

ПОДХОД К ФОРМИРОВАНИЮ АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ ОБРАЗОВ

В. М. Татьянкин

Введение

В настоящее время направление «распознавание образов» приобрело огромную популярность среди исследователей. Это связано с тем, что решение задач в рассматриваемой области находит всё большее прикладное применение в повседневной деятельности человека [4]. В качестве примера можно привести распознавание: штрих кодов, автомобильных номеров, лиц, речи, болезней. Особый интерес использования методов распознавания проявляется в исследовании социально-экономических систем. Это обусловлено тем, что задачи, решаемые в рассматриваемых областях, слабо формализованы и классические подходы не позволяют получить достоверные результаты, которые имели бы практическую пользу. Хотя область применения просто огромна, в подтверждение этому можно сказать, что задачи разработки и внедрения экспертно-аналитических систем прогнозирования социально-экономических процессов являются одними из основных в стратегии инновационного развития Российской Федерации на период до 2020 года. В связи с этим сейчас наблюдается устойчивая тенденция к росту популярности использования способов распознавания образов в социально-экономических системах.

Постоянное развитие теоретического, математического аппарата также является немаловажным фактором в популярности «распознавания образов». Среди основных подходов для решения задач распознавания образов стоит выделить: классификации с помощью решающих функций и функцией расстояния; различные алгоритмы кластеризации; машины опорных векторов; нейронные сети и различные статистические методы [3].

Нейронные сети глубокого доверия. Особый интерес представляют искусственные нейронные сети, в частности нейронные сети глубокого доверия. Нейронные сети глубокого доверия представляют собой многослойную нейронную сеть, представленную на рисунке 1. Основное отличие заключается в алгоритме обучения нейронной сети глубокого доверия и за счёт этих алгоритмов удаётся наращивать количество обрабатывающих слоёв нейронной сети при уменьшении ошибки обучения.

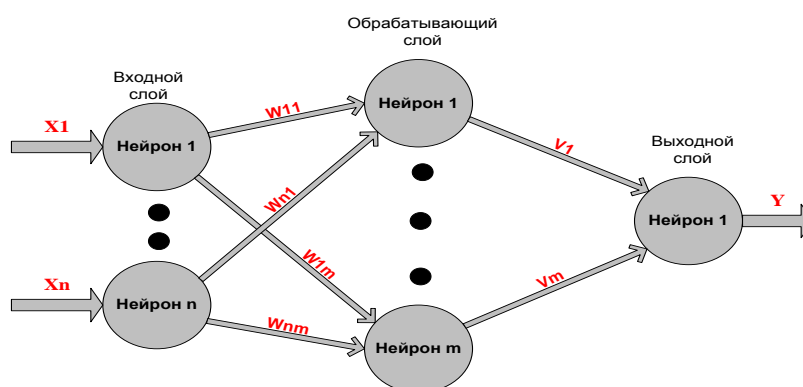


Рисунок 1 – Многослойная нейронная сеть

По мнению исследователей Массачусетского технологического института, нейронные сети глубокого доверия (deep belief neural networks) входят в список 10 наиболее прорывных высоких технологий, способных в ближайшее время значительно изменить повседневную жизнь человечества [2]. Данное мнение подтверждается и множественными практическими исследованиями в области распознавания образов, где в качестве тестовой выборки выступают открытые базы различных изображений.

Одной из самых популярных является база рукописных символов MNIST (Mixed National Institute of Standards and Technology) [1]. База состоит из 60000 изображений рукописных цифр для обучения и 10000 для тестирования. Размер изображения составляет 28 на 28 пикселей. Пример изображений представлен на рисунке 2.



Рисунок 2 – Пример рукописных цифр базы MNIST

Одни из самых лучших результатов, полученные при распознавании цифр для представленной базы, относятся к нейронным сетям глубокого доверия. Полученный результат позволяет говорить о том, что нейронные сети глубокого доверия являются эффективным инструментом для обнаружения сложных закономерностей в исследуемых данных. Что подтверждается использованием нейронных сетей глубокого доверия в повседневной жизни [6]. Однако есть сферы деятельности человека, где требуется 100 % точность идентификации, например в банковском секторе.

