

## НЕЙРОЭКСПЕРТНЫЙ АЛГОРИТМ ПОСЛЕДОВАТЕЛЬНОГО ОБНАРУЖЕНИЯ И ИСКЛЮЧЕНИЯ ПОСТОРОННИХ ВЫБРОСОВ ИЗ СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ ПРИ ПОСТРОЕНИИ НЕЙРОСЕТЕВЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Сичинава З.И.

*Пермский государственный гуманитарно-педагогический университет, Пермь, Россия (614990, г. Пермь, ул. Сибирская, 24), e-mail: zurabs@bk.ru*

---

Обобщен опыт Пермской научной школы искусственного интеллекта: указаны необходимые условия для разработки адекватной нейросетевой модели. Одним из этих условий является отсутствие в статистической информации посторонних выбросов – наблюдений, не удовлетворяющих закономерностям, которым подчиняется подавляющее большинство примеров поведения исследуемой предметной области. Причинами появления посторонних выбросов могут быть: не достаточно чисто проведенный эксперимент, ошибки измерений, сбой приборов и оборудования, искажения информации, а также влияние факторов, не учтенных при постановке задачи. Идея предлагаемого алгоритма обнаружения выбросов основана на том факте, что если выбросов в обучающем множестве сравнительно немного и если нейронная сеть имеет сравнительно небольшое количество синаптических весов, то после применения процедуры обучения нейронная сеть на примерах, являющихся выбросами, как правило, показывает более высокую погрешность обучения, чем на примерах, не являющихся выбросами. В отличие от других известных алгоритмов обнаружения и исключения выбросов предлагаемый алгоритм предполагает интерактивное участие эксперта, делающего заключение о правомерности удаления выявленного выброса, причем удаление выбросов производится на каждой итерации строго по одному.

---

Ключевые слова: моделирование, нейронные сети, алгоритм, выброс.

## NEURO-EXPERT ALGORITHM FOR EXCLUSION STATISTICAL FLUCTUATIONS FOR DESIGN OF NEURAL NETWORKS

Sichinava Z.I.

*Perm state humanitarian pedagogical university, Perm, Russia (614990, Perm, Sibirskaya St., 24), e-mail: zurabs@bk.ru*

---

Summed up the experience of the Perm school of artificial intelligence are necessary conditions for the development of adequate neural network model. One of these conditions is the lack of statistical information to unauthorized outliers - observations that do not satisfy the laws that govern the behavior of the vast majority of the examples studied the subject area. The causes of unauthorized outliers can be: not clean enough of the experiment, the measurement errors, failures of devices and equipment, noise and the impact of factors that are not included in the statement of the problem. The idea of the proposed algorithm for detecting outliers based on the fact that if the outliers in the training set are relatively few, and if the neural network has a relatively small number of synaptic weights, then after the application of the procedure of training a neural network on the examples which are outliers, generally, shows higher error learning than the examples that are not outliers. In contrast to other known algorithms for the detection and exclusion of the outliers, the proposed algorithm involves interactive participation of the expert, making the opinion on the legality of removing the detected outlier, and the removal of outliers in each iteration is strictly one.

---

Keywords: artificial intelligence, modeling, neural networks, outlier.

Несмотря на наблюдающийся в последнее десятилетие пик популярности нейронных сетей, обращает на себя внимание отсутствие общепринятых унифицированных технологий нейросетевого моделирования, что, по-видимому, объясняется несовершенством теоретической базы. В связи с этим многие специалисты в области нейросетевого моделирования

называют свою область не только наукой, но и в какой-то степени искусством [7]. Профессионалы-разработчики нейросетевых интеллектуальных систем, приступая к решению практически каждой новой задачи, применяют весьма внушительный арсенал различных, только им известных хитростей и ноу-хау, зачастую надеясь на интуицию, а иногда и просто – на везение. В России, как и за рубежом, существуют научные школы, которые развивают и применяют на практике свои собственные варианты нейросетевых парадигм и технологий, отмечают их преимущества, часто противопоставляя другим подходам, применяемым в других школах.

