

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ(12) **ЗАЯВКА НА ИЗОБРЕТЕНИЕ**

(21)(22) Заявка: 2017135085, 20.03.2015

Приоритет(ы):

(22) Дата подачи заявки: 20.03.2015

(43) Дата публикации заявки: 05.04.2019 Бюл. № 10

(85) Дата начала рассмотрения заявки РСТ на национальной фазе: 05.10.2017

(86) Заявка РСТ:
EP 2015/056008 (20.03.2015)(87) Публикация заявки РСТ:
WO 2016/150472 (29.09.2016)

Адрес для переписки:

129090, Москва, ул. Б. Спасская, 25, стр. 3, ООО
"Юридическая фирма Городисский и
Партнеры"

(71) Заявитель(и):

**ФРАУНХОФЕР-ГЕЗЕЛЛЬШАФТ ЦУР
ФЕРДЕРУНГ ДЕР АНГЕВАНДТЕН
ФОРШУНГ Е.Ф. (DE),
ТЕХНИШЕ УНИВЕРЗИТЕТ БЕРЛИН
(DE)**

(72) Автор(ы):

**ЛАПУШКИН Себастьян (DE),
САМЕК Войцех (DE),
МЮЛЛЕР Клаус-Роберт (DE),
БИНДЕР Александер (DE),
МОНТАВОН Грегуар (DE)**(54) **НАЗНАЧЕНИЕ ОЦЕНКИ РЕЛЕВАНТНОСТИ ДЛЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ**

(57) Формула изобретения

1. Устройство для назначения оценки релевантности набору элементов, причем оценка релевантности указывает релевантность в отношении применения искусственной нейронной сети (10), состоящей из нейронов (12), к набору (16) элементов (42), чтобы отображать набор (16) элементов (42) на выход (18) сети, причем устройство сконфигурировано, чтобы

перераспределять начальную оценку (R) релевантности, полученную из выхода (18) сети, на набор (16) элементов (42) путем обратного распространения начальной оценки релевантности через искусственную нейронную сеть (10), чтобы получить оценку релевантности для каждого элемента,

причем устройство сконфигурировано, чтобы выполнять обратное распространение таким образом, что, для каждого нейрона, предварительно перераспределенные оценки релевантности набора нисходящих соседних нейронов соответствующего нейрона распределяются по набору восходящих соседних нейронов соответствующего нейрона с использованием функции распределения.

2. Устройство по п. 1, причем устройство сконфигурировано таким образом, что функция распределения имеет свойство сохранения релевантности.

3. Устройство по п. 1, причем устройство сконфигурировано, чтобы выполнять обратное распространение с одинаковым использованием одной функции распределения

для всех нейронов искусственной нейронной сети.

4. Устройство по п. 1, причем устройство сконфигурировано таким образом, что функция распределения является функцией

весов искусственной нейронной сети, определяющих степень влияния соответствующего нейрона набором восходящих соседних нейронов соответствующего нейрона,

нейронных активаций набора восходящих соседних нейронов, проявляющихся при применении искусственной нейронной сети (10) к набору (16) элементов (42) и

суммы предварительно перераспределенных оценок релевантности набора нисходящих соседних нейронов соответствующего нейрона.

5. Устройство по п. 1, причем устройство сконфигурировано таким образом, что, для каждого нейрона j , функция распределения, обеспечивающая то, насколько релевантность перераспределяется как сообщение R_{ij} релевантности от

соответствующего нейрона j к восходящему соседнему нейрону i , представляет собой

$$R_{ij}=q(i) \cdot m(\{R_{jk}, k \text{ является нисходящим соседним нейроном для } j\}),$$

где $m(R^K)$ при K , являющемся числом нисходящих соседей соответствующего нейрона j , является монотонно возрастающей функцией для всех ее компонентов и дает предварительно перераспределенную оценку релевантности $R_j=m(\{R_{jk}, k \text{ является нисходящим соседним нейроном для } j\})$ соответствующего нейрона j , и $q(i)$ является функцией, зависящей от весов w_{ij} , соединяющих восходящий соседний нейрон i с соответствующим нейроном j , активации x_i восходящего соседнего нейрона i соответствующего нейрона j в результате применения искусственной нейронной сети (10) к набору (16) элементов (42) и, возможно, имеющего нулевое значение члена смещения b_j нейрона j .

6. Устройство по п. 5, в котором $m(\{R_{jk}, k \text{ является нисходящим соседним нейроном для } j\}) = \sum_k R_{jk}$.

