

Исследование сходимости алгоритма сокращения трафика в коммутаторной нейронной сети

С.Ю. Степанов

Аннотация. В статье дано краткое описание задачи сокращения трафика информации в коммутаторной нейронной сети, алгоритм фрагментации нейронной сети и доказательство сходимости данного алгоритма.

Ключевые слова: коммутаторная нейронная сеть, фрагментация, сходимость.

Введение

Коммутаторная нейронная сеть – новая технология построения нейронных сетей, позволяющая создавать большие искусственные нейронные сети для задач управления сложными техническими объектами и обработки информации [1].

В традиционном исполнении нейронных сетей существует ряд проблем, среди которых наиболее значимые – масштабирование сети, которое на определенном этапе приводит к такому увеличению размеров сети и количеству линий связи, что для успешной работы искусственной нейронной сети не будет достаточно самых современных вычислительных средств, – нагрузка на нейрон дополнительных вычислительных функций, необходимых для передачи информации с выхода на входы других нейронов. Существенно усложняется работа нейрона по решению транспортной задачи за счет увеличения объема передаваемой информации и увеличения внутренних таблиц связи. При аппаратной реализации нейронной сети квадратично увеличивается количество потенциально поддерживаемых линий передачи информации [2]. Для решения этих проблем была предложена архитектура, названная коммутаторной. В

такой архитектуре нейронная сеть произвольного вида представлена в виде системы передачи информации, имеющей древовидную структуру и состоящей из двух типов элементов:

- нейронов, обрабатывающих информацию, представляемых в виде «листьев» на древовидной структуре;
- коммутаторов, обеспечивающих передачу информации между нейронами и ее масштабирование.

Рассматривается нейрон, имеющий только один вход и один выход. Мультиплицирование входов и выходов обеспечивает коммутатор. Структура нейронной сети и масштабные коэффициенты передачи информации между нейронами хранятся в служебных таблицах коммутаторов.

Базовая нейронная сеть – нейронная сеть, которую обслуживает один коммутатор. Масштабирование нейронной сети осуществляется увеличением количества базовых нейронных сетей. Для интеграции отдельных базовых сетей в единую нейронную сеть используются коммутаторы верхнего уровня.

При связи нейронов между собой путь передаваемой информации (трафик) существенно зависит от взаимного расположения нейронов в

древовидной структуре. Чем дальше нейроны расположены друг от друга, тем больше трафик информации. Правомерна задача оптимизации трафика информации в нейронной сети за счет рационального расположения нейронов на коммутаторах. Внутри одной базовой нейронной сети сократить трафик невозможно. Переставляя нейроны между различными базовыми нейронными сетями, можно существенно снизить трафик в нейронной сети. Важным аспектом этой задачи является логическая группировка нейронов, по результатам которой можно создать отдельные группы, отвечающие за отдельные подзадачи общей задачи нейронной сети.

Рациональная группировка нейронов позволит по обученной нейронной сети определить, какие подмножества нейронов отвечают за отдельные подзадачи, поскольку существование более или менее изолированной группы нейронов позволяет говорить, что эта группа выполняет более или менее изолированную подзадачу общей задачи. При этом, чем меньше связей эта группа имеет с другими группами нейронов, тем более изолированную подзадачу она выполняет.

Для нейронных сетей небольшого размера данную задачу группирования можно решить методом перебора. Для нейронных сетей, размер которых превышает несколько тысяч нейронов, решение этой задачи будет затруднительно. Возникает необходимость в разработке алгоритма группирования нейронов для решения задачи сокращения трафика в коммутаторной нейронной сети.

Алгоритм фрагментации нейронной сети

Представим нейронную сеть в виде двух пересекающихся множеств: множество объектов, являющихся передающими информацию (имеющих выходы) и множество объектов, являющихся получателями информации (Рис. 1).