Пермской научной школой искусственного интеллекта ([www.PermAi.ru](http://www.PermAi.ru)) [5] на протяжении последних пятнадцати лет предпринимались попытки изобретения новых принципов построения и функционирования нейронных сетей. Осваивались пропагандируемые другими научными школами нейросетевые парадигмы неклассического типа. Однако до сих пор не нашлось практически значимых задач, для решения которых неклассические нейросети оказались бы более эффективны, чем перцептроны с сигмоидными активационными функциями, изобретенные в середине прошлого столетия.

У нас сложилось твердое убеждение, что, вне зависимости от применяемых нейросетевых парадигм, попытки разработки адекватной нейросетевой математической модели могут быть успешными только в том случае, если одновременно выполняются следующие условия.

1. При постановке задачи учтено достаточное количество признаков (входных параметров модели), оказывающих влияние на результат моделирования.
2. Сформировано достаточно репрезентативное множество примеров поведения предметной области (наблюдений).
3. Во множестве наблюдений не содержатся конфликтные (противоречащие друг другу) примеры.
4. Во множестве наблюдений не содержится посторонних выбросов, либо они незначительны, либо их количество мало по сравнению с общим количеством примеров.

Настоящая работа посвящена проблеме обнаружения и исключения посторонних выбросов из статистической информации, предназначенной для обучения нейронной сети.

Прежде всего, отметим, что под посторонними выбросами понимаются примеры поведения предметной области, по каким-либо причинам выпадающие из общих закономерностей этой предметной области. Другими словами – это наблюдения, не удовлетворяющие закономерностям, которым подчиняется подавляющее большинство примеров поведения ис-

следуемой предметной области. Причинами появления посторонних выбросов могут быть: не достаточно чисто проведенный эксперимент, ошибки измерений, сбои приборов и оборудования, искажения информации при формировании множества обучающих примеров. Причиной появления посторонних выбросов может быть также влияние факторов, не учтенных при постановке задачи и не включенных в качестве входных параметров модели.

Идея предлагаемого алгоритма обнаружения выбросов основана на том факте, что если выбросов в обучающем множестве сравнительно немного и если нейронная сеть имеет сравнительно небольшое число степеней свободы (небольшое количество синаптических весов), то после применения процедуры обучения нейронная сеть на примерах, являющихся выбросами, как правило, показывает более высокую погрешность обучения, чем на примерах, не являющихся выбросами. Используя этот эмпирический факт, алгоритм обнаружения и исключения посторонних выбросов предлагается в виде блок-схемы, представленной на рис. 1.

Согласно этой блок-схеме алгоритм включает следующие пункты.

1. Разбить множество примеров поведения предметной области на обучающее ( $L$ ), тестирующее ( $T$ ) и подтверждающее ( $P$ ) в отношении 80 : 15 : 5 (%).
2. Для множества примеров  $L \cup T$ , пользуясь формулами следствия теоремы Арнольда – Колмогорова – Хехт-Нильсена [7], рассчитать минимальное и максимальное количество скрытых нейронов двухслойного персептрона:

