



веб-сайта. // Научно-методический журнал «Информатизация образования и науки» № 3(27)/2015. ФГАУ ГНИИ ИТТ «Информика». С. 65-73.

2. Федеральное агентство по техническому регулированию и метрологии ГОСТ-28806-90: Качество программных средств. Термины и определения // Информационный портал по стандартизации. – Стандартиформ, 2017 – Режим доступа: <http://standard.gost.ru/> (дата обращения: 10.01.2017).

3. K. Sridevi, R. Umarani, V.Selvi. An Analysis of Web Document Clustering Algorithms. International Journal of Science and Technology. Volume 1 No.6, December 2011, pp.: 275 – 282.

4. Ngomo, A.C.N., Lyko, K., Christen, V.: Coala-correlation-aware active learning of link specifications. In: The Semantic Web: Semantics and Big Data, pp. 442–456. Springer (2013).

5. Alexander A. Sytnik, Sergey V. Papshev. Semantic Segmentation of Hypertext on the Basis of Automata Model. International Journal of Computing Anticipatory Systems, v. 28, 2014, D.M. Dubois (Ed.), CHAOS, Liège, Belgium, ISSN 1373-5411, ISBN 2-930396-17-2. P.109-115.

6. Сытник А.А., Шульга Т.Э. Математические модели адаптивных дискретных систем. Монография // Саратов: Сарат. гос. техн. ун-т, 2015. 272с. ISBN 978-5-433-2947-2.

7. Сытник А.А. Перечислимость при восстановлении поведения автоматов // Доклады РАН. 1993. Т.238. N1. С.25-26

А.А. Санталов, Д.А. Жуков

ДИАГНОСТИКА ТЕХНИЧЕСКОГО СОСТОЯНИЯ СИСТЕМЫ С ПРИМЕНЕНИЕМ НЕЙРОСЕТЕВЫХ МЕТОДОВ

(Ульяновский государственный технический университет)

Исследовалась эффективность применения нейросетей при диагностике технического состояния системы на примере станции водоочистки. В качестве исходных данных использовались семь показателей функционирования системы (физико-химические параметры водоисточника и дозы реагентов для очистки) и состояние системы: система исправна, если показатели качества очищенной (питьевой) воды лежат в допустимых пределах. Задача состоит в разработке нейронной сети, предсказывающей состояние системы по семи заданным показателям функционирования, подбором параметров сети и оценкой эффективности ее прогнозов с использованием метода кросс-валидации.

Сформулированная задача является задачей бинарной классификации, и для ее решения удобно использовать нейронную сеть, настраиваемую функцией `patternnet` из пакета MATLAB [1]. В качестве алгоритма обучения нейронной сети был использован метод сопряженных градиентов, отличающийся о высокой сходимостью и малыми затратами памяти [2,3].



Выбор количества слоев для нейронной сети можно обосновать эвристическим правилом, гласящим, что количество связей в ней не должно превышать 10% от размера выборки. Если из всех вариантов деления исходной выборки выбрать вариант с самой малой обучающей выборкой (45% от всех данных), и учитывать только ее размер, то с помощью несложных подсчетов можно получить, что количество связей в нейронной сети не должно превышать 15. Такому количеству связей удовлетворяет однослойная нейронная сеть с 10 нейронами в слое.

Для функции активации нейрона в сети с данным алгоритмом обучения важны следующие свойства: непрерывность, дифференцируемость на всей оси абсцисс и монотонным возрастанием. Такими свойствами обладает гиперболическая тангенциальная функция, помимо этих свойств она также — принимает значения от -1 до 1, усиливает слабые сигналы и насыщается от сильных, а также имеет простую для нахождения производную [2].

Поскольку при подготовке входных и выходных данных для функции `patternnet` пакетом MATLAB используется функция `mapminmax`, которая нормализует входящие данные и приводит их к значениям от -1 до 1, в последнем слое нейронной сети использовалась тангенциальная функция [4,5].

Так как техническое состояние системы характеризуется бинарной дискретной величиной, принимающей значения - 1 (исправное) или 0 (неисправное), а получаемое на выходе из нейросети значение непрерывное, для преобразования этого значения была использована функция Хевисайда. Эта функция не использована как функция активации в нейронной сети и лишь обрабатывает выведенное нейронной сетью значение.

Оценка качества обучения нейронной сети производилась с помощью метода кросс-валидации. Исходная выборка делилась на 10 подвыборок. В рамках исследования исходная выборка разбивалась на обучающую, валидационную и тестовую подвыборку с различными соотношениями исходных данных в них. При этом тестовая выборка в каждом варианте деления исходных данных составляла 10% от всего объема данных, изменялось лишь соотношение обучающей и валидационной выборок. Из-за отсутствия общего правила насчет деления данных для обучения нейронной сети, были опробованы несколько вариантов деления исходных данных. Для изучения влияния соотношения объемов выборок на качество обучения нейросети были выбраны следующие варианты деления на обучающую, валидационную и тестовую выборки: 90:0:10, 70:20:10, 60:30:10, 45:45:10 и 20:70:10.

