

ПОСТРОЕНИЕ И ИНИЦИАЛИЗАЦИЯ НЕЙРОСЕТЕВОГО АЛГОРИТМА РАСПОЗНАВАНИЯ ПО ПРОЕКЦИЯМ ЭЛЕМЕНТОВ ОБУЧАЮЩЕЙ ВЫБОРКИ.

А.Ю. Максимова

Институт прикладной математики и механики НАН Украины

Запропоновано спосіб створення економної архітектури нейронної мережі та її ініціалізації на основі попереднього аналізу маломірних проєкцій навчальної вибірки.

Введение. Нейросетевые алгоритмы распознавания обычно строятся по стандартной архитектуре. Если это сети прямого распространения, то предполагается, что они полносвязные, а количество нейронов в каждом слое подбирается на основе экспериментальных данных различными алгоритмами, например на основе использования генетических алгоритмов[1]. Инициализация сети также обычно проводится без учета особенностей обучающей выборки. Используя нечеткую базу знаний и нейронную сеть для ее настройки, можно объединять два подхода и учитывать знания о структуре данных при построении нейронной сети[2]. Предварительный анализ обучающей выборки может дать информацию для упрощения структуры нейронной сети и ускорения ее обучения за счет более точной инициализации нейронной сети. В работе предлагается способ такого упрощения и инициализации с использованием системы Matlab.

Система Matlab позволяет моделировать и настраивать нейронные сети, однако в системе имеется еще ряд дополнительных возможностей, в частности графические возможности, которые позволяют анализировать данные, просматривая двух и трехмерные диаграммы рассеивания [3]. На основании полученных в результате такого анализа данных строится структура сети и производится инициализация весовых коэффициентов и пороговых значений.

Метод формирования нейронной сети. Рассматривается классическая постановка задачи распознавания образов. Образы представляются векторами в n -мерном пространстве. Задано множество классов образов $V = \{v_i\}$, $i=1, \dots, p$ и обучающая выборка M как множество пар $M = \{(x^{(i)}, v^{(i)}), x^{(i)} \in R^n, v^{(i)} \in V, i=1, \dots, m\}$. В задаче не предполагается полная разделимость классов, что связано с природой данных и зашумленностью выборки.

Идея построения сети опирается на предварительный анализ двух и трехмерных проекций элементов обучающей выборки. На каждой проекции образуются визуально наблюдаемые кластеры. В зависимости от свойств исходных данных это могут быть монокластеры (рис. 1 а), которые состоят из представителей одного класса, или сгруппированные в один кластер представители нескольких классов (рис. 1 б).

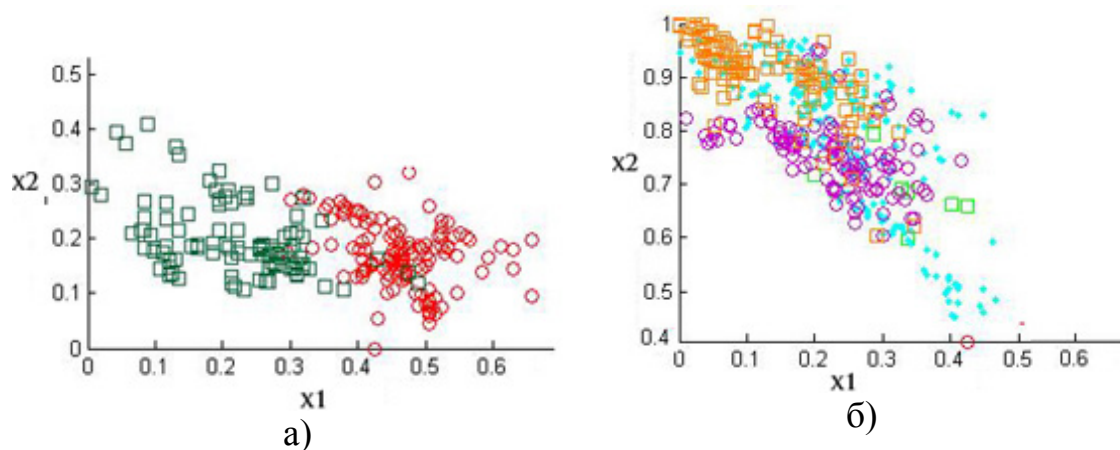


Рис. 1. Типы кластеров проекций элементов обучающей выборки

а) монокластеры; б) объединение нескольких классов образов в один кластер.

Для каждой проекции строится совокупность разделяющих прямых или плоскостей, разбивающая множество кластеров на области (рис. 2).

В результате этих шагов определяется структура внутреннего слоя нейронной сети вместе с начальной инициализацией весовых коэффициентов. Для выявления кластеров можно также использовать методы кластеризации.

Диаграмма рассеивания позволяет указать грубые значения параметров разделяющей прямой или плоскости. Используя свойство персептрона строить разделяющую гиперплоскость $\sum_{i=1}^n w_i x_i + b = 0$, можно получить эти значения автоматически.

В системе Matlab для работы с нейронными сетями используется пакет Neural Network Toolbox. В нем запрограммированы стандартные структуры сетей, а также алгоритмы их обучения, что было использовано для построения сети.

