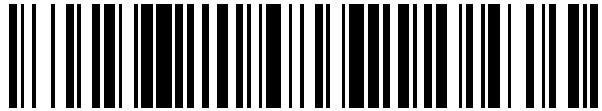


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 883 376**

51 Int. Cl.:

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **31.08.2015 PCT/JP2015/074720**

87 Fecha y número de publicación internacional: **08.12.2016 WO16194248**

96 Fecha de presentación y número de la solicitud europea: **31.08.2015 E 15894280 (5)**

97 Fecha y número de publicación de la concesión europea: **21.07.2021 EP 3306534**

54 Título: **Dispositivo de inferencia y método de inferencia**

30 Prioridad:

03.06.2015 JP 2015113440

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

07.12.2021

73 Titular/es:

**MITSUBISHI ELECTRIC CORPORATION (100.0%)
7-3, Marunouchi 2-chome, Chiyoda-ku
Tokyo 100-8310, JP**

72 Inventor/es:

**MATSUMOTO, WATARU;
YOSHIMURA, GENTA y
ZHAO, XIONGXIN**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 883 376 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Dispositivo de inferencia y método de inferencia

5 Campo técnico

La presente invención se refiere a un dispositivo de inferencia y a un método de inferencia que usa una red neuronal.

10 Técnica anterior

Como uno de los métodos de aprendizaje automático, las redes neuronales tienen altas capacidades en cuanto a la resolución de problemas y se usan en muchos tipos de procedimientos, tales como reconocimiento de imágenes, reconocimiento de sonidos, detección de anomalías y predicción de futuro.

15 Se conoce una red neuronal jerárquica como una de las estructuras de redes neuronales. Los métodos de aprendizaje de la misma incluyen principalmente dos tipos, que son aprendizaje supervisado y aprendizaje no supervisado.

20 El aprendizaje supervisado es un método de ajuste de un estado de conexión de una red neuronal de tal manera que la salida real y la salida objetivo coinciden entre sí después de facilitarse datos de entrada de una pluralidad de ejemplos de entrenamiento y la salida objetivo. Por otro lado, el aprendizaje no supervisado es un método de ajuste de un estado de conexión de una red neuronal para extraer una característica esencial que se presenta en ejemplos de entrenamiento sin proporcionar una salida objetivo.

25 Un método de retropropagación de errores (un algoritmo de retropropagación) que pertenece a métodos de aprendizaje supervisado puede provocar un problema de que los resultados de aprendizaje no convergen cuando aumenta el número de capas en una red neuronal.

30 Con el fin de resolver el problema anterior, hay un método de realizar entrenamiento previo en cada capa por medio de aprendizaje no supervisado, tal como el de un autocodificador o una máquina de Boltzmann restringida, para determinar un valor por defecto de un estado de conexión de una red neuronal. Después de eso, se realiza un ajuste fino del estado de conexión de la red neuronal usando un método de retropropagación de errores.

35 Este método puede ajustar el estado de conexión de la red neuronal de tal manera que la salida real y la salida objetivo coinciden entre sí sin provocar el problema de que los resultados de aprendizaje no converjan.

40 Una red neuronal jerárquica puede representarse mediante una estructura de gráficos que está constituida por una pluralidad de nodos (o puntos de unión) y bordes (o ramificaciones) que conectan entre los nodos. En una red neuronal que tiene cuatro capas, por ejemplo, una pluralidad de nodos están dispuestos en una capa de entrada, una primera capa intermedia, una segunda capa intermedia y una capa de salida. En este caso, no existe ningún borde entre nodos que pertenecen a la misma capa, mientras que solo existen bordes entre cada par de capas adyacentes. Obsérvese que las capas intermedias pueden denominarse "capas ocultas".

45 Cada uno de los bordes tiene un parámetro que indica un grado de conexión entre dos nodos conectados. Este parámetro se denomina "peso de borde".

50 Cuando se realiza aprendizaje o inferencia usando la red neuronal jerárquica, la cantidad de cálculo y la cantidad de memoria son proporcionales al número de bordes. En general, los nodos que pertenecen a cada una de las capas se conectan mediante bordes con todos los nodos que pertenecen a una capa adyacente a la misma. Por tanto, la cantidad de cálculo y la cantidad de memoria están directamente relacionadas con el número de nodos.

55 Por ejemplo, suponiendo que el número de nodos en la capa de entrada es N , el número de nodos en la primera capa intermedia es M_1 , el número de nodos en la segunda capa intermedia es M_2 y el número de nodos en la capa de salida es 1 , el número de bordes entre la capa de entrada y la primera capa intermedia es $N \times M_1$, el número de bordes entre la primera capa intermedia y la segunda capa intermedia es $M_1 \times M_2$ y el número de bordes entre la segunda capa intermedia y la capa de salida es M_2 . Por tanto, la cantidad de cálculo y la cantidad de memoria al aprender y realizar la inferencia son proporcionales a $(N \times M_1 + M_1 \times M_2 + M_2)$.

60 Especialmente, cuando el número de nodos en las capas intermedias es proporcional al número de nodos en la capa de entrada, el número de nodos en la capa de entrada es N , el número de nodos en la primera capa intermedia es $M_1 = a \times N$ y el número de nodos en la segunda capa intermedia es $M_2 = b \times N$. En este caso, el número total de bordes en la red neuronal es $N \times a \times N + a \times N \times b \times N + b \times N = (a + a \times b) \times N^2 + b \times N$. Por tanto, la cantidad de cálculo y la cantidad de memoria al aprender y realizar la inferencia son proporcionales a $(a + a \times b) \times N^2 + b \times N$.

65

Las redes neuronales jerárquicas tienen con frecuencia la configuración tal como se describió anteriormente y, por tanto, la cantidad de cálculo y la cantidad de memoria aumentan en proporción al cuadrado de N que es la cantidad de datos de entrada, es decir, el cuadrado de N que es el número de nodos en la capa de entrada. Por tanto, la cantidad de cálculo y la cantidad de memoria aumentan drásticamente a medida que aumenta la cantidad de datos de entrada, dando, por tanto, como resultado problemas tales como escasez de recursos informáticos, retraso de un procedimiento o aumento del coste de un dispositivo.

En el siguiente documento de patente 1, el número de bordes entre una capa de entrada y una capa intermedia o el número de bordes entre la capa intermedia y una capa de salida se reducen agrupando una pluralidad de conjuntos de datos de entrada basándose en la correlación entre la pluralidad de conjuntos de datos.

El documento EP 0 528 399 A1 describe un aparato para llevar a cabo una operación de aprendizaje de una red neuronal que tiene una capa de entrada, una capa intermedia y una capa de salida. Múltiples nodos de la capa de entrada están relacionados con múltiples nodos de la capa intermedia con múltiples pesos de conexión, mientras que los múltiples nodos de la capa intermedia también están relacionados con múltiples nodos de la capa de salida con múltiples pesos de conexión. En el mismo, durante un cálculo de error entre datos de salida a partir de la red neuronal y los datos de enseñanza, el aparato evalúa si cada uno de los componentes de los datos de enseñanza es eficaz o ineficaz, y errores de salida correspondientes a los componentes ineficaces se consideran como cero. Posteriormente se corrigen los pesos de conexión basándose en los errores calculados. Los niveles de actividad de las capas posteriores se calculan sumando las actividades de todos los nodos de la capa anterior.

Lista de referencias

Documento de patente 1: publicación japonesa n.º 2011-54200 A (figura 1)

Sumario de la invención

Dado que dispositivos de inferencia en la técnica convencional están configurados de la manera anterior, puede reducirse el número de bordes entre una capa de entrada y una capa intermedia o el número de bordes entre la capa intermedia y una capa de salida. Sin embargo, entre la capa de entrada y la capa intermedia que pertenecen al mismo grupo, cada nodo en la capa de entrada se conecta a todos los nodos en la capa intermedia y, por tanto, se limita la reducción del número de bordes. Por tanto, puede provocarse un problema de que la cantidad de cálculo y la cantidad de memoria con la inferencia todavía sean grandes.

La presente invención se ha concebido para resolver el problema anteriormente mencionado con el objetivo de obtener un método de inferencia y un dispositivo de inferencia que puedan reducir la cantidad de cálculo y la cantidad de memoria al realizar la inferencia. Un objetivo adicional es obtener un dispositivo de inferencia y un método de inferencia que tengan una alta precisión de inferencia.

Un dispositivo de inferencia según la invención incluye las características de las reivindicaciones de dispositivo independientes. Un método para clasificación de imágenes según la invención comprende las características de las reivindicaciones de método independientes.

Esta invención tiene el efecto de reducir la cantidad de cálculo y la cantidad de memoria al realizar la clasificación de imágenes. Otro efecto es que puede obtenerse una precisión de clasificación superior.

Breve descripción de los dibujos

La figura 1 es un diagrama que ilustra una configuración de un dispositivo de inferencia según la realización 1 de la invención.

La figura 2 es un diagrama que ilustra una configuración de hardware del dispositivo de inferencia según la realización 1 de la invención.

La figura 3 es un diagrama que ilustra una configuración de hardware cuando el dispositivo de inferencia está constituido por un ordenador.

La figura 4 es un diagrama de flujo que ilustra un método de inferencia que se refiere a contenido de procedimiento del dispositivo de inferencia según la realización 1 de la invención.

La figura 5 es un diagrama de flujo que ilustra contenido de procedimiento de un calculador de nivel de actividad de capa de entrada 1, un calculador de nivel de actividad de primera capa intermedia 5, un calculador de nivel de actividad de segunda capa intermedia 6, un calculador de nivel de actividad de tercera capa intermedia 7 y un calculador de nivel de actividad de capa de salida 9.

La figura 6 es un diagrama explicativo que ilustra una red neuronal jerárquica aplicada al dispositivo de inferencia según la realización 1 de la invención.

La figura 7 es un diagrama explicativo que ilustra un ejemplo de índices, peso de borde y valores de sesgo.

La figura 8 es un diagrama explicativo que ilustra ejemplos de un bucle formado por una pluralidad de bordes.

La figura 9 es un diagrama explicativo que ilustra una red neuronal a modo de ejemplo en la que datos de imagen facilitados a una capa de entrada se clasifican en diez tipos.

La figura 10 es un diagrama explicativo que ilustra una red de Elman como red neuronal recurrente (RNN) que tiene conexiones de borde desde una capa de entrada hasta una primera capa intermedia, desde la primera capa intermedia hasta una segunda capa intermedia (una capa de contexto), desde la segunda capa intermedia (una capa de contexto) hasta la primera capa intermedia y desde la primera capa intermedia hasta una capa de salida.

La figura 11 es un diagrama que ilustra una configuración de un dispositivo de inferencia según la realización 7 de la invención.

La figura 12 es un diagrama explicativo que ilustra una red de estado de eco a modo de ejemplo que es una red neuronal que tiene conexiones de borde entre nodos o una autoconexión dentro de la capa intermedia y que también tiene un borde que conecta una capa de entrada con una capa de salida omitiendo una capa intermedia.

La figura 13 es un diagrama explicativo que ilustra la red de estado de eco en la figura 12 mediante la capa.

La figura 14 es un diagrama que ilustra una configuración de un dispositivo de inferencia según una realización 8 de la invención.

Descripción de realizaciones

Para describir la invención en más detalle, a continuación se describirán realizaciones para llevar a cabo la invención junto con los dibujos adjuntos.

(Realización 1)

La figura 1 es un diagrama que ilustra una configuración de un dispositivo de inferencia según la realización 1 de la invención. La figura 2 es un diagrama que ilustra una configuración de hardware del dispositivo de inferencia según la realización 1 de la invención.

En la figura 1, se ilustra el dispositivo de inferencia a modo de ejemplo que usa una red neuronal jerárquica que tiene cinco capas, en la que una pluralidad de nodos están dispuestos en una capa de entrada, una primera capa intermedia, una segunda capa intermedia, una tercera capa intermedia y una capa de salida. En la figura 1, se supone que datos facilitados a la capa de entrada son datos de imagen.

En esta realización, aunque se ilustra el ejemplo de usar la red neuronal jerárquica de las cinco capas, no se limita a la red neuronal jerárquica de cinco capas. Puede usarse una red neuronal jerárquica de tres o cuatro capas o de seis o más capas.

Obsérvese que, en el caso de usar una red neuronal jerárquica de tres capas, solo se incluye una primera capa intermedia como capa intermedia. En este caso, no se necesitan un almacenamiento de segunda capa intermedia 3, un almacenamiento de tercera capa intermedia 4, un calculador de nivel de actividad de segunda capa intermedia 6 y un calculador de nivel de actividad de tercera capa intermedia 7, que se describirán más adelante.

En las figuras 1 y 2, se implementa un calculador de nivel de actividad de capa de entrada 1 mediante un conjunto de circuitos de nivel de actividad de capa de entrada 11 que está constituido por un circuito integrado de semiconductor o un microordenador de un chip equipado con una unidad de procesamiento central (CPU), por ejemplo. Cuando se facilitan los datos de imagen a cada nodo en la capa de entrada que constituye la red neuronal jerárquica, el calculador de nivel de actividad de capa de entrada 1 realiza un procedimiento de cálculo de nivel de actividad de cada nodo en la capa de entrada a partir de los datos de imagen.

Un almacenamiento de primera capa intermedia 2 como almacenamiento de capa intermedia se implementa mediante un dispositivo de almacenamiento de capa intermedia 12 que está compuesto por un medio de almacenamiento, tal como una RAM o un disco duro. El almacenamiento de primera capa intermedia 2 almacena un índice (o información de conexión) que representa una relación de conexión entre cada nodo en la primera capa intermedia y cada nodo en la capa de entrada, peso aplicado a cada uno de los bordes que conectan un nodo en la primera capa intermedia y un nodo en la capa de entrada, y un valor de sesgo dado a cada nodo en la

primera capa intermedia.

5 Un almacenamiento de segunda capa intermedia 3 como almacenamiento de capa intermedia se implementa mediante el dispositivo de almacenamiento de capa intermedia 12 que está compuesto por un medio de almacenamiento, tal como una RAM o un disco duro. El almacenamiento de segunda capa intermedia 3 almacena un índice que representa una relación de conexión entre cada nodo en la segunda capa intermedia y cada nodo en la primera capa intermedia, peso aplicado a cada uno de los bordes que conectan un nodo en la segunda capa intermedia y un nodo en la primera capa intermedia, y un valor de sesgo dado a cada nodo en la segunda capa intermedia.

10 Un almacenamiento de tercera capa intermedia 4 como almacenamiento de capa intermedia se implementa mediante el dispositivo de almacenamiento de capa intermedia 12 que está compuesto por un medio de almacenamiento, tal como una RAM o un disco duro. El almacenamiento de tercera capa intermedia 4 almacena un índice que representa una relación de conexión entre cada nodo en la tercera capa intermedia y cada nodo en la segunda capa intermedia, peso aplicado a cada uno de los bordes que conectan un nodo en la tercera capa intermedia y un nodo en la segunda capa intermedia, y un valor de sesgo dado a cada nodo en la tercera capa intermedia.

20 Un calculador de nivel de actividad de primera capa intermedia 5 como calculador de nivel de actividad de capa intermedia se implementa mediante un conjunto de circuitos de nivel de actividad de capa intermedia 13 que está constituido por un circuito integrado de semiconductor o un microordenador de un chip equipado con una unidad de procesamiento central (CPU), por ejemplo. El calculador de nivel de actividad de primera capa intermedia 5 consulta el índice almacenado en el almacenamiento de primera capa intermedia 2 y adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada 1 y el peso para los bordes respectivos y los valores de sesgo almacenados en el almacenamiento de primera capa intermedia 2, un nivel de actividad de cada nodo en la capa de entrada que se conecta con cada nodo en la primera capa intermedia, peso para cada uno de los bordes y un valor de sesgo. El calculador de nivel de actividad de primera capa intermedia 5 calcula un nivel de actividad de cada nodo en la primera capa intermedia usando el nivel de actividad adquirido de cada nodo en la capa de entrada, el peso adquirido para cada uno de los bordes y el valor de sesgo adquirido.

