(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 691 417**

(51) Int. CI.:

G06F 17/16 (2006.01)
G06F 17/30 (2006.01)
G06F 9/50 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **12.12.2011 PCT/US2011/064340**

(87) Fecha y número de publicación internacional: **28.06.2012 WO12087629**

(96) Fecha de presentación y número de la solicitud europea: **12.12.2011 E 11808992 (9)**

(97) Fecha y número de publicación de la concesión europea: **22.08.2018 EP 2656242**

(54) Título: **Sistemas y métodos para generar una matriz de productos cruzados en una sola pasada a través de datos utilizando nivelación de una sola pasada**

(30) Prioridad:

20.12.2010 US 972840

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
27.11.2018

(73) Titular/es:

SAS INSTITUTE INC. (100.0%)
Sas Campus Drive
Cary, NC 27513, US

(72) Inventor/es:

SCHABENBERGER, OLIVER y
GOODNIGHT, JAMES, HOWARD

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

DESCRIPCIÓN

Sistemas y métodos para generar una matriz de productos cruzados en una sola pasada a través de datos utilizando nivelación de una sola pasada

Campo técnico

La tecnología descrita en el presente documento se refiere, en general, a sistemas de procesamiento de datos y, más específicamente, a sistemas de procesamiento de datos que realizan análisis estadísticos.

Antecedentes

Las matrices de productos cruzados se generan frecuentemente por sistemas de procesamiento de datos que realizan análisis estadísticos, tales como los sistemas de procesamiento de datos que utilizan el método de mínimos cuadrados para ajustar los modelos lineales generales a los datos. En general, se puede formar una matriz ("matriz **X'X**") densa de productos cruzados, formando primero la fila **x** para la observación actual y luego agregando el producto externo a la matriz **X'X** calculada hasta el momento. Matemáticamente, esto se puede expresar como:

$$\mathbf{X'X} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'$$

donde *n* indica el número de observaciones, la matriz **X'X** es de orden (*p* x *p*) y el vector $\mathbf{x}_i$ es de orden (*p* x 1).

Los algoritmos de múltiples pasadas para resolver tales matrices pueden utilizarse en situaciones no limitativas, como cuando los elementos de $\mathbf{x}_i$ dependen de elementos en $\mathbf{x}_j$ (donde j es diferente de i). En este tipo de situaciones, es habitual calcular la matriz **X'X** en múltiples pasadas a través de los datos. Por ejemplo, en una primera pasada, se puede calcular la información necesaria para construir posteriormente el vector $\mathbf{x}_i$ para cualquier observación y luego calcular la matriz de productos cruzados en una segunda pasada.

Como otro escenario no limitativo, los algoritmos de múltiples pasadas se utilizan cuando las columnas de la matriz **X** dependen de variables de clasificación. Las variables de clasificación son variables cuyos valores sin procesar se asignan a una codificación de enteros. Por ejemplo, un estudio de una especie de pez podría incluir una variable de clasificación por género con tres categorías: *macho*, *hembra* y *no determinado*. Si un efecto de género está en un modelo estadístico con respecto al estudio (es decir, ocupa columnas en la matriz **X**), se requerirá el conocimiento de una serie de factores para construir la matriz **X**. Tales factores podrían incluir: (i) la cantidad de niveles del efecto del género que están representados en los datos; (ii) el orden adecuado para estos niveles; y (iii) la posición de la primera columna del efecto del género en la matriz **X** – es decir, qué otros términos preceden al efecto del género en el modelo y cuántas columnas ocupan.

El análisis estadístico con variables de clasificación en los efectos de modelo es común en una serie procedimientos de SAS/STAT®, tales como GLM, GENMOD, GLIMMIX, GLMSELECT, LOGISTIC, MIXED y PHREG. Estos procedimientos construyen las filas de **X** en hasta *tres* pasadas a través de los datos. En la primera pasada se determinan los valores únicos de las variables de clasificación y su orden de clasificación. En una segunda pasada, se determinan los niveles de los efectos en los que están involucradas las variables de clasificación. Finalmente, en una tercera pasada, se construye la fila de **X** (es decir, $\mathbf{x}_i$ para la *i*ª observación).

