

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО РАЗМЕРА НЕЙРОСЕТИ ОБРАТНОГО РАСПРОСТРАНЕНИЯ ЧЕРЕЗ СОПОСТАВЛЕНИЕ СРЕДНИХ ЗНАЧЕНИЙ МОДУЛЕЙ ВЕСОВ СИНАПСОВ

Царегородцев В.Г.

tsar@neuropro.ru www.neuropro.ru

Предлагается новая "кривая обучения" – график зависимости среднего модуля веса синапса от размера нейросети. Эксперименты показывают, что локальные минимумы и выходы на асимптоты данного индикатора хорошо соответствуют свойствам традиционных кривых обучения – зависимостей ошибок обучения и обобщения от размера нейросети. Индикатор может быть использован для определения оптимального размера сети при отсутствии тестовой выборки.

1. Задача определения оптимальной структуры нейросети

При использовании искусственных нейронных сетей важной задачей является нахождение оптимального размера (структуры) сети – такого числа скрытых слоев нейронов и нейронов в слоях, которые дадут максимум обобщающих способностей, т.е. минимум ошибки обобщения (generalization error), особенно в случае отсутствия независимой тестовой выборки или невозможностью искусственно разделить выборку данных на обучающую и тестовую части из-за недостаточности общего объема данных.

Поэтому широко используется парадигма "кривых обучения" (learning curves) – зависимостей ошибок обучения и обобщения от размеров нейросети и обучающей выборки [1-4]. Оптимуму соответствуют локальные минимумы или моменты выхода графиков на асимптоты. Формальные приемы экстраполяции таких графиков [2] позволяют также оценивать необходимые и достаточные для достижения максимума обобщающих способностей объемы обучающих выборок в случае первоначальной недостаточности объемов выборочных данных.

Иным классом кривых обучения являются зависимости "внутренних" свойств нейросети от её размера, сопоставляемые затем с динамикой ошибки обобщения. Варианты – анализ [1] внутреннего представления (internal representation) задачи, теоретическая связь [5] ошибки обучения и максимума суммы модулей весов синапсов, приходящих на нейрон сети, NIS-критерий [6], оперирующий с градиентами целевой функции и матрицей Гессе обученной сети и позволяющий оценить разницу между ошибками обучения и обобщения. Такие критерии позволяют обходиться без независимой тестовой выборки.

В работе предлагается новый вариант кривой обучения – зависимость среднего модуля веса синапса от размера нейросети. Точнее, в экспериментах далее будет использовано значение длины вектора весов синапсов сети (вычисленного в евклидовой норме), деленное на общее число синапсов, с целью увеличения влияния наибольших по модулю весов и следующей из этого перестраховке исходя из результатов [5] о нежелательности именно больших весов синапсов.

Этот критерий не является всеобъемлющим, т.к. имеется неоднородность наборов синапсов сети от слоя к слою (для сетей малого размера в часто наблюдалось статистическое различие средних модулей и дисперсий весов синапсов выходного и скрытого слоя сети). Структурная неоднородность слоистых сетей известна и уже учитывается алгоритмами обучения [7], но здесь влияние этого факта не исследуется.

2. Данные для экспериментальной проверки и результаты

Было взято 6 баз реальных данных, имеющих независимые тестовые выборки (чтобы не вносить погрешность в оценку ошибки обобщения способом разбиения обучающей выборки на обучающую и тестовую части). Взяты базы данных AnnThyroid, Opt digits, Pen digits, Satellite, Statlog shuttle из UCI KDD Database Repository <http://kdd.ics.uci.edu/>, и база данных Gong [8], доступная на страничке http://www-ee.uta.edu/eeweb/IP/training_data_files.htm. Все 6 задач представляют собой задачи классификации с учителем на то или иное число классов.

