Д. В. Гаруля

EQUILIBRIUM PROPAGATION КАК АЛГОРИТМ ОБУЧЕНИЯ АНАЛОГОВЫХ НЕЙРОСЕТЕЙ

Рассматривается принцип Equilibrium propagation применяемый для обучения нейронных сетей. Показана связь между Equilibrium propagation и backpropagation.

Введение

Алгоритм обратного распространения ошибки (backpropagation) для обучения нейронных сетей считается биологически неправдоподобным. Одна из основных причин заключается в том, что обратное распространение требует отдельной вычислительной схемы и особого вида вычислений на этапе обучения. В данной работе мы рассматривается алгоритм обучения, называемый Equilibrium Propagation, который требует только одну вычислительную схему и один тип вычислений и для этапа обучения, и для этапа предсказания. Подобно тому, как обратное распространение применяется к любому дифференцируемому вычислительному графу (а не только к обычной многослойной нейронной сети), EquilibriumPropagation применяется ко всему классу моделей, основанных на энергии (прототипом которых является нейронная сеть Хопфилда).

I. Модели нейросетей основанные на энергии

В [1] выдвинута гипотеза о том, что для заданного состояния сенсорной информации, нейроны коллективно выполняют некоторое «умозаключение»: они движутся к конфигурациям (здесь под конфигурацией нейронов подразумевается набор значений сигналов в узлах сети), которые лучше «объясняют» наблюдаемые данные. Можно рассматривать конфигурацию нейронов как «объяснение» (или «интерпретацию») наблюдаемых данных. В представленной здесь энергетической модели это означает, что элементы сети постепенно движутся к конфигурациям с более низкой энергией, которые более вероятны, учитывая входные данные и в соответствии с текущей «моделью мира», связанной с параметрами модели. Обозначим через u набор узлов сети, а через $(\theta) = (W, b)$ - набор свободных параметров, которые должны быть изучены, который включает синаптические веса W_{ij} и смещения нейронов b_i. Узлы сети принимают непрерывное значение и соответствуют усредненному потенциалу напряжения по времени, пикам и, возможно, нейронам в одной и той же миниколонке (кортекса). p - нелинейная активационная функция, $p(u_i)$ представляет собой скорость активации узла i. Мы рассматриваем следующую энергетическую функцию E, разновидность энергии Хопфилда [2,3] :

$$E(u) := \frac{1}{2}\sum_i u_i^2 - \frac{1}{2}\sum_{i\neq j} W_{ij}p(u_i)p(u_j) - \sum_i b_i p(u_i)$$

Рассматриваемая сеть рекуррентно связана симметричными связями, то есть $W_{ij} = W_{ji}$. Представленный здесь алгоритм применим к любой архитектуре (при условии, что соединения симметричны), даже к полносвязной сети. Однако, чтобы сделать связь с Backpropagation более очевидной, мы рассмотрим более конкретно многоуровневую архитектуру без соединений с пропуском уровней и без боковых соединений внутри уровня (рисунок 1). В архитектуре сети, изучаемой здесь, блоки сети разделяются в трех наборах: входы x, которые всегда фиксированы (как и в других моделях, таких как условная машина Больцмана), скрытые узлы h (которые сами могут быть разделены на несколько уровней) и выходные узлы. Мы используем обозначение y для целевого значения, которое не следует путать с выходным зычением γ. Набор всех узлов в сети равен $u = \{x, h, \gamma\}$. Как обычно в сценарии контролируемого обучения, выходные данные повторяют свои целевые значения. Расхождение между выходными узлами и целевыми значениями измеряется квадратичной функцией потерь.

$$C := \frac{1}{2}||y - \gamma||^2$$

Функция потерь также действует как внешняя потенциальная энергия для выходных узлов, которая «направляет» их к целевым значениям.

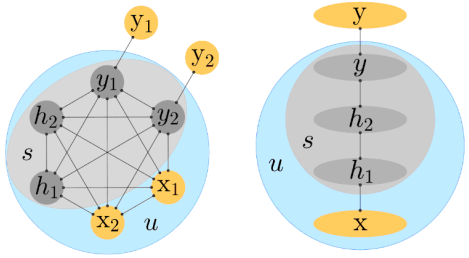

Рис. 1 – Архитектура сети. Слева: полносвязная сеть, справа: сеть, разделенная на 3 слоя (без соединений между узлами внутри слоя)

