

Модель нейрона с функцией активации на основе формулы Байеса

Г. И. Кожомбердиева

Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
kgi-liizht@yandex.ru

Аннотация. В докладе предлагается модель нейрона, принцип функционирования которого основан на применении формулы Байеса. При этом входные сигналы нейрона интерпретируются как свидетельства в пользу двух альтернативных гипотез: соответствия и несоответствия условию активации нейрона. Выходной сигнал нейрона формируется на основе рассчитанного по формуле Байеса апостериорного распределения вероятностей гипотез. Эффективность применения новой байесовской модели нейрона показана на примере.

Ключевые слова: нейронные сети; формула Байеса; апостериорное распределение вероятностей; функция активации; байесовская модель нейрона; распознавание и классификация

I. ВВЕДЕНИЕ

В настоящее время мир переживает очередную, третью волну популярности искусственного интеллекта и нейронных сетей. Может показаться даже, что все проблемы искусственного интеллекта сузились до обучения глубоких нейронных сетей (англ. *DNN – Deep Neural Network*) [1]. При этом все большее внимание уделяется применению байесовских техник в глубинном обучении (англ. *Bayesian Deep Learning*), своего рода «скрещиванию» байесовских методов с нейронными сетями [2, 3].

Интерес автора настоящего доклада, однако, связан не с применением байесовских методов в машинном обучении. В докладе предпринимается попытка вернуться, так сказать, к истокам, к первым моделям нейрона МакКаллоха – Питтса [4, с. 13–15] и персептрона Розенблатта [5, с. 120–123], и исследовать возможности использования формулы Байеса в самой структуре нейрона. Оптимизм автора основан на результатах работ, в частности, [6–12], выполненных при заинтересованном участии уважаемых соавторов, особенно Д. П. Буракова, и успешно прошедших апробацию на конференциях *SCM'2017* и *SCM'2019*.

В работах [6–12] рассматриваются нетрадиционные модели на основе формулы Байеса, объединенные общей ключевой идеей. Во всех моделях (оценивания качества решений, процессов или продуктов; группового экспертного рейтингового оценивания; нечеткого вывода) входные оценочные данные интерпретируются как свидетельства в пользу той или иной гипотезы из

множества возможных, определяемых спецификой решаемых задач. Свидетельства тем или иным способом трансформируются в набор байесовских условных вероятностей, вычисляемых в предположении истинности соответствующей гипотезы, а в качестве выходных данных используется апостериорное распределение вероятностей на множестве этих гипотез. Это распределение используется либо непосредственно как результат для принятия решения, либо как основа для вычисления итогового результата.

Можно ли распространить эту идею на структуру нейрона? По мнению автора – вполне возможно, если рассматривать входные сигналы нейрона как свидетельства в пользу двух альтернативных гипотез: соответствия и несоответствия условию активации нейрона, а при вычислениях условных вероятностей, оценивающих степени соответствия свидетельств предположениям об истинности байесовских гипотез, использовать вектор-маску нейрона (эталонный образ). Тогда, получив по формуле Байеса апостериорное распределение вероятностей гипотез, можно сформировать на его основе выходной сигнал нейрона.

II. БАЙЕСОВСКАЯ МОДЕЛЬ НЕЙРОНА

В рассуждениях о структуре нейрона с функцией активации на основе формулы Байеса оттолкнемся от модели МакКаллоха – Питтса [4, с. 13], структурная схема которой представлена на рис. 1. Входные сигналы x_i , $i = 1, \dots, n$, суммируются с учетом соответствующих весов $w_{j,i}$ в сумматоре, после чего результат сравнивается с пороговым значением $w_{j,0}$.