Для каждого элемента из множества A существует правило, по которому он связан с одним и более элементами из множества B . Это правило задается структурой нейронной сети и зависит от наличия и направления связи между нейронами. Для описания архитектуры коммутаторной нейронной сети, необходимо использовать множество C (множество коммутаторов)

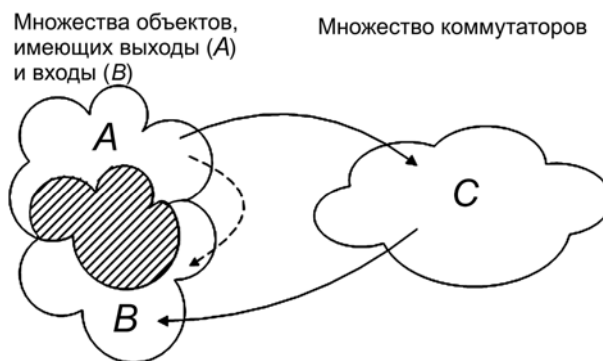


Рис. 1. Постановка задачи передачи информации

– элементов, передающих информацию. Таким образом для каждого правила связи элементов множеств A и B в нейронной сети будет существовать правило связи элемента множества A с последовательным рядом элементов из множества C , последний из которых связан с одним из элементов множества B . Последовательности связей элементов множества C между собой определяется древовидной структурой коммутаторной нейронной сети.

Требуется сократить общее количество элементов из множества C , необходимых для осуществления передачи информации от элементов множества A к элементам множества B .

Был разработан алгоритм для решения задачи группировки нейронов на группы заданного размера [3].

Пусть дана на плоскости координатная сетка, в узлах которой в произвольном порядке размещена группа объектов, соответствующих нейронам сети. Введем импульс притяжения $P_{прит}$ [кг \times м/с] между объектами таким образом, что если между двумя нейронами в сети существует связь, то эти две вершины притягиваются друг к другу. Импульс притяжения $P_{прит}$ прямо пропорционален весу связи $W(i,k)$ между объектами i и k в рассматриваемой нейронной сети и положительному коэффициенту притяжения A , используемому для масштабирования значений импульсов притяжений. Импульс притяжения должен быть постоянным для каждой пары объектов.

Для предотвращения слияния объектов в общее ядро введем импульс отталкивания $P_{отм}$ [кг \times м/с] между всеми объектами. Этот импульс обратно пропорционален расстоянию

между объектами $S(i, k)$ и прямо пропорционален положительному коэффициенту отталкивания B . Для того чтобы объекты могли объединяться в группы, следует уменьшить импульс отталкивания, если объекты сблизилась на определенное расстояние. Для этого используется коэффициент группирования C . В случае, когда расстояние между двумя объектами $S(i, k)$ на плоскости меньше заданного C , импульс отталкивания $P_{omm}=0$.

Время выполнения одного такта алгоритма t принято равным 1. В случаях, когда из-за дискретности вычислений затруднительно определить результирующее количество групп, величина t изменяется до нужного значения. Расстояние, на которое приблизятся объекты, определяется как $(P_{nrum}/m)t$. Масса одного объекта m принимается равной 1. Расстояние, на которое отдалятся объекты $= (P_{omm}/m)t$. Итоговое расстояние после одного такта алгоритма между парой объектов S_i находится как сумма начального расстояния S_0 и перемещения ΔS

$$S_i = S_0 - (P_{nrum}/m)t + (P_{omm}/m)t;$$

$$S_i = S_0 + (t/m)(B/S_0 - A)$$

Определив результирующий импульс, получим направление движения для каждого объекта и значение перемещения. Через некоторое время объекты, имеющие между собой импульсы притяжения, должны оказаться рядом на плоскости. Изменяя коэффициенты притяжения и отталкивания, можно варьировать количество и размеры получаемых элементарных сегментов.

Алгоритм группирования описывается следующим образом:

1 для каждой пары нейронов i и k провести следующие действия:

1.1. найти расстояние $S(i, k)$ между соответствующими объектами на плоскости;

1.2. рассчитать воздействующий на объекты импульс, как разность P_{omm} и P_{nrum} ;

1.3. рассчитать $\Delta x(i, k)$ и $\Delta x(k, i)$ – перемещение каждого объекта из пары;

1.4. рассчитать $\Delta y(i, k)$ и $\Delta y(k, i)$ – перемещение каждого объекта из пары;

2. Переместить каждый объект на плоскости в соответствии с рассчитанными $\Delta x(i)$ и $\Delta y(i)$;

3. Определить, пришла ли система из объектов на плоскости в состояние покоя, сравнив значения перемещений каждого объекта $\Delta x(i)$ и $\Delta y(i)$ с заданными минимальными величинами $\Delta x_{min}(i)$ и $\Delta y_{min}(i)$.

4. Повторить п.1-3 M раз до тех пор, пока система не придет в состояние равновесия.

