

ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ПРАКТИЧЕСКИХ ЗАДАЧ В ПАКЕТЕ «NEUROSOLUTIONS»

Цуриков А. Н.

Научный руководитель д-р техн. наук Гуда А. Н.
Ростовский государственный университет путей сообщения

Интерес вызывают новые направления имитации интеллектуальной деятельности человека, в частности искусственные нейронные сети (ИНС). Нередко использование ИНС для решения практических задач. Облегчить работу призваны программы для создания ИНС – нейропакеты – оболочки, эмулирующие среду нейрокомпьютера. Из нейропакетов рассмотрим «NeuroSolutions». Он содержит мастер нейросетевых архитектур («Neural Wizard»), с помощью которого задается архитектура, подбирается обучающая выборка, критерии обучения. По утверждениям разработчиков, мастер позволяет легко создать и обучить работоспособную ИНС.

Проверим это на примере решения практических задач, связанных с наблюдениями погоды. Приведем в таблице 1 фрагмент сводки погоды в Лондоне за март 1982 года.

Таблица 1. Фрагмент сводки погоды в Лондоне за март 1982 г.

День месяца	Минимальная температура, °С	Максимальная температура, °С	Уровень осадков, мм	Солнечное время дня, ч
1	9,4	11,0	7,5	3,2
2	4,2	12,5	4,1	6,7
3	7,6	11,2	7,7	1,1

Известно, что проблемой, с которой ИНС лучше всего умеют справляться является задача классификации. Сформулируем задачу. По имеющимся данным определить является ли рассматриваемый день дождливым.

Чтобы сформировать достаточное количество обучающих примеров дополним наблюдения данными о погоде за март последующих семи лет, получив $31 * 8 = 248$ записей таблицы. Будем считать, что погода в течение одного месяца не успевает претерпеть значительных сезонных изменений. Перемешаем записи с помощью программы на языке «Pascal», использующей генератор псевдослучайных чисел, и перепишем их в текстовый файл «TheWeather01.csv».

Щелчком на соответствующем ярлыке запускаем мастер «Neural Wizard», являющийся частью нейропакета «NeuroSolutions». На первом этапе программа предлагает выбрать тип решаемой задачи из четырех возможных: классификация, аппроксимация функции, прогнозирование, кластеризация. Выбираем классификацию. Далее указываем путь к файлу «TheWeather01.csv», отмечаем в качестве входов ИНС столбцы Temp_Min, Temp_Max, Time, в качестве выхода столбец Rain.

Определяем объем записей, выделяемых для тестового и экзаменационного наборов. Тестовый набор используется для предотвращения «переучивания» ИНС, экзаменационный – для оценки качества работы сети после обучения. Нормальным считается объем в 20 %. В итоге имеем следующее распределение данных: 150 записей (60 %) участвуют в процессе обучения (Training); 49 записей (20 %), тестовый набор данных (Cross Validation); 49 записей (20 %), экзаменационный набор данных (Testing).

Нажимаем «Finish». «NeuroSolutions» автоматически выстраивает ИНС, которую считает наиболее эффективной для решения описанной задачи. Эта сеть еще не обучена. Запускаем процесс обучения ИНС. Получаем результат, представленный на

кривой обучения (рис. 1).

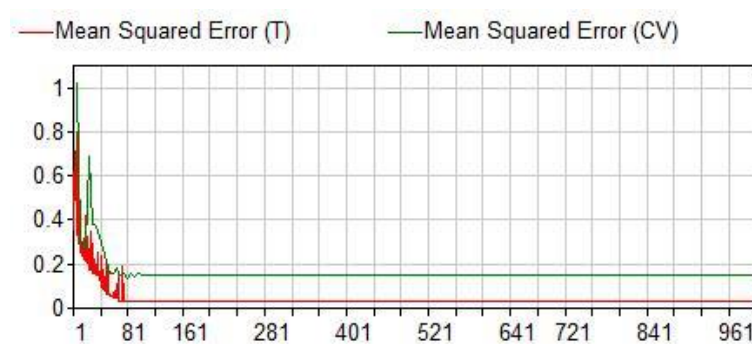


Рис. 1. Кривая обучения ИНС на первом этапе

Нижняя линия соответствует ошибке на обучающих данных, Верхняя – ошибке на тестовом наборе данных. При обучении ИНС следует стремиться, чтобы обе эти кривые располагались как можно ближе друг к другу и к оси абсцисс.

Из графика видно, что, начиная с 81-й эпохи обучения, эти линии перестают сходиться и идут параллельно, что может быть вызвано множеством причин, в частности попаданием в локальный минимум. «NeuroSolutions» имеет несколько встроенных средств для преодоления подобных проблем: возврат в исходное состояние (Reset); рандомизация (Randomize); перетряска (Jog). Возврат в исходное состояние – крайне нежелательный вариант, так как все результаты будут утрачены. Для выхода из локального минимума рекомендуется попробовать применить перетряску, а если это не помогло, то и рандомизацию. После применения указанных инструментов кривая обучения принимает вид, показанный на рисунке 2.