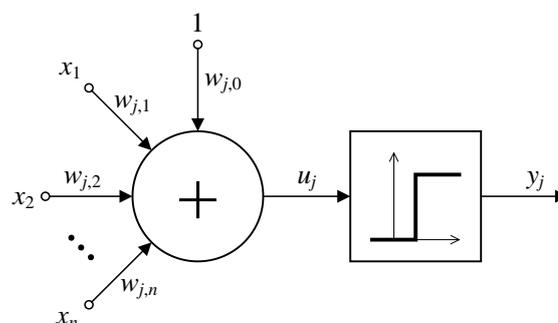


Рис. 1. Модель нейрона МакКаллоха – Питтса

Выходной сигнал нейрона y_j определяется зависимостью $y_j=f(u_j)$, где в качестве аргумента функции используется сумма:

$$u_j = \sum_{i=1}^n x_i \cdot w_{j,i} + w_{j,0}.$$

Функция $f(u_j)$ называется функцией активации нейрона. В модели МакКаллоха–Питтса в качестве функции активации используется пороговая функция вида

$$f(u_j) = \begin{cases} 1, & u_j > 0, \\ 0, & u_j \leq 0. \end{cases}$$

Веса синаптических связей $w_{j,i}$ подбираются, например, в процессе обучения нейрона.

Теперь, в соответствии с идеей, изложенной во Введении, и ориентируясь на обозначения, принятые в модели МакКаллоха–Питтса, на рис. 2 представим структуру байесовского нейрона.

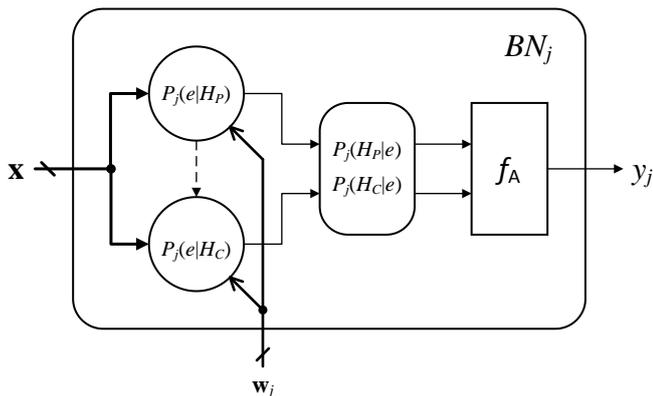


Рис. 2. Модель нейрона с функцией активации на основе формулы Байеса (BN – Bayesian Neuron)

Здесь двоичный входной вектор $\mathbf{x}=(x_1, \dots, x_n)$ рассматривается как свидетельство в пользу двух альтернативных гипотез H_P и H_C : соответствия (*pro* – «за») и несоответствия (*contra* – «против») условию активации нейрона.

Весовой вектор $\mathbf{w}_j=(w_{j,1}, \dots, w_{j,n})$ интерпретируется как двоичный вектор-маска (эталонный образ) нейрона.

На рис. 2 не показан вектор $\mathbf{p}_j=(P_j(H_P), P_j(H_C))$ априорных вероятностей гипотез H_P и H_C нейрона. Далее по умолчанию примем $P_j(H_P)=P_j(H_C)=1/2$.

Тогда условная вероятность свидетельства в предположении истинности гипотезы H_P (*pro* – «за»):

$$P_j(e | H_P) = \frac{\sum_{i=1}^n x_i \cdot w_{j,i}}{\sum_{i=1}^n w_{j,i}}. \quad (1)$$

Условная вероятность свидетельства в предположении истинности гипотезы H_C (*contra* – «против»):

$$P_j(e | H_C) = \frac{\sum_{i=1}^n x_i \cdot (1 - w_{j,i})}{\sum_{i=1}^n (1 - w_{j,i})}. \quad (2)$$

Апостериорные условные вероятности истинности гипотез H_P (*pro*) и H_C (*contra*) с использованием в формулах расчета значений соответствующих априорных вероятностей $P_j(H_P)$ и $P_j(H_C)$:

$$P_j(H_P | e) = \frac{P_j(H_P) \cdot P_j(e | H_P)}{P_j(H_P) \cdot P_j(e | H_P) + P_j(H_C) \cdot P_j(e | H_C)},$$

$$P_j(H_C | e) = \frac{P_j(H_C) \cdot P_j(e | H_C)}{P_j(H_P) \cdot P_j(e | H_P) + P_j(H_C) \cdot P_j(e | H_C)}.$$