Значение M зависит от количества нейронов N , от величины связей между ними W_{max} , от топологии сети и от выбранных коэффициентов A и B . Экспериментально определено, что значение коэффициента M сопоставимо с N^3 .

Трудоемкость алгоритма:

$$O(M)(O(N^2)+O(N)+O(N)) = O(M(N^2+N)).$$

Теоретическая сходимость алгоритма

Покажем, что через определенное количество тактов алгоритм, в соответствии с начальными условиями и заданными ограничениями, позволит получить результирующее положение объектов на плоскости, определяющее принадлежность объектов к тем или иным группам.

Каждый объект имеет координаты (x, y) на плоскости. Расстояние $S(i, k)$ между любыми двумя объектами i и k находится как:

$$S(i, k) = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}.$$

Для того чтобы система находилась в состоянии покоя, необходимо, чтобы импульсы притяжения и отталкивания были уравновешены, в таком случае суммарное перемещение всех объектов на плоскости будет равняться нулю.

$$\Delta S = \sum_{i=1}^N \Delta \bar{S}_i = \sum_{i=1}^N \sum_{k=1}^N \bar{S}_{oik} +$$

$$+ \left[\frac{B}{S_{oik}} - A |W(i, k)| \cdot \overline{1(i, k)} \right] = 0.$$

Суммарное перемещение всех объектов по осям координат будет равно:

$$\Delta x = \sum_{i=1}^N \sum_{k=1}^N \left[\left(\frac{B}{\sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}} - A |W(i, k)| \right) \frac{x_i - x_k}{\sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}} \right],$$

$$\Delta y = \sum_{i=1}^N \sum_{k=1}^N \left[\left(\frac{B}{\sqrt{(x_i - x_k)^2 + (y_i + y_k)^2}} - A|W(i, k)| \right) \frac{y_i - y_k}{\sqrt{(x_i - x_k)^2 + (y_i + y_k)^2}} \right].$$

Закон движения представлен системой функций вида: $y = f_x(x, y)$; $x = f_y(x, y)$.

Коэффициенты притяжения A , отталкивания B и группирования C в общем случае принимаются константами, вес связей является постоянным значением и не изменяется. Для пар объектов, не имеющих связей, вес связи между ними принят равным нулю. Для того чтобы суммарное перемещение объектов со временем стало нулевым, необходимо, чтобы на каждом шаге ΔS уменьшалось.

Таким образом, величина перемещения по каждой из осей координат должна уменьшаться с каждым тактом расчета действующих импульсов и перемещения объектов по плоскости. Объекты на плоскости могут располагаться произвольным образом с соблюдением единственного необходимого условия – изолированности, непринадлежности к группам.

В общем случае при соблюдении начальных условий задачи получим:

$$\frac{B}{\sqrt{(x_i - x_k)^2 + (y_i + y_k)^2}} - A|W(i, k)| < 0,$$

при $W(i, k) > 0$.

при $x_i > x_k$; $y_i > y_k$: $\Delta x_i < 0$; $\Delta y_i < 0$;

при $x_i > x_k$; $y_i < y_k$: $\Delta x_i < 0$; $\Delta y_i > 0$;

при $x_i < x_k$; $y_i > y_k$: $\Delta x_i > 0$; $\Delta y_i < 0$;

при $x_i < x_k$; $y_i < y_k$: $\Delta x_i > 0$; $\Delta y_i > 0$.

Следовательно, для любого соотношения координат (x_i, y_i) и (x_k, y_k) :

$$\begin{aligned} S(i, k)_2 &= S(i, k)_1 + \Delta S(i, k) = \\ &= \sqrt{(x_i - x_k + \Delta x_{ik})^2 + (y_i - y_k + \Delta y_{ik})^2} < S(i, k)_1. \end{aligned}$$

При выполнении условия $B/S > A$: $x_i \rightarrow x_k$; $y_i \rightarrow y_k$ и $\Delta S(i, k)_2 < \Delta S(i, k)_1$.

При нарушении этого условия получаем следующее:

при $x_i > x_k$; $y_i > y_k$: $\Delta x_i > 0$; $\Delta y_i > 0$;

при $x_i > x_k$; $y_i < y_k$: $\Delta x_i > 0$; $\Delta y_i < 0$;

при $x_i < x_k$; $y_i > y_k$: $\Delta x_i < 0$; $\Delta y_i > 0$;

при $x_i < x_k$; $y_i < y_k$: $\Delta x_i < 0$; $\Delta y_i < 0$.