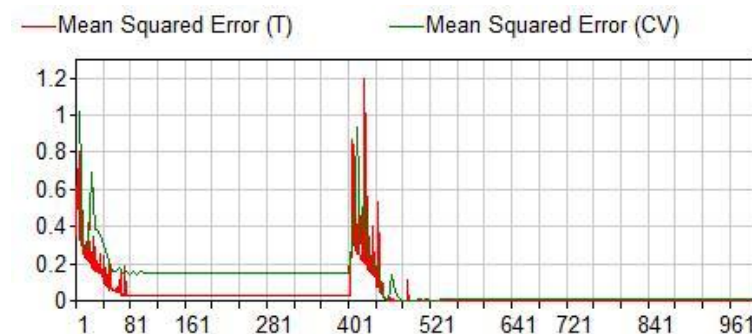


Рис. 2. Кривая обучения ИНС после применения средств вывода из локального минимума

Из графика видно, что, после применения этих средств на 401-й эпохе обучения, обе кривые идут вдоль оси абсцисс, ошибки на обучающих данных и на тестовом наборе данных равняются нулю.

Однако, теперь существует опасность переобучения ИНС, когда ИНС хорошо работает на примерах из обучающей выборки, но достаточно плохо работает на примерах из экзаменационной выборки. Проверим это, подав на вход ИНС 49 записей экзаменационного набора данных, которые были зарезервированы для этих целей и не участвовали в процессе обучения. Результат приведен в таблице 2.

Таблица 2. Результат проверки ИНС экзаменационным набором данных

№	Ожидаемые значения выхода ИНС		Реальные значения выхода ИНС		Результат
	No	Yes	No	Yes	
1	0.00000000	1.00000000	0.00374347	0.99625652	Верно
* * * * *					
5	1.00000000	0.00000000	0.99713714	0.00286285	Верно
6	1.00000000	0.00000000	0.02226600	0.97773399	Ошибка
7	1.00000000	0.00000000	1.00089697	-0.00089697	Верно
* * * * *					
36	1.00000000	0.00000000	1.00089695	-0.00089695	Верно
37	0.00000000	1.00000000	0.91788785	0.08211214	Ошибка
38	0.00000000	1.00000000	-0.00172970	1.00172970	Верно
* * * * *					
49	0.00000000	1.00000000	0.11754606	0.88245393	Верно

Из таблицы 2 видно, что ИНС совершила 2 ошибки. «NeuroSolutions» позволяет автоматически вычислить процент ошибочных результатов, что показано на рисунке 3. Из рисунка 3 следует, что процент ошибочных результатов приблизительно равен 4 %, что можно считать приемлемым результатом. Сохраним созданную и обученную ИНС.

	No	Yes
No	95.454545454545	4.545454545455
Yes	4.166666666667	95.833333333333

Рис. 3. Процент ошибочных результатов ИНС на экзаменационном наборе данных

Теперь усложним задачу. Введем лингвистическую переменную «Precipitation level» («Уровень осадков»), которая может принимать три значения: «None» («Нет»), «Medium» («Средний»), «High» («Высокий»), как это показано на рисунке 4.

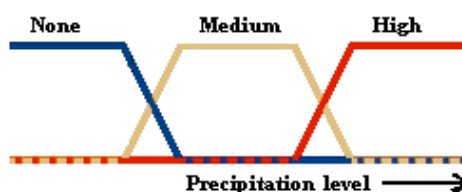


Рис. 4. Лингвистическая переменная «Уровень осадков»

ИНС должна по имеющимся входным данным определить уровень осадков в этот день. Процесс построения и обучения новой ИНС в «NeuroSolutions» аналогичен, описанному. В итоге получаем кривую обучения ИНС, приведенную на рисунке 5.

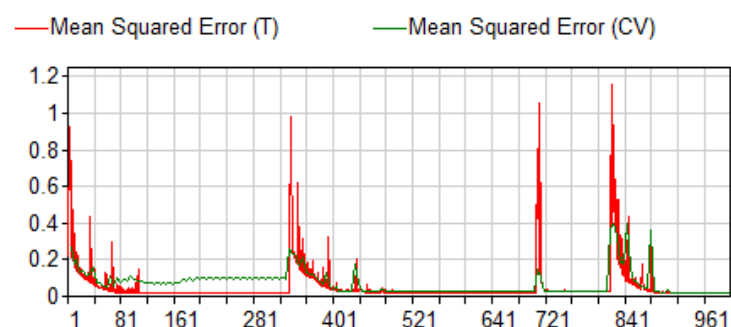


Рис. 5. Кривая обучения ИНС для задачи с лингвистической переменной

Из графика видно, что ошибка не достигает нуля, даже после 961 эпохи обучения. Подадим на вход ИНС записи экзаменационного набора данных. Получим результаты, представленные на рисунке 6.

	Medium	None	High
Medium	100.000000000000	0.000000000000	0.000000000000
None	4.347826086957	95.652173913043	0.000000000000
High	0.000000000000	7.692307692308	92.307692307692

Рис. 6. Процент ошибочных результатов ИНС на экзаменационном наборе данных

Из рисунка 6 видно, что созданная ИНС безошибочно идентифицировала дни со средним уровнем осадков и испытала некоторые трудности при распознавании сухих дней и дней с высоким уровнем осадков (ошибки составили 4,35 % и 7,69 % соответственно).

В итоге мы получили удовлетворительно функционирующую систему, основанную на ИНС, которую с некоторой долей оптимизма можно назвать «экспертной» в достаточно сложной области знаний о погоде. Описанные в статье приемы можно использовать при решении аналогичных задач по распознаванию образов, классификации.

В итоге можно сформулировать следующие выводы:

1. Сформулировано несколько практических задач, связанных с наблюдениями погоды. Показана возможность их решения при помощи технологии ИНС;
2. Получен опыт работы, создания ИНС и их обучения в современном нейропакете «Neurosolutions»;
3. Решено несколько практических задач, связанных с наблюдениями погоды. Результаты их решения можно считать вполне успешными.