El documento US 2008/162409 A1 está dirigido al particionamiento de grandes colecciones de datos para que las consultas puedan ejecutarse en paralelo en dichas particiones mediante la explotación del paralelismo en forma de árbol. Se indica la aplicabilidad a la multiplicación de matrices, pero no se proporcionan detalles.

SAS: "Building the SSCP Matrix", Guía del usuario de SAS/STAT(R) 9.2, segunda edición, 30 de abril de 2010 (2010-04-30), página 1, está dirigido a la construcción de la matriz de suma de cuadrados y de productos cruzados (SSCP) en una sola pasada a través de los datos.

SHAFER, J; AGRAWAL, R. Y MEHTA, M.: "SPRINT: A Scalable Parallel Classifier for Data Mining", PROCEDIMIENTOS DE LA 22ª CONFERENCIA VLDB, 3 de septiembre de 1996 (1996-09-03), Bombay, India, está dirigido a la clasificación paralela utilizando árboles de decisión y división de listas de atributos. A cada uno de los nodos se le asigna una partición de cada una de las listas de atributos. Los nodos se ordenan para que las listas combinadas permanezcan ordenadas. Cada uno de los nodos produce sus histogramas locales en paralelo, los cuales luego se combinan (para encontrar la mejor división).

ROBERT A. COHEN: "SAS Meets Big Iron: High Performance Computing in SAS Analytic Procedures", 27ª CONFERENCIA ANUAL INTERNACIONAL DEL GRUPO DE USUARIOS DE SAS, 1 de enero de 2002 (2002-01-

01), páginas 1-9, Cary, NC, está dirigido al cálculo multihilo de matriz de SSCP, donde se utiliza memoria contigua dentro de cada uno de los hilos.

BRYAN MARKER ET AL: “Toward Scalable Matrix Multiply on Multithreaded Architectures”, 28 de agosto de 2007 (2007-08-28), PROCESAMIENTO PARALELO EURO-PAR 2007; [NOTAS DE LA CONFERENCIA EN LA CIENCIA DE LA COMPUTADORA], SPRINGER BERLIN HEIDELBERG, BERLIN, HEIDELBERG, PÁGINA(S) 748 - 757, ISBN: 978-3-540-74465-8, está dirigido al particionamiento para actualizaciones de rango k multihilo.

El documento WO 2009/143073 A1 (MATHWORKS INC [US]; LUSZCZEK PIOTR R [US]; LITTLE JOHN N [US]; HICKLIN) 26 de noviembre de 2009 (2009-11-26), da a conocer un entorno de computación paralelo adaptado para realizar formulaciones matriciales y/o vectoriales que pueden utilizarse para el análisis de datos, etc. Estas formulaciones matriciales y/o vectoriales pueden utilizarse en muchas áreas, tal como la estadística. Las unidades de software de ejecución y/o laboratorios pueden ejecutar las porciones del programa y pueden proporcionar resultados al controlador. El controlador puede combinar los resultados en un solo resultado.

Resumen

De acuerdo con las enseñanzas proporcionadas en el presente documento, se proporcionan sistemas, métodos y medios de almacenamiento legibles por computadora para un sistema de procesamiento de datos, que tiene múltiples hilos ejecutables, que está configurado para generar una matriz de productos cruzados en una sola pasada a través de los datos a ser analizados. Un sistema de ejemplo comprende memoria para recibir los datos a ser analizados, uno o más procesadores que tienen una pluralidad hilos ejecutables para ejecutar código para analizar datos y un código del software para generar una matriz de productos cruzados en una sola pasada a través de los datos a ser analizados. El código del software incluye código de nivelación de variable en hilos para generar una pluralidad de árboles binarios específicos de hilo para una pluralidad de variables de clasificación, código de combinación de árbol de variable para combinar una pluralidad de árboles específicos de hilo en una pluralidad de árboles generales para la pluralidad de variables de clasificación, código de nivelación del efecto para generar una pluralidad de submatrices de la matriz de productos cruzados, utilizando la pluralidad de árboles generales para la pluralidad de variables de clasificación y código de generación de matriz de productos cruzados para generar la matriz de productos cruzados almacenando y ordenando los elementos de las submatrices en espacio de memoria contiguo.