7. Устройство по п. 5, в котором устройство сконфигурировано таким образом, что функция $q(i)$ является функцией p взвешенных активаций $z_{ij}=s(x_i, w_{ij}, b_j)$, которые вычисляются посредством функции s , так что $q(i)=p(\{z_{ij}|i \text{ является восходящим соседним нейроном для } j\})$.

8. Устройство по п. 7, в котором функция s выбрана так, что взвешенная активация z_{ij} задается как

$$z_{ij}=x_i w_{ij},$$

$$\text{или } z_{ij}=x_i w_{ij} + \frac{b_j}{I}$$

где I - число восходящих соседних нейронов i нейрона j .

9. Устройство по п. 5, причем устройство сконфигурировано таким образом, что функция $q(i)$ удовлетворяет, для каждого нейрона j , для которого $R_j > 0$, свойству упорядочения,

причем свойство упорядочения удовлетворяется, если

а) если $\sum_i z_{ij} > 0$, то для всех i_1 и i_2 , являющихся восходящими соседними нейронами нейрона j , для которых

$$z_{i_1 j} < z_{i_2 j}$$

справедливо, что $q(i_1) < q(i_2)$

б) или для всех i_1 и i_2 , являющихся восходящими соседними нейронами нейрона j , для которых

$$z_{kj} > 0 \text{ и } z_{ij} > 0 \text{ и } z_{kj} < z_{ij},$$

то справедливо, что $0 \leq q(i_1) \leq q(i_2)$.

10. Устройство по п. 5, причем устройство сконфигурировано таким образом, что функция $q(i)$ удовлетворяет свойству упорядочения,

причем свойство упорядочения удовлетворяется, если для всех i_1 и i_2 , являющихся восходящими соседними нейронами нейрона j , для которых

$$g(z_{i_1 j}) < g(z_{i_2 j})$$

справедливо, что $|q(i_1)| \leq |q(i_2)|$ для функции $g(\cdot)$, которая имеет свой минимум в нуле и которая монотонно убывает на интервале $(-\infty, 0)$ и монотонно возрастает на интервале $(0, +\infty)$.

11. Устройство по п. 10, причем устройство сконфигурировано таким образом, что функция $g(\cdot)$ задается следующим образом:

$$g(z) = \alpha \max(0, z) - \beta \min(0, z) \text{ при } \alpha > 0, \beta \geq 0.$$

12. Устройство по п. 5, причем устройство сконфигурировано таким образом, что функция $q(i)$ наследует или пропорциональна декомпозиции Тейлора функции нейронной сети нейронов.

13. Устройство по п. 5, причем устройство сконфигурировано таким образом, что сообщение R_{ij} релевантности пропорционально декомпозиции Тейлора функции, которая обучается на данных и которая отображает активации x_i восходящих соседей i нейрона j на значение $m(\{R_{jk}, k \text{ является нисходящим нейроном для } j\})$ до погрешности аппроксимации.

14. Устройство по п. 1, причем устройство сконфигурировано таким образом, что функция распределения представляет собой

$$R_{ij} = \frac{x_i w_{ij} + \frac{b_j}{n}}{h\left(\sum_r \left(x_r w_{rj} + \frac{b_j}{n}\right)\right)} \cdot m(\{R_{jk}, k \text{ является нисходящим соседним нейроном для } j\})$$

или

$$R_{ij} = \frac{x_i w_{ij}}{h\left(\sum_r \left(x_r w_{rj} + \frac{b_j}{n}\right)\right)} \cdot m(\{R_{jk}, k \text{ является нисходящим соседним нейроном для } j\})$$

где n - число восходящих соседних нейронов соответствующего нейрона j , R_{ij} - сообщение релевантности, перераспределенное от соответствующего нейрона j к восходящему соседнему нейрону i , и R_{jk} - сообщение релевантности, перераспределенное от нисходящего соседнего нейрона k к соответствующему нейрону j , x_i - активация восходящего соседнего нейрона i во время применения нейронной сети к набору (16) элементов (42), w_{ij} - вес, соединяющий восходящий соседний нейрон i с соответствующим нейроном j , w_{rj} - также вес, соединяющий восходящий соседний нейрон r с соответствующим нейроном j , и b_j - член смещения соответствующего нейрона j , и $h(\cdot)$ является скалярной функцией, при этом $m(R^K)$, где K является числом нисходящих соседей соответствующего нейрона j , является монотонно возрастающей функцией для

всех ее компонентов и дает предварительно перераспределенную оценку релевантности $R_j = m(\{R_{jk}, k \text{ является нисходящим нейроном для } j\})$ соответствующего нейрона j .

