

Применение байесовской модели нейрона в сети WTA для решения задач распознавания

Д. П. Бураков¹, Г. И. Кожомбердиева²

Петербургский государственный университет путей сообщения Императора Александра I

Санкт-Петербург, Россия

¹burakovdmitry8@gmail.com, ²kgi-liizht@yandex.ru

Аннотация. В докладе обсуждается способ применения модели нейрона, принцип функционирования которого основан на использовании формулы Байеса, в нейронной сети WTA для решения задач распознавания. Применимость предлагаемой модели нейрона показана на реальном примере распознавания цифр почтового индекса. Выявляются и уточняются особенности использования байесовской модели в сравнении с традиционно используемым методом распознавания. Предлагаются и демонстрируются подходы к решению проблемы неточного распознавания символов.

Ключевые слова: нейронные сети; формула Байеса; байесовская модель нейрона; апостериорное распределение вероятностей; нейронная сеть WTA; задачи распознавания образов.

I. ВВЕДЕНИЕ

Решение задач распознавания и классификации – одна из базовых сфер приложения группы методов применения вычислительной техники, называемых искусственным интеллектом [1, 2]. Наиболее широко применяемой для этого технологий являются искусственные нейронные сети [3], состоящие из совокупности простых вычислительных элементов, называемых (искусственными) нейронами.

В работе Г. И. Кожомбердиевой «Модель нейрона с функцией активации на основе формулы Байеса», представленной в данном сборнике, предложена новая модель искусственного нейрона, функция активации которого основана на использовании формулы Байеса, применяемой для вычисления апостериорных вероятностей гипотез о соответствии и несоответствии входных сигналов условию активации нейрона.

В данной работе демонстрируется возможность применения однослойных нейронных сетей, функционирующих по модели WTA (англ. *Winner Takes All*) [3] для решения задачи распознавания символа цифры почтового кода, приведенной в [4].

II. ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ РАСПОЗНАВАНИЯ

Следует отметить, что, несмотря на понятийное разделение, задачи распознавания образов и задачи классификации представляют собой одну и ту же задачу.

Согласно [5], классическая постановка задачи распознавания образов имеет следующий вид: дано множество объектов, относительно которых необходимо провести классификацию. Это множество объектов представлено семейством подмножеств, которые называются классами. При этом заданы информация о классах, описание всего множества объектов и описание информации об объекте, принадлежность которого к определенному классу неизвестна. Требуется по имеющейся информации о классах и описанию объекта установить, к какому из классов относится этот объект. Таким образом, понятия «образ» и «класс» отождествляются, поэтому в дальнейшем будем говорить о задаче распознавания, подразумевая под ней, в том числе, и задачу классификации.

Искусственные нейронные сети представляют собой математическую модель биологических нервных систем и предназначены для решения задач распознавания, т. е. отнесения наблюдаемой ситуации, информация о значениях факторов которой поступает на вход сети, к одному из классов ситуаций, известных сети. Формализуем задачу следующим образом [3, 4]:

1. Имеется множество возможных ситуаций S . Каждая ситуация $s \in S$ описывается вектором значений некоторых числовых характеристик $(f_1(s), f_2(s), \dots, f_n(s))$, отражающих выраженность при наблюдении тех или иных свойств, присущих среде в момент, когда наблюдается ситуация s . Обозначив $x_i = f_i(s)$, получим, что каждой ситуации $s_l \in S$ соответствует входной числовой вектор $\mathbf{x}^{(l)} = (x_1^{(l)}, \dots, x_n^{(l)})$, где n – число характеристик, используемых для описания (и различения) наблюдаемых ситуаций. Тогда множеству возможных ситуаций S ставится в соответствие равномошное множество возможных входных векторов X : $s_l \in S \Leftrightarrow \mathbf{x}^{(l)} \in X, l = 1, \dots, k, k = |S| = |X|$.

2. Имеется множество из m возможных образов (классов ситуаций) $R, |R| = m$.