Решать эту проблему можно несколькими направлениями: разработка новых способов обучения нейронных сетей; увеличение «объёма» нейронной сети, то есть наращивание слоёв и количества обрабатываемых нейронов; предварительная обработка обучающих данных; разработка новых способов формирования архитектуры нейронной сети.

Рассмотрим классический подход к формированию архитектуры нейронной сети на примере распознавания рукописных цифр MNIST: формируется входной слой размером в 784 (28*28) нейрона, то есть за каждый пиксель изображения отвечает один нейрон; формируются обрабатывающие слои, в зависимости от объёма обучающей выборки и от мощности электронной вычислительной машины; формируется выходной слой размером в 10 нейронов, то есть каждому нейрону соответствует одна цифра.

Далее в статье будет приведен альтернативный подход к формированию архитектуры нейронной сети.

Способ идентификации образов

Представленный далее способ будет базироваться на следующих личных рассуждениях. Человек при идентификации сложных, неизвестных образов руководствуется методом от противного, то есть он начинает перебирать все известные ему образы и производит сравнение. Спроецируем данный подход и на обучение нейронных сетей глубокого доверия. В качестве примера также рассмотрим базу рукописных цифр MNIST. Идея будет заключаться в том, что нейронная сеть будет обучаться распознаванию только двух цифр. Тогда понадобится 9 нейронных сетей для идентификации одной цифры (одна цифра будет сравниваться с каждой). А так как цифр 10, то всего понадобится 90 нейронных сетей для полноценного распознавания чисел.

Реализуем рассмотренную идею в виде численного эксперимента. Выберем из рукописной базы цифр MNIST все нули и единички, в нашем случае их получилось около 1000 нулей и 300 единиц. Так как их количество разное, то дополним повторяющимися изображениями единицы, чтоб тоже было около 1000. Сформируем обучающую выборку так, чтобы было чередование цифр. Входной слой оставим без изменения – 784 нейрона, обрабатывающие слои выберем следующего размера: 1000–500–300–100. Выходной слой соответственно будет равняться двум нейронам. Далее произведём обучение нейронной сети глубокого доверия в два этапа [2]:

1. Предобучение нейронной сети автоэнкодерным методом обучения, начиная с первого слоя, данный этап обучения осуществляется без учителя.

2. Настройка синаптических связей всей сети с использованием алгоритма обратного распространения ошибки.

При использовании алгоритма обратного распространения ошибки, могут возникнуть сложности в связи с тем, что рассматриваемый алгоритм не всегда сходится. Для этого предлагается использовать модифицированный алгоритм обратного распространения ошибки, который позволяет всегда завершить обучение многослойной нейронной сети [5, 7]:

1. Задаётся шаг обучения $alf \in (0:1)$ и время работы алгоритма t .

2. Случайным образом инициализируются весовые коэффициенты и пороговые значения нейронной сети.

3. Последовательно подаются векторы данных $\{x_1 \dots x_n\}^i$ на вход нейронной сети, при этом для каждого входного вектора выполняются следующие действия:

– вычисляется выходная активность всех нейронных элементов сети,

$$y_j = F\left(\sum_{i=1}^n w_{ij} \cdot y_i - T_j\right), \quad (1)$$

где индекс j характеризует нейроны следующего слоя по отношению к слою i ;

– определяется ошибка нейронных элементов сети для выходного слоя:

$$Es_j = Y_j - t_j \quad (2)$$

и скрытого:

$$Es_j = \sum_{i=1}^n Es_i \cdot F'(S_i) \cdot w_{ji}, \quad (3)$$

где индекс j характеризует нейроны следующего слоя по отношению к слою i , t_j – эталонное значение.