15. Устройство по п. 1, причем устройство сконфигурировано таким образом, что распределение на набор восходящих соседних нейронов i соответствующего нейрона j выполняется с использованием функции распределения, причем функция распределения представляет собой

$$R_{ij} = \left(\alpha \frac{\left(x_i w_{ij} + \frac{b_j}{n} \right)_+}{h \left(\sum_r \left(x_r w_{rj} + \frac{b_j}{n} \right)_+ \right)} - \beta \frac{\left(x_i w_{ij} + \frac{b_j}{n} \right)_-}{h \left(\sum_r \left(x_r w_{rj} + \frac{b_j}{n} \right)_- \right)} \right)$$

$\cdot m(\{R_{jk}, k \text{ является нисходящим соседним нейроном для } j\})$

или

$$R_{ij} = \left(\alpha \frac{(x_i w_{ij})_+}{h((b_j)_+ + \sum_r (x_r w_{rj})_+)} - \beta \frac{(x_i w_{ij})_-}{h((b_j)_- + \sum_r (x_r w_{rj})_-)} \right)$$

$\cdot m(\{R_{jk}, k \text{ является нисходящим соседним нейроном для } j\})$,

где $(z)_+ = \max(0, z)$, $(z)_- = \min(0, z)$, n - число восходящих соседних нейронов соответствующего нейрона, R_{ij} - сообщение релевантности, перераспределенное от соответствующего нейрона j к восходящему соседнему нейрону i , и R_{jk} - сообщение релевантности, перераспределенное от нисходящего соседнего нейрона k к соответствующему нейрону j , x_i - активация восходящего соседнего нейрона i во время применения нейронной сети к набору (16) элементов (42), w_{ij} - вес, соединяющий восходящий соседний нейрон i с соответствующим нейроном j , w_{rj} - также вес, соединяющий восходящий соседний нейрон r с соответствующим нейроном j , и b_j - член смещения соответствующего нейрона j , и $h()$ является скалярной функцией, и $\alpha > 0$, $\beta \geq 0$, $\alpha - \beta = 1$ и $m(R^K)$, где K является числом нисходящих соседей соответствующего нейрона j , является монотонно возрастающей функцией для всех ее компонентов и дает предварительно перераспределенную оценку релевантности $R_j = m(\{R_{jk}, k \text{ является нисходящим нейроном для } j\})$ соответствующего нейрона j .

16. Устройство по п. 14, в котором $m(\{R_{jk}, k \text{ является нисходящим нейроном для } j\}) = \sum_k R_{jk}$.

17. Устройство по п. 14, в котором $h()$ является стабилизирующей функцией $h(t) = t + \varepsilon \cdot \text{sign}(t)$.

18. Устройство по п. 1, причем устройство сконфигурировано, чтобы вычислять, для каждого элемента i , оценки R_i релевантности соответствующего элемента i путем суммирования сообщений релевантности нейронов, имеющих соответствующий элемент в качестве восходящего соседнего нейрона, перераспределенных на соответствующий элемент.

19. Устройство по п. 1, причем искусственная нейронная сеть непосредственно применяется к набору элементов, так что элементы набора (16) элементов (42) образуют восходящих соседей для поднабора искусственных нейронов искусственной нейронной сети, и выход сети соответствует нейронной активации нейрона на нисходящем конце искусственной нейронной сети.

20. Устройство по п. 1, причем выход (18) сети представляет собой скаляр с начальной оценкой релевантности, полученной из него, равной значению скаляра или полученной путем применения монотонно возрастающей функции к значению скаляра, или выход сети представляет собой вектор с начальным значением релевантности, равным значению одного или нескольких компонентов вектора, или полученным путем применения монотонно возрастающей функции к значению одного или нескольких компонентов вектора.

21. Устройство по п. 1, причем устройство сконфигурировано для осуществления обратного распространения, так что $0.95 \cdot R \leq f(\sum R_i) \leq 1.05 \cdot R$, где $\sum R_i$ обозначает сумму по оценкам релевантности всех элементов i набора (16) элементов (42), и f является монотонной функцией, зависящей только от $\sum R_i$.

22. Устройство по п. 21, причем устройство сконфигурировано таким образом, что f является функцией тождественности.

23. Устройство по п. 1, причем устройство сконфигурировано таким образом, что для каждого нейрона сумма значений сообщений релевантности, распределенных по набору восходящих соседних нейронов соответствующего нейрона с помощью функции распределения, равна $\xi(S_N)$ или отклоняется от этого не более чем на 5%, причем S_N обозначает сумму сообщений релевантности из набора нисходящих соседних нейронов соответствующего нейрона к соответствующему нейрону, и ξ обозначает монотонную функцию, зависящую только от S_N .

24. Устройство по п. 23, причем устройство сконфигурировано таким образом, что ξ является функцией тождественности.