Breve descripción de los dibujos

La Fig. **1** es un diagrama de bloques que muestra un entorno de ejemplo en el que los usuarios pueden interactuar con un entorno de computación que puede realizar análisis estadístico.

Las Fig. **2-4** son diagramas de bloques que muestran componentes de hardware y de software de ejemplo de sistemas de procesamiento de datos para generar una matriz de productos cruzados.

La Fig. **5** es un diagrama de bloques que muestra componentes de hardware de ejemplo de un sistema de procesamiento de datos que utiliza múltiples hilos de computación para realizar la nivelación de variable.

La Fig. **6** es un diagrama de flujo de proceso que muestra un escenario operacional de ejemplo que involucra un sistema de procesamiento de datos para realizar la nivelación de variable.

La Fig. **7** es un diagrama de flujo de proceso que muestra un escenario operacional de ejemplo que involucra un sistema de procesamiento de datos para combinar el análisis realizado por múltiples hilos de computación.

La Fig. **8** es un diagrama de flujo de proceso que muestra un escenario operacional de ejemplo que involucra un sistema de procesamiento de datos para realizar la nivelación de efecto.

La Fig. **9** es un diagrama de flujo de proceso que muestra un escenario operacional de ejemplo que involucra un sistema de procesamiento de datos para ensamblar una matriz de productos cruzados.

Las Fig. **10-11** son diagramas de bloques que ilustran componentes de hardware y de software de sistemas de procesamiento de datos para generar continuamente una matriz de productos cruzados a medida que se transmiten datos.

La Fig. **12** es un diagrama de bloques que muestra componentes de hardware de ejemplo y flujo de datos de ejemplo en un sistema de procesamiento de datos que realiza nivelación de variable.

La Fig. **13** es un diagrama de flujo de proceso que muestra un flujo de datos de ejemplo en un proceso para combinar el análisis realizado por dos hilos de computación.

Las Fig. **14-15** son diagramas de flujo de proceso que representan un flujo de datos de ejemplo en un proceso para realizar la nivelación de efecto.

La Fig. **16** es un diagrama de flujo de proceso que muestra un flujo de datos de ejemplo en un proceso para ensamblar una matriz de productos cruzados.

Descripción detallada

La Fig. **1** muestra en **30** un entorno de computación para procesar grandes cantidades de datos para muchos tipos diferentes de aplicaciones, tales como aplicaciones científicas, técnicas o comerciales. Una o más computadoras **32** de usuario pueden interactuar con el entorno **30** de computación a través de varias maneras, incluida una red **34**. Uno o más almacenes **36** de datos pueden estar acoplados al entorno **30** de computación para almacenar los datos a ser procesados por el entorno **30** de computación, así como para almacenar cualquier dato intermedio o final generado por el entorno de computación.

Una aplicación de ejemplo para el entorno **30** de computación implica el rendimiento del análisis estadístico. Con frecuencia, en el análisis estadístico, se generan modelos para conjuntos de datos y las matrices de productos cruzados (**"X'X"**) se generan durante el proceso de modelado por los sistemas de procesamiento de datos en el entorno **30** de computación que realizan el análisis estadístico. Los modelos involucran variables y los efectos de esas variables reflejados en los datos.

Los efectos en el contexto de la formación **X'X** son estructuras matemáticas lineales – es decir, un efecto está asociado con ciertas columnas de la matriz **X**. Excepto para tokens y palabras clave especialmente definidas (como "Interceptar"), los efectos dependen de las variables. Un efecto generalmente incluye una o más variables que contribuyen al efecto.

Dos tipos de variables que impactan en efectos son variables continuas y de clasificación. Una variable continua es una variable numérica y los valores sin procesar de la variable se utilizan en la construcción de los efectos. Por ejemplo, las alturas y los pesos de los sujetos son variables continuas.

Una variable de clasificación es una variable numérica o de carácter, cuyos valores sin procesar se utilizan indirectamente en la formación de la contribución del efecto. Los valores de una variable de clasificación se denominan ***niveles***. Por ejemplo, la variable de clasificación Sexo tiene los niveles "masculino" y "femenino". Durante la formación **X'X**, los valores de la variable de clasificación se asignan a valores enteros que representan niveles de la variable. El proceso de asignar los valores de la variable de clasificación a un nivel, en el presente documento se denomina nivelación de variable. Estos niveles de clasificación de las variables se utilizan para definir los niveles del efecto. El proceso de asignación de los niveles del efecto, en el presente documento se denomina nivelación de efecto.