35 Un calculador de nivel de actividad de segunda capa intermedia 6 como calculador de nivel de actividad de capa intermedia se implementa mediante el conjunto de circuitos de nivel de actividad de capa intermedia 13 que está constituido por un circuito integrado de semiconductor o un microordenador de un chip equipado con una unidad de procesamiento central (CPU), por ejemplo. El calculador de nivel de actividad de segunda capa intermedia 6 consulta el índice almacenado en el almacenamiento de segunda capa intermedia 3 y adquiere, a partir de los niveles de actividad de los nodos respectivos en la primera capa intermedia calculados mediante el calculador de nivel de actividad de primera capa intermedia 5 y el peso para los bordes respectivos y los valores de sesgo almacenados en el almacenamiento de segunda capa intermedia 3, un nivel de actividad de cada nodo en la primera capa intermedia que se conecta con cada nodo en la segunda capa intermedia, peso para cada uno de los bordes y un valor de sesgo. El calculador de nivel de actividad de segunda capa intermedia 6 calcula un nivel de actividad de cada nodo en la segunda capa intermedia usando el nivel de actividad adquirido de cada nodo en la primera capa intermedia, el peso adquirido para cada uno de los bordes y el valor de sesgo adquirido.

45 Un calculador de nivel de actividad de tercera capa intermedia 7 que es el calculador de nivel de actividad de capa intermedia se implementa mediante el conjunto de circuitos de nivel de actividad de capa intermedia 13 que está constituido por un circuito integrado de semiconductor o un microordenador de un chip equipado con una unidad de procesamiento central (CPU), por ejemplo. El calculador de nivel de actividad de tercera capa intermedia 7 consulta el índice almacenado en el almacenamiento de tercera capa intermedia 4 y adquiere, a partir de los niveles de actividad de los nodos respectivos en la segunda capa intermedia calculados mediante el calculador de nivel de actividad de segunda capa intermedia 6 y el peso para los bordes respectivos y los valores de sesgo almacenados en el almacenamiento de tercera capa intermedia 4, un nivel de actividad de cada nodo en la segunda capa intermedia que se conecta con cada nodo en la tercera capa intermedia, peso para cada uno de los bordes y un valor de sesgo. El calculador de nivel de actividad de tercera capa intermedia 7 calcula un nivel de actividad de cada nodo en la tercera capa intermedia usando el nivel de actividad adquirido de cada nodo en la segunda capa intermedia, el peso adquirido para cada uno de los bordes y el valor de sesgo adquirido.

60 Un almacenamiento de capa de salida 8 se implementa mediante un dispositivo de almacenamiento de capa de salida 14 que está compuesto por un medio de almacenamiento, tal como una RAM o un disco duro. El almacenamiento de capa de salida 8 almacena un índice (o información de conexión) que representa una relación de conexión entre cada nodo en la capa de salida y cada nodo en la tercera capa intermedia, peso para cada uno de los bordes que conectan un nodo en la capa de salida y un nodo en la tercera capa intermedia, y un valor de sesgo dado a cada nodo en la capa de salida.

65 Un calculador de nivel de actividad de capa de salida 9 se implementa mediante un conjunto de circuitos de nivel

de actividad de capa de salida 15 que está constituido por un circuito integrado de semiconductor o un microordenador de un chip equipado con una unidad de procesamiento central (CPU), por ejemplo. El calculador de nivel de actividad de capa de salida 9 consulta el índice almacenado en el almacenamiento de capa de salida 8 y adquiere, a partir de los niveles de actividad de los nodos respectivos en la tercera capa intermedia calculados mediante el calculador de nivel de actividad de tercera capa intermedia 7 y el peso para los bordes respectivos y los valores de sesgo almacenados en el almacenamiento de capa de salida 8, un nivel de actividad de cada nodo en la tercera capa intermedia que se conecta con cada nodo en la capa de salida, peso para cada uno de los bordes y un valor de sesgo. El calculador de nivel de actividad de capa de salida 9 calcula un nivel de actividad de cada nodo en la capa de salida usando el nivel de actividad adquirido de cada nodo en la tercera capa intermedia, el peso adquirido para cada uno de los bordes y el valor de sesgo adquirido.

En la figura 1, se supone que el calculador de nivel de actividad de capa de entrada 1, el almacenamiento de primera capa intermedia 2, el almacenamiento de segunda capa intermedia 3, el almacenamiento de tercera capa intermedia 4, el calculador de nivel de actividad de primera capa intermedia 5, el calculador de nivel de actividad de segunda capa intermedia 6, el calculador de nivel de actividad de tercera capa intermedia 7, el almacenamiento de capa de salida 8 y el calculador de nivel de actividad de capa de salida 9, cada uno de los cuales son componentes del dispositivo de inferencia, están constituidos de manera individual por hardware dedicado. Alternativamente, el dispositivo de inferencia puede estar constituido por un ordenador.

La figura 3 es un diagrama que ilustra una configuración de hardware cuando el dispositivo de inferencia está constituido por un ordenador.

Cuando el dispositivo de inferencia está constituido por un ordenador, el almacenamiento de primera capa intermedia 2, el almacenamiento de segunda capa intermedia 3, el almacenamiento de tercera capa intermedia 4 y el almacenamiento de capa de salida 8 se proporcionan en una memoria 21 del ordenador. Además, un programa, que describe contenido de procedimiento del calculador de nivel de actividad de capa de entrada 1, el calculador de nivel de actividad de primera capa intermedia 5, el calculador de nivel de actividad de segunda capa intermedia 6, el calculador de nivel de actividad de tercera capa intermedia 7 y el calculador de nivel de actividad de capa de salida 9, está almacenado en la memoria 21 del ordenador. Además, un procesador 22 del ordenador ejecuta el programa almacenado en la memoria 21.

La figura 4 es un diagrama de flujo que ilustra un método de inferencia que es contenido de procedimiento del dispositivo de inferencia según la realización 1 de la invención. La figura 5 es un diagrama de flujo que ilustra contenido de procedimiento del calculador de nivel de actividad de capa de entrada 1, el calculador de nivel de actividad de primera capa intermedia 5, el calculador de nivel de actividad de segunda capa intermedia 6, el calculador de nivel de actividad de tercera capa intermedia 7 y el calculador de nivel de actividad de capa de salida 9.

La figura 6 es un diagrama explicativo que ilustra la red neuronal jerárquica aplicada al dispositivo de inferencia según la realización 1 de la invención.

La red neuronal jerárquica en la figura 6 tiene cinco capas en las que una pluralidad de nodos están dispuestos en una capa de entrada, una primera capa intermedia, una segunda capa intermedia, una tercera capa intermedia y una capa de salida.

La figura 7 es un diagrama explicativo que ilustra un ejemplo de índices, peso de borde y valores de sesgo.

En la figura 7, el índice como información de conexión de un nodo indica que, por ejemplo, el nodo "N" en una primera capa intermedia se conecta con los nodos "0", "3" y "5" en una capa de entrada.

La figura 7 ilustra que, por ejemplo, se aplica un peso de "0,2" a un borde que conecta el nodo "N" en la primera capa intermedia y el nodo "0" en la capa de entrada, se aplica un peso de "-0,5" a un borde que conecta el nodo "N" en la primera capa intermedia y el nodo "3" en la capa de entrada, y se aplica un peso de "0,1" a un borde que conecta el nodo "N" en la primera capa intermedia y el nodo "5" en la capa de entrada.

La figura 7 ilustra además que, por ejemplo, se aplica un valor de sesgo de "1,8" al nodo "N" en la primera intermedia.

A continuación, se describirán operaciones.

Cuando se facilitan datos de imagen a cada nodo en la capa de entrada que constituye la red neuronal jerárquica, el calculador de nivel de actividad de capa de entrada 1 calcula, a partir de los datos de imagen, un nivel de actividad A_{IN} de cada nodo en la capa de entrada (etapa ST1 en la figura 4).

En un caso en el que los datos de imagen facilitados al calculador de nivel de actividad de capa de entrada 1 representan, por ejemplo, una imagen compuesta por píxeles que tienen un valor de píxel P que oscila entre 0 y

255, y el valor de píxel P de cada uno de los píxeles se proporciona a cada nodo en la capa de entrada, puede calcularse un nivel de actividad A_{IN} de cada nodo en la capa de entrada mediante la siguiente fórmula (1).

$$A_{IN} = \frac{P}{255} \quad (1)$$

5

En este ejemplo, se supone que se introducen datos de imagen y que se realiza normalización dividiendo un valor de píxel P de cada píxel entre 255, y se obtiene un valor de coma flotante (de 0,0 a 1,0) como nivel de actividad A_{IN} de cada nodo en la capa de entrada. Alternativamente, puede realizarse un procedimiento, tal como reducción de datos, cuantificación, o conversión, dependiendo del tipo de datos de entrada además de una simple normalización.

10

Después de que el calculador de nivel de actividad de capa de entrada 1 calcule el nivel de actividad A_{IN} de cada nodo en la capa de entrada, el calculador de nivel de actividad de primera capa intermedia 5 consulta el índice almacenado en el almacenamiento de primera capa intermedia 2, confirma cada nodo en la capa de entrada conectado con un nodo en la primera capa intermedia para cada nodo en la primera capa intermedia, y adquiere el nivel de actividad A_{IN} de cada nodo en la capa de entrada.

15

Específicamente, por ejemplo, en el caso del nodo "N" en la primera capa intermedia, el índice almacenado en el almacenamiento de primera capa intermedia 2 representa la conexión a los nodos "0", "3" y "5" en la capa de entrada. Por tanto, los niveles de actividad A_{IN-0} , A_{IN-3} y A_{IN-5} de los nodos "0", "3" y "5" en la capa de entrada se adquieren a partir de los niveles de actividad A_{IN} de los nodos respectivos en la capa de entrada que se han calculado mediante el calculador de nivel de actividad de capa de entrada 1.

20

El calculador de nivel de actividad de primera capa intermedia 5 consulta además el índice almacenado en el almacenamiento de primera capa intermedia 2, confirma un borde conectado con un nodo en la primera capa intermedia para cada nodo en la primera capa intermedia y adquiere el peso w para el borde a partir del almacenamiento de primera capa intermedia 2.

25

Específicamente, por ejemplo, en el caso del nodo "N" en la primera capa intermedia, el índice almacenado en el almacenamiento de primera capa intermedia 2 representa la conexión a los nodos "0", "3" y "5" en la capa de entrada. Por tanto, se adquiere un valor de "0,2" como peso w_{N-0} para el borde que conecta el nodo "N" en la primera capa intermedia y el nodo "0" en la capa de entrada, y se adquiere un valor de "-0,5" como peso w_{N-3} para el borde que conecta el nodo "N" en la primera capa intermedia y el nodo "3" en la capa de entrada. Además, se adquiere adicionalmente un valor de "0,1" como peso w_{N-5} para el borde que conecta el nodo "N" en la primera capa intermedia y el nodo "5" en la capa de entrada.

30

35

El calculador de nivel de actividad de primera capa intermedia 5 adquiere además, para cada nodo en la primera capa intermedia, un valor de sesgo B_{1M} de un nodo en la primera capa intermedia a partir del almacenamiento de primera capa intermedia 2.

40

Por ejemplo, en el caso del nodo "N" en la primera capa intermedia, se adquiere un valor de "1,8" como valor de sesgo B_{1M-N} .

Después de adquirir, para cada nodo en la primera capa intermedia, el nivel de actividad A_{IN} de cada nodo en la capa de entrada, el peso w y el valor de sesgo B_{1M} , el calculador de nivel de actividad de primera capa intermedia 5 calcula un nivel de actividad A_{1M} para cada nodo en la primera capa intermedia usando el nivel de actividad A_{IN} , el peso w y el valor de sesgo B_{1M} adquiridos (etapa ST2).

45

A continuación se describirá específicamente un cálculo a modo de ejemplo de un nivel de actividad A_{1M-N} del nodo "N" en la primera capa intermedia.

50

En primer lugar, el calculador de nivel de actividad de primera capa intermedia 5 lee el índice almacenado en el almacenamiento de primera capa intermedia 2 (etapa ST11 en la figura 5). Consultando el índice, el calculador de nivel de actividad de primera capa intermedia 5 adquiere, como parámetro usado para el cálculo de un nivel de actividad, los niveles de actividad A_{IN-0} , A_{IN-3} y A_{IN-5} en los nodos "0", "3" y "5" en la capa de entrada, el peso w_{N-0} , w_{N-3} y w_{N-5} para los bordes de los mismos, y el valor de sesgo B_{1M-N} del nodo "N" en la primera capa intermedia (etapa ST12).

55

El calculador de nivel de actividad de primera capa intermedia 5 realiza una operación de suma de productos, tal como se representa mediante la siguiente fórmula (2), entre los niveles de actividad A_{IN-0} , A_{IN-3} y A_{IN-5} de los nodos "0", "3" y "5" en la capa de entrada, y el peso w_{N-0} , w_{N-3} y w_{N-5} para los bordes (etapa ST13).

60

$$MADD = A_{IN-0} \times w_{N-0} + A_{IN-3} \times w_{N-3} + A_{IN-5} \times w_{N-5} \quad (2)$$

Después, el calculador de nivel de actividad de primera capa intermedia 5 añade el valor de sesgo B_{1M-N} del nodo

65

“N” en la primera capa intermedia al resultado de operación MADD de la operación de suma de productos, tal como se representa mediante la siguiente fórmula (3) (etapa ST14).

$$ADD = MADD + B_{1M-N} \quad (3)$$

5

Como función de activación F aplicada a la red neuronal, que se usa para el cálculo de un nivel de actividad, se prepara por adelantada una función lineal, una función sigmoidea, una función softmax, una función de unidad lineal rectificadora (ReLU) u otras funciones, para el calculador de nivel de actividad de primera capa intermedia 5. El calculador de nivel de actividad de primera capa intermedia 5 calcula, como nivel de actividad A_{1M-N} del nodo “N” en la primera capa intermedia, un valor resultante de la función de activación F con el resultado de adición ADD de la fórmula (3) como argumento de la función de activación F, tal como se representa mediante la siguiente fórmula (4) (etapa ST15).

10

$$A_{1M-N} = F(ADD) \quad (4)$$

15

En este caso se ha descrito el cálculo a modo de ejemplo del nivel de actividad A_{1M-N} del nodo “N” en la primera capa intermedia. Los niveles de actividad A_{1M} de otros nodos en la primera capa intermedia también se calculan de una manera similar.

20

Después de que el calculador de nivel de actividad de primera capa intermedia 5 haya calculado los niveles de actividad A_{1M} de los nodos respectivos en la primera capa intermedia, el calculador de nivel de actividad de segunda capa intermedia 6 calcula un nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia (etapa ST3 en la figura 4).

25

Un método de cálculo del nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia mediante el calculador de nivel de actividad de segunda capa intermedia 6 es similar al método de cálculo del nivel de actividad A_{1M} de cada nodo en la primera capa intermedia mediante el calculador de nivel de actividad de primera capa intermedia 5.

30

Es decir, el calculador de nivel de actividad de segunda capa intermedia 6 consulta el índice almacenado en el almacenamiento de segunda capa intermedia 3, confirma cada nodo en la primera capa intermedia conectado con un nodo en la segunda capa intermedia para cada nodo en la segunda capa intermedia, y adquiere el nivel de actividad A_{1M} de cada nodo en la primera capa intermedia.

35

El calculador de nivel de actividad de segunda capa intermedia 6 consulta el índice almacenado en el almacenamiento de segunda capa intermedia 3, confirma un borde conectado con un nodo en la segunda capa intermedia para cada nodo en la segunda capa intermedia, y adquiere el peso w para el borde a partir del almacenamiento de segunda capa intermedia 3.

40

El calculador de nivel de actividad de segunda capa intermedia 6 adquiere un valor de sesgo B_{2M} de un nodo en la segunda capa intermedia a partir del almacenamiento de segunda capa intermedia 3 para cada nodo en la segunda capa intermedia.

45

Después de adquirir, para cada nodo en la segunda capa intermedia, el nivel de actividad A_{1M} de cada nodo en la primera capa intermedia, el peso w y el valor de sesgo B_{2M} del borde, el calculador de nivel de actividad de segunda capa intermedia 6 calcula el nivel de actividad A_{2M} para cada nodo en la segunda capa intermedia usando el nivel de actividad A_{1M} , el peso w y el valor de sesgo B_{2M} adquiridos mediante un método de cálculo similar al del calculador de nivel de actividad de primera capa intermedia 5.