3. Искусственная нейронная сеть реализует отображение $S \rightarrow R$ следующим образом. На вход сети поступает входной вектор \mathbf{x} , представляющий собой количественное описание наблюдаемой ситуации s (это могут быть результаты измерений некоторых величин, вердикты о выраженности того или иного свойства, утверждения об истинности или ложности разных фактов), так, что $x_i \in [0, 1]$, а на выходе нейронной сети

формируется числовой вектор $\mathbf{y} = (y_1, y_2, \dots, y_m)$, компоненты которого показывают «уверенность» нейронной сети в том, что ситуация s , описываемая входным вектором \mathbf{x} , идентифицируется ею как каждый из образов $r_j \in R$, $j = 1, \dots, m$. Ситуация s считается распознанной сетью как образ r_j , если соответствующая компонента y_j выходного вектора \mathbf{y} имеет наибольшее значение: $(s \mapsto r_j) \Leftrightarrow \left(j = \arg \max_{i \in 1, m} y_i \right)$.

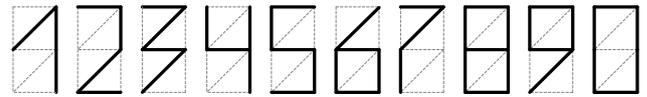
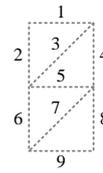


Рис. 1. Образы символов почтового кода

4. Таким образом, искусственная нейронная сеть множеству возможных входных векторов X ставит в соответствие множество возможных выходных векторов Y . При этом для произвольного входного вектора $\mathbf{x} \in X$, описывающего наблюдаемую ситуацию $s \in S$, нейронная сеть вычисляет выходной вектор $\mathbf{y} \in Y$, на основании значений которого по правилу, указанному выше, выбирается решение $r \in R$, которому наблюдаемая сетью ситуация соответствует в большей степени.

Внутренняя структура (топология) нейронной сети состоит из узлов, называемых нейронами, соединенных синаптическими связями.

В классических нейронных сетях, к которым, например, относится перцептрон [3], каждый нейрон сети N_j вычисляет взвешенную сумму сигналов, подаваемых на вход нейрона по синаптическим связям, и передает на выход значение функции активации, формирующей величину выходного сигнала нейрона на основании сравнения значения суммарного входного возбуждения с пороговым значением чувствительности нейрона:

$$y_j = f_A \left(\sum_{i=1}^n x_i \cdot w_{j,i} \cdot p_j \right).$$

Веса $w_{j,i}$ и пороги p_j , задаваемые для каждого нейрона N_j , $j = 1, \dots, t$, по отдельности, регулируют их восприимчивость к поданным входным сигналам и определяются в ходе обучения либо самообучения сети.

Сети типа WTA [3] предполагают дополнение структуры сети дополнительным механизмом сравнения, реализующего конкуренцию нейронов между собой: выходные сигналы нейронов сравниваются между собой и на выход сети поступает единичный сигнал, ассоциированный с номером «победившего» нейрона, а сигналы остальных нейронов сбрасываются в значение 0.

III. РАСПОЗНАВАНИЕ СИМВОЛОВ ПОЧТОВОГО КОДА

Цифры почтового индекса, для удобства их автоматического распознавания при сортировке почтовых отправлений, рисуются отправителями почтовой корреспонденции на специальном шаблоне. Схема шаблона для нанесения цифры почтового индекса и образцы десяти возможных цифр приведены на рис. 1. Каждая цифра индекса образуется путем закрашивания некоторого числа линий на шаблоне. Всего шаблон содержит 9 линий. Каждой цифре почтового индекса соответствует свой уникальный набор закрашенных линий.