Все эти задачи обладают значительным, от нескольких тысяч до нескольких десятков тысяч векторов, размером обучающей выборки – данное условие нужно для гарантирования представительности выборки (и, соответственно, наличия четкой асимптотики в ошибках обучения и обобщения после достижения и превышения нейронной сетью адекватного для задачи размера) и отсутствия возникновения эффекта переобучения при дальнейшем росте размера нейросети (шум и искажения в обучающей выборке, если таковые наличествуют, не смогут быть запомнены нейросетью из-за значительного, при большом объеме выборки, числа таких искажений, а не единичности случаев этих искажений).

Использовались сети с одним скрытым слоем, число нейронов в котором менялось от 1 до 25. В каждой задаче для каждого размера нейросети обучалось по 25 сетей (с разными начальными случайными значениями синапсов), свойства которых затем усреднялись при построении кривых обучения.

Далее на графиках представлены следующие кривые обучения:

- средние значения ошибок обучения и обобщения (выраженные в процентах доли неправильно решенных примеров в объеме соответствующей выборки);
- среднеквадратичный вес синапса в сети – предлагаемый индикатор;
- максимум среди понейронных сумм модулей весов синапсов – индикатор [5].

Вдоль осей ординат отложено число нейронов в скрытых слоях сетей. Значения индикаторов, отражающих свойства весов синапсов, перемасштабированы для приведения в диапазон значений величин ошибок обучения и обобщения, что было вызвано ограничениями программы построения графиков (невозможностью ввода двух шкал). Вокруг каждой точки отложена дисперсия соответствующей выборки из 25 экспериментальных значений.

Видно, что выход нового индикатора на асимптоту (и стабилизация – такое снижение дисперсии, что "усы" вокруг точки закрываются самой точкой) немножко отстает от выхода ошибок обучения и обобщения на асимптоты, т.е. немножко перестраховывается в плане требуемого размера сети, что можно только приветствовать исходя из теоретических результатов [5]: увеличение числа путей прохождения сигнала по сети может снизить максимальные веса синапсов за счет размножения каналов, где ранее требовалось усиление.

Также индикатор выявляет выход ошибки обобщения на оптимум во всех двух случаях возникновения переобучения (задачи AnnThyroid, Gong), когда с ростом размера сети с некоторого момента ошибка обобщения начинает снова возрастать – момент стабилизации и выход индикатора на асимптоту чуть запаздывает по сравнению с моментом достижения минимума ошибки в задаче AnnThyroid, а в задаче Gong локальный минимум при размере сети в 6 нейронов точно соответствует минимуму ошибки обобщения. Индикатор же [5] в задаче Gong не имеет четко выраженного экстремального поведения существенно нестабилен на всем диапазоне исследованных размеров нейросети – от 1 до 25 нейронов.

