

Филиал Московского Государственного Университета
имени М.В. Ломоносова в г. Ташкенте

Факультет прикладной математики и информатики

Кафедра прикладной математики и информатики

Мурадасилов Руслан Серверович

КУРСОВАЯ РАБОТА

на тему: «Проблема выразимости бинарных нейронных сетей»

по направлению 01.03.02 «Прикладная математика и
информатика»

Курсовая работа рассмотрена и рекомендована к защите

зав.кафедрой «ПМИИ», д.ф.-м.н., профессор _____ Кудрявцев В. Б.

Научный руководитель:

к.ф.-м.н. _____ Иванов И. Е.

«_____» _____ 2020 г.

Ташкент 2020

Аннотация

В данной работе рассматривается проблема выразимости бинарной нейронной сети. Рассматривается проблема полноты класса бинарных нейронных сетей, слоем которой является композиция линейной булевой функции и нелинейного бинарного оператора. Исследуется вопрос глубины таких нейронных сетей для разных нелинейностей, что впоследствии может быть применено на практике. Обозначена проблема аппроксимации булевых функций линейными булевыми функциями и предложена идея доказательства того, что все нелинейные булевы функции достаточно далеки от линейных.

Содержание

1	Введение	4
2	Историческая часть вопроса	5
3	Основные результаты	6
3.1	Полнота	6
3.2	Аппроксимация	9
4	Заключение	11
5	Список литературы	12

1 Введение

За последнее десятилетие сложность и способности нейронных сетей существенно выросли, однако их потенциал до сих пор ограничивают стоимость и энергия потребления. Как известно, нейронные сети состоят из нескольких слоев взвешенных сумм, которые предсказывают нужный результат. Хранение всех значений чисел с плавающей запятой значительно увеличивает время обучения и требует большое количество памяти, что вызывает необходимость использовать только специальное оборудование, которое выдержит подобную нагрузку.

Чтобы устройство с ограниченными ресурсами могло решать такие проблемы глубокого обучения, как распознавание лиц в реальном времени, необходимо использовать в качестве весов бинарные числа, а в качестве функций активации — бинарные аналоги функций активации в непрерывном случае, то есть «бинаризовать» нейронную сеть. Это позволит хранить гораздо больший объем данных, используя, например, 32-битный контроллер. Использование битовых операций сокращает время исполнения. Размеры бинарных нейронных сетей намного меньше, чем у их вещественных аналогов. Точность моделей также меньше, но эта разница в точности постепенно сокращается и бинарные нейронные сети становятся точнее на больших датасетах, как ImageNet.

В данной работе рассматривается вопрос полноты бинарной нейронной сети, то есть способность в точности выразить булеву функцию. А также ставится задача аппроксимации булевых функций, то есть способность выразить булеву функцию с некоторой допустимой погрешностью. С практической точки зрения устраивают оба варианта.

2 Историческая часть вопроса

Идея бинарной нейронной сети была впервые предложена Matthieu Courbariaux, где веса и функции активации используют только бинарные числа как и в обучении, так и в алгоритме обратного распространения ошибки с использованием метода градиентного спуска (SGD) [1].

Чуть позже Mohammad Rastegari в модели XNOR-Net [2] добавляет скалярное значение, чтобы компенсировать потерю информации во время бинаризации, который был получен из статистики весов и функций активации до бинаризации. Это улучшило общие показатели, но подсчет этого скалярного значения оказался дорогостоящим.

Victor Zhou попытался обобщить квантизацию и использовал преимущество битовых операций для фиксированной точки данных, варьируя ее размерность [3]. Zhou представил DoReFa-Net, модель с переменной размерностью весов, функций активации и даже вычисления градиента во время обратного распространения ошибки со значительно улучшенным временем обучения.

Другие модели также добились положительных результатов. Zheng Tang ускорил время обучения, изучив как влияет скорость обучения на показатели нейронной сети и на колеблемость бинарных значений [4]. Идея BNN+, созданная Sajad Darabi, также улучшает скорость обучения, используя другую эффективную функцию обратного распространения вместо импульсной [5].

Сравнение результатов всех этих моделей на разных датасетах были подробно описаны в работе Taylor Simons и Dah-Jye Lee [6]. Наглядно видно, что на MNIST, SVHN, CIFAR и ImageNet датасетах бинарные нейронные сети практически не уступают своим вещественным аналогам.

3 Основные результаты

3.1 Полнота

Введем основные определения.

Определение 1. *Бинарным оператором* называется конечный вектор $\bar{f} = (f_1, \dots, f_n)$, в котором каждая f_j является функцией алгебры логики от n переменных ($j = \overline{1, n}$).