50

Después de que el calculador de nivel de actividad de segunda capa intermedia 6 haya calculado los niveles de actividad A_{2M} de los nodos respectivos en la segunda capa intermedia, el calculador de nivel de actividad de tercera capa intermedia 7 calcula un nivel de actividad A_{3M} de cada nodo en la tercera capa intermedia (etapa ST4).

55

Un método de cálculo del nivel de actividad A_{3M} de cada nodo en la tercera capa intermedia mediante el calculador de nivel de actividad de tercera capa intermedia 7 es similar al método de cálculo del nivel de actividad A_{1M} de cada nodo en la primera capa intermedia mediante el calculador de nivel de actividad de primera capa intermedia 5.

60

Es decir, el calculador de nivel de actividad de tercera capa intermedia 7 consulta el índice almacenado en el almacenamiento de tercera capa intermedia 4, confirma cada nodo en la segunda capa intermedia conectado con un nodo en la tercera capa intermedia para cada nodo en la tercera capa intermedia, y adquiere el nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia.

65

El calculador de nivel de actividad de tercera capa intermedia 7 consulta el índice almacenado en el

almacenamiento de tercera capa intermedia 4, confirma un borde conectado con un nodo en la tercera capa intermedia para cada nodo en la tercera capa intermedia, y adquiere el peso w del borde a partir del almacenamiento de tercera capa intermedia 4.

5 El calculador de nivel de actividad de tercera capa intermedia 7 adquiere, para cada nodo en la tercera capa intermedia, un valor de sesgo B_{3M} de un nodo en la tercera capa intermedia a partir del almacenamiento de tercera capa intermedia 4.

10 Después de adquirir el nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia, el peso w y el valor de sesgo B_{3M} del borde para cada nodo en la tercera capa intermedia, el calculador de nivel de actividad de tercera capa intermedia 7 calcula el nivel de actividad A_{3M} para cada nodo en la tercera capa intermedia usando el nivel de actividad A_{2M} , el peso w y el valor de sesgo B_{3M} adquiridos mediante un método de cálculo similar al del calculador de nivel de actividad de primera capa intermedia 5.

15 Después de que el calculador de nivel de actividad de tercera capa intermedia 7 haya calculado los niveles de actividad A_{3M} de los nodos respectivos en la tercera capa intermedia, el calculador de nivel de actividad de capa de salida 9 calcula un nivel de actividad A_{OUT} de cada nodo en la capa de salida (etapa ST5).

20 Un método de cálculo del nivel de actividad A_{OUT} de cada nodo en la capa de salida mediante el calculador de nivel de actividad de capa de salida 9 es similar al método de cálculo del nivel de actividad A_{1M} de cada nodo en la primera capa intermedia mediante el calculador de nivel de actividad de primera capa intermedia 5.

25 Es decir, el calculador de nivel de actividad de capa de salida 9 consulta el índice almacenado en el almacenamiento de capa de salida 8, confirma cada nodo en la tercera capa intermedia conectado con un nodo en la capa de salida para cada nodo en la capa de salida, y adquiere el nivel de actividad A_{3M} de cada nodo en la tercera capa intermedia.

30 El calculador de nivel de actividad de capa de salida 9 consulta el índice almacenado en el almacenamiento de capa de salida 8, confirma un borde conectado con un nodo en la capa de salida para cada nodo en la capa de salida, y adquiere el peso w del borde a partir del almacenamiento de capa de salida 8.

El calculador de nivel de actividad de capa de salida 9 adquiere, para cada nodo en la capa de salida, un valor de sesgo B_{OUT} de un nodo en la capa de salida a partir del almacenamiento de capa de salida 8.

35 Después de adquirir el nivel de actividad A_{3M} de cada nodo en la tercera capa intermedia, el peso w y el valor de sesgo B_{OUT} del borde para cada nodo en la capa de salida, el calculador de nivel de actividad de capa de salida 9 calcula el nivel de actividad A_{OUT} para cada nodo en la capa de salida usando el nivel de actividad A_{3M} , el peso w y el valor de sesgo B_{OUT} adquiridos mediante un método de cálculo similar al del calculador de nivel de actividad de primera capa intermedia 5.

40 El nivel de actividad A_{OUT} para cada nodo en la capa de salida calculado mediante el calculador de nivel de actividad de capa de salida 9 se emite como resultado de inferencia del dispositivo de inferencia.

45 Específicamente, por ejemplo, en un caso de identificar un objeto captado en una imagen como uno de una persona, un perro, un gato y un coche, la capa de salida está formada por cuatro nodos. En este caso, se aprende que un nivel de actividad de cada nodo es un valor que representa una posibilidad de una persona, un perro, un gato o un coche.

50 Cuando se realiza la inferencia, se selecciona un nodo que tiene el mayor nivel de actividad en la capa de salida. Si, por ejemplo, el nodo seleccionado es el que emite la posibilidad de un gato, se emite "gato" como resultado de inferencia. Puede realizarse no solo el simple resultado de identificación, sino también procedimientos de calcular la fiabilidad usando un nivel de actividad o emitir un valor de estimación de regresión.

55 Tal como resulta evidente a partir de lo anterior, según la realización 1, el calculador de nivel de actividad de primera capa intermedia 5 consulta el índice almacenado en el almacenamiento de primera capa intermedia 2, adquiere, a partir de los niveles de actividad de nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada 1 y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de primera capa intermedia 2, un nivel de actividad de cada nodo en la capa de entrada que tiene conexión con cada nodo en la primera capa intermedia y peso para cada uno de los bordes y un valor de sesgo de los mismos, y calcula un nivel de actividad de cada nodo en la primera capa intermedia usando el nivel de actividad de cada nodo en la capa de entrada y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido. Esto puede lograr efectos de reducir la cantidad de cálculo y la cantidad de memoria al realizar la inferencia.

65 Dicho de otro modo, solo se requiere que el calculador de nivel de actividad de primera capa intermedia 5 realice el cálculo en cada nodo en la capa de entrada que tiene conexión con cada nodo en la primera capa intermedia

y, por tanto, puede reducirse drásticamente la cantidad de cálculo y la cantidad de memoria al realizar la inferencia.

5 Además, solo se requiere que el calculador de nivel de actividad de segunda capa intermedia 6 realice el cálculo en cada nodo en la primera capa intermedia que tiene conexión con nodos en la segunda capa intermedia y, por tanto, puede reducirse drásticamente la cantidad de cálculo y la cantidad de memoria al realizar la inferencia de manera similar al calculador de nivel de actividad de primera capa intermedia 5.

10 Asimismo, también se requiere únicamente que el calculador de nivel de actividad de tercera capa intermedia 7 realice el cálculo en cada nodo en la segunda capa intermedia que tiene conexión con nodos en la tercera capa intermedia y, por tanto, puede reducirse drásticamente la cantidad de cálculo y la cantidad de memoria al realizar la inferencia de manera similar al calculador de nivel de actividad de primera capa intermedia 5.

15 Además, también se requiere únicamente que el calculador de nivel de actividad de capa de salida 9 realice el cálculo en cada nodo en la tercera capa intermedia que se conecta con nodos en la capa de salida y, por tanto, puede reducirse drásticamente la cantidad de cálculo y la cantidad de memoria al realizar la inferencia de manera similar al calculador de nivel de actividad de primera capa intermedia 5.

20 En la realización 1, el calculador de nivel de actividad de primera capa intermedia 5, el calculador de nivel de actividad de segunda capa intermedia 6, el calculador de nivel de actividad de tercera capa intermedia 7 y el calculador de nivel de actividad de capa de salida 9 realizan la operación de suma de productos entre un nivel de actividad de cada nodo en una capa anterior y peso para un borde del mismo al calcular un nivel de actividad de cada nodo. Alternativamente, en vez del resultado de adición ADD en la fórmula (3), puede usarse el valor máximo o un valor promedio de niveles de actividad de los nodos respectivos en la capa anterior.

25 (Realización 2)

30 En la realización 1 descrita anteriormente, cada nodo en cada una de las capas que constituyen la red neuronal aplicada al dispositivo de inferencia no se conecta con todos los nodos en una capa anterior o una posterior, sino que se conecta con parte de los nodos.

35 Incluso en este caso en el que cada nodo en cada una de las capas se conecta con parte de los nodos en una capa anterior o una posterior, puede formarse un bucle mediante una pluralidad de bordes que conectan los nodos dependiendo de un modo de conexión entre los nodos.

En una red neuronal, una trayectoria específica se denomina bucle, que empieza desde un determinado nodo y vuelve a este nodo pasando a través de cada borde una única vez. El número de bordes que forman un bucle se denomina en el presente documento longitud del bucle.

40 La figura 8 es un diagrama explicativo que ilustra bucles a modo de ejemplo formados por una pluralidad de bordes.

45 Las figuras 8A y 8B ilustran un bucle a modo de ejemplo formado por cuatro bordes. La figura 8C ilustra un bucle a modo de ejemplo formado por seis bordes. La figura 8D ilustra un bucle a modo de ejemplo formado por ocho bordes.

50 En una red neuronal jerárquica, por ejemplo, como más corto puede formarse un bucle que tiene una longitud de 4. Especialmente, el bucle que tiene una longitud de 4 puede provocar deterioro de la precisión de inferencia porque la información de cálculo de gradiente tiende a propagarse mediante un método de retropropagación de errores al realizar el aprendizaje. También en un modelo que realiza inferencia mediante propagación de información bidireccional, tal como un método de propagación de probabilidad en una red bayesiana, un bucle corto provoca la circulación de información de propagación, induciendo de ese modo deterioro de la precisión de inferencia.

55 Por el motivo anterior, en la realización 2, una red neuronal aplicada a un dispositivo de inferencia se limita a una en la que cada bucle está formado por seis o más bordes cuando bordes que conectan nodos en capas respectivas que constituyen una red neuronal forman bucles. Tales bordes pueden ser bordes que conectan nodos en una primera capa intermedia y nodos en una capa de entrada, bordes que conectan nodos en una segunda capa intermedia y los nodos en la primera capa intermedia, bordes que conectan nodos en una tercera capa intermedia y los nodos en la segunda capa intermedia, y bordes que conectan nodos en una capa de salida y los nodos en la tercera capa intermedia.

60 Por consiguiente, en la realización 2, una red neuronal en la que un bucle está formado por cuatro bordes, tal como se ilustra en las figuras 8A y 8B, no se aplica a un dispositivo de inferencia. En cambio, una red neuronal en la que un bucle está formado por seis u ocho bordes, tal como se ilustra en las figuras 8C y 8D, puede aplicarse a un dispositivo de inferencia.

Esta configuración puede lograr efectos de suprimir el deterioro de la precisión de inferencia que puede atribuirse a un bucle corto en una red neuronal. Dicho de otro modo, puede reducirse la cantidad de cálculo y la cantidad de memoria al tiempo que se mantiene la precisión de inferencia.

5
(Realización 3)

10 En la realización 1 descrita anteriormente, cada nodo en cada una de las capas que constituyen la red neuronal aplicada al dispositivo de inferencia no se conecta con todos los nodos en una capa anterior o una posterior, sino que se conecta con parte de los nodos.

Cada nodo en la primera capa intermedia puede conectarse con parte de todos los nodos en la capa de entrada, que se seleccionan aleatoriamente de todos los nodos.

15 De manera similar, cada nodo en la segunda capa intermedia puede conectarse con parte de todos los nodos en la primera capa intermedia, que se seleccionan aleatoriamente de todos los nodos. Además, cada nodo en la tercera capa intermedia puede conectarse con parte de todos los nodos en la segunda capa intermedia, que se seleccionan aleatoriamente de todos los nodos.

20 De manera adicionalmente similar, cada nodo en la capa de salida puede conectarse con parte de todos los nodos en la tercera capa intermedia, que se seleccionan aleatoriamente de todos los nodos.

25 Para la selección aleatoria, puede establecerse un criterio de tal manera que un número promedio de conexiones por nodo en cada una de las capas (la capa de salida, la tercera capa intermedia, la segunda capa intermedia y la primera capa intermedia) con nodos en una capa anterior (la tercera capa intermedia, la segunda capa intermedia, la primera capa intermedia y la capa de entrada, respectivamente) es de cincuenta o menos.

30 Alternativamente, puede imponerse otro criterio de que un número promedio de conexiones por nodo en cada una de las capas (la capa de salida, la tercera capa intermedia, la segunda capa intermedia y la primera capa intermedia) con nodos en una capa anterior (la tercera capa intermedia, la segunda capa intermedia, la primera capa intermedia y la capa de entrada, respectivamente) es de una décima parte o menos del número de nodos en la capa anterior.

35 En un modo en el que cada nodo en cada capa se conecta con todos los nodos en una capa anterior, si se supone que el número de nodos en cada capa es M y el número de nodos en una capa anterior es N , la cantidad de cálculo de niveles de actividad y la cantidad de memoria en cada capa es del orden de $N \times M$. En cambio, estableciendo el criterio de que un número promedio de conexiones n ($n < N$) por nodo en cada una de las capas con nodos en una capa anterior es de cincuenta o menos o una décima parte o menos del número de nodos N en la capa anterior, es posible reducir la probabilidad de aparición de un bucle corto, suprimir el deterioro de la
40 precisión de inferencia y también reducir la cantidad de cálculo y la cantidad de memoria.

(Realización 4)

45 En la realización 1 descrita anteriormente, cada nodo en cada una de las capas que constituyen la red neuronal aplicada al dispositivo de inferencia no se conecta con todos los nodos en una capa anterior o una posterior, sino que se conecta con parte de los nodos.

50 Cada nodo en la primera capa intermedia puede conectarse con parte de todos los nodos en la capa de entrada, que no son adyacentes entre sí.

De manera similar, cada nodo en la segunda capa intermedia puede conectarse con parte de todos los nodos en la primera capa intermedia, que no son adyacentes entre sí. Además, cada nodo en la tercera capa intermedia puede conectarse con parte de todos los nodos en la segunda capa intermedia, que no son adyacentes entre sí.

55 De manera adicionalmente similar, cada nodo en la capa de salida puede conectarse con parte de todos los nodos en la tercera capa intermedia, que no son adyacentes entre sí.

60 Por ejemplo, en el caso del nodo "N" en la primera capa intermedia, se permite que el nodo "N" en la primera capa intermedia se conecte con los nodos "0" y "3" en la capa de entrada porque estos nodos no son adyacentes entre sí en la capa de entrada. Sin embargo, no se permite que el nodo "N" en la primera capa intermedia se conecte con un par de nodos "0" y "1" en la capa de entrada porque estos nodos son adyacentes entre sí en la capa de entrada.

65 En la realización 4, puede establecerse un criterio para conexiones permitidas de nodos de tal manera que un número promedio de conexiones por nodo en cada una de las capas (la capa de salida, la tercera capa intermedia, la segunda capa intermedia y la primera capa intermedia) con nodos en una capa anterior (la tercera

capa intermedia, la segunda capa intermedia, la primera capa intermedia y la capa de entrada, respectivamente) es de cincuenta o menos.

5 Alternativamente, puede establecerse otro criterio de tal manera que un número promedio de conexiones por nodo en cada una de las capas (la capa de salida, la tercera capa intermedia, la segunda capa intermedia y la primera capa intermedia) con nodos en una capa anterior (la tercera capa intermedia, la segunda capa intermedia, la primera capa intermedia y la capa de entrada, respectivamente) es de una décima parte o menos del número de nodos en la capa anterior.

10 Estableciendo los criterios anteriores, es posible reducir la probabilidad de aparición de un bucle corto, suprimir el deterioro de la precisión de inferencia y también reducir la cantidad de cálculo y la cantidad de memoria.

(Realización 5)

15 En las realizaciones 1 a 4 descritas anteriormente, la red neuronal aplicada al dispositivo de inferencia se muestra a modo de ejemplo mediante una red neuronal de un tipo predictivo jerárquico que tiene tres capas intermedias. Alternativamente, el número de capas intermedias puede ser mayor o menor de tres, y pueden no estar formadas capas intermedias tal como una estructura formada mediante un modelo de regresión logística.