Для решения задачи распознавания цифры на шаблоне при помощи нейронной сети будем считать ситуацией, предъявляемой сети, шаблон, закрашенный отправителем. В качестве факторов, описывающих цифру, как предъявляемую ситуацию, будем использовать сведения о том, какие из линий шаблонной сетки закрашены. Для этого перенумеруем их числами от 1 до 9. Тогда любой цифре, нарисованной на сетке шаблона, будет соответствовать входной вектор $\mathbf{x} = (x_1, \dots, x_9)$, где $x_i \in [0; 1]$ означает степень выраженности i -й линии шаблона (0 – линия не прорисована, 1 – линия прорисована абсолютно четко), $i = 1, \dots, 9$. Нейронная сеть должна отнести предъявленный образ цифры почтового индекса, описанный вектором \mathbf{x} , к одному из 10 классов $\{r_1, r_2, \dots, r_{10}\}$ по числу возможных цифр, которые могут быть изображены на шаблоне. Эталонные векторы цифр приведены в таблице ниже.

ТАБЛИЦА I Эталонные векторы образов

№ линии, i	Цифра почтового индекса									
	1	2	3	4	5	6	7	8	9	0
1	0	1	1	0	1	0	1	1	1	1
2	0	0	0	1	1	0	0	1	1	1
3	1	0	1	0	0	1	1	0	0	0
4	1	1	0	1	0	0	0	1	1	1
5	0	0	1	1	1	1	0	1	1	0
6	0	0	0	0	0	1	1	1	0	1
7	0	1	1	0	0	0	0	0	1	0
8	1	0	0	1	1	1	0	1	0	1
9	0	1	0	0	1	1	0	1	0	1

Нейронная сеть типа «перцептрон» для решения поставленной задачи распознавания будет иметь 9 нейронов входного слоя (по числу линий сетки шаблона) и 10 нейронов выходного слоя (по числу возможных цифр). Топология данной сети изображена на рис. 2.

Для корректного распознавания символов необходимо задать веса синаптических связей для каждого из нейронов. Запишем их в виде весовой матрицы $\mathbf{W} = (w_{j,i})$ размерности 19×19 (веса, внесённые в матрицу, полученные в ходе обучения сети на эталонных векторах из табл. I).

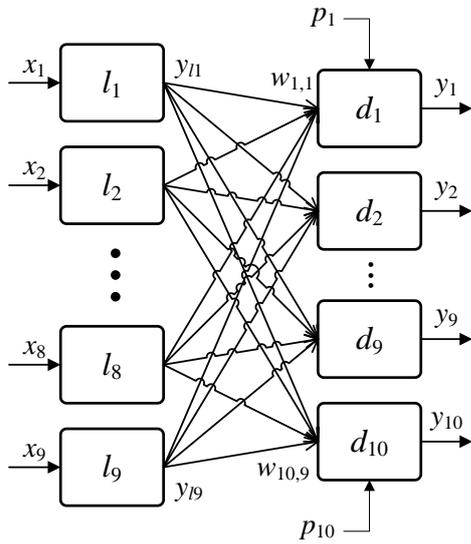


Рис. 2. Топология сети для распознавания цифры почтового индекса

Так как нейронная сеть на рис. 2 не содержит рекуррентных связей, матрицу удобно записать в блочном виде:

$$\mathbf{W} = \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{W}_{xy} & \mathbf{O} \end{pmatrix},$$

где блок \mathbf{W}_{xy} содержит веса связей, соединяющих нейроны входного слоя с нейронами выходного слоя:

$$\mathbf{W}_{xy} = \begin{pmatrix} -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -5 & 1 & 3 & 1 & -5 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -2 & 1 & -3 & 1 & -2 & 1 & 1 \end{pmatrix}$$

Пороги p для всех нейронов первого (входного) слоя заданы нулевыми, а для нейронов выходного слоя – равными 75% от ожидаемого эталонного возбуждения, когда на вход сети подан эталонный вектор \mathbf{x} , соответствующий идеальному изображению цифры почтового индекса на сетке шаблона. Снижение порога возбуждения ниже 100% позволяет сети распознавать цифры, изображения которых на шаблоне представлены линиями разной толщины, четкости и контрастности (в этом случае соответствующая компонента входного вектора будет меньше единицы).

Для демонстрации качества распознавания цифр почтового индекса обученной сетью подадим ей на вход два контрольных вектора: $\mathbf{x}' = (1; 0; 0,8; 0; 1; 0; 0,9; 0; 0)$ и $\mathbf{x}'' = (1; 0,9; 0,9; 1; 0,9; 0,9; 1; 0,9; 1)$.