Определение 2. Бинарный оператор называется *линейным*, если каждая его компонента является линейной. Соответственно, бинарный оператор называется *нелинейным*, если хотя бы одна из его компонент является нелинейной.

Определение 3. *Слоем* называется композиция линейной булевой функции и нелинейного бинарного оператора. *Входным слоем* будем называть вектор $\bar{X} = (x_1, \dots, x_n) \in \{0, 1\}^n$, а *выходным* — вектор $\bar{Y} = (y_1, \dots, y_m) \in \{0, 1\}^m$. Линейную часть слоя назовем *весовой функцией*. Она применяется ко всем выходам предыдущего слоя, включая входной слой. Нелинейный оператор назовем *функцией активации*. В качестве нее можно использовать и тождественный оператор.

Определение 4. *Нейронной сетью* называется композиция функций, каждая из которых является слоем.

Обозначим класс бинарных нейронных сетей с линейной булевой весовой функцией и нелинейным бинарным оператором в роли функции активации через \mathfrak{B} . Параметром этого класса будет семейство функций активации F . Справедлив следующий результат, относящийся к проблеме полноты бинарной нейронной сети с такой архитектурой.

Теорема 1. *Какое бы семейство функций активации F мы не взяли, класс бинарных нейронных сетей \mathfrak{B} будет полон в P_2 .*

Доказательство. Так как класс булевых линейных функций является предполным, то достаточно хотя бы одной нелинейной функции, чтобы

нейронная сеть могла порождать P_2 . По определению нелинейного бинарного оператора такая функция будет существовать, что доказывает теорему. \square

Рассмотрим вопрос существования полной бинарной нейросети с двумя слоями. Справедлива следующая теорема.

Теорема 2. *Существует бинарная нейронная сеть класса \mathfrak{B} с двумя слоями, которая полна в P_2 .*

Доказательство. Построим бинарную нейронную сеть, реализующую СДНФ функции. Во входном слое подаётся вектор $\bar{X} = (x_1, \dots, x_n) \in \{0, 1\}^n$, на котором хотим получить вектор значений. В первом слое будет бинарный оператор, состоящий из конъюнкций, то есть $f^1 = (f_1^1, \dots, f_{2^n}^1)$ $\forall f_j^1 = \wedge, j = \overline{1, 2^n}$. Весовая функция g первого слоя следующего вида:

$$g(x) = \begin{cases} x, & \text{х входит в конъюнкцию без отрицания} \\ 1 \oplus x, & \text{х входит в конъюнкцию с отрицанием} \end{cases}$$

Она позволит получать отрицания переменных при передаче их в конъюнкцию. Во втором слое бинарный оператор, принимающий на вход все конъюнкции СДНФ и реализующий их дизъюнкцию, то есть $f^2 = \vee$. Таким образом, функции активации (конъюнкция, дизъюнкция) — нелинейны, а весовая (исключающее «или») — линейна, и мы сохраняем структуру слоя бинарной нейронной сети класса \mathfrak{B} (Рис. 1).

Так как СДНФ функции конечна и единственна, то по ней мы сможем однозначно восстановить все функции алгебры логики, и такая нейронная сеть будет полна, что и требовалось доказать. \square

Исторически, проблема выразимости нейронных сетей с вещественными значениями прослеживается от 13-й проблемы Гильберта, ставящей вопрос "существует ли непрерывная функция трех переменных, которая не может быть представлена через композицию непрерывных функций двух переменных". Она была решена в 1957 г. В.А. Арнольдом; он показал, что любая непрерывная функция трех переменных представляется в виде композиции непрерывных функций двух переменных. В том