20 Además, puede implementarse una combinación con un método convencional de redes neuronales, tal como una capa que conecta todos los nodos entre capas, una capa de convolución o una capa de agrupación tal como en una red neuronal convolucional, o bloques de memoria larga a corto plazo (LSTM) en una red neuronal recurrente.

25 En la red neuronal convolucional, tiene una configuración en la que la capa de convolución y la capa de agrupación están alternadas. La capa de convolución se encarga de extraer características locales en una imagen y la capa de agrupación resume las características para cada parte local.

30 En las realizaciones 1 a 4 descritas anteriormente, la red neuronal aplicada al dispositivo de inferencia se muestra a modo de ejemplo mediante una red neuronal de un tipo predictivo jerárquico. La red neuronal puede incluir omitir la conexión de una capa, conexión entre nodos que pertenecen a la misma capa, una autoconexión en la que un destino de conexión y un origen de conexión son el mismo, o conexión circulante en la que los bordes forman un bucle (es decir, una red neuronal recurrente).

35 Además, puede emplearse una red neuronal que realiza inferencia usando otras estructuras de gráficos, tal como un mapa autoorganizado (SOM), un modelo direccionable por contenido, una red de Hopfield o una máquina de Boltzmann. Además, sin limitarse a una red neuronal, puede emplearse una red que realiza inferencia usando otros gráficos, tal como una red bayesiana.

40 En las realizaciones 1 a 4 descritas anteriormente, se asignan índices unidimensionales, tales como los nodos 0, 1, ..., N - 1 en la capa de entrada y los nodos N, N + 1, ..., N + M - 1 en la primera capa intermedia. Alternativamente, pueden asignarse índices bidimensionales, tales como (0, 0), (0, 1), ..., (0, N - 1) para nodos en la capa de entrada y (1, 0), (1, 1), ..., (1, M - 1) para nodos en la primera capa intermedia. Además, pueden usarse direcciones en la memoria como los índices y pueden asignarse otros índices.

45 En las realizaciones 1 a 4 descritas anteriormente, la red neuronal aplicada al dispositivo de inferencia se muestra a modo de ejemplo mediante el caso en el que el número de bordes y el número de valores de peso de borde son el mismo. Alternativamente, puede compartirse una pluralidad de valores de peso de borde tal como un coeficiente de filtro de convolución en una red convolucional.

50 En las realizaciones 1 a 4 descritas anteriormente, se describe un procedimiento de cálculo de un nivel de actividad en cada nodo en un orden. Alternativamente, los cálculos que no dependen unos de otros pueden realizarse en paralelo usando una pluralidad de CPU o GPU para acelerar el procedimiento de cálculo.

55 En las realizaciones 1 a 4 descritas anteriormente, se describe un ejemplo de un sistema de clasificación de imágenes que recibe datos de imagen y clasifica la imagen. Alternativamente, el dispositivo de inferencia puede aplicarse generalmente a un sistema de inferencia que emite cualquier resultado de inferencia en respuesta a datos de entrada siempre que se prepare una señal de instrucción correspondiente a los datos y, por tanto, puede realizarse aprendizaje supervisado.

60 Como ejemplos para el resultado de inferencia, puede emitirse una posición o un tamaño de una zona de objeto deseada después de introducir la imagen, o puede emitirse un texto que explica una imagen después de introducir la imagen. Además, después de introducir una imagen que incluye ruido, puede emitirse una nueva imagen, de la que se ha eliminado el ruido. Además, después de introducir una imagen y un texto, puede convertirse la imagen según el texto.

65

Además, después de introducir sonido vocal, puede emitirse un fonema y una palabra, o puede emitirse una palabra predicha que va a pronunciarse a continuación, o puede emitirse un sonido de respuesta apropiado. Alternativamente, después de introducir un texto con un determinado idioma, puede emitirse otro texto en otro idioma. De manera adicionalmente alternativa, tras introducir una serie temporal, puede emitirse una serie temporal predicha del futuro o puede emitirse un estado estimado de la serie temporal.

En las realizaciones 1 a 4 descritas anteriormente, se describe un sistema a modo de ejemplo para realizar inferencia mediante un modelo que ha aprendido mediante aprendizaje supervisado que usa una señal de instrucción correspondiente a datos. Alternativamente, puede emplearse otro sistema, que realiza inferencia mediante un modelo que ha aprendido mediante aprendizaje no supervisado que usa datos sin una señal de instrucción o aprendizaje semisupervisado.

En las realizaciones 1 a 4 descritas anteriormente, se ha descrito el ejemplo en cada uno de los cuales el dispositivo de inferencia recibe datos de imagen a partir de un dispositivo de introducción de datos (no ilustrado) y calcula un nivel de actividad de cada nodo en la primera capa intermedia. Alternativamente, el dispositivo de introducción de datos no ilustrado puede calcular un nivel de actividad de cada nodo en la primera capa intermedia, mientras que un nivel de actividad de cada nodo en la segunda capa intermedia, la tercera capa intermedia y la capa de salida puede calcularse mediante el dispositivo de inferencia. En este caso, cuando la dimensionalidad de la salida a partir del dispositivo de introducción de datos es menor que la dimensionalidad de la entrada, el dispositivo de introducción de datos tiene una función de compresión de datos.

En las realizaciones 1 a 4 descritas anteriormente, se ha descrito el ejemplo en el que un nivel de actividad se calcula únicamente una vez para cada nodo. Alternativamente, tal como en un método de propagación de probabilidad en una red bayesiana, los nodos pueden intercambiar de manera repetida información una pluralidad de veces para potenciar la precisión de inferencia.

(Realización 6)

En las realizaciones 1 a 4 descritas anteriormente, como red neuronal aplicada al dispositivo de inferencia, se ha descrito un ejemplo en el que todas las capas excluyendo la capa de entrada contienen índices de conexiones de borde. Alternativamente, solo parte de las capas pueden contener índices de conexiones de borde mientras que otras capas tienen conexiones de borde similares a las de una red neuronal normal.

Se pretende que el índice de conexiones de borde incluya peso de borde y un valor de sesgo tal como se ilustra en la figura 7.

Además, se pretende que las conexiones de borde similares a las de una red neuronal normal signifiquen conexiones de borde conectadas con todos los nodos en una capa de destino (conexión de borde de una capa de conexión completa) así como conexiones de borde de una red neuronal conocida, tal como una capa de convolución o una capa de agrupación, cada una de las cuales se conecta con nodos en una capa de destino o nodos adyacentes de la misma.

La figura 9 es un diagrama explicativo que ilustra una red neuronal un modo de ejemplo en la que datos de imagen facilitados a una capa de entrada se clasifican en diez tipos.

En el ejemplo en la figura 9, se conectan cinco capas intermedias entre una capa de entrada y una capa de salida, es decir, se conectan una primera capa intermedia, una segunda capa intermedia, una tercera capa intermedia, una cuarta capa intermedia y una quinta capa intermedia.

En el ejemplo en la figura 9, una capa de convolución 31 se especifica por la capa de entrada a la primera capa intermedia, una capa de agrupación 32 se especifica por la primera capa intermedia a la segunda capa intermedia, una capa de convolución 33 se especifica por la segunda capa intermedia a la tercera capa intermedia, y una capa de agrupación 34 se especifica por la tercera capa intermedia a la cuarta capa intermedia. Además, una capa que contiene índice 35 descrita en las realizaciones anteriores 1 a 4 se especifica por la cuarta capa intermedia a la quinta capa intermedia, y una capa de conexión completa 36 se especifica por la quinta capa intermedia a la capa de salida.

De manera similar a la primera capa intermedia ilustrada en la figura 7, cada nodo en la quinta capa intermedia contiene un índice que representa nodos en la cuarta capa intermedia como orígenes de conexión, y peso de borde y un valor de sesgo correspondientes a las conexiones.

Cuando los datos de imagen facilitados a la capa de entrada tienen, por ejemplo, "60 × 60" (altura × anchura) píxeles, se requiere una capa de entrada que incluye "3600" (= 60 × 60 × 1) nodos para la red neuronal de la figura 9.

En este caso, por ejemplo, cuando un tamaño de filtro de la capa de convolución 31 desde la capa de entrada

hasta la primera capa intermedia es de $5 \times 5 \times 1$, el número de mapas en la capa de convolución 31 es de "100", y la capa de agrupación 32 desde la primera capa intermedia hasta la segunda capa intermedia y la capa de agrupación 34 desde la tercera capa intermedia hasta la cuarta capa intermedia realiza la agrupación máxima que tiene un tamaño de filtro de $2 \times 2 \times 1$, el tamaño de la primera capa intermedia es igual a $56 \times 56 \times 100$ (= $(60 - 5 + 1) \times (60 - 5 + 1) \times 100$), y además el tamaño de la segunda capa intermedia es igual a $28 \times 28 \times 100$ (= $(56/2) \times (56/2) \times 100$).

Además, el tamaño de la tercera capa intermedia es igual a $24 \times 24 \times 200$ (= $(28 - 5 + 1) \times (28 - 5 + 1) \times 200$), el tamaño de la cuarta capa intermedia es igual a $12 \times 12 \times 200$ (= $(24/2) \times (24/2) \times 200$), el tamaño de la quinta capa intermedia es igual a $1 \times 1 \times 1000$ y el número de nodos en la capa de salida es igual a $1 \times 1 \times 10$.

Obsérvese que puede usarse una unidad lineal rectificada (ReLU) como función de activación para calcular un valor de propagación al propagar información desde la capa de entrada hasta la primera capa intermedia, una función de activación para calcular un valor de propagación al propagar información desde la segunda capa intermedia hasta la tercera capa intermedia, y una función de activación para calcular un valor de propagación al propagar información desde la cuarta capa intermedia hasta la quinta capa intermedia. Además, se usa una función softmax, que es una función exponencial normalizada, como función de activación para calcular un valor de propagación al propagar información desde la quinta capa intermedia hasta la capa de salida.

Las capas de convolución 31 y 33 y las capas de agrupación 32 y 34 desde la capa de entrada hasta la cuarta capa intermedia permiten que la red neuronal de la figura 9 extraiga de manera robusta la cantidad de características a partir de datos de imagen a pesar de un cambio en la posición de la imagen.

Además, la capa que contiene índice desde la cuarta capa intermedia hasta la quinta capa intermedia puede reducir drásticamente la cantidad de cálculo y la cantidad de memoria al realizar la inferencia, de manera similar a las realizaciones 1 a 4 descritas anteriormente.

En esta realización 6, se describe un ejemplo en el que se facilitan datos de imagen a la capa de entrada. Alternativamente, los datos facilitados a la capa de entrada no se limitan a datos de imagen, sino que pueden ser una señal de sensor que son datos observados por un sensor u otros datos, tales como sonido o textos, por ejemplo.

En esta realización 6, se describe un ejemplo en el que datos de imagen facilitados a la capa de entrada se clasifican en diez tipos. Alternativamente, inferencia distinta de para clasificar datos de imagen puede realizarse modificando la capa de salida que constituye la red neuronal.

Por ejemplo, puede realizarse inferencia tal como un procedimiento de eliminación de ruido para eliminar ruido a partir de datos de imagen o predicción regresiva o cálculo de probabilidades.

Además, el número de nodos en cada capa o el tamaño de filtro pueden modificarse dependiendo de un objeto de inferencia.

En la red neuronal en la figura 9, se describe el ejemplo en el que la capa de convolución 31, la capa de agrupación 32, la capa de convolución 33, la capa de agrupación 34, la capa que contiene índice 35 y la capa de conexión completa 36 se conectan mediante bordes en este orden. Sin embargo, las capas pueden conectarse mediante bordes en un orden distinto del orden mencionado anteriormente siempre que se conecte al menos una única capa que contiene índices descrita en las realizaciones 1 a 4. Alternativamente, las capas de agrupación 32 y 34 pueden no estar conectadas.

(Realización 7)

En las realizaciones 1 a 6, se describe el ejemplo en el que la red neuronal aplicada al dispositivo de inferencia es una red neuronal predictiva (FFNN) que no forma un bucle dirigido. Alternativamente, el dispositivo de inferencia puede emplear una red neuronal recurrente (RNN) en la que parte de la red forma el bucle dirigido. En este caso, parte de las capas en la red neuronal recurrente pueden ser la capa que contiene índice descrita en las realizaciones anteriores 1 a 4.

La figura 10 es un diagrama explicativo que ilustra una red de Elman que es la red neuronal recurrente (RNN) que tiene conexiones de borde desde una capa de entrada hasta una primera capa intermedia, desde la primera capa intermedia hasta una segunda capa intermedia (capa de contexto), desde la segunda capa intermedia (capa de contexto) hasta la primera capa intermedia y desde la primera capa intermedia hasta una capa de salida.

En la red de Elman de la figura 10, se supone que el número de nodos en la segunda capa intermedia (capa de contexto) es equivalente al número de nodos en la primera capa intermedia.

La figura 11 es un diagrama que ilustra una configuración de un dispositivo de inferencia según la realización 7 de la invención. En la figura 11, el mismo símbolo que el de la figura 1 representa la misma parte o una correspondiente y, por tanto, se omiten descripciones de la misma.

5 El calculador de nivel de actividad de segunda capa intermedia 41 que es un calculador de nivel de actividad de capa intermedia se implementa mediante el conjunto de circuitos de nivel de actividad de capa intermedia 13 que
 10 consiste, por ejemplo, en un circuito integrado de semiconductor o un microordenador de un chip equipado con una CPU. El calculador de nivel de actividad de segunda capa intermedia 41 realiza un procedimiento de copiado de un nivel de actividad de cada nodo en la primera capa intermedia como nivel de actividad de cada nodo en la segunda capa intermedia (capa de contexto).

En esta realización 7, se describe el ejemplo en el que el calculador de nivel de actividad de segunda capa intermedia 41 copia un nivel de actividad de cada nodo en la primera capa intermedia como nivel de actividad de cada nodo en la segunda capa intermedia (capa de contexto). Sin embargo, esto es simplemente un ejemplo.
 15 Alternativamente, un nivel de actividad de cada nodo en la segunda capa intermedia (capa de contexto) puede calcularse mediante la fórmula (1), de manera similar al calculador de nivel de actividad de capa de entrada 1.

Un almacenamiento de primera capa intermedia 42 que es un almacenamiento de capa intermedia se implementa mediante el dispositivo de almacenamiento de capa intermedia 12 que consiste en un medio de
 20 almacenamiento, tal como una RAM o un disco duro. El almacenamiento de primera capa intermedia 42 almacena un índice que representa una relación de conexión entre cada nodo en la primera capa intermedia y cada nodo en la segunda capa intermedia (capa de contexto), peso para cada uno de los bordes que conectan un nodo en la primera capa intermedia y un nodo en la segunda capa intermedia, y un valor de sesgo dado a cada nodo en la primera capa intermedia.

Un calculador de nivel de actividad de primera capa intermedia 43 que es el calculador de nivel de actividad de capa intermedia se implementa mediante el conjunto de circuitos de nivel de actividad de capa intermedia 13 que
 25 consiste en un circuito integrado de semiconductor o un microordenador de un chip equipado con una CPU, por ejemplo. El calculador de nivel de actividad de primera capa intermedia 43 consulta el índice almacenado en el almacenamiento de primera capa intermedia 42, adquiere, a partir de los niveles de actividad de los nodos respectivos en la segunda capa intermedia (capa de contexto) calculados mediante el calculador de nivel de actividad de segunda capa intermedia 41 y el peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de primera capa intermedia 42, un nivel de actividad de cada nodo en la segunda capa intermedia (capa de contexto) que tiene conexión con cada nodo en la primera capa intermedia, y también
 30 adquiere el peso para cada uno de los bordes y un valor de sesgo. El calculador de nivel de actividad de primera capa intermedia 43 realiza un procedimiento de cálculo de un nivel de actividad de cada nodo en la primera capa intermedia usando el nivel de actividad adquirido de cada nodo en la segunda capa intermedia (capa de contexto) y el peso adquirido para cada uno de los bordes y el valor de sesgo adquirido.

En la figura 11, se supone que cada uno del calculador de nivel de actividad de capa de entrada 1, el almacenamiento de primera capa intermedia 2, el calculador de nivel de actividad de primera capa intermedia 5, el calculador de nivel de actividad de segunda capa intermedia 41, el almacenamiento de primera capa intermedia 42, el calculador de nivel de actividad de primera capa intermedia 43, el almacenamiento de capa de salida 8 y el calculador de nivel de actividad de capa de salida 9, que son componentes del dispositivo de inferencia, está constituido por hardware dedicado. Alternativamente, el dispositivo de inferencia puede estar
 45 constituido por un ordenador.

