

ТЕОРЕТИЧЕСКИЕ И ПРАКТИЧЕСКИЕ ОСНОВЫ ПОСТРОЕНИЯ КОМПЛЕКСА ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ РУКОПИСНЫХ СИМВОЛОВ

А.А. Кочетов, к.т.н. А.В. Липанов
(представил д.т.н., проф. А.В. Королёв)

В статье описаны возможные пути построения нейронных сетей, применяемых для классификации изображений, а также приемы предварительной обработки входных данных с целью повышения качества работы системы. Приведено описание одного из возможных алгоритмов распознавания рукописных символов с применением набора нейронных сетей. Рассматриваемая математическая модель отличается высокой стабильностью работы при минимальных временных затратах в режиме работы комплекса нейронных сетей.

Введение. Нейронные сети в обработке изображений, в частности в задачах, связанных с классификацией образов, находят все более широкое применение. Задача распознавания печатного текста решается повсеместно при помощи мощных программных продуктов, имеющих огромные базы данных с информацией о печатных шрифтах и массой дополнительной информации. Однако задача распознавания символов, как известно, в независимости от шрифта, решается человеком за одно мгновение. Одним из замечательных свойств нейронных сетей является способность находить уникальные зависимости в совокупности входных данных, менять свою чувствительность в зависимости от характера их распределения [1].

Распознавание текста, в целом, достаточно сложный и трудоемкий процесс, связанный с грамотно построенной предварительной обработкой данных. Сначала образ с текстовыми данными подвергается сегментации на строки и слова текста, а затем происходит разбиение на символы. После завершения данного процесса можно приступить к распознаванию, т.е. отнесению символов к тому или иному классу. Проблема сегментации печатного и рукописного текста на строки и слова может быть решена одними и теми же алгоритмами по причине стационарности способов построения текстовых блоков. Процесс классификации не является стационарным, поскольку манера написания символа практически уникальна для каждого отдельно взятого человека.

При разработке систем распознавания рукописных символов целесообразно первоначально сузить класс распознаваемых символов. В связи с этим рассмотрим задачу распознавания рукописных цифр от 0 до 9. В целом проблему распознавания символов в контексте нейронных сетей можно представить состоящей из двух частей: обучения и распознавания. Обучение осуществляется путем подачи на вход отдельных образов с указанием их принадлежности тому или другому классу. В результате такого процесса распознающая система приобретает способность реагировать одинаковыми реакциями на все объекты одного класса и различными – на все объекты различных классов. За обучением следует процесс распознавания новых объектов, который характеризует действия уже обученной системы. Автоматизация этих процедур и составляет проблему обучения распознаванию образов.

Для распознавания символов на основе входных карт бит 8×8 точек используется сеть встречного распространения с двумя слоями. Сеть имеет также слой нейронов, не влияющих на ее работу и не участвующих в обучении. Этот слой является входным для сети и выполняет функцию распределения импульсов на нейроны следующего слоя. Все весовые коэффициенты нейронов входного слоя равны 1, т.е. входные данные не изменяются при передаче через слой. Второй слой – слой Кохонена [2]. Он включает 16 нейронов, на вход каждого из которых подается 16 импульсов с выходов нейронов сети распределительного слоя. Классический слой Кохонена работает по схеме «победитель забирает все». Однако для исследований более целесообразно использование слоя Кохонена в режиме интерполяции, т.е. когда на выходе слоя может быть несколько ненулевых импульсов с величинами от 0 до 1. Таким образом, выход слоя Кохонена можно интерпретировать как вероятность принадлежности входного вектора к тому или иному классу. В таком случае целая группа нейронов Кохонена, имеющих наибольшие выходы, может передавать свои выходные сигналы в слой Гроссберга [2]. Число нейронов в такой группе должно выбираться в зависимости от задачи, и убедительных данных относительно оптимального размера группы не имеется. Все нейроны вне группы имеют нулевые выходы. При этом задается некоторый радиус сходства нейронов. Эта величина, определяющая, какова максимальная абсолютная величина разности между выходами двух нейронов слоя Кохонена, при которой эти нейроны считаются достаточно близкими, чтобы попасть в группу с наибольшими выходами. После выбора максимального выхода среди нейронов слоя Кохонена производится сравнение абсолютных значений разностей выходов данного нейрона (с максимальным выходом) и остальных нейронов слоя. Для тех нейронов, выходы которых

попали в окрестность выхода максимального нейрона радиусом, равным заданному радиусу сходства, выходы не обнуляются.

