

REPÚBLICA DE COLOMBIA
SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

Resolución N° **31181**

Ref. Expediente N° NC2019/0011014

Por la cual se otorga una Patente de Invención

EL SUPERINTENDENTE DE INDUSTRIA Y COMERCIO

en ejercicio de sus facultades legales, en especial de las conferidas en el numeral 24 del artículo 3° del Decreto 4886 de 2011, y

CONSIDERANDO

Que mediante escrito radicado en esta Superintendencia el 4 de octubre de 2019, con el N° NC2019/0011014, por MICROSOFT TECHNOLOGY LICENSING, LLC, entró a fase nacional la solicitud de patente de invención titulada “PROCESADOR DE RED NEURONAL QUE UTILIZA COMPRESIÓN Y DESCOMPRESIÓN DE DATOS DE ACTIVACIÓN PARA REDUCIR LA UTILIZACIÓN DE ANCHO DE BANDA DE MEMORIA”, que corresponde a la solicitud internacional PCT/US2018/027840 del 16 de abril de 2018.

Que la solicitud fue publicada en la Gaceta de la Propiedad Industrial N° 875 el 21 de octubre de 2019, sin que se hubieran presentado oposiciones por parte de terceros.

Que en virtud de lo dispuesto en el artículo 14 de la Decisión 486 expedida por la Comisión de la Comunidad Andina “*Los países miembros otorgarán patentes para las invenciones, sean de producto o de procedimiento, en todos los campos de la tecnología, siempre que sean nuevas, tengan nivel inventivo y sean susceptibles de aplicación industrial*”.

Que en el presente caso las reivindicaciones 1 a 15 incluidas en el radicado bajo el N° NC2019/0011014 el 4 de octubre de 2019, cumplen los requisitos indicados en el considerando anterior, toda vez que se refieren a un procesador de red neuronal y un método implementado por computadora que difieren del estado de la técnica más cercano, US20120130928, en que la invención incluye generar una porción de máscara de un fragmento de salida comprimido, la porción de máscara comprende una cantidad de bits equivalente a la cantidad fija de bytes en el fragmento de datos descomprimido, cada bit en la porción de máscara corresponde a un byte en el fragmento de datos descomprimido, y en donde cada bit en la porción de máscara se establece en cero lógico donde un byte correspondiente en el fragmento de datos descomprimido es cero y se establece en uno lógico donde un byte correspondiente en el fragmento de datos descomprimido es distinto de cero, generar un fragmento de datos del fragmento de salida comprimido, determinar, con base en la cantidad de bytes distintos de cero en el fragmento de datos descomprimido, una cantidad de bits en la porción de datos del fragmento de salida comprimido disponible para almacenar bytes distintos de cero truncados del fragmento de datos descomprimido y truncar los bytes distintos de cero en el fragmento de datos descomprimido a la cantidad determinada de bits.

Adicionalmente, estas diferencias no se encuentran sugeridas en el estado de la técnica y, como consecuencia de ello, se evidencia el efecto técnico de comprimir el fragmento de datos descomprimido, codificar la presencia y ubicación de los bytes cero y no cero en el fragmento de datos descomprimido y generar la porción de datos del fragmento de salida comprimido, permitiendo reducir la utilización del ancho de banda del bus de



Resolución N° **31181**

Ref. Expediente N° NC2019/0011014

memoria en un módulo de DNN, permitir que un módulo de DNN complete las operaciones de procesamiento más rápidamente y reducir el consumo de energía.

En consecuencia, las reivindicaciones 1 a 15 cumplen los requisitos de novedad, nivel inventivo y aplicación industrial establecidos en la normatividad citada en precedencia y este Despacho encuentra procedente conceder para las misma la patente solicitada.