$$N_{w\min} = \frac{N_y Q}{1 + \log_2(Q)},$$

$$N_{w\max} = N_y \left( \frac{Q}{N_x} + 1 \right) (N_x + N_y + 1) + N_y,$$

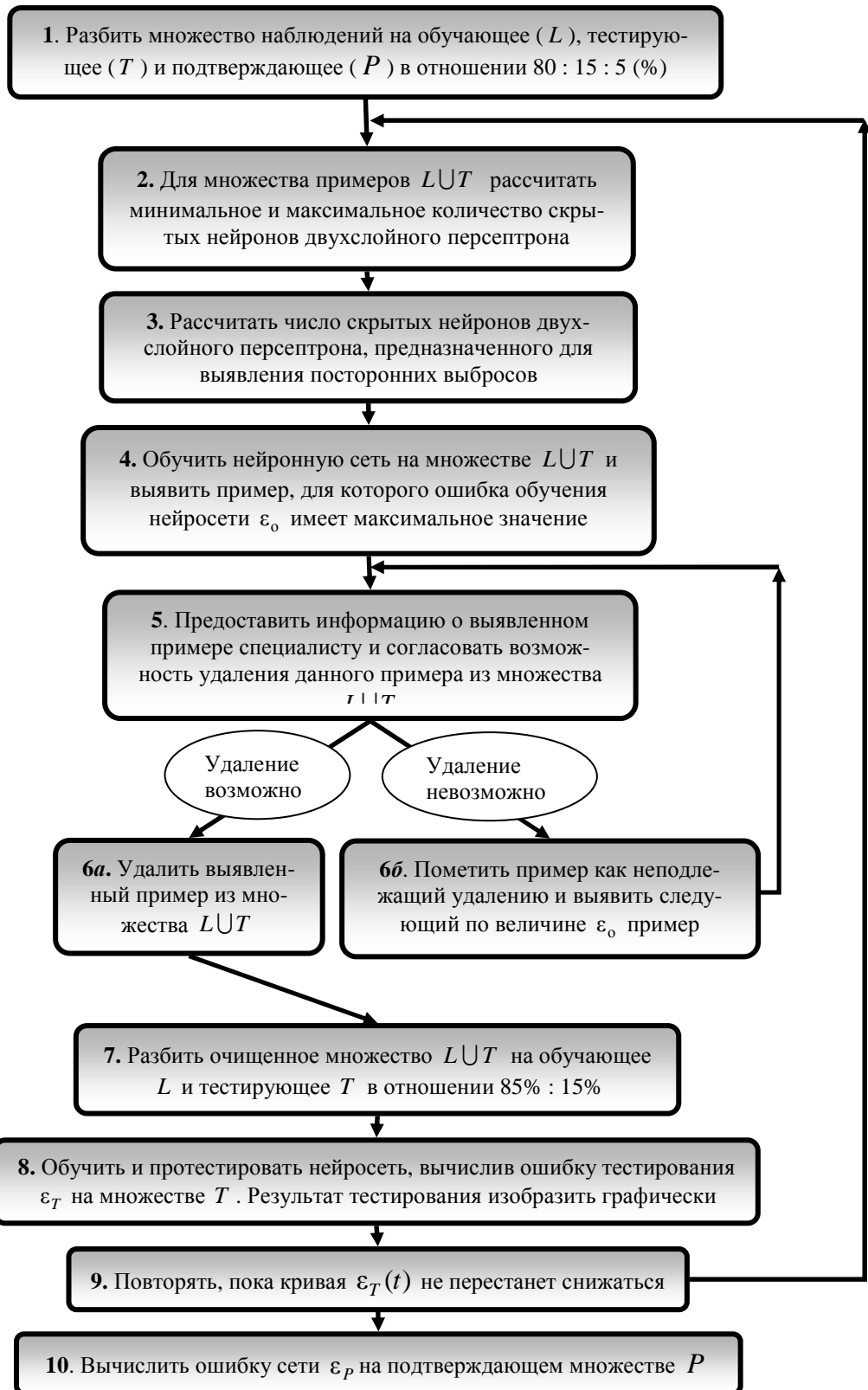


Рис. 1. Блок-схема нейроэкспертного алгоритма последовательного обнаружения и исключения посторонних выбросов

$$N_{\min} = \frac{N_{w\min}}{N_x + N_y},$$

$$N_{\max} = \frac{N_{w\max}}{N_x + N_y}.$$

Здесь:  $N_{\min}$  и  $N_{\max}$  – минимальное и максимальное количество скрытых нейронов;  $N_{w\min}$  и  $N_{w\max}$  – минимальное и максимальное количество сил синаптических связей;  $N_x$  – количество нейронов входного слоя;  $N_y$  – количество нейронов выходного слоя;  $Q$  – число элементов множества  $L \cup T$ .

3. Рассчитать число скрытых нейронов двухслойного персептрона, предназначенного для выявления посторонних выбросов с помощью предлагаемой в настоящей статье эмпирической формулы:

$$N = N_{\min} + \xi(N_{\max} - N_{\min}),$$

в которой  $\xi$  – эмпирический коэффициент, значение которого, в зависимости от решаемой задачи, мы рекомендуем принимать от 0 до 0,2 (рис. 2).

4. Обучить нейронную сеть на множестве  $L \cup T$  и выявить пример, для которого ошибка обучения нейросети  $\varepsilon_0$  имеет максимальное значение.