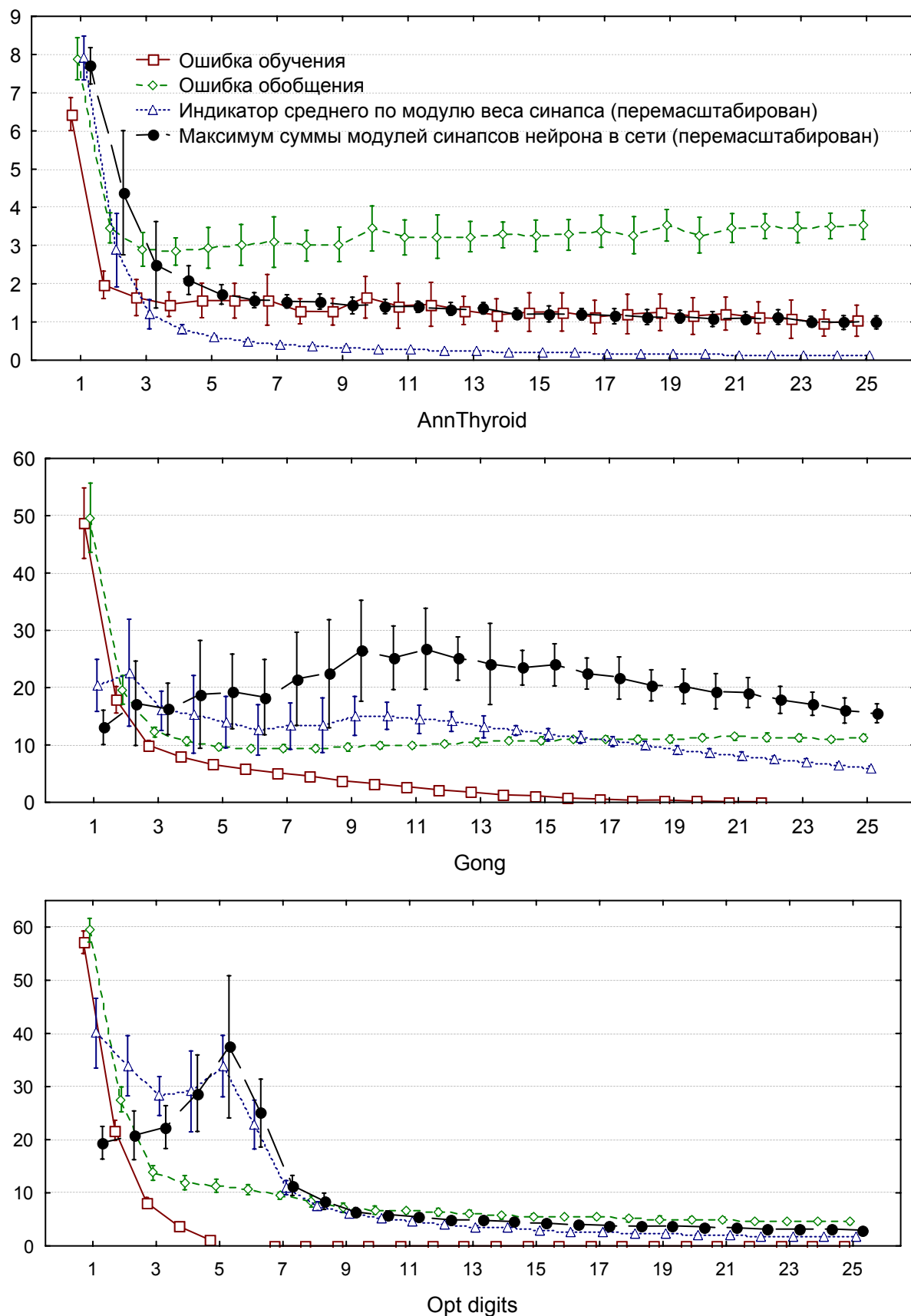


Рис. Зависимости величин ошибок обучения и обобщения (процент неправильно решенных примеров в соответствующей выборке) и индикаторов внутренних свойств нейросетей от числа нейронов в скрытом слое сети (от 1 до 25 нейронов).