Que el solicitante en el escrito radicado bajo el N° NC2019/0011014 el 4 de octubre de 2019, tituló la invención como: “PROCESADOR DE RED NEURONAL QUE UTILIZA COMPRESIÓN Y DESCOMPRESIÓN DE DATOS DE ACTIVACIÓN PARA REDUCIR LA UTILIZACIÓN DE ANCHO DE BANDA DE MEMORIA”. Sin embargo, no se acepta porque no relaciona una de las modalidades referente al método implementado por computadora de la materia reclamada y de acuerdo con el objeto concedido, el título de la invención quedará de la siguiente manera: “PROCESADOR DE RED NEURONAL Y MÉTODO IMPLEMENTADO POR COMPUTADORA QUE UTILIZA COMPRESIÓN Y DESCOMPRESIÓN DE DATOS DE ACTIVACIÓN PARA REDUCIR LA UTILIZACIÓN DE ANCHO DE BANDA DE MEMORIA”.

Con fundamento en las anteriores consideraciones, el Superintendente de Industria y Comercio,

RESUELVE

ARTÍCULO 1: Otorgar patente de invención a la solicitud que entró en fase nacional en virtud del Tratado de Cooperación en Materia de Patentes (PCT), para la creación titulada:

“PROCESADOR DE RED NEURONAL Y MÉTODO IMPLEMENTADO POR COMPUTADORA QUE UTILIZA COMPRESIÓN Y DESCOMPRESIÓN DE DATOS DE ACTIVACIÓN PARA REDUCIR LA UTILIZACIÓN DE ANCHO DE BANDA DE MEMORIA”

Clasificación IPC: G08B 13/196, G06N 3/063, H03M 7/30, H03M 7/46.

Reivindicación(es): 1 a 15 incluidas en el radicado bajo el N° NC2019/0011014 el 4 de octubre de 2019, de acuerdo con el anexo 1.

Titular(es): MICROSOFT TECHNOLOGY LICENSING, LLC.

Domicilio(s): ATTN: INTERNATIONAL PATENTS, ONE MICROSOFT WAY, REDMOND, WASHINGTON, 98052-6399, ESTADOS UNIDOS DE AMÉRICA.

Inventor(es): Joseph Leon CORKERY, Benjamin Eliot LUNDELL, Larry Marvin WALL, Chad Balling MCBRIDE, Amol Ashok AMBARDEKAR, George PETRE, Kent D. CEDOLA y Boris BOBROV.

Prioridad(es) N°: 17/04/2017, US (ESTADOS UNIDOS DE AMÉRICA), 62/486,432.
13/04/2018, US (ESTADOS UNIDOS DE AMÉRICA), 15/953,356.

Solicitud internacional N°: PCT/US2018/027840

Fecha: 16 de abril de 2018.

Vigente desde: 16 de abril de 2018

Hasta: 16 de abril de 2038.



Resolución N° **31181**

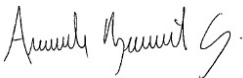
Ref. Expediente N° NC2019/0011014

ARTÍCULO 2: El titular tendrá los derechos y las obligaciones establecidos en la Decisión 486 de la Comisión de la Comunidad Andina y en las demás disposiciones legales vigentes sobre propiedad industrial, precisando que para mantener vigente la patente se deberá cancelar la tasa anual de mantenimiento, conforme lo dispone el artículo 80 de la referida norma comunitaria.

ARTÍCULO 3: Notificar el contenido de la presente resolución a MICROSOFT TECHNOLOGY LICENSING, LLC, advirtiéndole que contra ella procede el recurso de reposición, ante el Superintendente de Industria y Comercio, el cual podrá ser interpuesto en el momento de la notificación o dentro de los diez (10) días hábiles siguientes a ella.

NOTIFÍQUESE Y CÚMPLASE

Dada en Bogotá D.C., el 23 de mayo de 2022



ANDRÉS BARRETO GONZÁLEZ
SUPERINTENDENTE DE INDUSTRIA Y COMERCIO

Firmado digitalmente
por: ANDRÉS
BERNARDO BARRETO
GONZÁLEZ
Fecha: 2022.05.23
09:00:35 COT
Razón: Validez jurídica.
Ubicación: Bogotá,
Colombia.



