

Лабораторная работа №5

«Нейротехнологии. Решение задач классификации, кластеризации и распознавания образов с помощью нейропакетов»

1. Цель работы

Изучить базовые понятия нейротехнологий, ознакомиться с возможностями современных нейропакетов и получить навыки работы с ними.

2. Подготовка к работе

Изучить основные определения и понятия в области искусственных нейронных сетей (ИНС), а также проанализировать возможности современных нейропакетов (НП), предназначенных для решения задач классификации, кластеризации и распознавания образов [1].

Установить НП (например, НП *NeuroSolutions* или НП *Neuro Office*). Для установки НП *NeuroSolutions* необходимо перейти по следующей ссылке [NeuroSolutions](#), загрузить демонстрационную версию с сайта производителя и выполнить инсталляцию программы. Для установки НП *Neuro Office* необходимо распаковать архив *NeuroOffice.zip* в папке «8 - Дополнительные материалы» и запустить для инсталляции программы файл *Neuro_Office_rus_instal.exe*. Системные требования *Neuro Office*: операционная система *Microsoft Windows XP*.

3. Лабораторное задание

1. Решить задачу классификации (кластеризации или распознавания образов) с использованием НП.
2. Выбрать НП для решения поставленной задачи;
3. Разработать собственную ИНС;
4. Создать топологию ИНС;
5. Обучить ИНС;
6. Выполнить эмуляцию сети (проверить качество решения задачи на примерах, которые не вошли в обучающую выборку);
7. В отчет по лабораторной работе включить информацию по выбранному НП, описание разработанной ИНС (топология, функции активации нейронов, алгоритм обучения сети и т.д.) и результаты тестирования обученной ИНС.
8. Подготовить отчет для защиты лабораторной работы №5.

4. Требования

- Ограничения для задачи классификации:
 - Количество признаков должно быть не менее 7, на выходе 3 класса (центра группирования).

- ИНС должна состоять, как минимум, из 4 слоев, и в каждом слое не менее чем 3 нейрона;
- Точность обучения 5 % (не более 1000 эпох).

5. Методические указания

Лабораторная работа может быть выполнена с использованием различных НП (например, *NeuroSolutions*, *Neuro Office* и др.).

Перед использованием НП необходимо подготовить обучающую и тестовую выборки данных для последующего обучения и тестирования нейронной сети. Формат для представления указанных данных выбирается в зависимости от используемого НП.

При выполнении лабораторной работы средствами универсального НП *NeuroSolutions* для построения нейронной сети, имеющей одну из стандартных топологий, но с возможностью настройки параметров, выбирается компонент *NeuralBulder* (рис. 1).

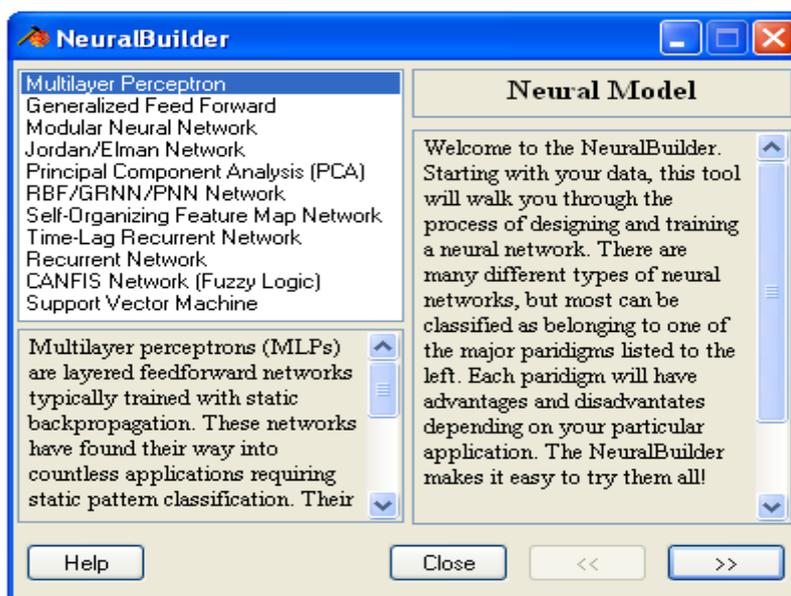


Рис. 1. Окно *NeuralBulder* для выбора структуры ИНС

Далее необходимо указать путь к данным, используемым в последующем обучении и тестировании сети (рис. 2). Входные данные необходимо пометить идентификатором *Input*, предполагаемые выходные данные – *Desired*. Если имеется символьная информация, она отмечается идентификатором *Symbol*. Если к входным данным применяется оптимизация, то ее тип выбирается из списка *Input Optimization*.