Los efectos que involucran variables de clasificación, ocupan una o más columnas en la matriz **X**. El número exacto de columnas para un efecto de clasificación depende de los valores de las variables involucradas, de las reglas de asignación para la nivelación de variable y de cualquier operador de efecto.

Para un efecto principal, los niveles del efecto son típicamente los niveles de la variable de clasificación, a menos que todas las observaciones asociadas con un nivel particular de la variable no sean utilizables en el análisis. Para un efecto que contiene más de una variable de clasificación, los efectos del nivel dependen de los niveles de las variables de clasificación que se producen juntos en los datos.

Aunque, en muchos escenarios, la nivelación de una variable puede realizarse sin conocer niveles o valores de otras variables, la nivelación de efectos, sin embargo, no puede realizarse sin conocer los niveles de todas las variables en el efecto. A diferencia de muchos sistemas actuales de procesamiento de datos que implementan algoritmos de nivelación que requieren que los datos se lean varias veces, el entorno **30** de computación incluye un sistema de procesamiento de datos que puede realizar la nivelación de variable y de efecto en una sola pasada a través de los datos.

La Fig. **2** muestra un ejemplo de sistema de procesamiento de datos para construir una Matrix **100 X'X** en una sola pasada a través de datos, que incluyen datos de variable de clasificación. El sistema de procesamiento de datos de ejemplo, incluye uno o más procesadores de datos (no mostrados), que tienen una cantidad de hilos de ejecución que son capaces de realizar independientemente pasos de análisis de datos, un búfer **102** de datos para recibir

datos de un almacén **36** de datos y un motor **110** de nivelación de una sola pasada. El motor **110** de nivelación de una sola pasada, en este ejemplo, incluye un componente o código **112** de software de nivelación de variable en hilos, un componente o código **114** de software de combinación de árbol de variable, un componente o código **116** de software de nivelación de efecto, instrucciones **117** de decisión y un ensamblaje de matriz **X'X** de componente de software o un código **118** de generación de matriz de productos cruzados.

En funcionamiento, el motor **110** de nivelación de una sola pasada puede generar una **matriz 100 X'X** en una sola pasada a través de los datos en el búfer **102** de datos. Después de leerse los datos del búfer **102** de datos, uno o más hilos de ejecución ejecutan instrucciones desde el componente **112** de software de nivelación de variable en hilos. Al utilizar múltiples hilos, el tiempo de procesamiento de los componentes o el código del software de nivelación de variable en hilos puede reducirse en plataformas de múltiples núcleos o de híper hilos, porque múltiples hilos pueden ejecutarse simultáneamente. La ganancia de rendimiento de la ejecución de múltiples hilos debe superar el costo computacional de combinar los árboles específicos de hilo en un árbol general al final de la nivelación de variable. Los resultados generados por el componente **112** de software de nivelación de variable en hilos se proporcionan como entrada para el componente **114** de software de combinación de árbol de variable. Los resultados generados al ejecutar las instrucciones del componente **114** de software de combinación de árbol de variable se proporcionan, a su vez, como entrada para el componente **116** de software de nivelación de efecto. Las instrucciones **117** de decisión se ejecutan, las cuales determinan si existen datos adicionales para procesar en el búfer **102** de datos, antes de proceder a ensamblar una matriz **X'X**. Si existen datos adicionales, los datos se leen del búfer **102** de datos y el control del proceso se devuelve al componente **112** de software de nivelación de variable en hilos. Si no existen datos adicionales, los resultados generados al ejecutar las instrucciones del componente **116** de software de nivelación de efecto se proporcionan al componente **118** de software de ensamblaje de matriz **X'X**, el cual ensambla una matriz **100 X'X**.