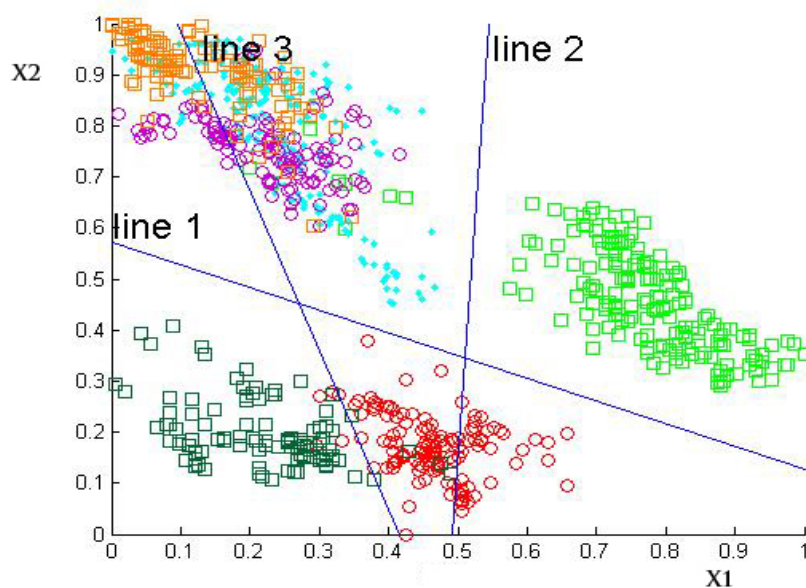


Рис. 2 Совокупность разделяющих прямых, разбивающая множество кластеров на области.

Структура нейронной сети. Как известно, универсальной нейронной сетью, реализующей функциональную зависимость, является двухслойная нейронная сеть[4]. Такую архитектуру и будем рассматривать в дальнейшем.

Количество нейронов во входном слое обусловлено размерностью обучающей выборки, в выходном слое сети каждому классу соответствует свой нейрон.

Каждая разделяющая прямая или плоскость задана весовыми коэффициентами активных связей каждого нейрона скрытого слоя (далее по тексту будем говорить о прямых, подразумевая, что эта информация распространяется и на плоскости). Активными являются связи от входов сети, определяющих проекцию, для которой задана прямая. По количеству таких разделяющих прямых определяется количество нейронов скрытого слоя. В качестве функции активации нейронов скрытого слоя можно брать биполярную сигмоиду или симметричную пороговую функцию. Важно, что область значений этой функции $(-1,1)$ или $[-1,1]$. Разделяющие прямые определяют области, характеризующие кластеры. Нейроны выходного слоя агрегируют информацию о принадлежности областям. Для настройки нейронов выходного слоя строится прямая, отделяющая одну точку бинарного куба (значения координат вершин куба 1 и -1). Функцией активации нейронов второго слоя может быть униполярная сигмоида или пороговая функция.

Данный способ формирования структуры и инициализации нейронной сети был апробирован на реальной задаче распознавания производителей топлива и их видов. Входной вектор представлял из себя набор из пяти измеряемых показателей, таких как октановое число, суммарные ароматические соединения, ароматические бензолы, олефины и метил-трет-бутиловый эфир (МТБЭ). Число классов было выбрано равным шести. По диаграммам рассеивания обучающей выборки (объемом около тысячи векторов) было определено 6 базовых двумерных проекций, и определены 17 нейронов, соответствующих прямым, разделяющим найденные от двух до пяти кластеров по каждой проекции. Инициализация определялась обучением частичных нейронных сетей. В результате была построена двухслойная нейронная сеть (5,17,6) и задана ее начальная инициализация. Эта начальная инициализация существенно ускорила процесс обучение сети по сравнению с инициализацией нулевыми значениями.

Выводы

В работе предложен метод формирования архитектуры и инициализации нейронной сети для задачи распознавания. Метод основан на визуальном выявлении кластеров по двумерным и(или) трех мерным проекциям элементов обучающей выборки с использованием графических возможностей Matlab. Полученная в результате предварительного анализа данных информация о внутренней структуре данных может быть использована для повышения скорости сходимости процесса обучения и качества распознавания.

Библиографический список

1. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы. / Д. Рутковская, М. Пилиньский, Л. Рутковский // Перевод с польского И. Д. Рудинского. – М. : Горячая линия - Телеком, 2006. – 383 с.: ил.
2. Ротштейн А.П. Нейролингвистическая идентификация нелинейных зависимостей. / А.П. Ротштейн, Ю.И. Митюшкин. // Кибернетика и системный анализ. – 2000. – № 2. – С. 37-44.
3. Медведев В.С. Нейронные сети. MATLAB 6. / В.С. Медведев, В.Г. Потемкин // Под общей редакцией к.т.н. В.Г. Потемкина. – М.: ДИАЛОГ-МИФИ, 2002. – 496 с. – (Пакеты прикладных программ; Кн. 4)
4. Осовский С. Нейронные сети для обработки информации. / С. Осовский // Перевод с польского И.Д. Рудинского. – М. : Финансы и статистика, 2002. –344 с.: ил.