5. Предоставить информацию о выявленном примере специалисту в исследуемой предметной области и согласовать с ним вопрос о возможности удаления данного примера из множества  $L \cup T$ .

6. В зависимости от решения эксперта: либо удалить выявленный пример из множества  $L \cup T$ , либо пометить его как неподлежащий удалению, выявить следующий по величине  $\varepsilon_0$  пример и перейти на п. 5.

7. Разбить очищенное множество  $L \cup T$  на обучающее  $L$  и тестирующее  $T$  в отношении 85% : 15%.

8. Обучить и протестировать нейросеть, вычислив ошибку тестирования  $\varepsilon_T$  на множестве  $T$ . Результат тестирования изобразить графически, как на рис. 2.

9. Повторять пп. 2 – 8, пока кривая на рис. 2 не перестанет снижаться.

10. Вычислить ошибку сети  $\varepsilon_P$  на подтверждающем множестве  $P$ .

В заключение отметим, что применение предлагаемого алгоритма обнаружения и исключения посторонних выбросов в ряде случаев, при работе с сильно зашумленными статистическими данными, позволило уменьшить погрешности нейросетевых моделей в десятки раз. Особенно эффективным применение этого алгоритма оказалось при создании нейросетевого детектора лжи [2; 11; 12; 14], а также в задачах, относящихся к сфере экологии [8; 10] и *медицины* [9]. Менее заметное снижение погрешности – от одного до 10 процентов, зафиксировано в задачах, относящихся к сферам: экономики и бизнеса [13], политологии и социологии [15], педагогики [3], исторических наук [1], туризма [6].

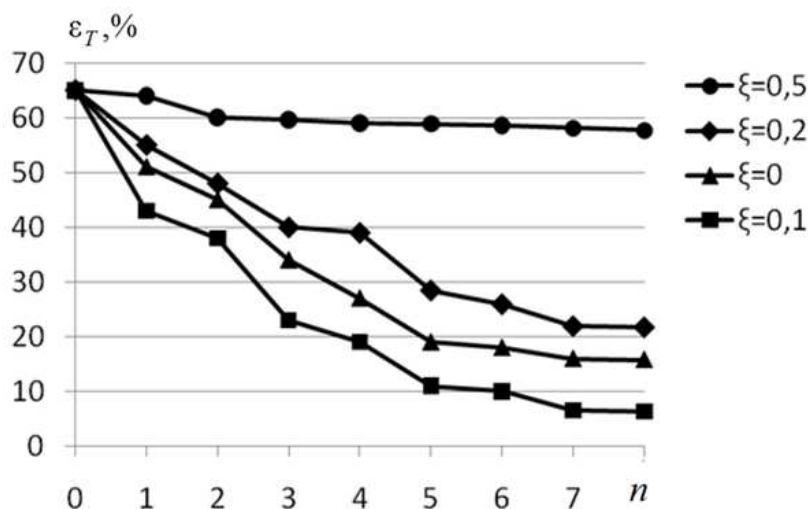


Рис. 2. Примерные зависимости погрешности тестирования  $\varepsilon_T$  от эмпирического коэффициента  $\xi$  и от числа итераций  $n$  по пп. 2-8 предлагаемого алгоритма

Отметим также, что, в отличие от других известных алгоритмов обнаружения и исключения выбросов, например [4], предлагаемый алгоритм предполагает интерактивное участие эксперта, анализирующего правомерность удаления выявленного в качестве выброса наблюдения. Кроме того, удаление выбросов производится не за один прием, как это регламентируется алгоритмом [4], а в результате итерационного процесса строго по одному на каждой итерации.

Применение этих ноу-хау делает предлагаемый алгоритм более эффективным, позволяющим строить нейросетевые математические модели, как правило, на 1–15% более точными, чем известный алгоритм [4], преследующий аналогичные цели.

Отметим, что в нашей практике встречались случаи, как например [2; 8; 10-12; 14], когда попытки построения нейросетевых, а также регрессионных моделей, без предварительной очистки статистической информации от выбросов, осуществляемой с помощью

предлагаемого алгоритма, вообще не приводили к положительным результатам, т.е. погрешности математических моделей не удавалось снизить до сколько-нибудь приемлемых для практического применения значений.