La Fig. 3 muestra, con más detalle, un sistema de procesamiento de datos de ejemplo para construir una matriz **100 X'X** en una sola pasada a través de datos que incluyen datos de variable de clasificación. Este sistema de procesamiento de datos de ejemplo, también incluye un procesador de datos (no mostrado) que tiene una serie de hilos de ejecución que son capaces de realizar independientemente pasos de análisis de datos, un búfer **102** de datos para recibir datos desde un almacén **36** de datos y un componente **110** de software de nivelación de una sola pasada. El motor **110** de nivelación de una sola pasada, en este ejemplo, incluye una pluralidad de subcomponentes **112a - 112n** de software de nivelación de variable en hilos, un componente **114** de software de combinación de árbol de variable, un componente **116** de software de nivelación de efecto, instrucciones **117** de decisión y un ensamblaje de matriz **X'X** de componente **118** de software. Con el uso de instrucciones de decisión, no es necesario procesar el conjunto de datos de entrada entero de una vez. En cambio, el sistema puede operar en búferes de datos separados, por ejemplo, en una aplicación donde los datos se proporcionan en ciertos puntos de tiempo. Los componentes de nivelación de variable pueden construir la información necesaria para los pasos aguas abajo, tal como la formación de la matriz **X'X** – un búfer cada vez. No es necesario mantener todos los datos en memoria.

En el ejemplo de la Fig. 3, las instrucciones **120** de procesador se proporcionan para leer un nuevo búfer de datos desde el búfer **102** de datos y para ingresar esos datos en los subcomponentes **112a - 112n** de software de nivelación de variable en hilos. Cada uno de los subcomponentes de software de nivelación de variable en hilos se ejecuta mediante un hilo de ejecución independiente del procesador, el cual resulta en la generación de árboles **122a - 122n** binarios específicos de hilo que describen características de las variables de clasificación encontradas en los datos. La formación de árboles binarios específicos de hilo separados, tiene las ventajas de que el árbol en un hilo puede formarse independientemente de los árboles en otros hilos. Si se formara un árbol común (en la etapa 122 de la Figura 3), entonces el árbol tendría que bloquearse cada vez que un hilo quisiera agregar un nuevo valor al árbol. Este bloqueo, esencialmente, serializaría el trabajo y reduciría la ventaja obtenida al permitir que los hilos operen sobre datos de forma independiente. Después de la generación, estos árboles **122a - 122n** binarios específicos de hilo se combinan mediante el componente **114** de software de combinación de árbol de variable para generar árboles **124** binarios generales para cada una de las variables de clasificación.

Después de completarse la nivelación de variable y de generarse los árboles **124** binarios generales para cada una de las variables de clasificación, los árboles **124** binarios se procesan mediante el componente **116** de software de nivelación de efecto, que genera submatrices **126a - 126m** parciales de la matriz general de productos cruzados utilizando los árboles **124** binarios generales. Se ejecutan instrucciones **117** de decisión que determinan si existen datos adicionales a ser procesados en el búfer **102** de datos antes de proceder a ensamblar una matriz **X'X**. Si existen datos adicionales, los datos se leen desde el búfer **102** de datos y el control del proceso se devuelve a las instrucciones **120** del procesador. Si no existen datos adicionales, entonces las submatrices **126a - 126m** parciales se proporcionan al componente **118** de software de ensamblaje de matriz **X'X**, que ensambla una **matriz 100 X'X**. El almacenamiento de los componentes de la matriz **X'X** eventual en submatrices parciales ofrece varias ventajas. Cuando las matrices se almacenan en una memoria de computadora separada, es fácil agregar filas y columnas a las submatrices. Es más complicado insertar filas/columnas en una matriz. El algoritmo construye las submatrices en

el orden en que aparecen los valores únicos de las variables en los datos. Si se recupera un nuevo búfer de datos, los nuevos valores llevarán a agregar filas/columnas a las submatrices, pero no conducirán a una inserción de filas o de columnas.

La Fig. **4** muestra otro sistema de procesamiento de datos de ejemplo para construir una matriz **100 X'X** en una sola pasada a través de datos que incluyen datos de variable de clasificación. Este sistema de procesamiento de datos de ejemplo contiene elementos similares al sistema de ejemplo mostrado en la Fig. **3**. El sistema de ejemplo mostrado en la Fig. **4**, sin embargo, incluye un motor de nivelación de efecto que podría ejecutarse mediante múltiples hilos. En el ejemplo mostrado, el hilo 0 de ejecución del procesador ejecuta el subcomponente **116a** de software de nivelación de efecto, el hilo 1 de ejecución del procesador ejecuta el subcomponente **116b** de software de nivelación de efecto y el hilo n de ejecución del procesador ejecuta el subcomponente **116n** de software de nivelación de efecto. Cada uno de los hilos de ejecución del procesador ejecuta instrucciones que resultan en la generación de una o más submatrices de la matriz **X'X** general.