В первом случае сеть формирует выходной вектор $\mathbf{y}' = (0; 0; 3,7; 0; 0; 0; 0; 0; 0)$, соответствующий цифре 3, а во втором случае – вектор $\mathbf{y}'' = (0; 0; 0; 0; 0; 0; 0; 0; 0)$, не соответствующий ни одной из возможных цифр.

Теперь построим нейронную сеть WTA составленную из байесовских нейронов, решающую ту же задачу распознавания. В отличие от сети, представленной на рис. 2, она будет содержать всего 10 байесовских нейронов и блок сравнения WTA, как показано на рис. 3.

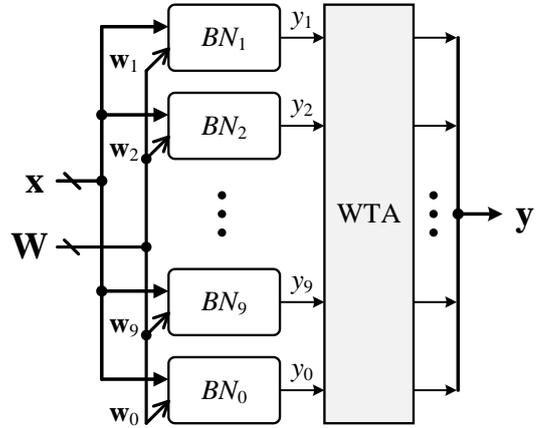


Рис. 3. Сеть WTA, составленная из байесовских нейронов

Здесь на вход сети поступают те же векторы $\mathbf{x} = (x_1, \dots, x_n)$, играющие для каждого из нейронов BN_j роль свидетельств в пользу каждой из цифр, изображенных на рис. 1. Через \mathbf{W} обозначена матрица, содержащая весовые маски нейронов:

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_0 \end{pmatrix}, \quad \mathbf{w}_j = (w_{j,1}, \dots, w_{j,n}).$$

Так как веса в данном случае играют роль маски, определяющей для каждого байесовского нейрона BN_j , какие компоненты входного вектора \mathbf{x} используются для вычисления условных вероятностей «за» и «против» $P_j(H_p|e)$ и $P_j(H_c|e)$ цифры 1, 2, ..., 0, за распознавание которой ответственен этот нейрон, в качестве векторов $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_0$ принимаются векторы, совпадающие с эталонными векторами из табл. I. Так, например,

$$\mathbf{w}_1 = (0, 0, 1, 1, 0, 0, 0, 1, 0).$$

Для вычисления всех необходимых условных вероятностей каждого нейрона BN_j будем использовать формулы, приведенные в работе Г. И. Кожомбердиевой «Модель нейрона с функцией активации на основе формулы Байеса», представленной в данном сборнике:

- Для вычисления вероятностей $P_j(e|H_p)$ и $P_j(e|H_c)$ – формулы (1) и (6) соответственно;
- Для вычисления вероятностей $P_j(H_p|e)$ и $P_j(H_c|e)$ – формулы (3) и (4).

Для вычисления выходного значения нейрона BN_j воспользуемся формулой (5) из этой же работы.

Априорные вероятности обеих гипотез для каждого из десяти байесовских нейронов примем одинаковыми: $P_j(H_P) = P_j(H_C) = 1/2$.

Таким образом, видно, что в отличие от классической сети типа «перцептрон», сеть WTA, построенная с использованием байесовских нейронов, имеет существенно меньшее число параметров, подлежащих настройке, и состоит из меньшего числа нейронов.

Проверим качество распознавания символов почтового индекса при помощи такой сети, приведенной на рис. 3. Для этого подадим на вход сети входной вектор \mathbf{x} , компоненты которого $x_i \in [0, 1]$, $i = 1, \dots, 9$ будут содержать интенсивности выделения соответствующих линий на сетке шаблона символа, приведенной на рис. 1:

$$\mathbf{x} = (0,75; 0,8; 0; 0,95; 0; 1; 0; 0,8; 0,8).$$

Указанный вектор соответствует символу цифры «0», часть линий контура которой изображена недостаточно четко.