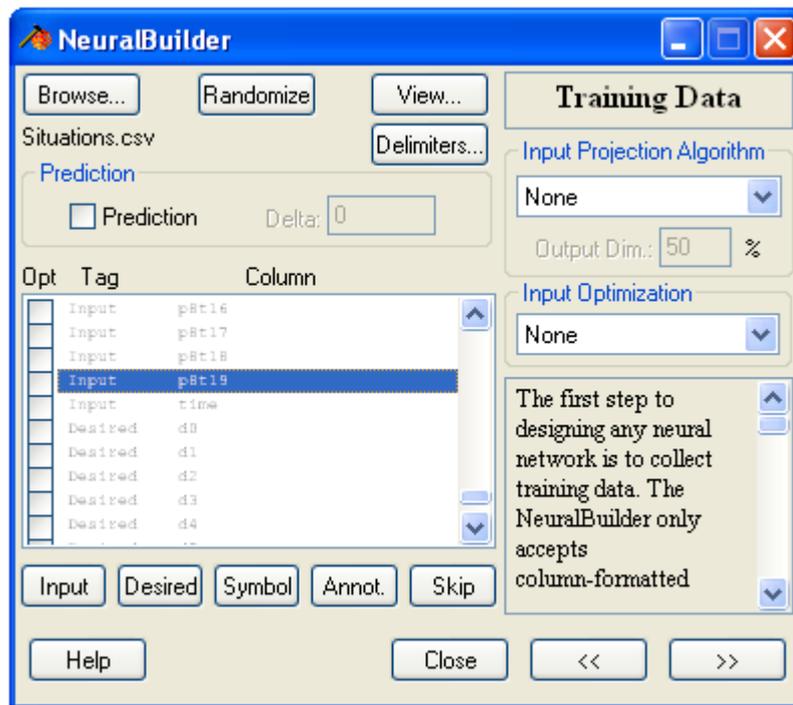


Рис. 2. Окно *NeuralBuilder* для настройки обучающей выборки данных

Далее необходимо выбрать тестовый набор данных, являющийся необязательным. Эти данные могут быть извлечены из обучающей выборки. Тестовый набор данных используется для тестирования производительности сети.

Далее переходим к конфигурированию ИНС (например, многослойного персептрона (рис. 3), использующего при обучении метод обратного распространения ошибки).

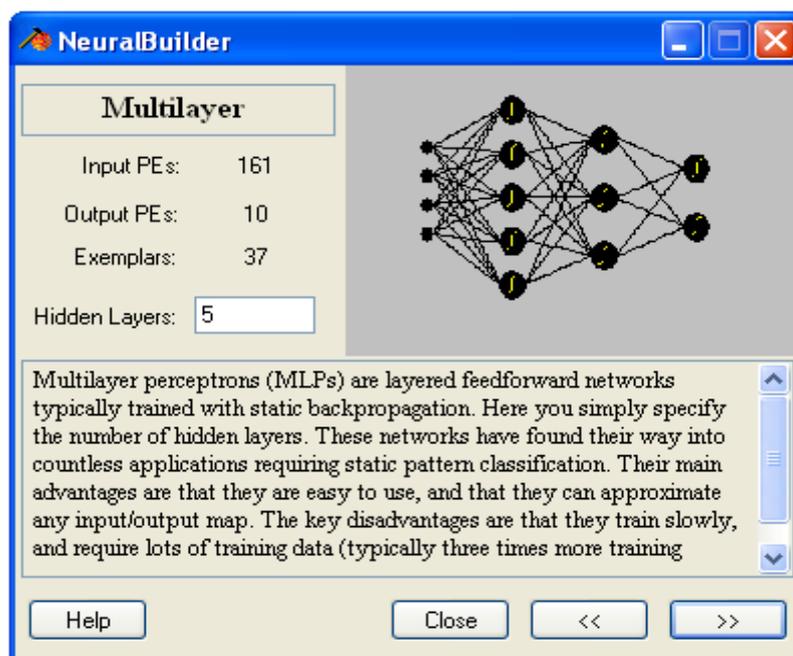


Рис. 3. Окно *NeuralBuilder* для настройки структуры ИНС

Далее задаются параметры каждого слоя (рис. 4). В частности, определяется число нейронов в слое и активационная функция (например, гиперболический тангенс).