Настроенная описанным образом нейронная сеть выдает показатели эффективности (в качестве критерия используется средняя ошибка кросс-валидации), отраженные в таблице 1. Как видно из этой таблицы, при разных способах деления выборки средняя ошибка прогнозов нейронной сети получается примерно одинаковой.

В ходе исследования исходная выборка разбивалась несколькими способами, каждый из которых показал свое значение средней ошибки, причем каждый раз он лежал в пределах 21.81-23.30%. Процент ошибок достаточно высок,



что обуславливается небольшим размером исходной выборки. Также вполне ожидаемо, что чем меньше данных использовалось для обучения, тем хуже справлялась нейронная сеть с задачей классификации.

Если сравнивать среднюю ошибку при использовании нейросетевого метода классификации со средней ошибкой при использовании метода опорных векторов, полученной на том же примере в [6] (средняя ошибка равна 0.238), то нейронная сеть лучше справляется с задачей. Однако она проигрывает классификации с агрегированием по нескольким методам бинарной классификации (в случае агрегирования средняя ошибка составляет 0.196). В исследовании [7-8] результаты кросс-валидации имеют близкие значения (20.13% для метода опорных векторов и 23.96% для дискриминантного анализа).

Таблица 1. Средние ошибки при разных соотношениях объемов обучающей, валидационной и тестовой выборок

№ блока, используемого для тестирования	Соотношения обучающей, валидационной и тестовой выборок, %				
	90:0:10	70:20:10	60:30:10	45:45:10	20:70:10
1	25.71%	25.71%	25.71%	25.71%	20.00%
2	28.57%	22.86%	22.86%	17.14%	22.86%
3	26.47%	11.76%	14.71%	20.59%	23.53%
4	31.43%	28.57%	28.57%	28.57%	22.86%
5	20.00%	17.14%	17.14%	17.14%	17.14%
6	14.29%	22.86%	17.14%	14.29%	14.29%
7	17.14%	25.71%	25.71%	25.71%	28.57%
8	23.53%	20.59%	26.47%	23.53%	29.41%
9	14.29%	17.14%	20.00%	20.00%	22.86%
10	17.14%	25.71%	25.71%	31.43%	31.43%
Среднее значение ошибки:	21.86%	21.81%	22.40%	22.41%	23.30%

Проведенное исследование показало, что качество классификации состояния системы не слишком существенно зависит от способа разбиения исходных данных, однако демонстрирует, что обучающая выборка не должна быть слишком малой, что может быть недостаточно для обучения нейросети.

Литература

1. Медведев В. С., Потемкин В. Г. Нейронные сети. MATLAB 6 / Под общ. ред. В. Г. Потемкина. – М.: ДИАЛОГ-МИФИ, 2002. – 496 с.



2. Круглов В. В., Борисов В. В. Искусственные нейронные сети. Теория и практика. – М.: Горячая линия-Телеком, 2002. – 382 с.
3. Jose Orozco, Carlos A. Reyes Garcia Detecting Pathologies from Infant Cry Applying Scaled Conjugate Gradient Neural Networks // European Symposium on Artificial Neural Networks, Bruges (Belgium), 23-25 April 2003, d-side publ. – Pp. 349-354.
4. Mark Hudson Beale, Martin T. Hagan, Howard B. Demuth Neural Network Toolbox. User's Guide. – 2017.
5. Клячкин В.Н., Кувайскова Ю.Е., Алексеева В.А. Статистические методы анализа данных. – М. : Финансы и статистика, 2016. – 240 с.
6. Клячкин В. Н., Кувайскова Ю. Е., Жуков Д. А. Использование агрегированных классификаторов при технической диагностике на базе машинного обучения // Сборник трудов III международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ-2017). – Самара: Новая техника, 2017. – С. 1763-1766.
7. Жуков Д.А., Клячкин В.Н., Кувайскова Ю.Е. Сравнительный анализ методов машинного обучения при прогнозировании состояния технического объекта // Радиоэлектронная техника. – 2017. – №1(10). – С. 189-192.
8. Жуков Д. А. Повышение эффективности машинного обучения при решении задач технической диагностики // IN MEMORIAM: Султан Галимзянович Валеев : сборник памяти С.Г. Валеева. – Ульяновск : УлГТУ, 2016. – С. 139-143.

А.А. Тарасов, И.В. Лёзина

АППРОКСИМАЦИЯ ПЛОТНОСТЕЙ ВЕРОЯТНОСТИ НЕЧЁТКИМ ПЕРСЕПТРОНОМ

(Самарский университет)

При обработке статистического материала часто приходится решать вопрос о том, как подобрать для данного статистического ряда теоретическую кривую распределения, выражающую лишь существенные черты статистического материала [1]. Эта задача относится к классу задач аппроксимации.

Существует множество способов аппроксимации, одним из которых является использование нейронных сетей в качестве универсального аппроксиматора. Это следует из теоремы Хета-Нильсена о представимости любой многомерной функции нескольких переменных с помощью нейронной сети фиксированной размерности [3]. Таким образом, любую сложную функцию можно заменить более простой нейросетевой аппроксимацией, причём требуется только нелинейность функции активации нейрона.

В качестве обучающей последовательности может использоваться файл со сгенерированной случайной последовательностью. На основании этого файла будет построена гистограмма. Выделив на этой гистограмме M крайних ле-