Результат работы сети приведен в табл. II.

ТАБЛИЦА II Состояния байесовских нейронов сети WTA

Номер нейрона, j	Вероятности				Выход Y_j
	$P_j(e H_P)$	$P_j(e H_C)$	$P_j(H_P e)$	$P_j(H_C e)$	
1	0,583	0,621	0,484	0,516	0,000
2	0,625	0,600	0,510	0,490	0,510
3	0,188	0,892	0,174	0,826	0,000
4	0,638	0,592	0,519	0,481	0,519
5	0,630	0,590	0,516	0,484	0,516
6	0,520	0,700	0,426	0,574	0,000
7	0,583	0,621	0,484	0,516	0,000
8	0,729	0,333	0,686	0,314	0,686
9	0,500	0,720	0,410	0,590	0,000
0	0,850	0,250	0,773	0,227	0,773

Таким образом, совокупно байесовские нейроны формируют выходной вектор:

$$\mathbf{y}' = (0; 0,510; 0; 0,519; 0,516; 0; 0; 0,686; 0; 0,773),$$

который подается на вход блока сравнения, и в соответствии с принципом WTA преобразуется в итоговый выходной вектор сети:

$$\mathbf{y} = (0; 0; 0; 0; 0; 0; 0; 0; 0; 1).$$

Вектор \mathbf{y} сигнализирует, что входной вектор \mathbf{x} , поданный на вход сети, был интерпретирован как свидетельство в пользу гипотезы «цифра 0».

Обратим внимание на то, что входной вектор \mathbf{x} не противоречит гипотезе «цифра 8», (гипотеза H_P нейрона BN_8), хотя и не в полной мере ей соответствует. Для того, чтобы уменьшить апостериорную вероятность этой гипотезы, повышается вероятность альтернативной гипотезы H_C этого нейрона, путем трактовки недостатка свидетельств в пользу 8 как дополнительного свидетельства «против» активного состояния нейрона BN_8 .

IV. ЗАКЛЮЧЕНИЕ

В докладе продемонстрирована применимость байесовской модели нейрона, предложенной в докладе Г.И. Кожомбердиевой «Модель нейрона с функцией активации на основе формулы Байеса», представленном в данном сборнике, для построения нейронных сетей WTA, предназначенных для решения задач распознавания. Указанная применимость проиллюстрирована на решении модельной задачи распознавания символа почтового индекса, взятой из [4].

Сравнение структур классической нейронной сети «перцептрон» и сети WTA, составленной из байесовских нейронов, спроектированных для решения одной и той же задачи распознавания, позволяет отметить следующие преимущества последней:

- Структура сети получается более простой по сравнению со структурой классической сети за счет уменьшения числа слоев;
- Обучение сети производится проще за счет использования меньшего числа весов, в простейшем случае веса, соответствующие эталонным векторам распознаваемых образов, обеспечивают приемлемое качество работы сети.

СПИСОК ЛИТЕРАТУРЫ

- [1] Поспелов Д.А. Фантазия или наука: на пути к искусственному интеллекту. М: Наука. Главная редакция физико-математической литературы. 1982. 224 с.
- [2] Горелик А.Л., Скрипкин В.А. Методы распознавания: учеб. пособие для вузов. 3-е изд., перераб. и доп. М.: Высшая школа, 1989. 232 с.
- [3] Осовский С. Нейронные сети для обработки информации; [пер. с польск.]. 2-е изд., перераб. и доп. М.: Горячая линия – Телеком, 2018. 448 с.
- [4] Бураков Д.П. Логические основы интеллектуальных систем. Ч. 2: учеб. пособие. СПб.: ФГБОУ ВО ПГУПС, 2020. 72 с. (в печати).
- [5] Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики. М.: Наука, 1978, вып. 33. С. 5–68.