La Fig. **5** muestra un sistema de ejemplo y la Fig. **6** muestra un proceso de ejemplo para generar árboles **122a -122c** específicos de hilo, utilizando el componente **112** de software de nivelación de variable en hilos (mostrado en las Fig. 2-4). El proceso comienza (paso **200** en la Fig. **6**) con un búfer **100** (Fig. 5) de datos sin procesar que contienen *k* observaciones que se pasan al código de nivelación. Si la nivelación se realiza en múltiples hilos, la memoria **100** de búfer se asigna a los hilos **130a -130c** (Fig. 5), de tal manera que cada uno de los hilos **130a -130c** procesa aproximadamente el mismo número de observaciones. En este ejemplo, la nivelación se realiza con tres hilos **130a - 130c** y cada uno de los hilos **130a -130c** procesa aproximadamente 1/3 de las observaciones. Este reparto en este ejemplo, incluye establecer el puntero de lectura de cada uno de los hilos en la posición correcta en el búfer **100**.

Cada uno de los hilos **130a -130c** examina cada una de las filas en el área **132a - 132c** de búfer asignada (paso **202**) y determina si la observación se utiliza para el análisis (paso **204**). Si se va a utilizar la observación, los valores sin procesar únicos para cada una de las variables se dividen en un árbol **122a - 122c** binario, que también contiene información auxiliar en cada uno de los nodos del árbol (paso **206**). Cada vez que se encuentra un nuevo valor sin procesar (paso **208**), se deriva un valor con formato (paso **210**), se deriva el número de observación en la aplicación general (paso **212**) y se actualiza la frecuencia con la que se produce el valor (paso **214**).

Alternativamente, en el paso **206**, los valores formateados se derivan para cada una de las observaciones, independientemente del valor sin procesar. En este ejemplo alternativo, se omite el paso **208**. Cada una de las observaciones utilizada en el análisis se asigna a un valor formateado, pero no se deriva un nuevo valor formateado para cada uno de los valores sin procesar único. Esta variación es útil cuando el número de valores sin procesar es mucho mayor que el número de valores formateados; por ejemplo, cuando una variable continua se agrupa en intervalos.

Después de leer y procesar la fila de datos asignada, se realiza una verificación para determinar si existen filas de datos asignadas adicionales que no se hayan procesado (paso **216**). Si es así, entonces se lee la fila adicional de datos (paso **218**) y se examina (paso **202**). Si no es así, los árboles binarios específicos de hilo para cada una de las variables de clasificación están completos (paso **220**).

La Fig. 7 muestra un proceso de ejemplo para generar un árbol **124** general para cada una de las variables de clasificación utilizando el componente **114** de software de combinación de árbol de variable. Después de que todos los hilos **130a -130c** (Fig. 5) hayan completado de agregar al árbol de las observaciones en su búfer, los árboles **122a -122c** específicos de hilo para cada una de las variables de clasificación se combinan en un árbol **124** general. Se pueden utilizar múltiples maneras para lograr esto, tal como acumular árboles en el árbol construido por el primer hilo.

Los árboles generales para cada una de las variables de clasificación retienen información con respecto al orden en que se vieron los valores sin procesar/formateados. En este ejemplo, para cada uno de los valores de una variable de clasificación, el nivel asociado de la variable corresponde al orden de los datos, es decir, los niveles de variables se organizan por el orden en que aparecen en los datos.

La Fig. **8** representa un proceso de ejemplo para realizar la nivelación de efecto. Los árboles generales para cada una de las variables de clasificación (**230**) generada por el componente de software de combinación de árbol de variable se utilizan para determinar los niveles para cada uno de los efectos (paso **232**). Debido a que los niveles de variable se organizaron según el orden en que aparecieron en los datos, los niveles de efecto también se organizarán según el orden en que aparecen en los datos (paso **232**). Además de determinar los niveles para cada uno de los efectos en el orden de los datos, se construyen submatrices parciales de la matriz **X'X** general (paso **234**).