Cuando el dispositivo de inferencia está constituido por un ordenador, el almacenamiento de primera capa intermedia 2, el almacenamiento de primera capa intermedia 42 y el almacenamiento de capa de salida 8 se proporcionan en la memoria 21 del ordenador ilustrada en la figura 3. Además, un programa, que describe contenido de procedimiento del calculador de nivel de actividad de capa de entrada 1, el calculador de nivel de actividad de primera capa intermedia 5, el calculador de nivel de actividad de segunda capa intermedia 41, el calculador de nivel de actividad de primera capa intermedia 43 y el calculador de nivel de actividad de capa de salida 9, está almacenado en la memoria 21 del ordenador ilustrada en la figura 3. Además, el procesador 22 del
 50 ordenador ejecuta el programa almacenado en la memoria 21.

A continuación, se describirán operaciones.

Obsérvese que las operaciones de componentes distintos del calculador de nivel de actividad de segunda capa intermedia 41, el almacenamiento de primera capa intermedia 42 y el calculador de nivel de actividad de primera capa intermedia 43 son similares a las de la realización 1 tal como se describió anteriormente. Por tanto, solo se describirán las operaciones del calculador de nivel de actividad de segunda capa intermedia 41, el almacenamiento de primera capa intermedia 42 y el calculador de nivel de actividad de primera capa intermedia 43.
 65

El calculador de nivel de actividad de primera capa intermedia 5 calcula un nivel de actividad A_{1M} de cada nodo

5 en la primera capa intermedia usando un nivel de actividad de cada nodo en la capa de entrada que se ha calculado mediante el calculador de nivel de actividad de capa de entrada 1, de manera similar a la realización 1. Después de eso, el calculador de nivel de actividad de segunda capa intermedia 41 copia el nivel de actividad A_{1M} de cada nodo en la primera capa intermedia como nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia (capa de contexto).

Como resultado, el nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia en un tiempo t pasa a ser equivalente al nivel de actividad A_{1M} de cada nodo en la primera capa intermedia en el tiempo t .

10 Después de que el calculador de nivel de actividad de segunda capa intermedia 41 haya obtenido los niveles de actividad A_{2M} de los nodos respectivos en la segunda capa intermedia (capa de contexto), el calculador de nivel de actividad de primera capa intermedia 43 calcula un nivel de actividad A'_{1M} de cada nodo en la primera capa intermedia.

15 Un método de cálculo para el nivel de actividad A'_{1M} de cada nodo en la primera capa intermedia usado por el calculador de nivel de actividad de primera capa intermedia 43 es similar al método de cálculo del nivel de actividad A_{1M} de cada nodo en la primera capa intermedia usado por el calculador de nivel de actividad de primera capa intermedia 5.

20 Es decir, el calculador de nivel de actividad de primera capa intermedia 43 consulta el índice almacenado en el almacenamiento de primera capa intermedia 42, confirma cada nodo en la segunda capa intermedia (capa de contexto) conectado con un nodo en la primera capa intermedia para cada nodo en la primera capa intermedia, y adquiere el nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia (capa de contexto).

25 Además, el calculador de nivel de actividad de primera capa intermedia 43 consulta el índice almacenado en el almacenamiento de primera capa intermedia 42, confirma un borde conectado con un nodo en la primera capa intermedia (borde conectado con un nodo en la segunda capa intermedia) para cada nodo en la primera capa intermedia, y adquiere el peso w para el borde a partir del almacenamiento de primera capa intermedia 42.

30 El calculador de nivel de actividad de primera capa intermedia 43 adquiere además un valor de sesgo B_{1M} de un nodo en la primera capa intermedia a partir del almacenamiento de primera capa intermedia 42 para cada nodo en la primera capa intermedia.

35 Después de adquirir el nivel de actividad A_{2M} de cada nodo en la segunda capa intermedia (capa de contexto), el peso w del borde y el valor de sesgo B_{1M} para cada nodo en la primera capa intermedia (capa de contexto), el calculador de nivel de actividad de primera capa intermedia 43 calcula el nivel de actividad A'_{1M} para cada nodo en la primera capa intermedia usando el nivel de actividad A_{2M} adquirido, el peso w adquirido del borde y el valor de sesgo B_{1M} adquirido mediante un método de cálculo que es similar al del calculador de nivel de actividad de primera capa intermedia 5.

40 Después de que el calculador de nivel de actividad de primera capa intermedia 43 haya calculado el nivel de actividad A'_{1M} de los nodos respectivos en la primera capa intermedia, el calculador de nivel de actividad de capa de salida 9 calcula un nivel de actividad A_{OUT} de cada nodo en la capa de salida usando el nivel de actividad A'_{1M} de cada nodo en la primera capa intermedia.

45 Un método de cálculo del nivel de actividad A_{OUT} de cada nodo en la capa de salida usado por el calculador de nivel de actividad de capa de salida 9 es similar al de la realización 1 anterior.

50 Tal como resulta evidente a partir de lo anterior, según la realización 7, el calculador de nivel de actividad de primera capa intermedia 43 consulta el índice almacenado en el almacenamiento de primera capa intermedia 42, adquiere, a partir de los niveles de actividad de nodos respectivos en la segunda capa intermedia (capa de contexto) calculados mediante el calculador de nivel de actividad de segunda capa intermedia 41 y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de primera capa intermedia 42, un nivel de actividad de cada nodo en la segunda capa intermedia (capa de contexto) que tiene conexión con cada
 55 nodo en la primera capa intermedia y peso para cada uno de los bordes y un valor de sesgo de los mismos, y calcula un nivel de actividad de cada nodo en la primera capa intermedia usando el nivel de actividad de cada nodo en la segunda capa intermedia (capa de contexto) y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido. Esto puede lograr efectos de reducir la cantidad de cálculo y la cantidad de memoria al realizar la inferencia incluso en el caso de emplear una red neuronal recurrente (RNN) parte de la cual forma
 60 un bucle dirigido.

Dicho de otro modo, solo se requiere que el calculador de nivel de actividad de primera capa intermedia 43 realice el cálculo en cada nodo en la segunda capa intermedia (capa de contexto) que tiene conexión con cada
 65 nodo en la primera capa intermedia y, por tanto, puede reducirse drásticamente la cantidad de cálculo y la cantidad de memoria al realizar la inferencia.

En la realización 7, se describe el ejemplo en el que la red neuronal recurrente (RNN) aplicada al dispositivo de inferencia es la red de Elman. Sin embargo, esto es simplemente un ejemplo. Alternativamente, por ejemplo, puede emplearse una red de Jordan, una red neuronal recurrente que tiene bloques de memoria larga a corto plazo (LSTM), una red neuronal recurrente jerárquica, una red neuronal recurrente bidireccional o una red neuronal recurrente de tiempo continuo.

Obsérvese que, en la realización 7, parte de las capas contienen índices para conexiones de borde independientemente de los tipos de la red neuronal recurrente.

(Realización 8)

En las realizaciones 1-7 descritas anteriormente, la red neuronal aplicada al dispositivo de inferencia se muestra a modo de ejemplo mediante una red neuronal predictiva (FFNN) o una red neuronal recurrente (RNN), en la que no se incluye una conexión de borde entre nodos que pertenecen a la misma capa, o no se incluye una autoconexión en la que un nodo de origen de conexión es igual a un nodo de destino de conexión. Alternativamente, puede emplearse una red neuronal predictiva (FFNN) o una red neuronal recurrente (RNN), en la que se incluye la conexión de borde o la autoconexión. Además, puede emplearse una red neuronal predictiva (FFNN) o una red neuronal recurrente (RNN), en la que se incluye una conexión que omita una determinada capa.

La figura 12 es un diagrama explicativo que ilustra una red de estado de eco un modo de ejemplo que es una red neuronal que incluye conexiones de borde entre nodos dentro de la capa intermedia o una autoconexión y también incluye un borde que conecta una capa de entrada a una capa de salida al tiempo que omita una capa intermedia.

En la figura 12, la conexión de borde entre los nodos dentro de la capa intermedia o la autoconexión pueden considerarse como una conexión de borde desde la capa intermedia hasta esta capa intermedia. Por tanto, cada capa de la red de estado de eco puede representarse por la figura 13.

La figura 14 es un diagrama que ilustra una configuración de un dispositivo de inferencia según una realización 8 de la invención. En la figura 14, el mismo símbolo que el de la figura 1 representa la misma parte o una correspondiente y, por tanto, se omitirán descripciones de la misma.

Un almacenamiento de capa intermedia 51 se implementa mediante el dispositivo de almacenamiento de capa intermedia 12 que consiste en un medio de almacenamiento, tal como una RAM o un disco duro. El almacenamiento de capa intermedia 51 almacena un índice que representa una relación de conexión entre cada nodo en la capa intermedia y cada nodo en la capa de entrada o la capa de salida, peso para cada uno de los bordes que conectan un nodo en la capa intermedia y un nodo en la capa de entrada o la capa de salida, y un valor de sesgo dado a cada nodo en la capa intermedia.

El almacenamiento de capa intermedia 51 almacena además un índice que representa una relación de una conexión de borde entre los nodos en la capa intermedia o una autoconexión, peso para cada uno de los bordes de la conexión de borde entre los nodos en la capa intermedia o la autoconexión, y un valor de sesgo dado a cada nodo en la capa intermedia.

Un calculador de nivel de actividad de capa intermedia 52 se implementa mediante el conjunto de circuitos de nivel de actividad de capa intermedia 13 constituido por un circuito integrado de semiconductor o un microordenador de un chip equipado con una CPU, por ejemplo. El calculador de nivel de actividad de capa intermedia 52 consulta el índice almacenado en el almacenamiento de capa intermedia 51, de manera similar al calculador de nivel de actividad de primera capa intermedia 5 en la figura 1, adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante un calculador de nivel de actividad de capa de entrada 1 o niveles de actividad de los nodos respectivos en la capa de salida calculados mediante un calculador de nivel de actividad de capa de salida 54 y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de capa intermedia 51, un nivel de actividad de cada nodo en la capa de entrada o la capa de salida que tiene conexión con cada nodo en la capa intermedia y peso para cada uno de los bordes y un valor de sesgo de los mismos, y realiza un procedimiento de cálculo de un nivel de actividad de cada nodo en la capa intermedia usando el nivel de actividad de cada nodo en la capa de entrada o la capa de salida y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido.

El calculador de nivel de actividad de capa intermedia 52 consulta además el índice almacenado en el almacenamiento de capa intermedia 51, adquiere, a partir de los niveles de actividad de nodos respectivos en la capa intermedia que se han calculado y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de capa intermedia 51, un nivel de actividad de cada nodo de un origen de conexión en la capa intermedia que se conecta con cada nodo de un destino de conexión en la capa intermedia y peso para cada uno de los bordes y un valor de sesgo de los mismos, y calcula un nivel de actividad de cada nodo de un destino de conexión en la capa intermedia usando el nivel de actividad de cada nodo de un origen de conexión en la capa

intermedia y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido.

Obsérvese que un nodo de un destino de conexión en la capa intermedia significa un nodo conectado con otro nodo en la capa intermedia o un nodo conectado consigo mismo en la capa intermedia.

5 Un nodo de un destino de conexión en la capa intermedia significa el otro nodo conectado con el nodo del origen de conexión en la capa intermedia o el nodo de autoconexión en la capa intermedia.

10 Un almacenamiento de capa de salida 53 se implementa mediante el dispositivo de almacenamiento de capa de salida 14 que consiste en un medio de almacenamiento, tal como una RAM o un disco duro. El almacenamiento de capa de salida 53 almacena un índice que representa una relación de conexión (información de conexión) entre cada nodo en la capa de salida y cada nodo en la capa de entrada o la capa intermedia.

15 Cuando un nodo en la capa de salida se conecta con un nodo en la capa de entrada, el almacenamiento de capa de salida 53 almacena además peso para cada uno de los bordes que conectan el nodo en la capa de salida y el nodo en la capa de entrada y un valor de sesgo dado al nodo en la capa de salida conectado con el nodo en la capa de entrada.

20 Cuando un nodo en la capa de salida se conecta con un nodo en la capa intermedia, el almacenamiento de capa de salida 53 almacena además peso para cada uno de los bordes que conectan el nodo en la capa de salida y el nodo en la capa intermedia y un valor de sesgo dado al nodo en la capa de salida conectado con el nodo en la capa intermedia.

25 El calculador de nivel de actividad de capa de salida 54 se implementa mediante el conjunto de circuitos de nivel de actividad de capa de salida 15 constituido por un circuito integrado de semiconductor o un microordenador de un chip equipado con una CPU, por ejemplo. Cuando un nodo conectado con un nodo en la capa de salida está en la capa de entrada, el calculador de nivel de actividad de capa de salida 54 adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada 1 y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de capa de salida 53, un nivel de actividad de cada nodo en la capa de entrada que tiene conexión con cada nodo en la capa de salida y peso para cada uno de los bordes y un valor de sesgo de los mismos. Cuando un nodo conectado con un nodo en la capa de salida está en la capa intermedia, el calculador de nivel de actividad de capa de salida 54 adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa intermedia calculados mediante el calculador de nivel de actividad de capa intermedia 52 y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de capa de salida 53, un nivel de actividad de cada nodo en la capa intermedia que tiene conexión con cada nodo en la capa de salida y peso para cada uno de los bordes y un valor de sesgo de los mismos.

40 El calculador de nivel de actividad de capa de salida 54 realiza además un procedimiento de cálculo de un nivel de actividad de cada nodo en la capa de salida usando el nivel de actividad de cada nodo en la capa de entrada o la capa intermedia y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido.

45 En la figura 14, se supone que cada uno del calculador de nivel de actividad de capa de entrada 1, el almacenamiento de capa intermedia 51, el calculador de nivel de actividad de capa intermedia 52, el almacenamiento de capa de salida 53 y el calculador de nivel de actividad de capa de salida 54 que son componentes del dispositivo de inferencia está constituido por hardware dedicado. Sin embargo, el dispositivo de inferencia puede estar constituido por un ordenador.

50 Cuando el dispositivo de inferencia está constituido por un ordenador, solo se requiere que el almacenamiento de capa intermedia 51 y el almacenamiento de capa de salida 53 estén constituidos en la memoria 21 del ordenador ilustrada en la figura 3, que un programa que describe contenido de procedimiento del calculador de nivel de actividad de capa de entrada 1, el calculador de nivel de actividad de capa intermedia 52 y el calculador de nivel de actividad de capa de salida 54 esté almacenado en la memoria 21 del ordenador ilustrada en la figura 3, y que el procesador 22 del ordenador ejecute el programa almacenado en la memoria 21.

55 A continuación, se describirán operaciones.

60 Después de que el calculador de nivel de actividad de capa de entrada 1 calcule el nivel de actividad de cada nodo en la capa de entrada, de manera similar a la realización 1 descrita anteriormente, el calculador de nivel de actividad de capa intermedia 52 consulta el índice almacenado en el almacenamiento de capa intermedia 51, confirma, a partir de los nodos respectivos en la capa intermedia, un nodo que tiene conexión con un nodo en la capa de entrada y un nodo que tiene conexión con un nodo en la capa de salida.

65 Con respecto a un nodo en la capa intermedia, que tiene conexión con un nodo en la capa de entrada, se calcula un nivel de actividad mediante el calculador de nivel de actividad de capa intermedia 52 usando el nivel de actividad de cada nodo en la capa de entrada, de manera similar al calculador de nivel de actividad de primera

capa intermedia 5 en la figura 1.

Con respecto a un nodo en la capa intermedia, que tiene conexión con un nodo en la capa de salida, se calcula un nivel de actividad usando el nivel de actividad de la capa de salida calculado mediante el calculador de nivel de actividad de capa de salida 54.