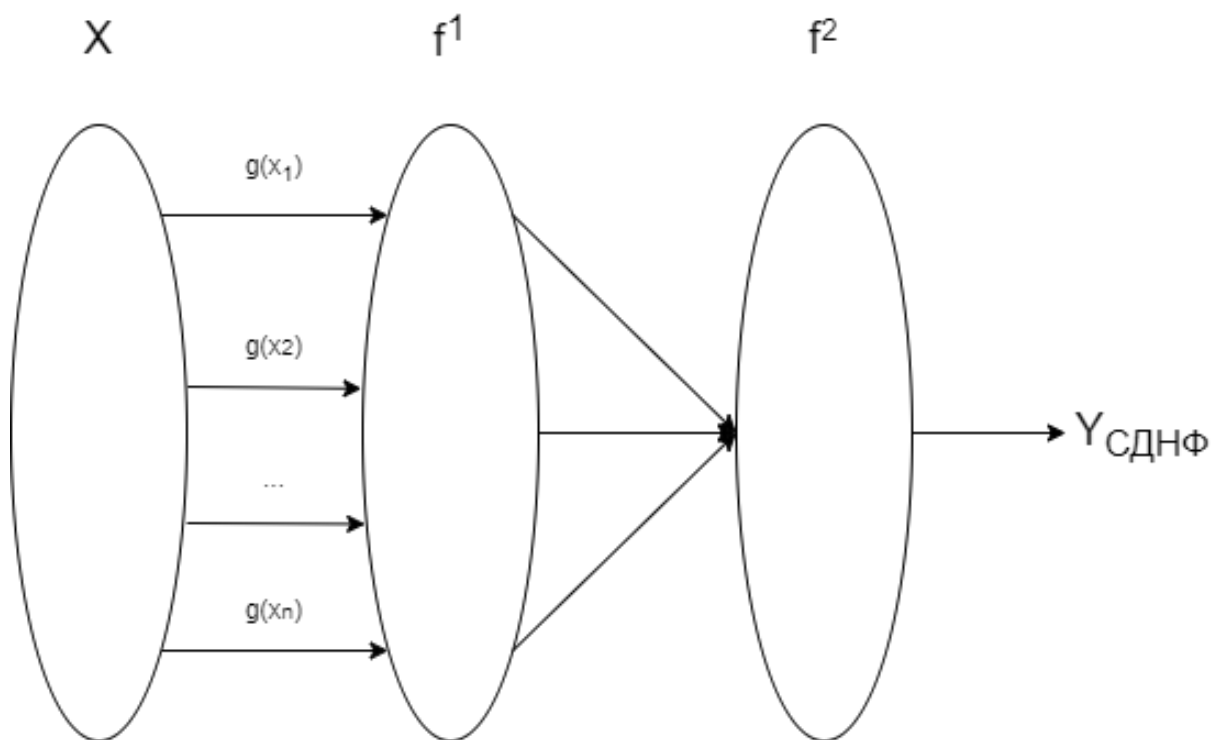


Рис. 1

же 1957 г. А.Н. Колмогоров доказал более сильную теорему: для реализации функций многих переменных достаточно операций суммирования и композиции функций одной переменной. Эта теорема в 1987 году была переложена Хехт–Нильсеном для нейронных сетей: любая функция нескольких переменных может быть представлена двухслойной нейронной сетью с прямыми полными связями с N нейронами входного слоя, $(2N + 1)$ нейронами скрытого слоя с ограниченными функциями активации (например, сигмоидальными) и M нейронами выходного слоя с неизвестными функциями активации [7, 8, 9, 10].

В бинарных нейронных сетях класса \mathfrak{B} справедлив аналогичный результат.

Теорема 3. *Какое бы семейство функций активации F мы не взяли, класс бинарных нейронных сетей \mathfrak{B} с четырьмя слоями будет полон в P_2 .*

Доказательство. Из доказательства критерия Поста [11] известно, что имея нелинейную функцию, инвертор и константу 1, получим конъюнк-

цию. Имея конъюнкцию и инвертор, получим дизъюнкцию. Такая система будет полна. Среди линейных функций есть константа 1 и инвертор, в роли которого будет выступать исключающее «или» в виде $\bar{x} = 1 \oplus x$. Тогда на первом слое мы получим конъюнкцию, на втором — дизъюнкцию, а на третьем и четвертом — мы построим СДНФ нужной функции по **Теореме 2**. \square

3.2 Аппроксимация

В практике расчетов, связанных с обработкой экспериментальных данных, вычислением $f(x)$, разработкой вычислительных методов, встречаются следующие две ситуации:

1. Как установить вид функции $y = f(x)$, если она неизвестна? Предполагается при этом, что задана таблица ее значений, которая получена либо из экспериментальных измерений, либо из сложных расчетов.
2. Как упростить вычисление известной функции $f(x)$ или же её характеристик, если $f(x)$ слишком сложная?

Ответы на эти вопросы даются теорией аппроксимации функций, основная задача которой состоит в нахождении функции $\varphi(x)$, близкой (т.е. аппроксимирующей) в некотором нормированном пространстве к исходной функции $y = f(x)$. Функцию $\varphi(x)$ при этом выбирают такой, чтобы она была максимально удобной для последующих расчетов.