ANEXO 1

REIVINDICACIONES CONCEDIDAS

1. Un procesador de red neuronal, que comprende:
 - una o más neuronas; y
 - una unidad de compresión configurada para recibir un fragmento de datos descomprimido generado por al menos una de las neuronas en el procesador de red neuronal, el fragmento de datos descomprimido comprende una cantidad fija de bytes;
 - generar una porción de máscara de un fragmento de salida comprimido, la porción de máscara comprende una cantidad de bits equivalente a la cantidad fija de bytes en el fragmento de datos descomprimido, cada bit en la porción de máscara corresponde a un byte en el fragmento de datos descomprimido, y en donde cada bit en la porción de máscara se establece en cero lógico donde un byte correspondiente en el fragmento de datos descomprimido es cero y se establece en uno lógico donde un byte correspondiente en el fragmento de datos descomprimido es distinto de cero;
 - generar un fragmento de datos del fragmento de salida comprimido al determinar una cantidad de bytes distintos de cero en el fragmento de datos descomprimido,
 - determinar, con base en la cantidad de bytes distintos de cero en el fragmento de datos descomprimido, una cantidad de bits en la porción de datos del fragmento de salida comprimido disponible para almacenar bytes distintos de cero truncados del fragmento de datos descomprimido,
 - truncar los bytes distintos de cero en el fragmento de datos descomprimido a la cantidad determinada de bits, y
 - almacenar los bytes distintos de cero truncados en la porción de datos del fragmento de salida comprimido; y
 - producir el fragmento de salida comprimido, el fragmento de salida comprimido comprende la porción de máscara y la porción de datos.
2. El procesador de red neuronal de acuerdo con la reivindicación 1, en donde el procesador de red neuronal comprende, además, una unidad de descompresión configurada para:
 - recibir el fragmento de salida comprimido;
 - determinar la cantidad de bytes distintos de cero en la porción de datos del fragmento de datos descomprimido con base en la porción de máscara del fragmento de salida comprimido;
 - determinar ubicaciones de los bytes distintos de cero en el fragmento de datos descomprimido con base en la porción de máscara del fragmento de salida comprimido;
 - determinar la cantidad de bits utilizados por la unidad de compresión para almacenar los bytes distintos de cero truncados en la porción de datos del fragmento de salida comprimido;
 - para cada posición en la porción de máscara del fragmento de salida comprimido que es un cero lógico, insertar un byte cero en la posición correspondiente de un fragmento de salida descomprimido; y
 - para cada posición en la porción de máscara que es un uno lógico, insertar el byte distinto de cero truncado desde la posición correspondiente del fragmento de entrada comprimido en una posición correspondiente del fragmento de salida descomprimido y una cantidad de bits cero equivalente a la cantidad de bits truncados durante la compresión del

Resolución N° **31181**

Ref. Expediente N° NC2019/0011014

fragmento de salida comprimido.

3. El procesador de red neuronal de acuerdo con la reivindicación 1, en donde la unidad de compresión está configurada, además, para:
determinar una cantidad de bits adicionales en la porción de datos del fragmento de salida comprimido disponible para almacenar bytes distintos de cero truncados del fragmento de datos descomprimido; y
asignar los bits adicionales a uno o más bytes distintos de cero en el fragmento de datos descomprimido antes de truncar el uno o más bytes distintos de cero.

4. El procesador de red neuronal de acuerdo con la reivindicación 3, en donde la unidad de descompresión está configurada, además, para determinar la cantidad de bits adicionales asignados al uno o más de los bytes distintos de cero almacenados en la porción de datos del fragmento de salida comprimido.