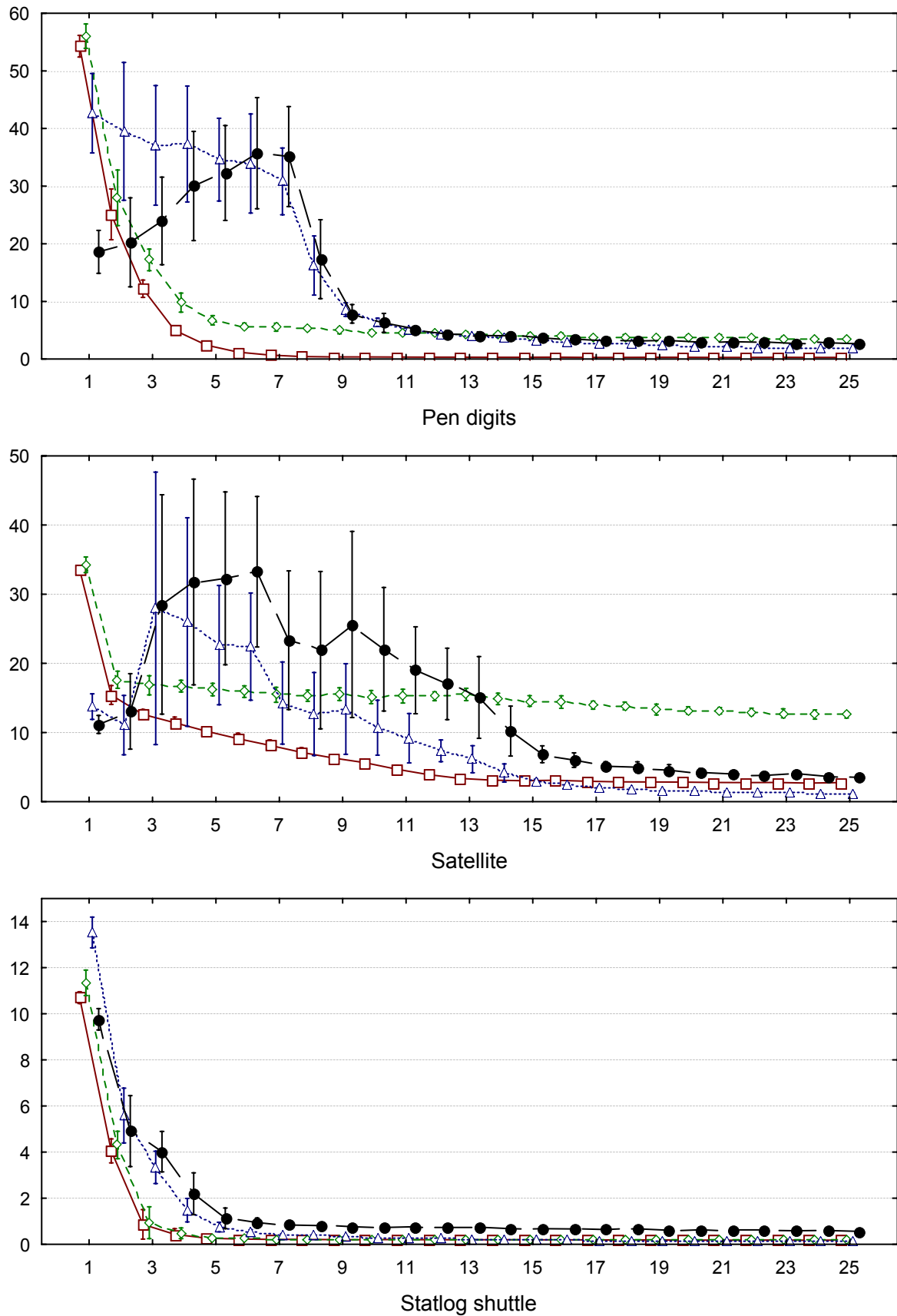


Рис. Зависимости величин ошибок обучения и обобщения (процент неправильно решенных примеров в соответствующей выборке) и индикаторов внутренних свойств нейросетей от числа нейронов в скрытом слое сети (от 1 до 25 нейронов).

Локальные минимумы индикатора (шесть нейронов для задачи Gong, три для задачи Opt digits, два для задачи Satellite) также могут указывать на оптимум ошибки обобщения (задача Gong) или на структурные уровни сложности задачи (последнее совпадает с изломами графиков ошибок обучения и обобщения). Последнее может позволять идентифицировать моменты перехода от области адекватности малопараметрических моделей классической статистики (линейная регрессия, линейный дискриминант или байесовский классификатор на основе оценок ковариационных матриц для каждого класса) к областям адекватности многопараметрических моделей (нейронные сети, полиномиальные аппроксимации) или непараметрических методов (непараметрическая статистика на основе ядерных аппроксимаций плотностей вероятности, метод потенциальных функций).