Для выражения

$$\Delta S_{ik} = \left[\frac{B}{S_{0ik}} - A|W(i, k)| \cdot \overline{1(i, k)} \right]$$

$$\Delta S(i, k)_2 < \Delta S(i, k)_1.$$

Вне зависимости от расстояния при применении законов движения $x = f(x, y)$; $y = f(x, y)$ на каждом такте получим:

$$\lim_{t \rightarrow \infty} \overrightarrow{\Delta S_i} = \frac{BN}{\sum_{k=1}^N \overline{S_k}} - A \sum_{k=1}^N |W(i, k)| \cdot \overline{1(i, k)} = 0;$$

$$\lim_{t \rightarrow \infty} \Delta x = \sum_{i=1}^N \sum_{k=1}^N \left[\left(\frac{B}{S(i, k)} - A|W(i, k)| \right) \left(\frac{x_i - x_k}{S(i, k)} \right) \right] = 0.$$

$$\lim_{t \rightarrow \infty} \Delta y = \sum_{i=1}^N \sum_{k=1}^N \left[\left(\frac{B}{S(i, k)} - A|W(i, k)| \right) \left(\frac{y_i - y_k}{S(i, k)} \right) \right] = 0.$$

По условию задачи дано, что все объекты являются нейронами нейронной сети, соответственно, у каждого объекта имеется хотя бы одна связь с другим объектом, все объекты образуют цепь таким образом, что от каждого объекта можно найти такую последовательность связанных с ним объектов, которая позволяет построить соединение каждый-каждый. В противном случае будет рассматриваться задача группирования двух и более независимых нейронных сетей. В этом случае $\Delta S(i, k)_2 < \Delta S(i, k)_1$, но итоговое $\Delta S(i, k)$ никогда не примет значение, равное нулю. Итоговое суммарное расстояние между каждой парой объектов на плоскости на последнем шаге исполнения алгоритма определяется так:

$$\lim_{t \rightarrow \infty} \sum_{k=1}^N \overline{S_{i, k}} = \frac{BN}{\sum_{k=1}^N (A|W(i, k)| \cdot \overline{1(i, k)})}.$$

Таким образом, при соблюдении начальных условий (при группировании нейронной сети) алгоритм является детерминированным, понятным, завершаемым, массовым и в общем смысле результативным. Результативность алгоритма будет определяться соотношением коэффициентов A , B и C . Путем регулирования величин коэффициентов нейронная сеть группируется на группы разного размера.

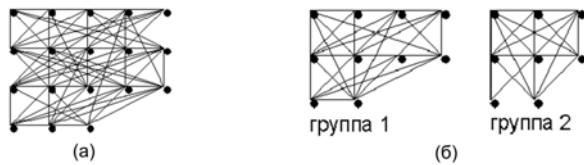


Рис. 2. Пример группирования
а) исходная сеть, б) результат группирования

В [4] показана эффективность работы алгоритма при группировании нейронной сети с количеством связей приблизительно на один порядок больше количества объектов. Всего в рассматриваемой в примере сети 18 нейронов и 126 связей между ними. Количество нейронов в каждой группе не должно превышать 10. Нейроны произвольно разделим на 2 подгруппы по 9 нейронов в каждой. Подключив эти подгруппы к коммутаторам первого уровня, получим следующий результат: в первой подгруппе будет 45 связей между элементами, во второй 29, а через коммутатор второго уровня пройдет 52 связи.

Затем подключим подгруппы, полученные в процессе работы алгоритма группирования. На первом коммутаторе первого уровня будет находиться 10 элементов и 65 связей между ними. На втором коммутаторе первого уровня будет находиться 8 элементов и 48 связей между ними. Через коммутатор второго уровня пройдет 13 связей (Рис.2).

Рассчитаем G - суммарное количество линий передачи информации коммутатор-коммутатор и коммутатор-нейрон, пройденных потоком информации:

для первого случая: $G_1 = (45+29) \cdot 2 + 52 \cdot 4 = 356$.

для второго случая: $G_2 = (65+48) \cdot 2 + 13 \cdot 4 = 278$.

Таким образом, для рассматриваемого примера сокращение трафика информации при использовании предложенного алгоритма оптимизации составит 21,91% относительно первого варианта подключения.