Un método de cálculo del nivel de actividad del nodo que tiene conexión con un nodo en la capa de salida usado por el calculador de nivel de actividad de capa intermedia 52, es similar al método de cálculo del nivel de actividad de un nodo que tiene conexión con un nodo en la capa de entrada, excepto porque un nodo que tiene conexión con un nodo seleccionado como objetivo, para el que se calcula el nivel de actividad, no es un nodo en la capa de entrada sino en la capa de salida.

Después de calcular el nivel de actividad de cada nodo en la capa intermedia que tiene conexión con un nodo en la capa de entrada o la capa de salida, el calculador de nivel de actividad de capa intermedia 52 consulta el índice almacenado en el almacenamiento de capa intermedia 51 y confirma, a partir de los nodos respectivos en la capa intermedia, un nodo de un origen de conexión (un nodo conectado con otro nodo dentro de la capa intermedia, o un nodo conectado consigo mismo en la capa intermedia) que tiene conexión con un nodo de un destino de conexión.

Después de confirmar un nodo de un origen de conexión que tiene conexión con un nodo de un destino de conexión, el calculador de nivel de actividad de capa intermedia 52 adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa intermedia que se han calculado y el peso para bordes respectivos y los valores de sesgo almacenados en el almacenamiento de capa intermedia 51, un nivel de actividad de cada nodo de un origen de conexión en la capa intermedia que tiene conexión con un nodo de un destino de conexión en la capa intermedia y peso para cada uno de los bordes y un valor de sesgo de los mismos.

Cuando el nodo de un origen de conexión en la capa intermedia tiene conexión con un nodo en la capa de entrada o en la capa de salida, un nivel de actividad del nodo de origen de conexión ya se calcula tal como se describió anteriormente. Por tanto, es posible tratar cada nodo en la capa intermedia como un nodo seleccionado como objetivo (un nodo de destino de conexión), para el que se calcula un nivel de actividad, en un orden más próximo a un nodo en la capa intermedia que tiene conexión con un nodo en la capa de entrada o en la capa de salida.

El calculador de nivel de actividad de capa intermedia 52 calcula un nivel de actividad de cada nodo como destino de conexión en la capa intermedia usando el nivel de actividad de cada nodo como origen de conexión en la capa intermedia y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido.

Un método de cálculo del nivel de actividad del nodo como destino de conexión en la capa intermedia mediante el calculador de nivel de actividad de capa intermedia 52 es similar al método de cálculo del nivel de actividad de un nodo conectado con un nodo en la capa de entrada, excepto porque un nodo de un origen de conexión no es un nodo en la capa de entrada sino en la capa intermedia.

El calculador de nivel de actividad de capa de salida 54 consulta el índice almacenado en el almacenamiento de capa de salida 53, confirma cada nodo en la capa de entrada o la capa intermedia conectado con cada nodo en la capa de salida.

Cuando un nodo conectado con un nodo en la capa de salida es el de la capa de entrada, el calculador de nivel de actividad de capa de salida 54 adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada que se han calculado mediante el calculador de nivel de actividad de capa de entrada 1 y el peso para bordes respectivos y los valores de sesgo almacenados en el almacenamiento de capa de salida 53, un nivel de actividad de cada nodo en la capa de entrada conectado con cada nodo en la capa de salida y peso para cada uno de los bordes y un valor de sesgo de los mismos. Mientras tanto, cuando un nodo conectado con un nodo en la capa de salida es el de la capa intermedia, el calculador de nivel de actividad de capa de salida 54 adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa intermedia que se han calculado mediante el calculador de nivel de actividad de capa intermedia 52 y el peso para bordes respectivos y los valores de sesgo almacenados en el almacenamiento de capa de salida 53, un nivel de actividad de cada nodo en la capa intermedia conectado con cada nodo en la capa de salida y peso para cada uno de los bordes y un valor de sesgo de los mismos.

Después de adquirir el nivel de actividad de cada nodo en la capa de entrada o la capa intermedia y el peso para cada uno de los bordes y el valor de sesgo, el calculador de nivel de actividad de capa de salida 54 calcula un nivel de actividad de cada nodo en la capa de salida usando el nivel de actividad de cada nodo en la capa de entrada o la capa intermedia y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido.

Tal como resulta evidente a partir de lo anterior, según la realización 8, el calculador de nivel de actividad de capa intermedia 52 consulta el índice almacenado en el almacenamiento de capa intermedia 51, adquiere, a

partir de los niveles de actividad de los nodos respectivos en la capa intermedia que se han calculado y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de capa intermedia 51, un nivel de actividad de cada nodo como origen de conexión en la capa intermedia que se conecta con un nodo como destino de conexión en la capa intermedia y peso para cada uno de los bordes y un valor de sesgo de los mismos, y calcula un nivel de actividad de cada nodo como destino de conexión en la capa intermedia usando el nivel de actividad de cada nodo como origen de conexión en la capa intermedia y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido. Por tanto, esto logra efectos de reducir la cantidad de cálculo y la cantidad de memoria al realizar la inferencia incluso en el caso de emplear una red neuronal en la que se incluyen conexiones de borde entre nodos en la capa intermedia o una autoconexión.

Según la realización 8, además, cuando un nodo conectado con un nodo en la capa de salida está en la capa de entrada, el calculador de nivel de actividad de capa de salida 54 adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada 1 y peso para bordes respectivos y valores de sesgo almacenados en el almacenamiento de capa de salida 53, un nivel de actividad de cada nodo en la capa de entrada que tiene conexión con cada nodo en la capa de salida y peso para cada uno de los bordes y un valor de sesgo de los mismos y calcula un nivel de actividad de cada nodo en la capa de salida usando el nivel de actividad de cada nodo en la capa de entrada y el peso para cada uno de los bordes y el valor de sesgo que se han adquirido. Por tanto, esto logra efectos de reducir la cantidad de cálculo y la cantidad de memoria al realizar la inferencia incluso en el caso de emplear una red neuronal en la que se incluye un borde que conecta la capa de entrada y la capa de salida al tiempo que se omite la capa intermedia.

En la realización 8, se describe el ejemplo en el que la red neuronal aplicada al dispositivo de inferencia es una red de estado de eco. Sin embargo, esto es simplemente un ejemplo. Alternativamente, puede emplearse una red neuronal recurrente con conexión completa, una red de Hopfield o una máquina de Boltzmann.

Obsérvese que, en la realización 8, parte de las capas contienen índices para conexiones de borde independientemente de qué tipo de la red neuronal se emplee.

(Realización 9)

Las realizaciones 1 a 6 descritas anteriormente se muestran a modo de ejemplo por el hecho de que la red neuronal aplicada al dispositivo de inferencia es una red neuronal predictiva (FFNN), y las realizaciones 7 y 8 descritas anteriormente se muestran a modo de ejemplo por el hecho de que la red neuronal aplicada al dispositivo de inferencia es una red neuronal recurrente (RNN). Sin embargo, estos son simplemente ejemplos. Alternativamente, el dispositivo de inferencia puede emplear la siguiente red neuronal. Obsérvese que parte de las capas contienen índices para conexiones de borde independientemente de qué tipo de la red neuronal se emplee.

Los ejemplos de una red neuronal aplicada a un dispositivo de inferencia incluyen una red de función de base radial (RBF), un mapa autoorganizado (SOM), un método de cuantificación de vector de aprendizaje (LVQ), una red neuronal modular, una red neuronal de impulsos, una red neuronal dinámica, una red neuronal en cascada u otro tipo de red neuronal distinto de FFNN o RNN, tal como el de una memoria temporal jerárquica (HTM).

(Realización 10)

En las realizaciones 1 a 9 descritas anteriormente, se describe un ejemplo en el que un método de aprendizaje del dispositivo de inferencia es uno de aprendizaje supervisado que usa una señal de instrucción correspondiente a datos, aprendizaje no supervisado que usa datos sin una señal de instrucción y un aprendizaje semisupervisado. Alternativamente, un método de aprendizaje del dispositivo de inferencia puede ser aprendizaje por refuerzo.

El aprendizaje por refuerzo es un método de aprendizaje de un modelo para que un agente en una determinada situación observe un estado actual y determine una acción que debe tomarse. Se pretende que el agente sea una función para realizar una operación apropiada mediante recopilación autónoma de información y determinación de situaciones sin operación continua por un usuario de un ordenador.

Cuando el agente selecciona una acción, se obtiene una recompensa a partir del entorno. En el aprendizaje por refuerzo, se aprende una política para poder maximizar la recompensa mediante una serie de acciones.

En el aprendizaje por refuerzo, se usa una función de valor de estado $V(s)$, que representa el valor de un estado "s", o una función de valor de acción $Q(s, a)$, que representa una recompensa obtenida a partir del entorno al seleccionar una acción "a" en el estado "s", como índice para medir la calidad de un estado o acción actual. Como algoritmo de aprendizaje por refuerzo, se usa aprendizaje por diferencia temporal (TD), tal como un SARSA o aprendizaje Q.

5 Cuando un método de aprendizaje del dispositivo de inferencia es aprendizaje por refuerzo, se aprende una red neuronal que recibe el estado "s" y emite una función de valor de estado V(s) o una función de valor de acción Q(s, a), y se realiza aprendizaje por TD usando esas funciones. Es decir, la función de valor de estado V(s) o la función de valor de acción Q(s, a) se calcula usando una red neuronal cuya parte de las capas contienen índices para conexiones de borde, y de ese modo se realiza aprendizaje por refuerzo. Las realizaciones 9 y 10 no son según la invención y se presentan únicamente con fines de ilustración.

10 Obsérvese que, dentro del alcance de la presente invención, la presente invención puede incluir una combinación flexible de las realizaciones respectivas, una modificación

10 Un dispositivo de inferencia según la presente invención es adecuado para quienes tengan una alta necesidad de reducir la cantidad de cálculo o cantidad de memoria al realizar la inferencia.

15 **Lista de signos de referencia**

15 1: calculador de nivel de actividad de capa de entrada, 2: almacenamiento de primera capa intermedia (almacenamiento de capa intermedia), 3: almacenamiento de segunda capa intermedia (almacenamiento de capa intermedia), 4: almacenamiento de tercera capa intermedia (almacenamiento de capa intermedia), 5: calculador de nivel de actividad de primera capa intermedia (calculador de nivel de actividad de capa intermedia),
 20 6: calculador de nivel de actividad de segunda capa intermedia (calculador de nivel de actividad de capa intermedia), 7: calculador de nivel de actividad de tercera capa intermedia (calculador de nivel de actividad de capa intermedia), 8: almacenamiento de capa de salida, 9: calculador de nivel de actividad de capa de salida, 11: conjunto de circuitos de nivel de actividad de capa de entrada, 12: dispositivo de almacenamiento de capa intermedia, 13: conjunto de circuitos de nivel de actividad de capa intermedia, 14: dispositivo de almacenamiento
 25 de capa de salida, 15: conjunto de circuitos de nivel de actividad de capa de salida, 21: memoria, 22: procesador, 31: capa de convolución, 32: capa de agrupación, 33: capa de convolución, 34: capa de agrupación, 35: capa que contiene índice, 36: capa de conexión completa, 41: calculador de nivel de actividad de segunda capa intermedia (calculador de nivel de actividad de capa intermedia), 42: almacenamiento de primera capa intermedia (almacenamiento de capa intermedia), 43: calculador de nivel de actividad de primera capa intermedia
 30 (calculador de nivel de actividad de capa intermedia), 51: almacenamiento de capa intermedia, 52: calculador de nivel de actividad de capa intermedia, 53: almacenamiento de capa de salida, 54: calculador de nivel de actividad de capa de salida.

REIVINDICACIONES

1. Dispositivo de inferencia configurado para clasificar imágenes, que comprende:

5 un calculador de nivel de actividad de capa de entrada (1) para calcular (ST1), cuando se facilitan datos de imagen a cada nodo de una capa de entrada que constituye una red neuronal, un nivel de actividad de cada nodo de la capa de entrada a partir de los datos facilitados;

10 un almacenamiento de capa intermedia (2) para almacenar un índice que representa una relación de conexión entre cada nodo en una capa intermedia que constituye la red neuronal y cada nodo en la capa de entrada, peso aplicado a un borde que conecta un nodo de la capa intermedia y un nodo de la capa de entrada;

15 un calculador de nivel de actividad de capa intermedia (5) para

consultar el índice almacenado en el almacenamiento de capa intermedia, confirmar cada nodo en la capa de entrada conectado con un nodo en la capa intermedia, y confirmar un borde conectado con un nodo en la capa intermedia para cada nodo de la capa intermedia,

20 adquirir, consultando el índice, a partir de los niveles de actividad de los nodos respectivos de la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada (1) y el peso para los bordes respectivos almacenados en el almacenamiento de capa intermedia (2), un nivel de actividad de cada nodo en la capa de entrada que tiene conexión con un nodo en la capa intermedia y peso para un borde correspondiente, para realizar únicamente el cálculo en nodos en la
25 capa de entrada que tienen una conexión con un nodo en la capa intermedia, y

30 calcular (ST2, ST3, ST4) un nivel de actividad del nodo en la capa intermedia usando el nivel de actividad adquirido del nodo en la capa de entrada y el peso adquirido para el borde correspondiente; y

35 un calculador de nivel de actividad de capa de salida (9) para calcular (ST5) un nivel de actividad de cada nodo en una capa de salida que constituye la red neuronal usando el nivel de actividad de cada nodo en la capa intermedia calculado mediante el calculador de nivel de actividad de capa intermedia (5), en el que el nivel de actividad de cada nodo en la capa de salida se emite como valor que representa una posibilidad de un tipo de un objeto en la imagen,

en el que

40 cada nodo en la capa intermedia se conecta con parte de todos los nodos en la capa de entrada, que se seleccionan aleatoriamente de todos los nodos, y

cada nodo en la capa de salida se conecta con parte de todos los nodos en la capa intermedia, que se seleccionan aleatoriamente de todos los nodos.

45 2. Dispositivo de inferencia configurado para clasificar imágenes, que comprende:

50 un calculador de nivel de actividad de capa de entrada (1) para calcular (ST1), cuando se facilitan datos de imagen a cada nodo de una capa de entrada que constituye una red neuronal, un nivel de actividad de cada nodo de la capa de entrada a partir de los datos facilitados;

55 un almacenamiento de capa intermedia (2) para almacenar un índice que representa una relación de conexión entre cada nodo en una capa intermedia que constituye la red neuronal y cada nodo en la capa de entrada, peso aplicado a un borde que conecta un nodo de la capa intermedia y un nodo de la capa de entrada;

un calculador de nivel de actividad de capa intermedia (5) para

60 consultar el índice almacenado en el almacenamiento de capa intermedia, confirmar cada nodo en la capa de entrada conectado con un nodo en la capa intermedia, y confirmar un borde conectado con un nodo en la capa intermedia para cada nodo de la capa intermedia,

65 adquirir, consultando el índice, a partir de los niveles de actividad de los nodos respectivos de la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada (1) y el peso para los bordes respectivos almacenados en el almacenamiento de capa intermedia (2), un nivel de actividad de cada nodo en la capa de entrada que tiene conexión con un nodo en la capa intermedia y peso para un borde correspondiente, para realizar únicamente el cálculo en nodos en la

capa de entrada que tienen una conexión con un nodo en la capa intermedia, y

calcular (ST2, ST3, ST4) un nivel de actividad del nodo en la capa intermedia usando el nivel de actividad adquirido del nodo en la capa de entrada y el peso adquirido para el borde correspondiente; y

un calculador de nivel de actividad de capa de salida (9) para calcular (ST5) un nivel de actividad de cada nodo en una capa de salida que constituye la red neuronal usando el nivel de actividad de cada nodo en la capa intermedia calculado mediante el calculador de nivel de actividad de capa intermedia (5), en el que el nivel de actividad de cada nodo en la capa de salida se emite como valor que representa una posibilidad de un tipo de un objeto en la imagen

en el que

cada nodo en la capa intermedia se conecta con parte de todos los nodos en la capa de entrada, que no son adyacentes entre sí, y

cada nodo en la capa de salida se conecta con parte de todos los nodos en la capa intermedia, que no son adyacentes entre sí.