Cada una de las submatrices se almacena por separado en la memoria y, a medida que se encuentran niveles adicionales en los datos, se pueden agregar nuevas filas y columnas al final del espacio de memoria utilizado, asignado a las submatrices. Por ejemplo, una submatriz C puede ser una matriz de 3x3 después de procesar un cierto número de observaciones y se convierte en una matriz de 4x4 después de procesar la siguiente observación. La información agregada a la 4ª fila y la 4ª columna se almacena en el espacio de memoria asignado a la submatriz C, después de la información que conforma las primeras tres filas y columnas en la submatriz C. Al almacenar las submatrices en memoria separada, las submatrices pueden crecer a medida que se detectan niveles adicionales en los datos.

Las submatrices parciales pueden ensamblarse como se ilustra en el siguiente ejemplo. Si, por ejemplo, hay tres efectos en un modelo, E_1, E_2 y E_3, la matriz **X'X** se puede construir a partir de seis submatrices de acuerdo con la siguiente tabla:

$$\mathbf{X'X} = \begin{bmatrix} \mathbf{X}'_{E1}\mathbf{X}_{E1} & & \\ \mathbf{X}'_{E1}\mathbf{X}_{E2} & \mathbf{X}'_{E2}\mathbf{X}_{E2} & \\ \mathbf{X}'_{E1}\mathbf{X}_{E3} & \mathbf{X}'_{E2}\mathbf{X}_{E3} & \mathbf{X}'_{E3}\mathbf{X}_{E3} \end{bmatrix}$$

Incluso si los efectos están en el orden de datos, la posición de la submatriz diagonal para $\mathbf{X'_{E2}X_{E2}}$ no puede determinarse sin conocer la dimensión de la submatriz $\mathbf{X'_{E1}X_{E2}}$ (o al menos sin saber el número de niveles en el efecto E_1). Sin embargo, si los niveles de variable y de efecto están en el orden de datos, un nuevo nivel de efecto E_2 llevará a la adición de una nueva fila/columna al final de la submatriz $\mathbf{X'_{E2}X_{E2}}$. El componente de software de nivelación de efecto mantiene las submatrices de la tabla **X'X** en una memoria no contigua y agrega filas y columnas al final, a medida que se encuentran nuevos niveles. En una sola realización, las submatrices son dispersas y el componente de software de nivelación de efecto hace que las submatrices se almacenen dispersamente, de tal manera que la memoria pueda crecer fácilmente, por ejemplo, manteniendo filas de matrices simétricas o rectangulares en subárboles binarios balanceados.

Una vez que se han construido las submatrices parciales, se realiza una verificación para determinar si hay un nuevo búfer de datos disponible para el análisis (paso **236**). Si se recibe un nuevo búfer de datos, el proceso comienza nuevamente con la nivelación de hilo de las variables (paso **238**). Si se determina que se han recibido todos los datos, la matriz **X'X** puede ensamblarse en un proceso de múltiples pasadas (paso **240**).

Ilustrado en la Fig. **9,** está un proceso de ejemplo para el ensamblaje de una matriz **X'X** a partir de las submatrices parciales. En este proceso, se reordenan los elementos de las submatrices parciales (**242**) generadas por el proceso de nivelación de efecto. Esto implica, determinar el orden del nivel de efecto en base a una solicitud de la aplicación cliente que inició el análisis de datos (paso **244**) y reordenar los niveles de las variables de clasificación y los efectos para cumplir con el orden solicitado especificado por la aplicación cliente. Si la aplicación cliente solicitó que las variables se ordenaran en el orden de datos, no es necesario reordenarlas. Sin embargo, si la aplicación cliente especificaba un orden diferente, los árboles de variable y los árboles de efecto deben reorganizarse adecuadamente para que coincidan con el orden especificado.