Чтобы распределять весовые векторы в соответствии с плотностью входных векторов, которые должны быть разделены, помещая тем самым больше весовых векторов в окрестности большого числа входных векторов, выбран метод с использованием радиуса коррекции. Обучение сети начинается со случайных весов, но на начальной стадии обучающего процесса подстраиваются все веса, а не только связанные с выигравшим нейроном Кохонена. Тем самым весовые векторы перемещаются ближе к области входных векторов. В процессе обучения коррекция весов начинает производиться лишь для ближайших к победителю нейронов Кохонена. Этот радиус коррекции постепенно уменьшается, и корректируются только веса, связанные с выигравшим нейроном Кохонена [3].

Входные векторы желательно нормализовать перед тем, как осуществлять их подачу на вход их сети, следующим образом (1).

$$x'_i = \frac{x_i}{\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}}. \quad (1)$$

Это превращает входной вектор в единичный вектор с тем же самым направлением, т.е. в вектор единичной длины в n -мерном пространстве.

На рис. 1 такой двумерный вектор V представлен в координатах x - y ,

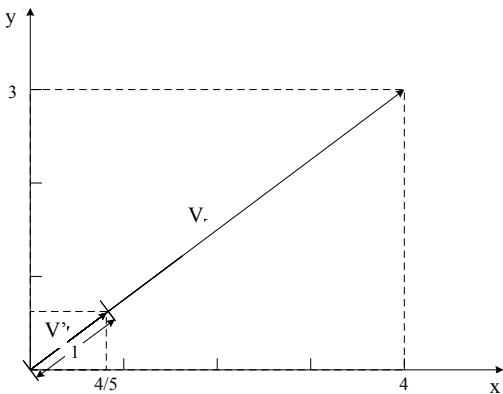


Рис. 1. Единичный входной вектор

причем координата x равна четырем, а координата y – трем. Квадратный корень из суммы квадратов этих компонент равен пяти. Деление каждой компоненты V на пять дает вектор V' с компонентами $4/5$ и $3/5$, где V' указывает в том же направлении, что и V , но имеет единичную длину.

На рис. 2 показано несколько единичных векторов. Они оканчиваются в точках единичной окружности (окружности единичного радиуса), что имеет место, когда у сети лишь два входа. В случае трех входов векторы представлялись бы стрелками, оканчива-

ющимися на поверхности единичной сферы. Эти представления могут быть перенесены на сети, имеющие произвольное число входов, где каждый входной вектор является стрелкой, оканчивающейся на поверхности единичной гиперсферы [3].

При обучении слоя Кохонена на вход подается входной вектор и вычисляются его скалярное произведение с векторами весов, связанными со всеми нейронами Кохонена. Нейрон с максимальным значением скалярного произведения объявляется «победителем» и его веса подстраиваются. Так как скалярное произведение,

используемое для вычисления величин NET, является мерой сходства между входным вектором и вектором весов, то процесс обучения состоит в выборе нейрона Кохонена с весовым вектором, наиболее близким к входному вектору, и дальнейшем приближении весового вектора к входному. Данный процесс является самообучением. Сеть самоорганизуется таким образом, что данный нейрон Кохонена имеет максимальный выход для данного входного вектора. Процесс обучения описывается следующим образом

$$w_n = w_c + \alpha(x - w_c), \quad (2)$$

где w_n – новое значение веса, соединяющего входную компоненту с выигравшим нейроном; w_c – предыдущее значение этого веса; α – коэффициент скорости обучения, который может варьироваться в процессе обучения.

Каждый вес, связанный с выигравшим нейроном Кохонена, изменяется пропорционально разности между его величиной и величиной входа, к которому он присоединен. Направление изменения минимизирует разность между весом и его входом.