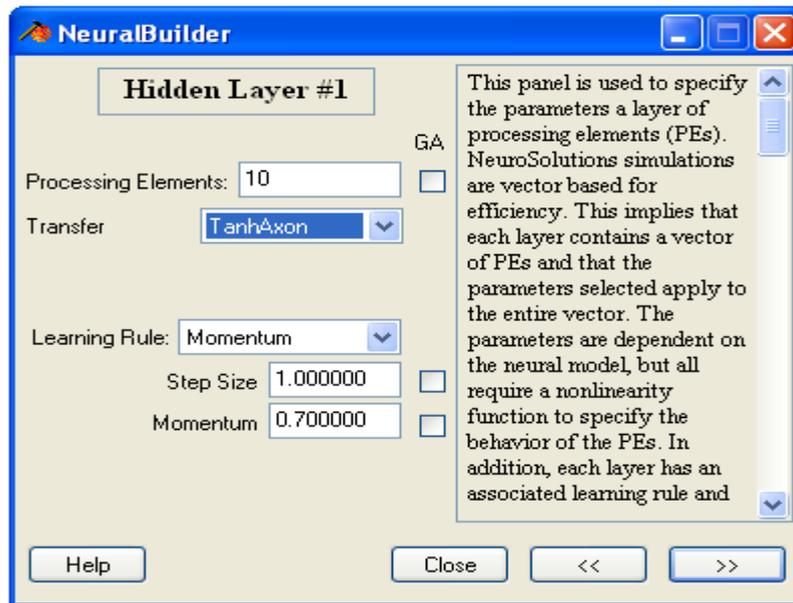


Рис. 4. Окно *NeuralBuilder* для настройки параметров каждого слоя ИНС

Далее задаются параметры обучения нейросети (рис. 5).