Также индикатор чуть быстрее снижает свою дисперсию по набору проб, чем максимальная понейронная сумма модулей весов синапсов [5], что в реальной работе позволит обойтись меньшим числом попыток обучения для каждого размера нейросети, или даже вообще без необходимости статистического усреднения свойств нескольких нейросетей одного размера для получения четкой картины на графиках наподобие приведенных в данной работе.

Как видно из экспериментальных графиков, при выборе оптимального размера сети опираться только на значение ошибки обучения недостаточно – нельзя выявить возникновение переобучения нейросети, поэтому сопоставление поведения нескольких индикаторов (как было сделано на приведенных графиках) позволяет либо более обоснованно подтвердить выбор размера нейросети, либо увидеть возможное существование проблем (например, неадекватности модели из-за возникновения переобучения). Возможность же обойтись без проверки на тестовой выборке позволяет обучать нейросеть на всем доступном наборе примеров, без деления его на обучающий и тестовый фрагменты, и ожидать, что с ростом числа обучающих примеров снизится и риск переобучения нейросети [2].

3. Заключение

Предложен новый вариант кривой обучения – зависимость среднего значения модуля веса синапса в сети от размера нейросети. Экспериментально показано, что с его помощью возможно достаточно надежное определение оптимального размера сети, обеспечивающего минимум ошибки обобщения. Индикатор позволяет обходиться без расчета ошибки обобщения на независимой тестовой выборке, допускает вариации путем выбора нормы (модуль веса, среднее квадратичное значение,...) и учета структурной неоднородности сети для максимизации прогностических способностей.

Также данный критерий может быть применен и при обучении растущих нейросетей, наподобие нейросетей каскадной корреляции, причем как на этапе отбора обученного нейрона-кандидата для вставки в нейросеть (наряду с использованием значения целевой функции для этого нейрона), так и после вставки отобранного нейрона в сеть и коррекции последней (не единственный отобранный нейрон-кандидат вставляется в нейросеть, а несколько лучших возможных нейронов вставляются каждый в свою копию нейросети, и уже эти доученные копии сравниваются между собой как по значению целевой функции, так и по предлагаемому индикатору). Подобные эксперименты автор планирует провести в дальнейшем.

Работа поддержана грантом 15G277 Красноярского краевого фонда науки.

Литература

1. Watanabe E., Shimizu H. Relationships between internal representation and generalization ability in multi layered neural network for binary pattern classification problem / Proc. IJCNN'1993, Nagoya, Japan, 1993. Vol.2. – pp.1736-1739.
2. Cortes C., Jackel L.D., Solla S.A., Vapnik V., Denker J.S. Learning curves: asymptotic values and rate of convergence / Advances in Neural Information Processing Systems 6 (1993). Morgan Kaufmann, 1994. – pp.327-334.
3. Cortes C., Jackel L.D., Chiang W.-P. Limits on learning machine accuracy imposed by data quality / Advances in Neural Information Processing Systems 7 (1994). MIT Press, 1995. – pp.239-246.
4. Gu B., Hu F., Liu H. Modelling classification performance for large data sets: an empirical study / Lecture Notes in Comp. Sci., 2001. Vol.2118. – pp.317-328.
5. Bartlett P.L. For valid generalization, the size of the weights is more important than the size of the network / Advances in Neural Information Processing Systems 9 (1996). MIT Press, 1997. – pp.134-140.
6. Murata T., Yoshizawa S., Amari S. Learning curves, model selection and complexity of neural networks / Advances in Neural Information Processing Systems 5 (1992). Morgan Kaufmann, 1993. – pp.607-614.
7. Schraudolph N.N., Sejnowski T.J. Tempering backpropagation networks: not all weights are created equal / Advances in Neural Information Processing Systems 8 (1995). MIT Press, 1996. – pp.563-569.
8. Gong W., Yau H.C., Manry M.T. Non-gaussian feature analysis using a neural network / Progress in Neural Networks, 1994. Vol. 2. – pp.253-269.