На рис. 3 этот процесс показан геометрически в двумерном виде. Сначала находится вектор $x - w_c$, для этого проводится отрезок из конца w в конец x . Затем этот вектор укорачивается умножением его на скалярную величину α , меньшую единицы, в результате чего получается вектор изменения δ . Окончательно новый весовой вектор w_n является отрезком, направленным из начала координат в конец вектора δ . Отсюда

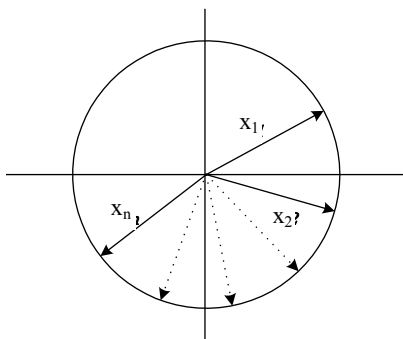


Рис. 2. Двумерные единичные векторы на единичной окружности

можно видеть, что эффект обучения состоит во вращении весового вектора в направлении входного вектора без существенного изменения его длины.

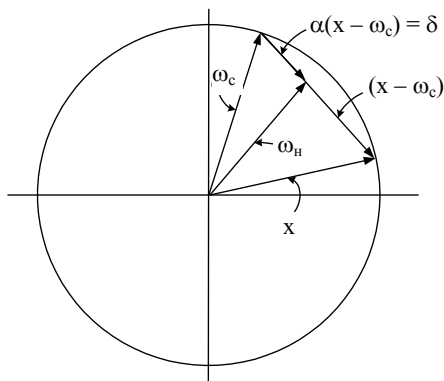


Рис. 3. Вращение весового вектора в процессе обучения (ω_n – вектор новых весовых коэффициентов, ω_c – вектор старых весовых коэффициентов)

Переменная α является коэффициентом скорости обучения, который вначале принят $\sim 0,5$ и может постепенно уменьшаться в процессе обучения. Это позволяет делать большие начальные шаги для быстрого грубого обучения и меньшие шаги при подходе к окончательной величине.

Если бы с каждым нейроном Кохонена ассоциировался один входной вектор, то слой Кохонена мог бы быть обучен

с помощью одного вычисления на вес. Веса нейрона-победителя приравнивались бы к компонентам обучающего вектора ($\alpha = 1$). Как правило, обучающее множество включает много сходных между собой входных векторов, и сеть должна быть обучена активировать один и тот же нейрон Кохонена для каждого из них. В этом случае веса этого нейрона должны получаться усреднением входных векторов, которые должны его активировать. Постепенное уменьшение величины α уменьшает воздействие каждого обучающего шага, так что окончательное значение будет средней величиной от входных векторов, на которых происходит обучение. Таким образом, веса, ассоциированные с нейроном, примут значение вблизи «центра» входных векторов, для которых данный нейрон является «победителем».

Когда сеть обучена, приложение входного вектора приводит к требуемому выходному вектору. Обобщающая способность сети позволяет получать правильный выход даже при приложении входного вектора, который является неполным или слегка неверным. Это позволяет использовать данную сеть для распознавания образов, восстановления образов и усиления сигналов.

Для решения задач обработки изображений и классификации образов, в частности распознавания рукописных символов, подача на вход сети (или комплекса сетей) необработанных предварительных данных зачастую не дает желаемого результата. Это связано не со спецификой распознаваемых

образов, а скорее, со спецификой возможных отличий среди образов одного или разных классов. Для символов, написанных человеком, такой спецификой являются наклон, размах написания и т.п. Очевидно, что образ одной и той же цифры, полученный на основе написания разными людьми, могут существенно отличаться. В данном случае необходим алгоритм, способный акцентировать общности входных данных.

В качестве алгоритма, позволяющего идентифицировать особенности каждого символа, можно использовать схему, основанную на методе выделения границ Кирша. Для каждого входного изображения выбираются лишь те сегменты линий образа, которые имеют определенное направление: выделяются горизонтальные, вертикальные, лево- и праводиагональные особенности изображения символа. Для этого в двумерной апертуре 3×3 , используемой в методе Кирша, рассматриваются лишь те элементы, которые характеризуют наличие сегмента линии в данном направлении.