Если использовать значения по умолчанию $P_j(H_P)=P_j(H_C)=1/2$, формулы для расчета апостериорных условных вероятностей истинности гипотез H_P (*pro*) и H_C (*contra*) приобретают более изящный вид:

$$P_j(H_P | e) = \frac{P_j(e | H_P)}{P_j(e | H_P) + P_j(e | H_C)}, \quad (3)$$

$$P_j(H_C | e) = \frac{P_j(e | H_C)}{P_j(e | H_P) + P_j(e | H_C)}. \quad (4)$$

Функция активации нейрона BN_j , значение которой используется в качестве выходного сигнала нейрона:

$$f_A = \begin{cases} 1, & \text{если } P_j(H_P | e) \gg P_j(H_C | e), \\ 0, & \text{иначе,} \end{cases} \quad (5)$$

$$i = \overline{1, n}, \quad j = \overline{1, m}.$$

Здесь знак « \gg » нуждается в дополнительной интерпретации. Понятно, что идеальным случаем является сочетание вероятностей $P_j(H_P|e)=1$ и $P_j(H_C|e)=0$. Но можно допустить такую нечеткую трактовку: разность $P_j(H_P|e) - P_j(H_C|e) > \Delta P$, т.е. некоторого порогового значения, близкого к единице.

Далее, проверим работоспособность байесовского нейрона на примере решения задачи распознавания графических примитивов. Как будет показано ниже в п. III, с целью предотвращения возможных коллизий при распознавании полезно уточнить выражение (2). Этот факт отмечен на рис. 2 пунктирной линией.

III. ПРИМЕР РАСПОЗНАВАНИЯ ГРАФИЧЕСКИХ ПРИМИТИВОВ

На рис. 3 представлены шаблон для графических примитивов и эталоны распознаваемых образов. Нанесенному на сетку шаблону примитиву будет соответствовать входной вектор $\mathbf{x}=(x_1, x_2, x_3, x_4, x_5, x_6)$, где двоичный сигнал x_i отражает степень выраженности i -й

линии шаблона (0 – линия не прорисована, 1 – линия четко прорисована). Однослойная нейронная сеть включает четыре выходных нейрона по числу возможных примитивов, которые могут быть нанесены на шаблон, и должна отнести предъявленный образ, описанный вектором x , к одному из четырех эталонных классов изображений.

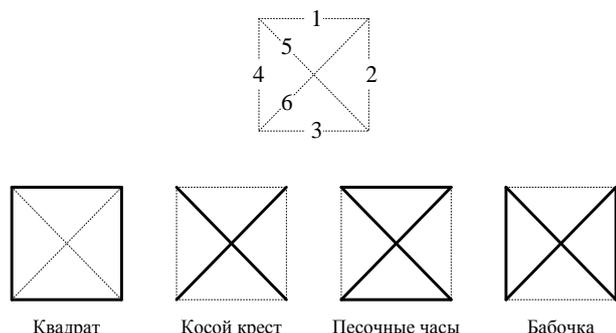


Рис. 3. Шаблон графического примитива и эталоны распознаваемых образов

В табл. I приводятся весовые векторы эталонных образов.

ТАБЛИЦА I ВЕСОВЫЕ ВЕКТОРЫ ЭТАЛОННЫХ ОБРАЗОВ

№ линии, i	Эталонный образ			
	Квадрат	Косой крест	Песочные часы	Бабочка
1	1	0	1	0
2	1	0	0	1
3	1	0	1	0
4	1	0	0	1
5	0	1	1	1
6	0	1	1	1

В табл. II–V представлены итоговые результаты несложных расчетов по формулам (1)–(4) для всех выходных нейронов сети при распознавании ими эталонных образов. Нетрудно убедиться, что нейроны сети с функцией активации (5) на основе формулы Байеса правильно распознают графические примитивы, относящиеся к эталонным образам.

