

УДК 004.855

ОБЗОР АКТУАЛЬНОГО СОСТОЯНИЯ ИССЛЕДОВАНИЙ В ОБЛАСТИ ВЫЧИСЛЕНИЯ ОПТИМАЛЬНОГО КОЛИЧЕСТВА НЕЙРОННЫХ СЛОЁВ И НЕЙРОНОВ В СЛОЕ

Барабанов М.Ю., Абрамов И.О. магистрант гр.167001

Белорусский государственный университет информатики и радиоэлектроники

г. Минск, Республика Беларусь

Астровский И.И. – канд. техн. наук

Аннотация. В данной статье приводится обзор актуальных исследований по теме влияния количества нейронных слоев и нейронов в слое на общую характеристику искусственной нейронной сети. Найдены наиболее распространенные подходы.

Ключевые слова. Искусственная нейронная сеть, нейронный слой, машинное обучение, обзор.

Искусственные нейронные сети являются мощным инструментом для решения различных задач, включая распознавание образов, классификацию, регрессию, распознавание речи, машинный перевод и т.д. Одним из ключевых параметров нейронной сети является количество нейронных слоев и нейронов в каждом слое. Оптимальный выбор количества слоев и нейронов может существенно повлиять на производительность искусственной нейронной сети и качество ее работы. Об этом напрямую может свидетельствовать улучшение лингвистической нейронной модели GPT с каждой последующей версией [1]: увеличение количества слоев продемонстрировано на рисунке 1.

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or "GPT-3"	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

Рисунок 1 – Данные об улучшениях лингвистической модели GPT

Многие исследования были проведены для определения оптимального количества нейронных слоев и нейронов в каждом слое. Некоторые из этих исследований описывают определенные методы и эвристики для выбора количества слоев и нейронов, в то время как другие исследования сравнивают различные архитектуры ИНС, чтобы определить, какая из них является наилучшей для данной задачи.

Существует несколько подходов к определению оптимального количества слоев и нейронов. Один из таких подходов - это эмпирический метод, основанный на экспериментальном тестировании различных конфигураций сетей. Однако этот метод является довольно затратным по времени и ресурсам, так как в некотором смысле основан на удаче подбора данных.

Другой подход основан на математических моделях, таких как теория информации и теория сложности. В этих моделях оптимальное количество слоев и нейронов определяется с учетом статистических свойств данных, на которых будет работать сеть. Однако эти модели могут быть достаточно сложными и трудными для понимания.

Одним из основных критериев количества нейронных слоев долгое время была теорема Цыбенко, она же универсальная теорема аппроксимации. Данная теорема, доказанная Джорджем Цыбенко в 1989 году, утверждает, что искусственная нейронная сеть прямой связи (в которых связи не образуют циклов) с одним скрытым слоем может аппроксимировать любую непрерывную функцию многих переменных с любой точностью. Условиями выполнения являются: достаточное

количество нейронов скрытого слоя, удачный подбор весов нейронов скрытого и выходного нейрона, а также смещения весов входного слоя. [2]

Однако, более новые исследования показали, что использование нескольких скрытых слоев может привести к лучшим результатам для некоторых задач. [3] Хотя эта теорема утверждает, что достаточно одного скрытого нейронного слоя для любой задачи, она не говорит ничего о времени требуемое на обучение этой системы и количество нейронов внутри скрытого слоя, число которых может потребоваться значительно больше, чем в архитектуре с двумя и более скрытыми нейронными слоями.

Долгое время использовался только один скрытый нейронный слой по причине технической ограниченности в обучении большого количества слоев. Изменение произошло в 2006 году после выхода работы Hinton и Osindero, основной особенностью которой является алгоритм быстрого обучения глубоких сетей. [4]

Далее, в статье [2] было проведено исследование оптимального количества нейронов в скрытых слоях для классификации изображений. Авторы рассмотрели наборы данных MNIST и CIFAR-10 и исследовали производительность нейронных сетей с различным количеством нейронов в скрытых слоях. Они показали, что оптимальное количество нейронов может варьироваться в зависимости от конкретной задачи и набора данных.

Другие исследования сравнивают различные архитектуры искусственных нейронных сетей, такие как перцептрон, многослойный перцептрон, сверточная нейронная сеть и рекуррентная нейронная сеть. Например, в статье "ImageNet Classification with Deep Convolutional Neural Networks" авторы использовали сверточные нейронные сети для классификации изображений на датасете ImageNet и достигли значительного улучшения результатов по сравнению с традиционными методами, включая многослойные перцептроны. [6] Таким образом, важно понимать не только сложность работы нейронной сети для определения количества слоев, но также и её архитектуру, так как разные нейронные слои могут по-разному обрабатывать нужные типы данных.

Наиболее частые рекомендации, упомянутые в сети Интернет, по определению количества нейронных слоев для создания нейронной сети продемонстрированы в таблице 1.