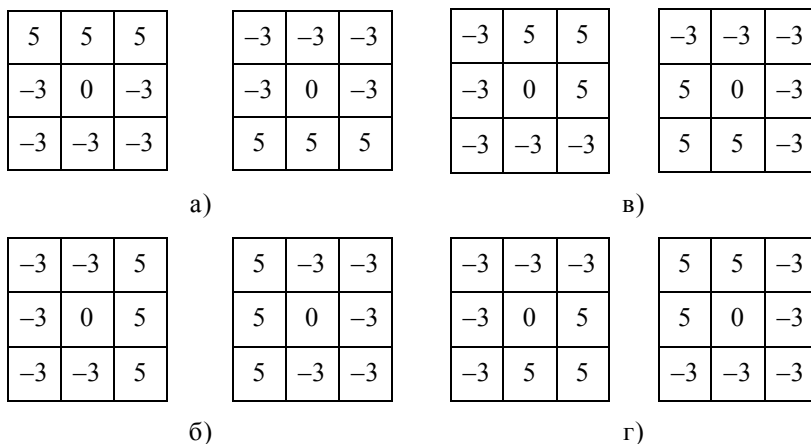


Рис. 4. Анализ образа на основе масок Кирша:

- а) горизонтальные характеристики;
- б) вертикальные характеристики;
- в) праводиагональные характеристики;
- г) леводиагональные характеристики

Формально процесс можно выразить следующими соотношениями:

$$G(i, j)_H = \max (|5S_0 - 3T_0|, |5S_4 - 3T_4|); \quad (3)$$

$$G(i, j)_V = \max (|5S_2 - 3T_2|, |5S_6 - 3T_6|); \quad (4)$$

$$G(i, j)_R = \max (|5S_1 - 3T_1|, |5S_5 - 3T_5|); \quad (5)$$

$$G(i, j)_L = \max (|5S_3 - 3T_3|, |5S_7 - 3T_7|). \quad (6)$$

Величины S_i и T_i определяются из соотношений:

$$S_k = A_k + A_{k+1} + A_{k+2}; \quad (7)$$

$$T_k = A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + A_{k+7}. \quad (8)$$

При этом индексы A вычисляются по модулю 8, и

$$A_k = (k = 0, 1, \dots, 7) - \quad (9)$$

8 соседних точек по отношению к точке с индексами (i, j) определенные, как показано на рис. 5.

A_0	A_1	A_2
A_7	(i, j)	A_3
A_6	A_5	A_4

Рис. 5. Определение 8 соседей по отношению к точке с индексами (i, j)

Таким образом, на основе указанных формул формируется 4 изображения – маски, которые в дальнейшем накладываются на входное изображение 16×16 точек. В результате из исходного изображения удастся получить еще 4 карты бит, характеризующих горизонтальные, вертикальные, лево- и праводиагональные особенности изображения символа.

Проведенные исследования показали, что использование при распознавании карт особенностей написания символа позволяет повысить процент правильного распознавания символов искусственной нейронной сетью (сетями) более чем на 20% по сравнению с обработкой первичных образов.

В зависимости от подбора обучающей и тестовой выборки описанная схема организации нейронной сети позволяет получить от 40 до 60% точности распознавания.

Исходя из того, что входной образ подвергается обработке для определения четырех дополнительных образов-характеристик написания символа, необходимо производить анализ 5 входных образов. Для этого используется набор сетей, основанных на модели встречного распространения со слоями Кохонена и Гроссберга. На вход одной из них подается так называемая глобальная характеристика – изображение символа без обработки алгоритмами выделения направленных сегментов линий изображения. На другие 4 сети подаются изображения, полученные после выделения горизонтальных, вертикальных, лево- и праводиагональных особенностей изображения символа.