3. Dispositivo de inferencia según la reivindicación 1 o 2,

en el que la red neuronal está constituida para incluir una pluralidad de capas intermedias,

en el que, para cada una de las capas intermedias, el almacenamiento de capa intermedia (2, 3, 4) almacena peso para un borde que conecta un nodo en la presente capa intermedia y un nodo en la capa de entrada cuando se encuentra dicha conexión de los nodos, y también almacena peso para un borde que conecta un nodo en la presente capa intermedia y un nodo en otra capa intermedia cuando se encuentra dicha conexión de los nodos,

cuando un nodo en la presente capa intermedia se conecta con un nodo en la capa de entrada, el calculador de nivel de actividad de capa intermedia (5, 6, 7) adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada (1) y el peso para los bordes respectivos almacenados en el almacenamiento de capa intermedia (2, 3, 4), un nivel de actividad del nodo en la capa de entrada conectado con el nodo en la presente capa intermedia y peso para un borde correspondiente,

cuando un nodo en la presente capa intermedia se conecta con un nodo en otra capa intermedia, el calculador de nivel de actividad de capa intermedia (5, 6, 7) adquiere, a partir de los niveles de actividad de los nodos respectivos en dicha otra capa intermedia y el peso para los bordes respectivos almacenados en el almacenamiento de capa intermedia (2, 3, 4), un nivel de actividad del nodo en dicha otra capa intermedia conectado con el nodo en la presente capa intermedia y peso para un borde correspondiente, y

el calculador de nivel de actividad de capa intermedia (5, 6, 7) calcula un nivel de actividad del nodo en la presente capa intermedia usando el nivel de actividad adquirido del nodo en la capa de entrada o dicha otra capa intermedia y el peso adquirido para el borde correspondiente.

4. Dispositivo de inferencia según la reivindicación 1 o 2, que comprende además un almacenamiento de capa de salida (8) para almacenar peso para un borde que conecta un nodo en la capa de salida y un nodo en la capa intermedia,

en el que el calculador de nivel de actividad de capa de salida (9)

adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa intermedia calculados mediante el calculador de nivel de actividad de capa intermedia (5, 6, 7) y el peso para los bordes respectivos almacenados en el almacenamiento de capa de salida (8), un nivel de actividad de un nodo en la capa intermedia conectado con un nodo en la capa de salida y peso para un borde correspondiente, y

calcula un nivel de actividad del nodo en la capa de salida usando el nivel de actividad adquirido del nodo en la capa intermedia y el peso adquirido para el borde correspondiente.

5. Dispositivo de inferencia según la reivindicación 1 o 2, que comprende además un almacenamiento de capa de salida (8) para almacenar peso para un borde que conecta un nodo en la capa de salida y un nodo en la capa de entrada cuando se encuentra dicha conexión de los nodos, y también almacenar

peso para un borde que conecta un nodo en la capa de salida y un nodo en la capa intermedia cuando se encuentra dicha conexión de los nodos,

5 en el que, cuando un nodo en la capa de salida se conecta con un nodo en la capa de entrada, el calculador de nivel de actividad de capa de salida (9) adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada (1) y el peso para los bordes respectivos almacenados en el almacenamiento de capa de salida (8), un nivel de actividad del nodo en la capa de entrada que tiene conexión con el nodo en la capa de salida y peso para un borde correspondiente,

10 cuando un nodo en la capa de salida se conecta con un nodo en la capa intermedia, el calculador de nivel de actividad de capa de salida (9) adquiere, a partir de los niveles de actividad de los nodos respectivos en la capa intermedia calculados mediante el calculador de nivel de actividad de capa intermedia (5, 6, 7) y el peso para los bordes respectivos almacenados en el almacenamiento de capa de salida (8), un nivel de actividad del nodo en la capa intermedia que tiene conexión con el nodo en la capa de salida y peso para un borde correspondiente, y el calculador de nivel de actividad de capa de salida (9) calcula un nivel de actividad del nodo en la capa de salida usando el nivel de actividad adquirido del nodo en la capa de entrada o la capa intermedia y el peso adquirido para el borde correspondiente.

15 6. Dispositivo de inferencia según la reivindicación 1 o 2, en el que el almacenamiento de capa intermedia (2, 3, 4) almacena un valor de sesgo dado a cada nodo en la capa intermedia además del peso para el borde, y

25 el calculador de nivel de actividad de capa intermedia (5, 6, 7) calcula el nivel de actividad del nodo en la capa intermedia usando el nivel de actividad del nodo en la capa de entrada y el peso para el borde y el valor de sesgo.

30 7. Dispositivo de inferencia según la reivindicación 6, en el que, para cada nodo en la capa intermedia, el calculador de nivel de actividad de capa intermedia (5, 6, 7)

35 realiza (ST13) una operación de suma de productos entre niveles de actividad respectivos de nodos respectivos en la capa de entrada conectados con el presente nodo en la capa intermedia y peso para bordes respectivos que conectan el presente nodo en la capa intermedia y los nodos respectivos en la capa de entrada,

40 añade (ST14) un valor de sesgo del presente nodo en la capa intermedia a un resultado de la operación de suma de productos, y

calcula (ST15), usando el resultado de adición como argumento de una función de activación de la red neuronal, un valor resultante de la función de activación como un nivel de actividad del presente nodo en la capa intermedia.

45 8. Dispositivo de inferencia según la reivindicación 4, en el que el almacenamiento de capa de salida (8) almacena un valor de sesgo dado a cada nodo en la capa de salida además del peso para el borde, y

50 el calculador de nivel de actividad de capa de salida (9) calcula el nivel de actividad del nodo en la capa de salida usando el nivel de actividad del nodo en la capa intermedia y el peso para el borde y el valor de sesgo.

55 9. Dispositivo de inferencia según la reivindicación 8, en el que, para cada nodo en la capa de salida, el calculador de nivel de actividad de capa de salida (9)

60 realiza (ST13) una operación de suma de productos entre niveles de actividad respectivos de nodos respectivos en la capa intermedia conectados con el presente nodo en la capa de salida y peso para bordes respectivos que conectan el presente nodo en la capa de salida y los nodos respectivos en la capa intermedia,

añade (ST14) un valor de sesgo del presente nodo en la capa de salida a un resultado de la operación de suma de productos, y

65 calcula (ST15), usando el resultado de adición como argumento de una función de activación de la red neuronal, un valor resultante de la función de activación como un nivel de actividad del presente nodo

en la capa de salida.

- 5 10. Dispositivo de inferencia según la reivindicación 1 o 2, en el que los bordes que conectan los nodos en la capa de entrada y la capa intermedia o los bordes que conectan los nodos en la capa intermedia y la capa de salida incluyen seis o más bordes que forman un bucle.
11. Dispositivo de inferencia según la reivindicación 3, en el que los bordes que conectan los nodos en la pluralidad de capas intermedias incluyen seis o más bordes que forman un bucle.
- 10 12. Dispositivo de inferencia según la reivindicación 1, en el que cada nodo en la capa intermedia se conecta con parte de todos los nodos en la capa de entrada u otra capa intermedia, que se seleccionan aleatoriamente de todos los nodos.
- 15 13. Dispositivo de inferencia según la reivindicación 2, en el que cada nodo en la pluralidad de capas intermedias se conecta con parte de todos los nodos en la capa de entrada u otra capa intermedia, que no son adyacentes entre sí.
- 20 14. Dispositivo de inferencia según la reivindicación 1 o 2, en el que un número promedio de conexiones por nodo en la capa intermedia con nodos en la capa de entrada es de cincuenta o menos, y un número promedio de conexiones por nodo en la capa de salida con nodos en la capa intermedia es de cincuenta o menos.
- 25 15. Dispositivo de inferencia según la reivindicación 1 o 2, en el que un número promedio de conexiones por nodo en la pluralidad de capas intermedias con nodos en la capa de entrada u otra capa intermedia es de cincuenta o menos.
- 30 16. Dispositivo de inferencia según la reivindicación 1 o 2, en el que un número promedio de conexiones por nodo en la capa intermedia con nodos en la capa de entrada es de una décima parte o menos del número de nodos en la capa de entrada, y un número promedio de conexiones por nodo en la capa de salida con nodos en la capa intermedia es de una décima parte o menos del número de nodos en la capa intermedia.
- 35 17. Dispositivo de inferencia según la reivindicación 1 o 2, en el que un número promedio de conexiones por nodo en la pluralidad de capas intermedias con nodos en la capa de entrada u otra capa intermedia es de una décima parte o menos del número de nodos en la capa de entrada o dicha otra capa intermedia.
- 40 18. Método para clasificación de imágenes que comprende las etapas de:
- 45 almacenar, en un almacenamiento de capa intermedia (2), un índice que representa una relación de conexión entre cada nodo en una capa intermedia y cada nodo en una capa de entrada, peso aplicado a un borde que conecta nodos respectivos de la capa intermedia y la capa de entrada, cada uno de los cuales constituye una red neuronal;
- 50 calcular (ST1), mediante un calculador de nivel de actividad de capa de entrada (1) cuando se facilitan datos de imagen a cada nodo en la capa de entrada, un nivel de actividad de cada nodo en la capa de entrada a partir de los datos facilitados;
- realizar, mediante un calculador de nivel de actividad de capa intermedia (5), un procedimiento de
- 55 consultar el índice almacenado en el almacenamiento de capa intermedia, confirmar cada nodo en la capa de entrada conectado con un nodo en la capa intermedia, y confirmar un borde conectado con un nodo en la capa intermedia para cada nodo de la capa intermedia,
- 60 adquirir, consultando el índice, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada (1) y el peso para los bordes respectivos almacenados en el almacenamiento de capa intermedia (2), un nivel de actividad de cada nodo en la capa de entrada que tiene una conexión con un nodo en la capa intermedia y peso para un borde correspondiente, y
- 65 calcular (ST2, ST3, ST4) un nivel de actividad del nodo en la capa intermedia usando el nivel de actividad adquirido del nodo en la capa de entrada y el peso adquirido para el borde

correspondiente, para realizar únicamente el cálculo en nodos en la capa de entrada que tienen una conexión con un nodo en la capa intermedia; y

5 calcular (ST5), mediante un calculador de nivel de actividad de capa de salida (9), un nivel de actividad de cada nodo en una capa de salida que constituye la red neuronal usando el nivel de actividad de cada nodo en la capa intermedia calculado mediante el calculador de nivel de actividad de capa intermedia (5, 6, 7), en el que el nivel de actividad de cada nodo en la capa de salida se emite como valor que representa una posibilidad de un tipo de un objeto en la imagen,

10 en el que

cada nodo en la capa intermedia se conecta con parte de todos los nodos en la capa de entrada, que se seleccionan aleatoriamente de todos los nodos, y

15 cada nodo en la capa de salida se conecta con parte de todos los nodos en la capa intermedia, que se seleccionan aleatoriamente de todos los nodos.

19. Método de inferencia para clasificación de imágenes que comprende las etapas de:

20 almacenar, en un almacenamiento de capa intermedia (2), un índice que representa una relación de conexión entre cada nodo en una capa intermedia y cada nodo en una capa de entrada, peso aplicado a un borde que conecta nodos respectivos de la capa intermedia y la capa de entrada, cada uno de los cuales constituye una red neuronal;

25 calcular (ST1), mediante un calculador de nivel de actividad de capa de entrada (1) cuando se facilitan datos de imagen a cada nodo en la capa de entrada, un nivel de actividad de cada nodo en la capa de entrada a partir de los datos facilitados:

30 realizar, mediante un calculador de nivel de actividad de capa intermedia (5), un procedimiento de

consultar el índice almacenado en el almacenamiento de capa intermedia, confirmar cada nodo en la capa de entrada conectado con un nodo en la capa intermedia, y confirmar un borde conectado con un nodo en la capa intermedia para cada nodo de la capa intermedia,

35 adquirir, consultando el índice, a partir de los niveles de actividad de los nodos respectivos en la capa de entrada calculados mediante el calculador de nivel de actividad de capa de entrada (1) y el peso para los bordes respectivos almacenados en el almacenamiento de capa intermedia (2), un nivel de actividad de cada nodo en la capa de entrada que tiene una conexión con un nodo en la capa intermedia y peso para un borde correspondiente, y

40 calcular (ST2, ST3, ST4) un nivel de actividad del nodo en la capa intermedia usando el nivel de actividad adquirido del nodo en la capa de entrada y el peso adquirido para el borde correspondiente, para realizar únicamente el cálculo en nodos en la capa de entrada que tienen una conexión con un nodo en la capa intermedia; y

45 calcular (ST5), mediante un calculador de nivel de actividad de capa de salida (9), un nivel de actividad de cada nodo en una capa de salida que constituye la red neuronal usando el nivel de actividad de cada nodo en la capa intermedia calculado mediante el calculador de nivel de actividad de capa intermedia (5, 6, 7), en el que el nivel de actividad de cada nodo en la capa de salida se emite como valor que representa una posibilidad de un tipo de un objeto en la imagen,

50 en el que

55 cada nodo en la capa intermedia se conecta con parte de todos los nodos en la capa de entrada, que no son adyacentes entre sí, y

cada nodo en la capa de salida se conecta con parte de todos los nodos en la capa intermedia, que no son adyacentes entre sí.