Одним из замечательных свойств нейронных сетей с вещественными значениями является способность аппроксимировать и, более того, быть универсальными аппроксиматорами. С их помощью можно аппроксимировать сколь угодно точно непрерывные функции многих переменных. [12]

Что касается бинарных нейронных сетей, здесь этот вопрос ещё хорошо не изучен. Важно доказать, насколько хорошо можно аппроксимировать любую булеву функцию линейными, то есть важен вопрос линейной аппроксимации в бинарном случае.

Определение 1. Пусть f — булева функция от n переменных, то есть $f : V^n \rightarrow \{0, 1\}$, где V^n — булевый вектор размерности n . Ее *весом* $w(f)$ будем называть количество наборов, на которых функция принимает значение 1.

Определение 2. Пусть f, g — булевы функции от n переменных. *Расстоянием* от булевой функции f до булевой функции g называется величина

$$\text{dist}(f, g) = w(f \oplus g),$$

то есть количество наборов, на которых значения f и g различаются. Расстоянием от f до множества M булевых функций от n переменных назовем величину

$$\text{dist}(f, M) = \min_{g \in M} \text{dist}(f, g).$$

Гипотеза. Всегда найдется булева функция f из класса нелинейных функций такая, что расстояние от нее до класса линейных будет больше числа $k \in \mathbb{N}$, то есть $\text{dist}(f, L) > k$.

Задача. Найти максимальное число k , при котором гипотеза будет верна.

Основная идея. Известно, что линейных булевых функций от n переменных 2^{n+1} , а нелинейных — $2^{2^n} - 2^{n+1}$. То есть при $n \geq 3$ количество нелинейных функций начинает преобладать над количеством линейных. Необходимо рассмотреть систему шаров радиусами k , центрами которых будут линейные функции. Внутри этих шаров будут нелинейные функции. Начиная с $k = 1$, увеличивать значение радиусов шаров на 1, пока все нелинейные функции не попадут в них. Итоговый радиус и будет порогом, когда линейные функции получают способность аппроксимировать нелинейные.

Получив значение k , мы определим, насколько далеки нелинейные булевы функции от линейных.

4 Заключение

В данной работе исследована проблема полноты бинарной нейронной сети. Доказывается, что для полноты бинарных нейронных сетей, слоем которой является композиция линейной булевой функции и нелинейного бинарного оператора, в некоторых случаях достаточно двух слоев, а в общем случае — четырёх. В дальнейшем планируется получить результаты по глубине для разных функций активации.

Обозначена важность проблемы аппроксимации булевых функций линейными функциями. Предложена идея доказательства плохой аппроксимации линейными функциями. Ставится цель изучить аппроксимацию в бинарном случае нейронной сетью, в слое которой будут линейные функции и некоторые нелинейности, а также применять подобную архитектуру в практических задачах.

Автор выражает благодарность И. Е. Иванову за научное руководство.

5 Список литературы

- [1] Courbariaux, M.; Bengio, Y. BinaryNet: Training Deep Neural Networks with Weights and Activations Constrained to $+1$ or -1 . arXiv:1602.02830.
- [2] Rastegari, M.; Ordonez, V.; Redmon, J.; Farhadi, A. XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 525–542.32.
- [3] Zhou, S.; Ni, Z.; Zhou, X.; Wen, H.; Wu, Y.; Zou, Y. DoReFa-Net: Training Low Bitwidth Convolutional Neural Networks with Low Bitwidth Gradients. arXiv 2016, arXiv:1606.06160.
- [4] Tang, W.; Hua, G.; Wang, L. How to Train a Compact Binary Neural Network with High Accuracy? In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017
- [5] Darabi, S.; Belbahri, M.; Partovi Nia, V.; Courbariaux, M. Regularized Binary Network Training. arXiv:1812.11800
- [6] Simons, T.; Lee, D. A Review of Binarized Neural Networks. Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602, USA.
- [7] Колмогоров А. Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных. ДАН СССР, 1956, Т.108, №2, С. 179-182
- [8] Арнольд В.И. О функции трех переменных. ДАН СССР, 1957, Т.114, №4, С. 679-681
- [9] Колмогоров А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения. ДАН СССР, 1957, Т.114, №5, С. 953-956

- [10] Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem. IEEE First Annual Int. Conf. on Neural Networks, San Diego, 1987. Vol. 3. — P. 11—13.
- [11] Алексеев В. Б. Дискретная математика (II семестр). — М., МГУ, 2002. — 44 с.
- [12] Cybenko, G. V. Approximation by Superpositions of a Sigmoidal function // Mathematics of Control Signals and Systems. — 1989. — Т. 2, № 4. — С. 303—314.