Каждая из рассматриваемых нейронных сетей обучается независимо, параллельно. В дальнейшем результаты работы с каждой из сетей также

могут быть подвергнуты анализу для определения точной классификации. Существует множество различных методов анализа результатов работы нескольких нейронных сетей одного типа и архитектуры. Например, для режима аккредитации слоя Кохонена (когда только один нейрон слоя имеет единичный вход), можно применять метод голосования, при котором окончательной классификацией берется та, которая сгенерирована большинством сетей комплекса. В режиме интерполяции можно усреднять выходы сетей, применять элементы нечеткой логики, принимая в расчет вес каждой из нейронных сетей (выбранный, например, на основе статистических данных о том, насколько точна в результатах классификации та или иная нейронная сеть). Альтернативным, но вполне логичным является применение еще одной нейронной сети для произведения классификации. В процессе обучения такая сеть сможет производить коррекцию параметров таким образом, чтобы наибольшие величины семантических весов соответствовали выходам той нейронной сети, которая наиболее точна в оценках. И наоборот – для нейронных сетей, дающих больше других ошибки классификации, чувствительность будет ниже.

Сеть представляет собой достаточно простую модель. Составлена комбинирующая сеть из 10 нейронов. Число 10 определяет количество выходов сетей классификации (их число соответствует количеству классов распознаваемых символов – в данном случае числа от 0 до 9). Каждый нейрон имеет по 4 семантических веса. Каждый вес определяет, насколько точна работа той нейронной сети классификации, которая соответствует данному входу (или весу). Таким образом, процесс обучения комбинирующей сети в некотором смысле представляет собой процесс выставления оценок точности распознавания не только конкретной сети классификации, но и точности распознавания каждого символа каждой сетью. Выходов сети также 10. Из них определяется максимальный выход, а к классу, соответствующему максимальному выходу комбинирующей сети, относится входное изображение.

Использование набора независимо обучаемых нейронных сетей классификации в комбинации с нейронной сетью, анализирующей их выходы, позволяет повысить процент распознавания до 75% для выборки рукописных символов MNIST (Американского Национального Института Информации и Технологий).

Для практических исследований предлагаемого алгоритма разработана программа, включающая реализацию основных компонентов. Разработанная иерархия классов включает модели нейрона, слоя нейронов, персептрона, сети встречного распространения и сети распознавания с числом входных нейронов, равным 16, и числом нейронов, равным

10, для слоя Гроссберга. Приложение осуществляет процессы обучения сети на основе упомянутых выборок MNIST. Состояния комплекса сетей могут быть сохранены для поэтапного обучения. В любой момент времени качество работы комплекса нейронных сетей может быть проверено путем распознавания тестовых выборок, с возможностью анализа не только правильности работы в процентном соотношении, но и того, на каких цифрах ошибок больше всего и в пользу какой цифры была произведена классификация в момент распознавания. В дальнейшем в целях распознавания более широкого объема классов символов возможен переход к распознаванию букв алфавита.

Дальнейшие исследования проблемы распознавания рукописных символов планируется направить на усовершенствование архитектуры нейронных сетей классификации. Как известно, вопрос выбора оптимального количества нейронов в каждом слое нейронной сети и ее архитектуры чрезвычайно сложен. Однако свойства и особенности работы слоя Кохонена, входящего в состав сети встречного распространения, делают возможным производить его реорганизацию, т.е. создавать модель нейронной сети, способной подбирать оптимальное количество нейронов, а также семантические веса [4]. Для реализации подобной модели в частном случае использования сетей встречного распространения необходим лишь дополнительный шаг в обучении нейронной сети – определение тех нейронов, на место которых нужно вставить дополнительные элементы, и тех, которые необходимо удалить из архитектуры.

ЛИТЕРАТУРА

1. Горбань А.Н. и др. *Нейроинформатика*. – Электронная публикация.
2. Ф. Уоссермен. *Нейрокомпьютерная техника: Теория и практика / Перевод на русский язык Ю.А. Зуева, В.А. Точенова*. – М.: Мир, 1992. – 344 с.
3. Grossberg S. *Classical and instrumental learning by neural networks. Progress in theoretical biology // New York : Academic Press. – 1974. – Vol. 3. – P. 51 – 141.*
4. Kohonen T. *Self-organization and associative memory / Series in Information Sciences, vol. 8. – Berlin: Springer verlag, 1984.*

Поступила 22.04.2004

КОЧЕТОВ Анатолий Анатольевич, факультет ПММ Харьковского национального университета радиоэлектроники.

ЛИПАНОВ Александр Витальевич, канд. техн. наук, факультет ПММ Харьковского национального университета радиоэлектроники.