FIG.1

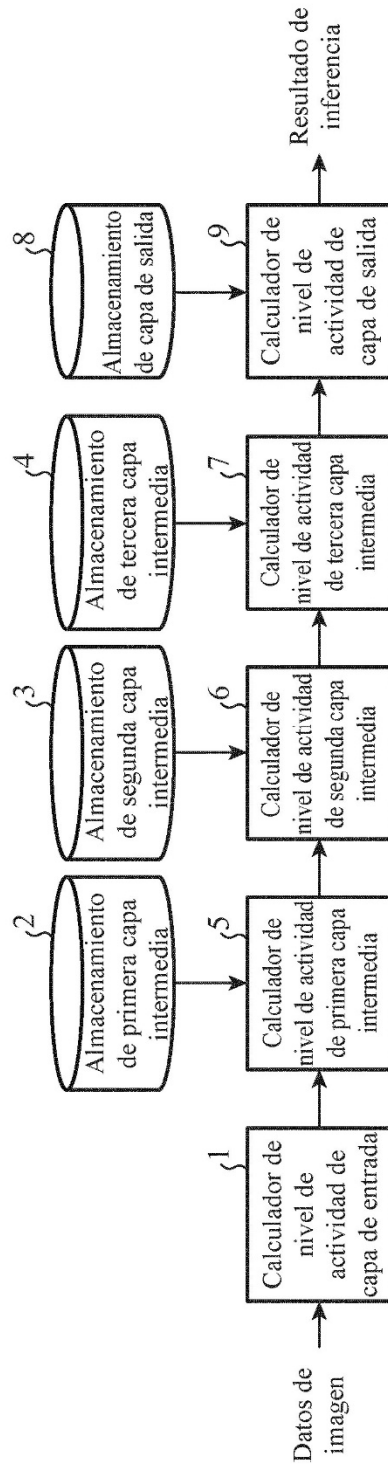


FIG.2

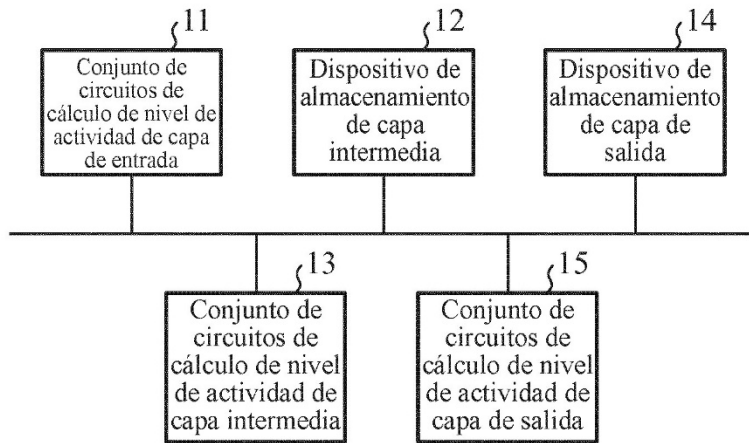


FIG.3

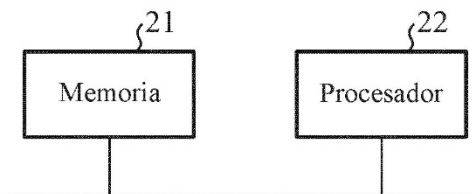


FIG.4

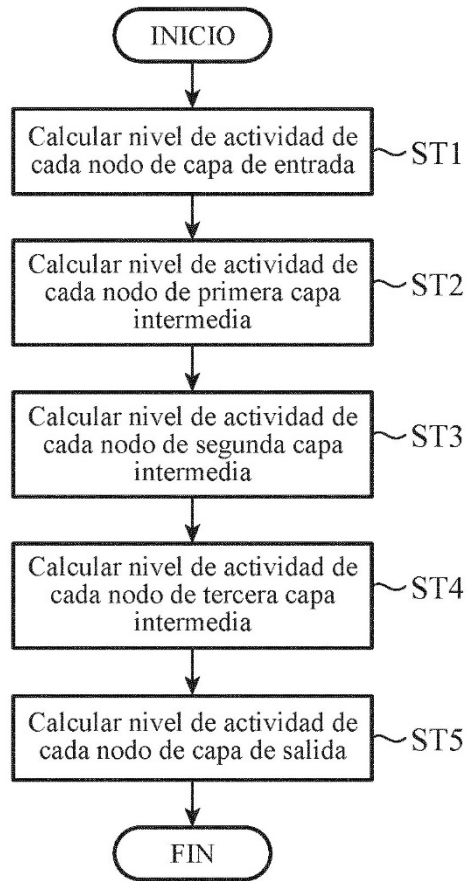


FIG.5

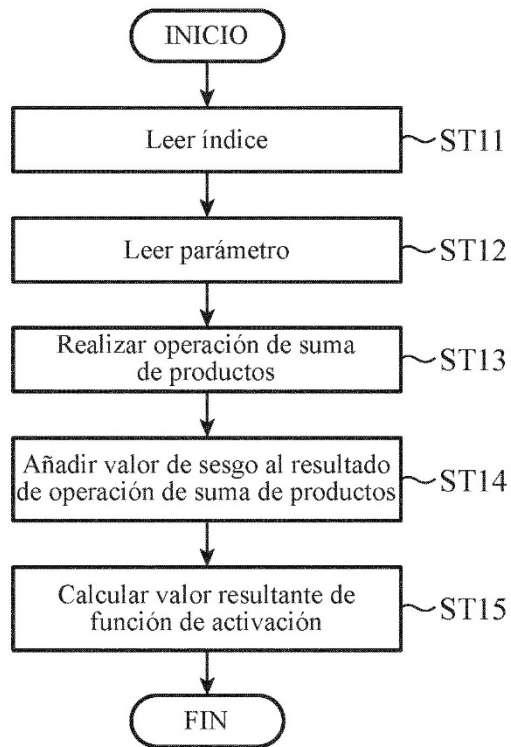


FIG.6

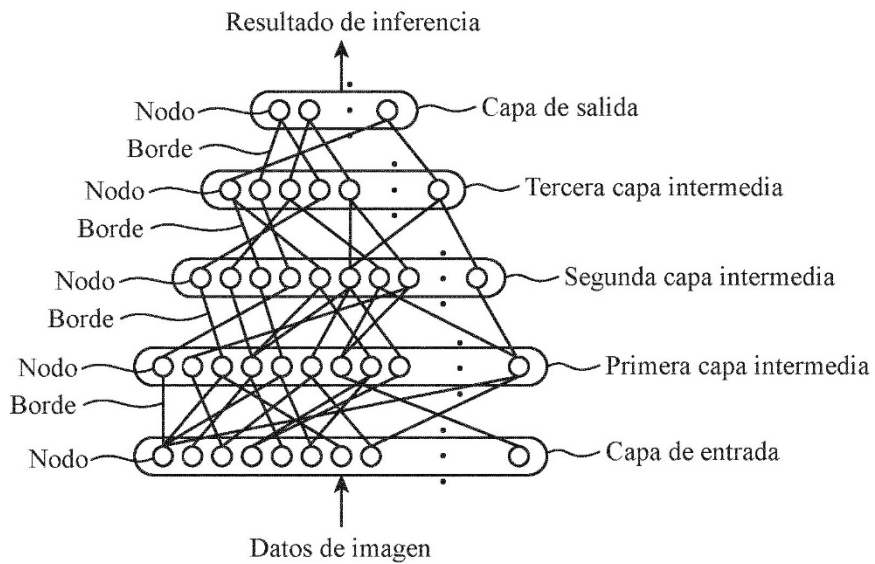


FIG.7

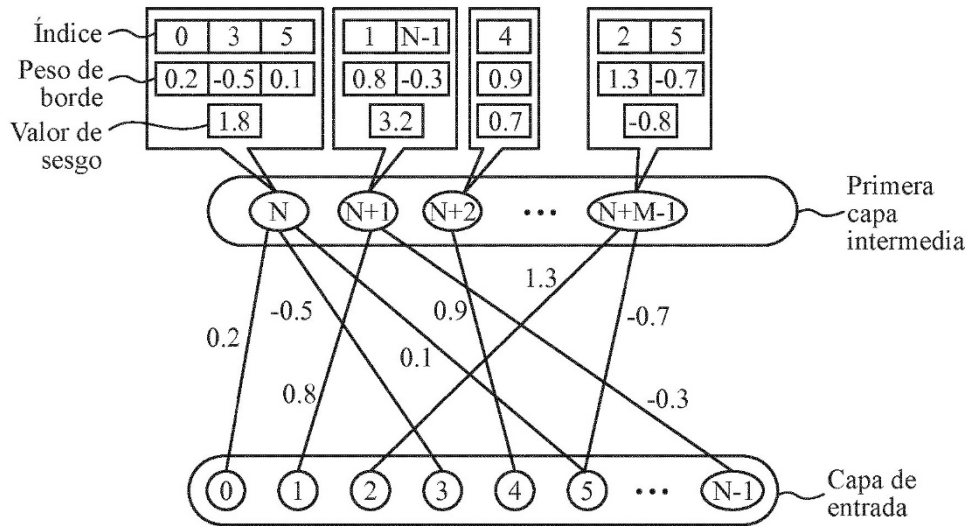


FIG.8A

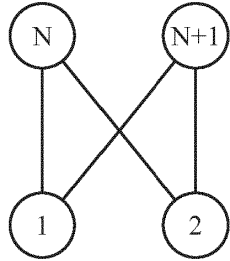


FIG.8B

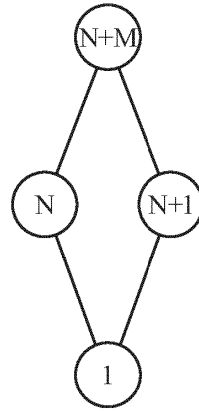


FIG.8C

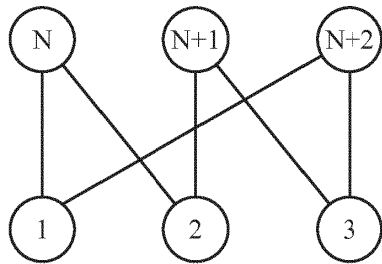


FIG.8D

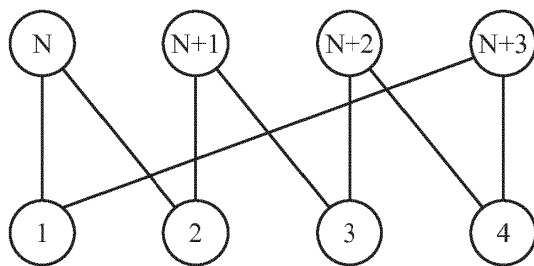


FIG.9

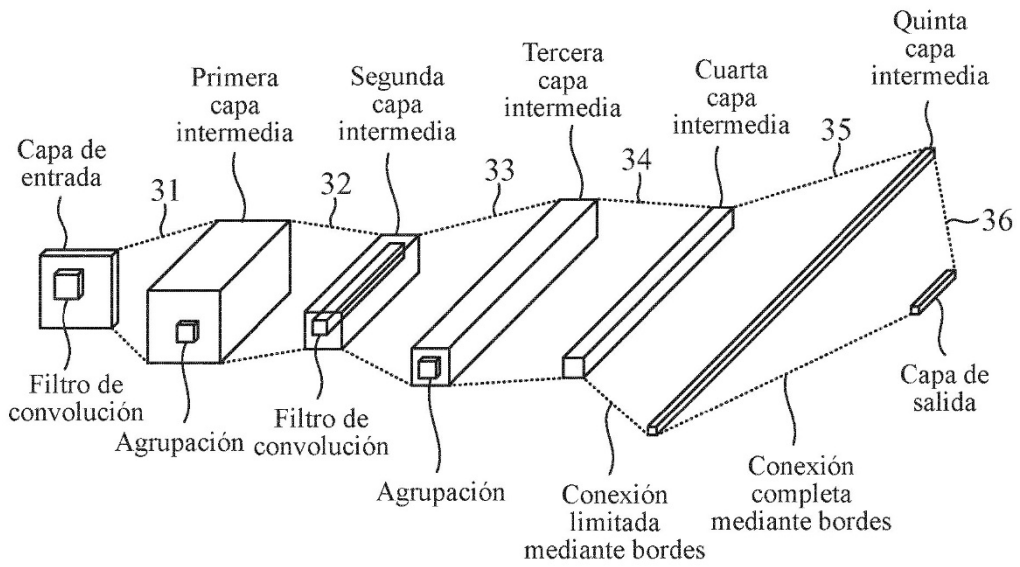


FIG.10

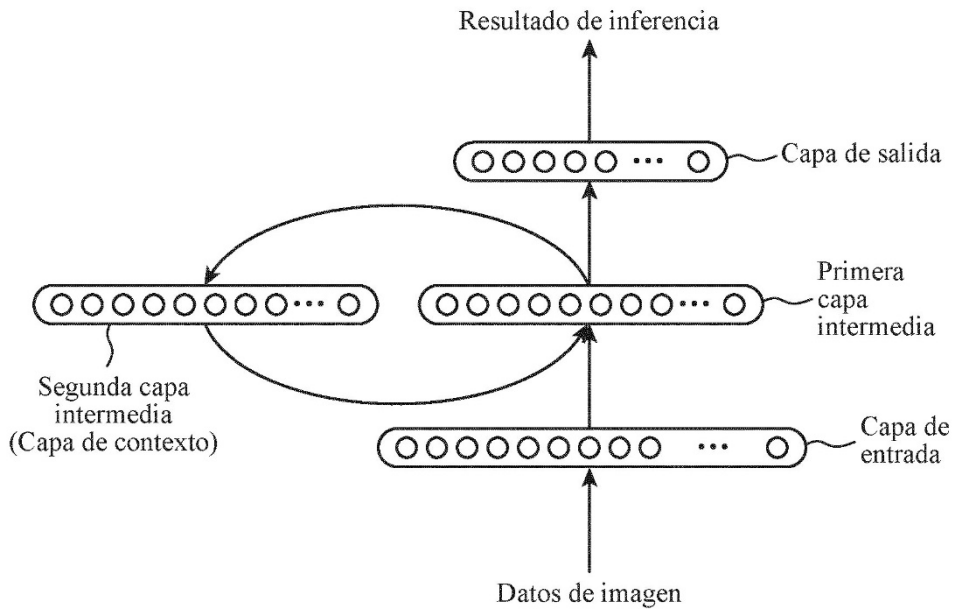


FIG.11

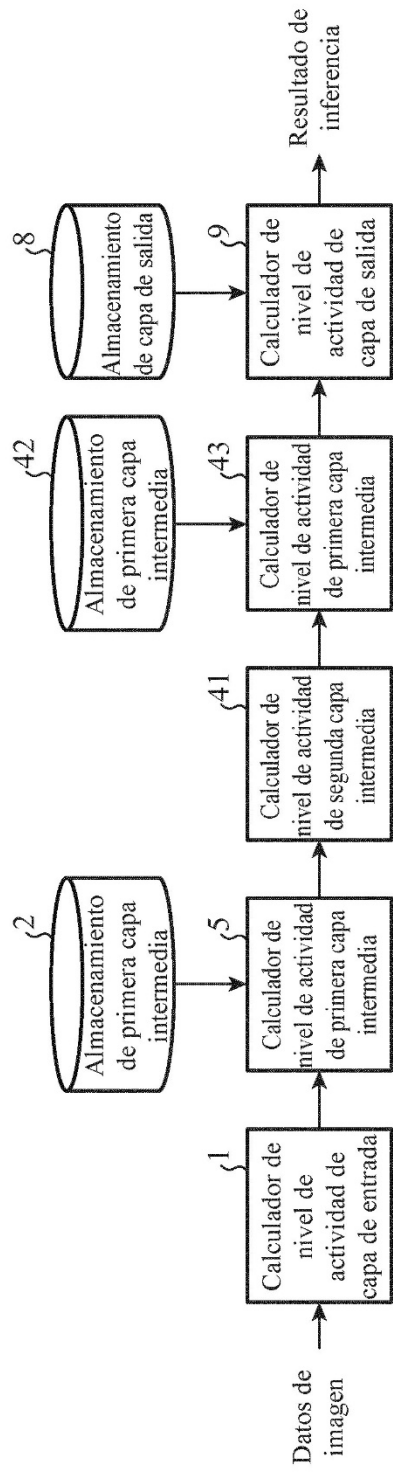


FIG.12

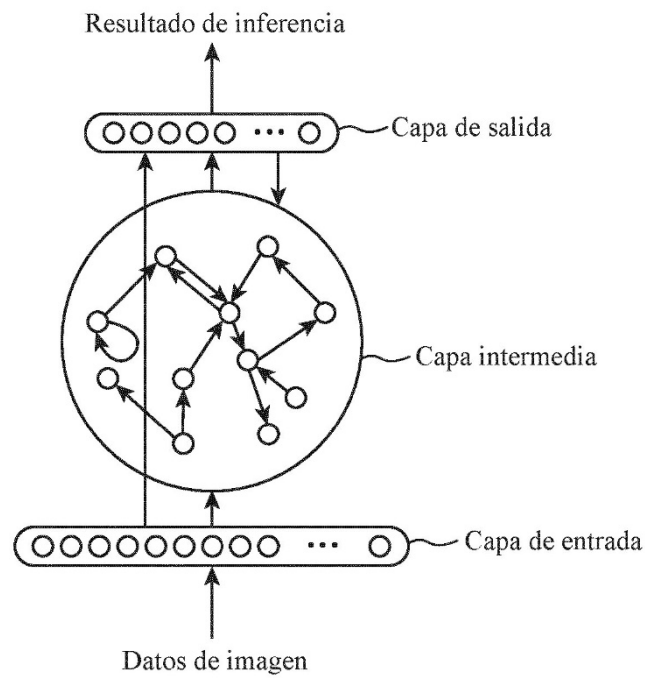


FIG.13

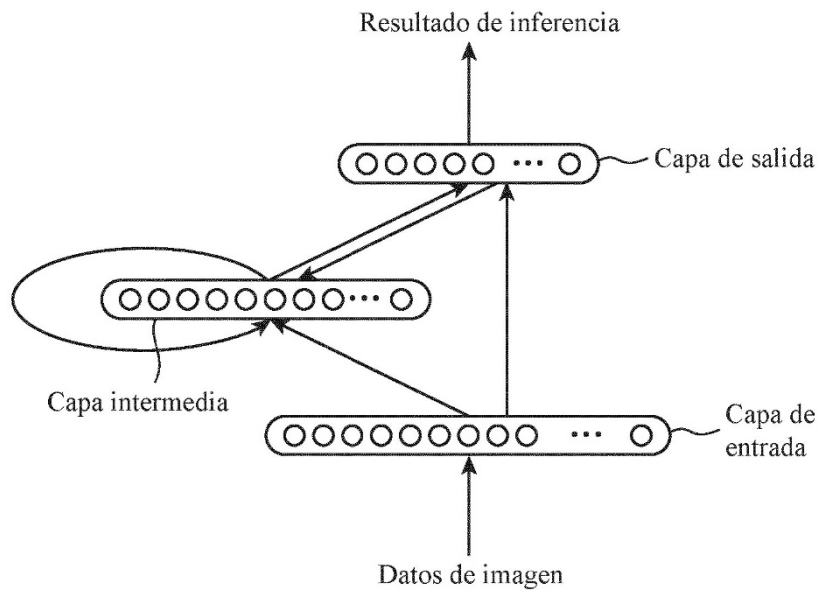


FIG.14