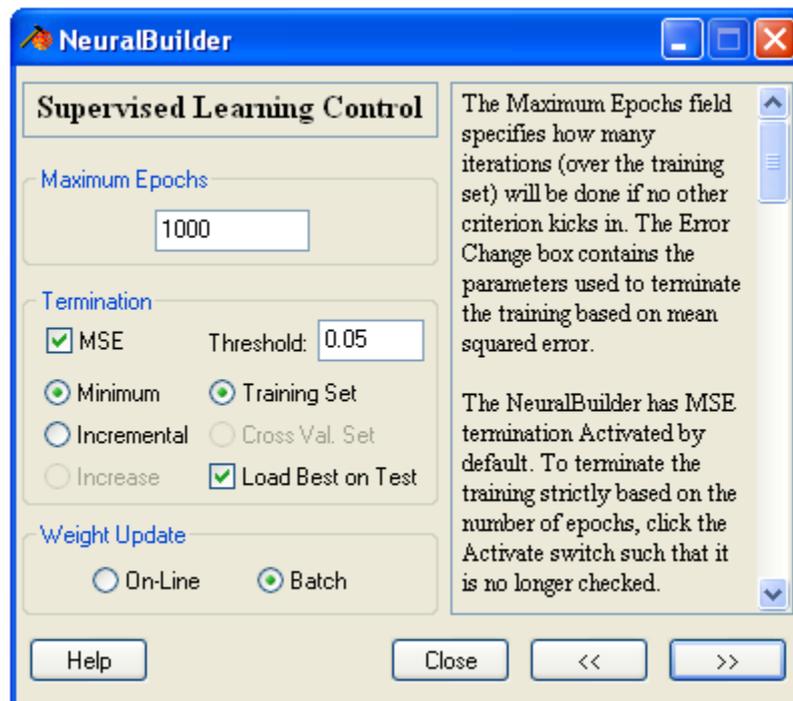


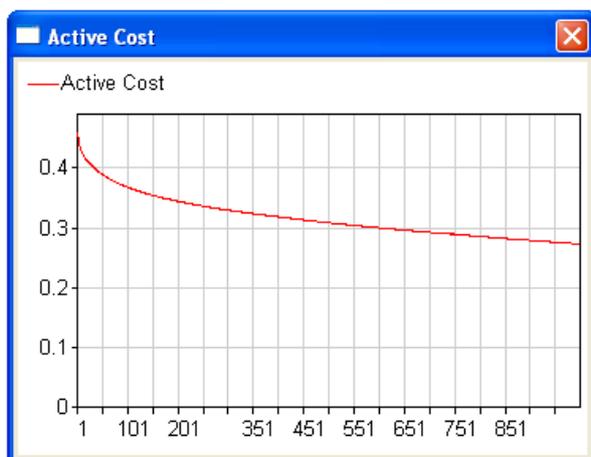
Рис. 5. Настройка параметров обучения ИНС

Максимальное число эпох определяет количество итераций при обучении ИНС, если нет другого критерия останова обучения. *MSE* означает, что обучение прекращается в случае достижения необходимого значения средне квадратичной ошибки (например, 5 %). На этой панели можно выбрать *On-line* обучение, когда каждый экземпляр выборки обучается отдельно, либо *Batch* – пакетное обучение ИНС.

Далее можно выполнить обучение ИНС, указав расположение файлов с входными и ожидаемыми выходными данными, и тестирование ИНС с выдачей соответствующих результатов (рис. 6).

Out d0	Out d1	Out d2	Out d3
0.851227300176	0.617155384892	0.526979450326	0.317283482932
0.830386447137	0.538552463242	0.496842602859	0.249014810735
0.877012818855	0.611598631108	0.538477614840	0.315269993896
0.856985871054	0.581563579679	0.503347371711	0.302465131664
0.770767716934	0.459846487792	0.445306714309	0.220129723512
0.856756577617	0.582451038491	0.503896715005	0.302735074210
0.856754056357	0.582461252223	0.503903076999	0.302738213634
0.832468306202	0.751194318539	0.637579186017	0.349546620145
0.856754400164	0.582459858863	0.503902209045	0.302737785314
0.832468301157	0.751195230916	0.637579964964	0.349546693699

Des d0	Des d1	Des d2	Des d3
0.000000000000	0.000000000000	1.000000000000	0.000000000000
0.000000000000	0.000000000000	0.000000000000	0.000000000000
0.000000000000	0.000000000000	0.000000000000	0.000000000000
0.000000000000	0.000000000000	1.000000000000	0.000000000000
0.000000000000	0.000000000000	0.000000000000	0.000000000000
0.000000000000	1.000000000000	0.000000000000	0.000000000000
0.000000000000	1.000000000000	0.000000000000	0.000000000000
0.000000000000	0.000000000000	0.000000000000	0.000000000000
0.000000000000	1.000000000000	0.000000000000	0.000000000000
0.000000000000	0.000000000000	1.000000000000	0.000000000000
0.000000000000	0.000000000000	0.000000000000	0.000000000000



Active Performance	
MSE	0.934607056851
NMSE	3.997454414861
r	-0.010388065043
% Error	5.328036317650
AIC	4337.430094272957
MDL	3944.211077576086

Рис. 6. Результаты тестирования ИНС

MSE – средняя квадратичная ошибка, $NMSE$ – нормализованная средняя квадратичная ошибка, r – коэффициент корреляции, $\% Error$ – значение ошибки в процентах, AIC – информационный критерий Экейка, MDL – длина минимального описания Риссанена.

Литература

1. Башмаков А.И., Башмаков И.А. Интеллектуальные информационные технологии: Учеб. пособие. –М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. – 304 с.

Контрольные вопросы

1. На какой парадигме основан нейрокомпьютинг?
2. Что понимается под обучением ИНС? Какую роль оно играет в нейротехнологиях?
3. Какие существуют подходы к представлению результатов обучения ИНС?
4. Перечислите основные классы НП.
5. Назовите основные модули, входящие в архитектуру универсального НП.
6. По каким критериям сравнивают универсальные НП?