ТАБЛИЦА II РАСПОЗНАВАНИЕ ОБРАЗА «КВАДРАТ»

Вероятности гипотез	Нейроны, которым предъявлен образ «Квадрат»			
	Квадрат	Косой крест	Песочные часы	Бабочка
$P(H_{p e})$	1	0	0,33	0,33
$P(H_{c e})$	0	1	0,67	0,67

ТАБЛИЦА III РАСПОЗНАВАНИЕ ОБРАЗА «КОСОЙ КРЕСТ»

Вероятности гипотез	Нейроны, которым предъявлен образ «Косой крест»			
	Квадрат	Косой крест	Песочные часы	Бабочка
$P(H_{p e})$	0	1	0,67	0,67
$P(H_{c e})$	1	0	0,33	0,33

ТАБЛИЦА IV РАСПОЗНАВАНИЕ ОБРАЗА «ПЕСОЧНЫЕ ЧАСЫ»

Вероятности гипотез	Нейроны, которым предъявлен образ «Песочные часы»			
	Квадрат	Косой крест	Песочные часы	Бабочка
$P(H_{p e})$	0,33	0,67	1	0,67
$P(H_{c e})$	0,67	0,33	0	0,33

ТАБЛИЦА V РАСПОЗНАВАНИЕ ОБРАЗА «БАБОЧКА»

Вероятности гипотез	Нейроны, которым предъявлен образ «Бабочка»			
	Квадрат	Косой крест	Песочные часы	Бабочка
$P(H_{p e})$	0,33	0,67	0,67	1
$P(H_{c e})$	0,67	0,33	0,33	0

IV. ВОЗМОЖНЫЕ КОЛЛИЗИИ ПРИ РАСПОЗНАВАНИИ И УТОЧНЕНИЕ МОДЕЛИ НЕЙРОНА ДЛЯ ИХ УСТРАНЕНИЯ

Рассмотренная выше нейронная сеть не должна допускать распознавания графических примитивов, не относящихся к эталонным образам. Примеры таких примитивов представлены на рис. 4.

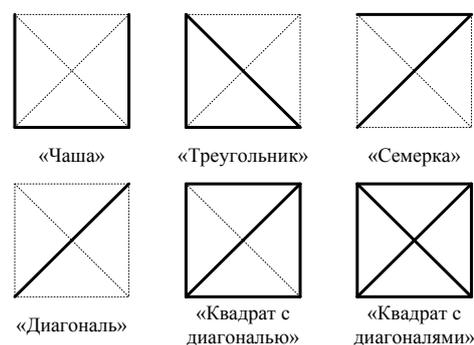


Рис. 4. Примеры графических примитивов, не являющихся эталонными образами

В основном, проблем не возникает, и в случае предъявления сети некорректного образа ни один нейрон не активируется. В качестве примера можно привести попытку распознавания, описанную в табл. VI.

ТАБЛИЦА VI РАСПОЗНАВАНИЕ ОБРАЗА «КВАДРАТ С ДИАГОНАЛЬЮ»

Вероятности гипотез	Нейроны, которым предъявлен образ «Квадрат с диагональю»			
	Квадрат	Косой крест	Песочные часы	Бабочка
$P(H_{p e})$	0,67	0,33	0,43	0,43
$P(H_{c e})$	0,33	0,67	0,57	0,57

В то же время может возникнуть неприятная ситуация, когда компоненты входного вектора x недостаточно свидетельствуют «за», но при этом не свидетельствуют «против». Пример такого распознавания представлен в табл. VII.

ТАБЛИЦА VII РАСПОЗНАВАНИЕ ОБРАЗА «ЧАША»

Вероятности гипотез	Нейроны, которым предъявлен образ «Чаша»			
	Квадрат	Косой крест	Песочные часы	Бабочка
$P(H_P e)$	1	0	0,2	0,5
$P(H_C e)$	0	1	0,8	0,5

По формулам (1) и (2) для некорректного распознаваемого образа «Чаша» на нейроне с весовым вектором эталона «Квадрат» получаем: $P_j(e|H_P) = 3/4$ и $P_j(e|H_C) = 0/2 = 0$, что при расчете по формулам (3) и (4) очевидно дает: $P_j(H_P|e) = 1$ и $P_j(H_C|e) = 0$. На выходе нейрона в соответствии с функцией активации (5) формируется ложный сигнал 1.

Устранить подобную коллизию можно, уточнив формулу (2) и приведя ее к следующему виду:

$$P_j(e|H_C) = \frac{\sum_{i=1}^n x_i \cdot (1 - w_{j,i}) + (1 - [P_j(e|H_P)])}{\sum_{i=1}^n (1 - w_{j,i}) + (1 - [P_j(e|H_P)])}. \quad (6)$$

По сути, в модифицированной формуле (6) с целью устранения возможных коллизий при распознавании образа учитывается следующее важное обстоятельство: при отсутствии свидетельств «против» в виде компонентов входного вектора \mathbf{x} недостаток свидетельств «за» трактуется как дополнительный компонент, свидетельствующий «против» активного состояния нейрона.