4. Вычисляется суммарная среднеквадратическая ошибка нейронной сети:

$$E = \frac{1}{2} \sum_{k=1}^L \sum_j (y_j^k - t_j^k)^2, \quad (4)$$

где L – количество обучающих векторов.

5. Создаём переменную $Error = E$.

6. Включаем счётчик времени $t1$.

7. Если $t > t1$, переходим к пункту восемь, иначе алгоритм заканчивается.

8. Весовые коэффициенты и пороги нейронных элементов изменяются по следующим правилам:

$$w_{ij}(t+1) = w_{ij}(t) - alf \cdot Es_j \cdot F'(S_j) \cdot y_i \quad (5)$$

$$T_j(t+1) = T_j(t) + alf \cdot Es_j \cdot F'(S_j). \quad (6)$$

9. Выполняем пункты 3 и 4.

10. Если получившаяся ошибка меньше $Error$, то $Error = E$, запоминаем весовые коэффициенты W и порог T , переходим к пункту 7, иначе переходим в пункт 7.

Особенность предложенного алгоритма заключается в использовании времени, как критерия остановки алгоритма.

После того как произвели обучение нейронной сети для распознавания нуля или единицы, повторим вышепредложенную процедуру для нуля и двойки, нуля и тройки и так далее. В итоге у нас есть укрупненная группа нейронных сетей для идентификации нуля. Аналогично сделаем и для других цифр.

Теперь можем проверить получившуюся нейронную сеть, с целью точности распознавания, на всей обучающей выборке. В нашем случае получаем одну неверно идентифицированную цифру.

Заключение

В ходе численной апробации предложенного способа формирования архитектуры нейронной сети, на примере базы рукописных изображений, была получена близко к нулевому значению ошибка распознавания в ходе обработки обучающей выборки. Проведённое исследование создаёт предпосылки для создания экспертно-аналитических систем прогнозирования социально-экономических процессов, что позволит повысить эффективность управления государственного аппарата при снижении финансирования.

Литература

1. База изображений рукописных цифр MNIST [Электронный ресурс]. – Режим доступа : <http://yann.lecun.com/exdb/mnist/> (дата обращения 1.10.2015).
2. Головкин, В. А. Применение нейронных сетей глубокого доверия для выделения семантически значимых признаков [Электронный ресурс]. – Режим доступа: <http://libeldoc.bsuir.by/bitstream/123456789/4138/1/Применение%20нейронных%20сетей.PDF>. (дата обращения 1.10.2015).
3. Лепский, А. Е. Математические методы распознавания образов. Курс лекций [Текст] / А. Е. Лепский, А. Г. Броневиц [Электронный ресурс]. – Режим доступа: http://lepskiy.ucoz.com/lect_Lepskiy_Bronevich_pass.pdf (дата обращения 1.10.2015).
4. Карминская, Т. Д. Использование кластерного анализа и нейронных сетей в задаче управления региональным рынком труда [Текст] / Т. Д. Карминская, В. М. Татьянакин, О. Д. Тей [и др.] // Доклады Томского государственного университета систем управления и радиоэлектроники: сб. науч. тр. – 2013. – № 3(30). – С. 201–205.
5. Татьянакин, В. М. Модифицированный алгоритм обратного распространения ошибки [Электронный ресурс] / В. М. Татьянакин // Приоритетные направления развития науки и образования : материалы III междунар. науч.–практ. конф. (Чебоксары, 04 дек. 2014 г.) / редкол. : О. Н. Широков [и др.] – Чебоксары: ЦНС «Интерактив плюс», 2014. – Режим доступа: <http://interactive-plus.ru/e-articles/collection-20141204/collection-20141204-5263.pdf> (дата обращения 1.10.2015).
6. Третье поколение нейросетей: «Глубокие нейросети» [Электронный ресурс]. – Режим доступа : <https://www.mql5.com/ru/articles/1103> (дата обращения 1.10.2015).