5. El procesador de red neuronal de acuerdo con la reivindicación 2, en donde la unidad de descompresión está configurada, además, para agregar un desplazamiento a uno o más de los bytes distintos de cero truncados almacenados en el fragmento de salida descomprimido.

6. El procesador de red neuronal de acuerdo con la reivindicación 1, en donde se truncan uno o más bits menos significativos (LSB) de los bytes distintos de cero.

7. Un procesador de red neuronal, que comprende:
una o más neuronas; y
una unidad de descompresión configurada para
recibir un fragmento de datos comprimido que comprende una porción de máscara y una porción de datos;
determinar una cantidad de bytes distintos de cero en un fragmento de datos descomprimido con base en bits en la porción de máscara;
determinar, con base, al menos en parte, en la cantidad de bytes distintos de cero, una cantidad de bits utilizados para almacenar bytes distintos de cero truncados en la porción de datos del fragmento de datos de salida comprimido;
para cada posición de bit en la porción de máscara del fragmento de datos comprimido que es un cero lógico, insertar un byte cero en la posición correspondiente del fragmento de datos descomprimido; y
para cada posición en la porción de máscara del fragmento de datos comprimido que es un uno lógico, insertar un byte distinto de cero truncado desde la posición correspondiente en la porción de datos del fragmento de datos comprimido en una posición correspondiente en el fragmento de datos descomprimido y una cantidad de cero bits equivalente a una cantidad de bits truncados durante la compresión del fragmento de datos comprimido.

8. El procesador de red neuronal de acuerdo con la reivindicación 7, que comprende, además, una unidad de compresión configurada para:
recibir un fragmento de datos descomprimido generado por al menos una de las neuronas en el procesador de red neuronal, el fragmento de datos descomprimido comprende una cantidad fija de bytes;
generar la porción de máscara del fragmento de datos comprimido, la porción de máscara comprende una cantidad de bits equivalente a la cantidad fija de bytes en el fragmento de datos descomprimido, cada bit en la porción de máscara corresponde a un byte en el fragmento de datos descomprimido, y en donde cada bit en la porción de máscara

Resolución N° **31181**

Ref. Expediente N° NC2019/0011014

comprende un cero lógico donde un byte correspondiente en el fragmento de datos descomprimido es cero y comprende un uno lógico donde un byte correspondiente en el fragmento de datos descomprimido es distinto de cero;
generar la porción de datos del fragmento de datos comprimido al
determinar una cantidad de bytes distintos de cero en el fragmento de datos descomprimido,
determinar, con base en la cantidad de bytes distintos de cero en el fragmento de datos descomprimido, una cantidad de bits en la porción de datos del fragmento de datos comprimido disponible para almacenar bytes distintos de cero truncados del fragmento de datos descomprimido,
truncar los bytes distintos de cero en el fragmento de datos descomprimido a la cantidad determinada de bits, y
almacenar los bytes distintos de cero truncados en la porción de datos del fragmento de datos comprimido; y
generar el fragmento de datos comprimido, el fragmento de datos comprimido comprende la porción de máscara y la porción de datos.

9. El procesador de red neuronal de acuerdo con la reivindicación 8, en donde la unidad de compresión está configurada, además, para almacenar los bytes distintos de cero en el fragmento de datos descomprimido en la porción de datos del fragmento de datos comprimido sin truncamiento si la cantidad de bytes distintos de cero en el fragmento de datos descomprimido es menor o igual que una cantidad de bytes en la porción de datos del fragmento de datos comprimido.

10. El procesador de red neuronal de acuerdo con la reivindicación 8, en donde la unidad de compresión está configurada, además, para:
determinar una cantidad de bits adicionales en la porción de datos del fragmento de salida comprimido disponible para almacenar bytes distintos de cero truncados del fragmento de datos descomprimido; y
asignar los bits adicionales a uno o más bytes distintos de cero en el fragmento de datos descomprimido antes de truncar el uno o más bytes distintos de cero.