V. ЗАКЛЮЧЕНИЕ

В докладе предложена модель нейрона, принцип функционирования которого основан на применении формулы Байеса. При этом входные бинарные сигналы нейрона интерпретируются как свидетельства в пользу двух альтернативных гипотез: соответствия и несоответствия условию активации нейрона. Получены выражения (1) и (6) для вычисления условных вероятностей – оценок соответствия свидетельств предположениям об истинности байесовских гипотез. Выходной сигнал нейрона формируется на основе рассчитанного по формуле Байеса апостериорного распределения вероятностей гипотез (3), (4) и функции активации (5). Эффективность применения новой модели нейрона показана на простом примере распознавания графических примитивов.

В представленном в настоящем сборнике докладе Д. П. Буракова и Г. И. Кожомбердиевой «Применение байесовской модели нейрона в сети WTA для решения задач распознавания» реализуемость предложенной модели демонстрируется на реальном примере распознавания цифр почтового индекса.

СПИСОК ЛИТЕРАТУРЫ

- [1] Николенко С. Три революции нейронных сетей. Инвест-Форсайт, июнь 22, 2018 [Электронный ресурс]. URL: https://www.if24.ru/tri-revolutsii-nejronnyh-setej/?utm_source=fb&utm_medium=cpc&utm_campaign=pod_nejron (дата обращения 30.03.2020)
- [2] Ветров Д. Нейробайесовский подход к задачам машинного обучения. ITnan, 08.02.17 [Электронный ресурс]. URL: <https://itnan.ru/post.php?c=1&p=321434> (дата обращения 30.03.2020)
- [3] Bayesian Deep Learning. NeurIPS 2019 Workshop, December 13, 2019 [Электронный ресурс]. URL: <http://bayesiandeeplearning.org/> (дата обращения 30.03.2020)
- [4] Осовский С. Нейронные сети для обработки информации; [пер. с польск.]. 2-е изд., перераб. и доп. М.: Горячая линия – Телеком, 2018. 448 с.
- [5] Поспелов Д.А. Фантазия или наука: на пути к искусственному интеллекту. М: Наука. Главная редакция физико-математической литературы, 1982. 224 с.
- [6] Кожомбердиева Г.И., Красковский А.Е. Способ определения условных вероятностей при байесовском оценивании качества управленческих решений на железнодорожном транспорте // Интеллектуальные системы на транспорте: матер. IV Междунар. науч.-практич. конф. «ИнтеллектТранс-2014». СПб: Изд-во ПГУПС, 2014. С. 412–418.
- [7] Кожомбердиева Г.И., Бураков Д.П. Об использовании формулы Байеса в задачах оценивания качества // Международная конференция по мягким вычислениям и измерениям. 2017. Т. 1. С. 31–34.
- [8] Кожомбердиева Г.И., Бураков Д.П. Байесовский подход к решению задач оценивания качества // Мягкие измерения и вычисления. 2018. № 5. С. 15–26.
- [9] Кожомбердиева Г.И. Байесовская логико-вероятностная модель нечеткого вывода // Международная конференция по мягким вычислениям и измерениям. 2019. Т. 1. С. 35–38.
- [10] Кожомбердиева Г.И., Бураков Д.П. Байесовская логико-вероятностная модель нечеткого вывода: этапы вывода заключений и дефазификации // Нечеткие системы и мягкие вычисления. 2019. Т. 14. № 2. С. 92–110.
- [11] Кожомбердиева Г.И., Бураков Д.П., Хамчиев Г.А. Использование формулы Байеса при групповом экспертном рейтинговом оценивании // Международная конференция по мягким вычислениям и измерениям. 2019. Т. 1. С. 43–46.
- [12] Кожомбердиева Г.И., Бураков Д.П., Хамчиев Г.А. Реализация группового экспертного рейтингового оценивания на основе формулы Байеса // Мягкие измерения и вычисления. 2019. № 11 (24). С. 4–16.