Эффективность алгоритма существенно зависит от структуры и размеров нейронной сети, количества и величин связей. В ходе проведения экспериментальных исследований полностью связанных нейронных сетей, включающих порядка 10^2 нейронов, результирующее количество линий передачи данных составляло 0,97-0,8 от первоначальной величины. Для больших нейронных сетей, состоящих из сотен и тысяч ней-

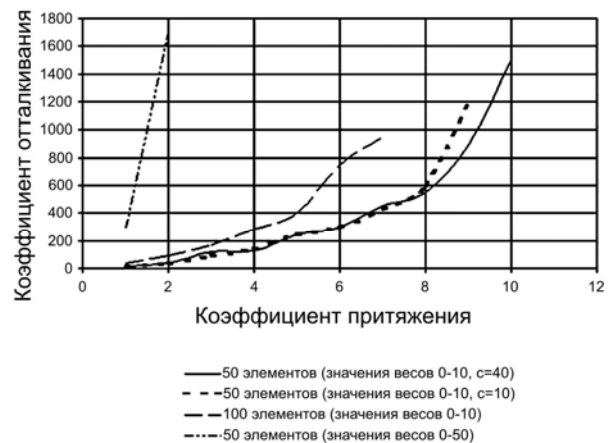


Рис. 3. Экспериментальная зависимость сходимости от коэффициентов

ронов, величина сокращения количества связей в коммутаторной нейронной сети зависит от соотношения количества связей и количества элементов. И при равном соотношении при увеличении размеров сети эффективность увеличивается.

Экспериментальная сходимость алгоритма

Проверим экспериментально полученные выводы. При моделировании алгоритма на персональном компьютере вычисления являются дискретными. Были проведены эксперименты по определению экспериментальной сходимости и получены следующие результаты (Рис. 3). Кривая отмечает границу экспериментальной сходимости.

Так, как из условия задачи известно:

$$S_1 = S_0 + (t/m)(B/S_0 - A).$$

Уменьшив величину такта, компенсируем дискретность вычислений. Это аналогично кратному сокращению коэффициентов A и B . В таком случае пропорционально уменьшаются величины Δx и Δy , и для любой пары коэффициентов алгоритм сходится.

Заключение

На основе разработанного алгоритма фрагментации нейронной сети решается задача декомпозиции нейронной сети на автономные группы, необходимая для перехода от структуры коммутаторной нейронной сети к доменной

нейронной сети [5,6]. Разработанный алгоритм фрагментации нейронной сети может применяться для задачи сокращения трафика в коммутаторной архитектуре нейронных сетей. Использование данного алгоритма позволяет при построении искусственных нейронных сетей большого размера на основе коммутаторной архитектуры сократить затраты на передачу информации между нейронами и количество информационных связей, используя переход к доменной архитектуре нейронной сети.

Литература

1. Кабак И.С., Суханова Н.В. Нейронная сеть. Патент на ПМ № 75247
2. Теория нейронных сетей. Кн. 1: Учеб пособия для вузов / Общая ред. А.И. Галушкина. - М.: ИПРЖР, 2000.- 416 с.
3. Степанов С.Ю. Снижение трафика информации в коммутаторной нейронной сети на основе ее фрагментации. Труды XVI международной научно-технической конференции "Информационные средства и технологии." В 3 томах, Т. 3. - М.: Издательский дом МЭИ, 2008 - стр. 236...241.
4. Степанов С.Ю. Группирование нейронов в двухуровневой коммутаторной нейронной сети. Труды международной научно-технической конференции "Информационные средства и технологии." 16-18 октября 2007г., в 3-х т., Т. 3. - М.:МЭИ, 2007. стр.178...181.
5. Кабак И.С. Доменная организация коммутаторных нейронных сетей. Труды международной научно-технической конференции "Информационные средства и технологии" 16-18 октября 2007г., в 3-х т., Т. 3 - М.:МЭИ, 2007. стр.128...131.
6. Кабак И.С., Суханова Н.В, Патент на ПМ № 72084 РФ. Доменная нейронная сеть.

Степанов Сергей Юрьевич. Аспирант кафедры Компьютерные системы управления МГТУ «Станкин». Окончил ГОУ ВПО МГТУ «Станкин» в 2007 году. Имеет 5 публикаций.