II. Принцип Equilibrium propagation

Отличием в этой работы по сравнению с ранее изученными моделями, основанными на энергии, является то, что сдесь введена «общая энергетическаю функция» F, которая принимает вид

$$F := E + \beta C$$

где $\beta \geq 0$ - это скаляр с действительными значениями, который управляет тем, будет ли выходной вектор «подталкиваться» к целевому вектору или нет и насколько сильно. Мы называем «параметр влияния» или «ограничивающий фактор». Полная энергия F представляет собой сумму двух потенциальных энергий: внутреннего потенциала E, который моделирует взаимодействия внутри сети, и внешнего потенциала βC, который определяет влияние целевых значений на выходные узлы. В отличие от машин Больцмана, где видимые(входные и выходные) узлы либо полностью свободны, либо полностью закреплены, здесь параметр с действительным знаком β позволяет слабо закреплять выходные узлы. Мы обозначаем переменную состояния сети как $s = \{h, \gamma\}$, которая не включает входные узлы x, поскольку они всегда фиксированны. Мы предполагаем, что временная эволюция переменных состояния определяется градиентной динамикой

$$\frac{ds}{dt} = -\frac{\vartheta F}{\vartheta s}$$

В отличие от более традиционных искусственных нейронных сетей, изучаемая здесь модель представляет собой динамическую систему с непрерывным временем, описываемую дифференциальным уравнением движения. Полная энергия системы уменьшается с течением времени (x и y остаются фиксированными) следовательно

$$\frac{dF}{dt} = \frac{\vartheta F}{\vartheta s} \cdot \frac{ds}{dt} = -||\frac{ds}{dt}||^2 \leq 0$$

Энергия перестает уменьшаться, когда сеть достигает фиксированной точки:

$$\frac{dF}{dt} = 0 \iff \frac{ds}{dt} = 0 \iff \frac{\vartheta F}{\vartheta s} = 0$$

Дифференциальное уравнение движения можно рассматривать как производную по времени от s:

$$\frac{ds}{dt} = -\frac{\vartheta E}{\vartheta s} - \beta\frac{\vartheta C}{\vartheta s}$$

«Внутренняя сила», индуцированная внутренним потенциалом (энергия Хопфилда) в n-м узле, равна

$$-\frac{\vartheta E}{\vartheta s_i} = p\prime(s_i)(\sum_{j \neq i} W_{ij} p(u_j) + b_i) - s_i$$

Форма данного уравнения напоминает модель нейрона как интегратора с утечкой (The leaky-integrator), нейроны которой рассматриваются как выполняющие текущую интеграцию своих прошлых входов. Обратите внимание, что предположение симметричных связей ($W_{ij} = W_{ji}$) использовалась для вывода данного уравнения. Как обсуждалось в [2], коэффициент $p\prime(s_i)$ обуславливает тот факт что когда нейрон насыщен (активируется с максимальной или минимальной скоростью, так что $p\prime(s_i) \approx 0$), его состояние нечувствительно к внешним входам, в то время как утечка выводит его из режима насыщения. В то время как «внешняя сила», индуцированная внешним потенциалом на h_i и y_i, соответственно $-\beta\frac{\vartheta C}{\vartheta h_i} = 0$ и $-\beta\frac{\vartheta C}{\vartheta y_i} = \beta(y_i - \gamma_i)$

Форма данного уравнения предполагает, что когда $\beta = 0$, блоки вывода нечувствительны к внешнему миру. В этом случае мы говорим, что сеть находится в свободной фазе (или первой фазе). Напротив, когда $\beta > 0$, «внешняя сила» движет выходной узел к цели y_i. В этом случае мы говорим, что сеть находится в слабо закрепленной фазе (или второй фазе). Наконец, более вероятная динамика будет включать в себя некоторый шум.

III. Выводы

В данной модели стационарное представление функции ошибок заменяется развернутым во времени распределением «энергии» системы. Таким образом, задача минимизации функции ошибок (в данном случае минимизации «энергии») сводится к задаче поиска неподвижной точки, которая затем заменяется представлением из стационарного распределения. Данный подход приводит к значительным вычислительным затратам если речь идет о цифровых вычислениях с дискретным временем, что делает equilibrium propagation неэффективным по сравнению с backpropagation. Тем не менее, рассматриваемый алгоритм является перспективным для аналоговых нейросетей, где каждый нейрон представлен как интегратор с утечкой (квазиинтегратор, leaky integrator). Кроме того, требование симметричных связей позволяет рассматривать представленную модель как подобие автоэнкодера.

1. G. E. Hinton, and T. J. Sejnowski / Parallel distributed processing: Explorations in the microstructure of cognition, 1986, v.1, pp. 282–317.
2. Y. Bengio and A. Fischer/TechnicalReport , 2015, arXiv:1510.02777, Universite de Montreal.
3. Bengio et al./ Neural Computation, 2017, pp 1–23.

Гаруля Дмитрий Владимирович, магистрант кафедры информационных технологий автоматизированных систем БГУИР, dimagarul58@gmail.com.

Научный руководитель: Навроцкий Анатолий Александрович, заведующий кафедрой информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент, navrotsky@bsuir.by.