систем, базирующихся на торцевых произведениях матриц// Кибернетика и системный анализ. – 1999.- Том 35; № 4.- С. 636 - 643. DOI: 10.1007/BF02835859.
5. Thomas D. Ahle, Jakob Bæk Tejs Knudsen. Almost Optimal Tensor Sketch. - 2019. Mathematics, Computer Science, ArXiv.
6. Слюсар В.І. Тензорно-матрична версія LeNet5.// IV Міжнародна науково-практична конференція «Інтеграція інформаційних систем і інтелектуальних технологій в умовах трансформації інформаційного суспільства», що присвячена 50-ій річниці кафедри інформаційних систем та технологій, 21-22 жовтня 2021, Полтавський державний аграрний університет. - С. 114 - 119. DOI: 10.32782/978-966-289-562-9.
7. Vadym Slyusar, Mykhailo Protsenko, Anton Chernukha, Pavlo Kovalov, Pavlo Borodych, Serhii Shevchenko, Oleksandr Chernikov, Serhii Vazhynskiy, Oleg Bogatov, Kirill Khrustalev. Improvement of the object recognition model on aerophotos using deep conventional neural network.// Eastern-European Journal of Enterprise Technologies. - 2021, Vol. 5, No. 2 (113). Pp. 6 – 21. DOI: 10.15587/1729-4061.2021.243094.
8. Fjodor Van Veen. The neural network zoo. - <https://www.asimovinstitute.org/neural-networkzoo/>.
9. Слюсар В.И. Мультимодальные квазифрактальные нейросети.// Нейромережні технології та їх застосування НМТіЗ-2021: збірник наукових праць XX Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2021». - Краматорськ: Донбаська державна машинобудівна академія. – 8 - 9 грудня 2021. - С. 134 -137.