11. El procesador de red neuronal de acuerdo con la reivindicación 8, en donde la unidad de descompresión está configurada, además, para determinar la cantidad de bits adicionales asignados al uno o más de los bytes distintos de cero almacenados en la porción de datos del fragmento de salida comprimido.

12. Un método implementado por computadora, que comprende:
recibir, en una unidad de compresión de un procesador de red neuronal, un fragmento de datos descomprimido generado por al menos una neurona en el procesador de red neuronal, el fragmento de datos descomprimido que comprende una cantidad fija de bytes;
generar una porción de máscara de un fragmento de salida comprimido, la porción de máscara comprende
una cantidad de bits equivalente a la cantidad fija de bytes en el fragmento de datos descomprimido,
cada bit en la porción de máscara corresponde a un byte en el fragmento de datos descomprimido, y
en donde cada bit en la porción de máscara comprende un cero lógico donde un byte correspondiente en el fragmento de datos descomprimido es cero y comprende un uno lógico donde un byte correspondiente en el fragmento de datos descomprimido es distinto de cero;
generar una porción de datos del fragmento de salida comprimido al



Resolución N° **31181**

Ref. Expediente N° NC2019/0011014

determinar una cantidad de bytes distintos de cero en el fragmento de datos descomprimido,
determinar, con base en la cantidad de bytes distintos de cero en el fragmento de datos descomprimido, una cantidad de bits en la porción de datos del fragmento de salida comprimido disponible para almacenar bytes distintos de cero truncados del fragmento de datos descomprimido,
truncar los bytes distintos de cero en el fragmento de datos descomprimido a la cantidad determinada de bits, y
almacenar los bytes distintos de cero truncados en la porción de datos del fragmento de salida comprimido; y
almacenar el fragmento de salida comprimido en una memoria del procesador de red neuronal, el fragmento de salida comprimido comprende la porción de máscara y la porción de datos.

13. El método implementado por computadora de acuerdo con la reivindicación 12, que comprende, además:
determinar una cantidad de bits adicionales en la porción de datos del fragmento de salida comprimido disponible para almacenar bytes distintos de cero truncados del fragmento de datos descomprimido; y
asignar los bits adicionales a uno o más bytes distintos de cero en el fragmento de datos descomprimido antes de truncar uno o más bytes distintos de cero.

14. El método implementado por computadora de acuerdo con la reivindicación 12, que comprende, además, almacenar los bytes distintos de cero en el fragmento de datos descomprimido en la porción de datos del fragmento de datos comprimido sin truncamiento si la cantidad de bytes distintos de cero en el fragmento de datos descomprimido es menor o igual que una cantidad de bytes en la porción de datos del fragmento de datos comprimido.

15. El método implementado por computadora de acuerdo con la reivindicación 12, que comprende, además:
recibir, en una unidad de descompresión de un procesador de red neuronal, el fragmento de salida comprimido;
determinar la cantidad de bytes distintos de cero en la porción de datos del fragmento de datos descomprimido con base en la porción de máscara del fragmento de salida comprimido;
determinar las ubicaciones de los bytes distintos de cero en el fragmento de datos descomprimido con base en la porción de máscara del fragmento de salida comprimido;
determinar la cantidad de bits utilizados por la unidad de compresión para almacenar los bytes distintos de cero truncados en la porción de datos del fragmento de salida comprimido;
para cada posición de bit en la porción de máscara del fragmento de salida comprimido que es un cero lógico, insertar un byte cero en la posición correspondiente de un fragmento de salida descomprimido; y
para cada posición en la porción de máscara del fragmento de salida comprimido que es un uno lógico, insertar el byte distinto de cero truncado desde la posición correspondiente del fragmento de salida comprimido en una posición correspondiente del fragmento de salida descomprimido y una cantidad de bits cero equivalente a la cantidad de bits truncados durante la compresión del fragmento de salida comprimido.