25. Устройство по п. 1, причем искусственная нейронная сеть выполнена многослойной, так что каждый нейрон (12) принадлежит к одному из последовательности слоев, и устройство сконфигурировано, чтобы выполнять обратное распространение с одинаковым использованием одной функции распределения для всех нейронов искусственной нейронной сети.

26. Устройство по п. 1, причем искусственная нейронная сеть выполнена многослойной, так что каждый нейрон (12) принадлежит одному из последовательности слоев, и устройство сконфигурировано, чтобы выполнять обратное распространение так, что, для каждого уровня, сумма значений сообщений релевантности, распределенных для нейронов соответствующего уровня, равна $\zeta(S_L)$ или отклоняются от этого не более чем на 5%, причем S_L обозначает сумму предварительно перераспределенных оценок релевантности нейронов слоя, нисходящего относительно соответствующего слоя, и ζ обозначает монотонную функцию, зависящую только от S_L .

27. Устройство по п. 1, в котором набор (16) элементов представляет собой комбинацию

изображения с каждым из элементов (42) набора (16) элементов (42), соответствующих одному или нескольким пикселям или субпикселям изображения, и/или

видео с каждым из элементов (42) набора (16) элементов (42), соответствующих одному или нескольким пикселям или субпикселям изображений видео, изображениям видео или последовательностям изображений видео, и/или

аудиосигнала с каждым элементом (42) набора (16) элементов (42), соответствующих одной или нескольким аудиовыборкам аудиосигнала, и/или

карты признаков локальных признаков или преобразования, локально или глобально извлеченных из изображения, видео или аудиосигнала с элементами (42) набора (16) элементов (42), соответствующих локальным признакам, и/или

текста с элементами (42) набора (16) элементов (42), соответствующих словам,

предложениям или абзацам текста, и/или

графа, такого как граф отношений социальных сетей, с элементами (42) набора (16) элементов (42), соответствующих узлам или ребрам или наборам узлов или набору ребер или подграфам.

28. Система (100) для обработки данных, содержащая устройство (50) для назначения оценки релевантности набору элементов в соответствии с любым из предыдущих пунктов, и

устройство (102) для обработки набора (16) элементов или данных, подлежащих обработке (106) и полученных из набора элементов с адаптацией обработки в зависимости от оценок релевантности.

29. Система по п. 28, в которой обработка представляет собой обработку с потерями, и устройство для обработки сконфигурировано для уменьшения потери обработки с потерями для элементов, имеющих более высокие оценки релевантности, назначенные им, по сравнению с элементами, имеющими более низкие оценки релевантности, назначенные им.

30. Система по п. 28, в которой обработка представляет собой визуализацию, причем устройство для адаптации сконфигурировано, чтобы выполнять выделение в визуализации в зависимости от оценок релевантности.

31. Система по п. 28, в которой обработка представляет собой пополнение данных путем считывания из памяти или выполнение дополнительного измерения, причем устройство (102) для обработки сконфигурировано, чтобы фокусировать пополнение данных в зависимости от оценок релевантности.

32. Система (110) для выделения области, представляющей интерес, содержащая устройство (50) для назначения оценки релевантности набору элементов по п. 1 и устройство (112) для генерирования графа (114) релевантности в зависимости от оценок релевантности.

33. Система (120) для оптимизации нейронной сети, содержащая устройство (50) для назначения оценки релевантности набору элементов по п.1; устройство (122) для применения устройства для назначения к множеству различных наборов элементов; и

устройство (124) для обнаружения части повышенной релевантности (128) в нейронной сети путем накопления релевантностей, назначенных нейронам сети во время применения устройства для назначения к множеству различных наборов элементов, и оптимизации искусственной нейронной сети в зависимости от части повышенной релевантности.

34. Способ назначения оценки релевантности набору элементов, причем оценка релевантности указывает релевантность в отношении применения искусственной нейронной сети (10), состоящей из нейронов (12), к набору (16) элементов (42), чтобы отображать набор (16) элементов (42) на выход (18) сети, причем устройство сконфигурировано, чтобы

перераспределять начальную оценку (R) релевантности, полученную из выхода (18) сети, на набор (16) элементов (42) путем обратного распространения начальной оценки релевантности через искусственную нейронную сеть (10), чтобы получить оценку релевантности для каждого элемента,

причем обратное распространение выполняется таким образом, что, для каждого нейрона, предварительно перераспределенные оценки релевантности набора нисходящих соседних нейронов соответствующего нейрона распределяются по набору восходящих соседних нейронов соответствующего нейрона с использованием функции распределения.

35. Компьютерная программа, имеющая программный код для выполнения, при исполнении на компьютере, способа по п.34.