Таблица 1 – Общие сведения о количестве

Количество скрытых слоев	Применение
0	Только способно представлять линейные разделимые функции или решения
1	Может аппроксимировать любую функцию, которая содержит непрерывное отображение из одного конечного пространства в другое
2	Может представлять произвольную границу решения с произвольной точностью с рациональными функциями активации и может аппроксимировать любое гладкое отображение с любой точностью.
Более 2	Дополнительные слои могут изучать сложные представления (своего рода автоматическая разработка признаков) для слоев слоев

Важно понимать, что определение числа слоев нейронной сети недостаточно для построения нейронной сети, важно также определить количество нейронов в каждом слое. Чаще всего в задачах регрессии и распознавания, количество нейронов с каждым последующим слоем уменьшается, что следует из работы сети. Однако не все задачи сводятся к такому рода работе. Из наиболее частых рекомендаций по определению количества нейронов можно найти следующие:

- Количество скрытых нейронов должно быть между размером входного слоя и размером выходного слоя.;
- Количество скрытых нейронов должно составлять 2/3 размера входного слоя плюс размер выходного слоя.;
- Количество скрытых нейронов должно быть менее чем в два раза больше размера входного слоя.

Результаты, которые могут быть получены на основе нейронных сетей с разным количеством слоев могут сильно отличаться, так как сети получив одинаковое количество обучающего материала по-разному смогут адаптироваться к нему. Опасным может быть не только недообучение сети, но и переобучение, когда построенная модель хорошо объясняет примеры из обучающей выборки [7][8],

но относительно плохо работает на примерах, не участвовавших в обучении. Нагляднее процесс переобучения можно увидеть на рисунке 2.

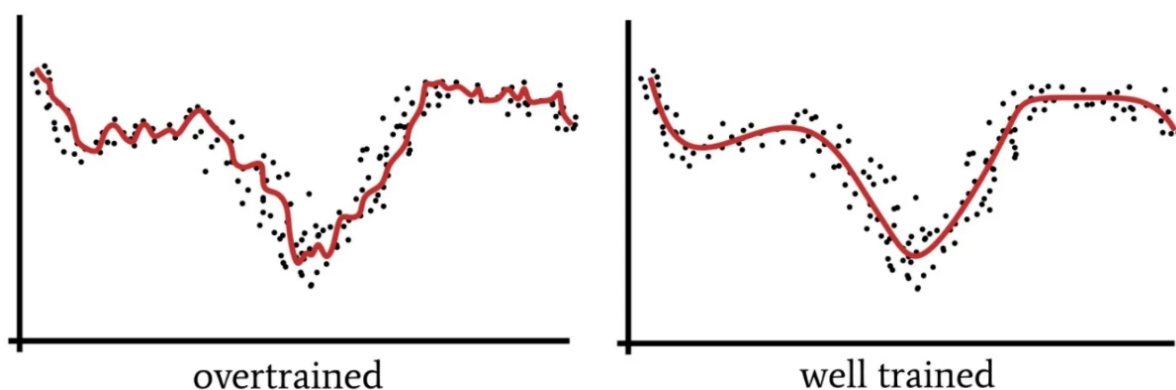


Рисунок 2 – Результаты переобученной и хорошо натренированной ИНС

Именно поэтому так важно определение количества нейронных слоев и нейронов в слое для дальнейшего создания и тренировки в сети, ведь как малое, так и чрезмерное количество слоев и нейронов могут привести к неудачным результатам.

В целом, определение оптимальной конфигурации нейронной сети - это сложная задача, которая требует сбалансированного подхода и учета множества факторов. Из многих исследований, проведенных в области определения оптимального количества нейронных слоев и нейронов, можно сделать вывод, что нет универсального правила для выбора архитектуры ИНС. Несмотря на это, исследователи продолжают работать над различными методами и подходами для решения этой задачи. Оптимальное количество слоев и нейронов зависит от конкретной задачи, объема данных и других факторов, поэтому необходимо проводить тщательный анализ с регулярной проверкой системы эмпирическими результатами.

Список использованных источников:

1. Tom B. Brown, Benjamin Mann, Nick Ryder // Language Models are Few-Shot Learners - 2020. P. 6
2. Approximation by superpositions of a sigmoidal function [Электронный ресурс] – режим доступа: <https://web.archive.org/web/20151010204407/http://deeplearning.cs.cmu.edu/pdfs/Cybenko.pdf>
3. LeCun, Y., Bengio, Y., & Hinton, G. // Deep learning. Nature 521(7553), – 2015 - P. 436-444.
4. A fast learning algorithm for deep belief nets * [Электронный ресурс] – режим доступа: <https://www.cs.toronto.edu/~hinton/absps/fastnc.pdf>
5. Srivastava, N. // Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. - 2014 – P.15
4. Zhang, W., & Du, Y. // An empirical study on the optimal number of hidden neurons in neural network. IEEE Access – 2018 – P. 4850-4862.
6. ImageNet Classification with Deep Convolutional Neural Networks [Электронный ресурс] – режим доступа: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
7. Simonyan, K., & Zisserman, A. // Very deep convolutional networks for large-scale image recognition. – 2015 – P. 1409-1556.
8. Muhammad Uzair. Effects of Hidden Layers on the Efficiency of Neural networks Muhammad Uzair, Noreen Jamil. // IEEE 23rd International Multitopic Conference – 2020