Таким образом, применение предлагаемого в настоящей работе алгоритма обнаружения и исключения посторонних выбросов из статистической информации позволяет не только повышать точность нейросетевых и регрессионных моделей, но и расширяет круг задач, для которых применение данного вида моделирования вообще возможно.

### Список литературы

1. Корниенко С.И., Айдаров Ю.Р., Гагарина Д.А., Черепанов Ф.М., Ясницкий Л.Н. Программный комплекс для распознавания рукописных и старопечатных текстов // Информационные ресурсы России. – 2011. – № 1. – С. 35-37.
2. Петров А.М., Сичинава З.И., Ясницкий Л.Н. Анкетный способ построения нейросетевого детектора лжи // Вестник Пермского университета. Серия: Математика. Механика. Информатика. – 2010. – № 1. – С. 84-87.
3. Семакин И.Г., Ясницкий Л.Н. Искусственный интеллект и школьный курс информатики // Информатика и образование. – 2010. – № 9. – С. 48-54.
4. Черепанов Ф.М., Ясницкий Л.Н. Нейросетевой фильтр для исключения выбросов в статистической информации // Вестник Пермского университета. Серия: Математика. Механика. Информатика. – 2008. – № 4. – С. 151-155.
5. Ясницкий Л.Н., Богданов К.В., Черепанов Ф.М. Технология нейросетевого моделирования и обзор работ Пермской научной школы искусственного интеллекта // Фундаментальные исследования. – 2013. – № 1-3. – С. 736-740.
6. Ясницкий Л.Н., Бржевская А.С., Черепанов Ф.М. О возможностях применения методов искусственного интеллекта в сфере туризма // Сервис plus. – 2010 – № 4. – С. 111-115.
7. Ясницкий Л.Н. Введение в искусственный интеллект. - М. : Издательский центр «Академия», 2005. – 176 с.
8. Ясницкий Л.Н., Гусев А.Л., Шур П.З. О возможностях применения метода нейросетевого математического моделирования для выявления целесообразных действий Роспотребнадзора // Вестник Пермского университета. Серия: Биология. – 2012. – № 3. – С. 49-53.

9. Ясницкий Л.Н., Думлер А.А., Полещук А.Н., Богданов К.В., Черепанов Ф.М. Нейросетевая система экспресс-диагностики сердечно-сосудистых заболеваний // Пермский медицинский журнал. – 2011. – № 4. – С. 77-86.
10. Ясницкий Л.Н., Зайцева Н.В., Гусев А.Л., Шур П.З. Нейросетевая модель региона для выбора управляющих воздействий в области обеспечения гигиенической безопасности // Информатика и системы управления. – 2011. - № 3. – С. 51-59.
11. Ясницкий Л.Н., Петров А.М., Сичинава З.И. Сравнительный анализ алгоритмов нейросетевого детектирования лжи // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2010. – № 1. – С. 64-72.
12. Ясницкий Л.Н., Петров А.М., Сичинава З.И. Технологии построения детектора лжи на основе аппарата искусственных нейронных сетей // Информационные технологии. – 2010. – № 11. – С. 66-70.
13. Ясницкий Л.Н., Порошина А.М., Тавафиев А.Ф. Нейросетевые технологии как инструмент для прогнозирования успешности предпринимательской деятельности // Российское предпринимательство. – 2010. – № 4-2. – С. 8-13.
14. Ясницкий Л.Н., Сичинава З.И. Нейросетевые алгоритмы анализа поведения респондентов // Нейрокомпьютеры: разработка и применение. – 2011. – № 10. – С. 59-64.
15. Ясницкий Л.Н., Черепанов Ф.М. О возможностях применения нейросетевых технологий в политологии // Нейрокомпьютеры: разработка и применение. – 2010. – № 8. – Вып. 4. – С. 47-53.

**Рецензенты:**

Шварц Константин Григорьевич, доктор физико-математических наук, доцент, профессор кафедры прикладной математики и информатики, Пермский государственный национальный исследовательский университет, г. Пермь.

Ясницкий Леонид Нахимович, доктор технических наук, профессор, заведующий кафедрой прикладной информатики, Пермский государственный гуманитарно-педагогический университет, г. Пермь.