Con los niveles de variable y de efecto reasignados, en el paso **246**, la matriz **X'X 248** se forma en forma densa copiando elementos de las submatrices en la posición correcta del orden de nivel de la matriz **X'X** general. Como resultado, la matriz **X'X** se puede formar en una sola pasada a través de los datos en el búfer de datos. Como alternativa al paso **246**, la matriz **X'X 248** podría ensamblarse en una matriz dispersa utilizando cualquier número de métodos para la representación de la matriz dispersa.

Las Fig. **10** y **11** representan sistemas de ejemplo adicionales para formar una matriz **X'X** en una sola pasada a través de los datos. En estos ejemplos, los datos pueden transmitirse al sistema. El sistema puede generar una matriz **X'X** de manera similar a la descrita en los ejemplos anteriores. Los sistemas de ejemplo de las Fig. **10** y **11**, además, pueden recalcular la matriz **X'X** si se reciben datos adicionales después de la generación inicial de la matriz **X'X**.

Estos sistemas de ejemplo tienen instrucciones **140** que hacen que estos sistemas verifiquen periódicamente los nuevos datos en el búfer **102** de datos. Si se encuentran nuevos datos, se leen los nuevos datos, se actualizan los árboles específicos de hilo, se actualizan los árboles generales y se actualizan las submatrices generadas previamente. Debido a que las submatrices formadas en el proceso de nivelación de efecto se mantienen en espacios de memoria no contiguos y los niveles se mantienen en el orden de datos, a medida que se procesan nuevos datos, las filas y columnas de las submatrices pueden actualizarse y, nuevas filas y columnas pueden agregarse al final, para reflejar los nuevos datos. Después de actualizar las submatrices, los elementos de las submatrices parciales se reordenan, si es necesario. Con los niveles de variable y de efecto reasignados, la **matriz**

100 X'X se vuelve a formar en forma densa o dispersa copiando elementos de las submatrices en la posición correcta del orden de nivel en la matriz **X'X** general. Como resultado, la matriz **100 X'X** puede volver a formarse continuamente en una sola pasada, a medida que se transmiten nuevos datos al búfer **102** datos.

La Fig. **12** muestra un flujo de datos de ejemplo en un proceso de nivelación de variable de hilo. Como se ilustra, se proporciona una tabla **300** de ejemplo que contiene nueve observaciones a la memoria **100** de búfer. Las nueve observaciones incluyen datos relacionados con las variables de clasificación Género y Droga, y una variable de respuesta (Y). Las nueve observaciones en este ejemplo están asignadas a dos hilos. Las primeras cinco observaciones se asignan al primer hilo, como se muestra en **302a**, y las últimas cuatro observaciones se asignan al segundo hilo, como se muestra en **302b**. En este ejemplo, el hilo 1 maneja todas las observaciones con Género = "M", y el hilo 2 maneja todas las observaciones para Género = "F". Una intercepción de modelo, la cual se ha agregado en este ejemplo, se representa mediante una columna de "1".

Cuando se aplica el proceso de nivelación de variable de hilo, especialmente por los subcomponentes 112 [y/o los componentes 112a - 112n] de software de nivelación de variable, cada uno de los hilos genera un árbol binario específico del hilo en forma de un árbol de droga y de un árbol de género a partir de las observaciones asignadas a ello como se ilustra en **304a-d**. La codificación de nivel es independiente para cada uno de los hilos y el orden de los niveles para cada uno de los hilos es el orden en que se encontraron los niveles por el hilo particular. Las tablas mostradas en **304a-d** representan la información almacenada y gestionada en árboles binarios por el código.

Después del proceso de nivelación de variable de hilo, los árboles **304a-d** específicos de hilo se combinan, especialmente, mediante el componente 114 de software de combinación de árbol de variable, en árboles binarios generales en forma de un solo árbol **306 a-b** general para cada una de las variables de clasificación, como se ilustra en la Fig. **13**. El orden de los niveles en los árboles **306 a-b** generales, es en el orden de datos con respecto a todo el conjunto de datos. En este ejemplo, los dos hilos asignaron un nivel diferente para el valor "A" de la variable Droga. Debido a que el valor "A" tenía un número de observación más bajo en el hilo 1 que el valor "B" en el hilo 2, el valor "A" se asignó a un nivel más bajo en el árbol general para la variable Droga.

Las Fig. **14** y **15** continúan el flujo de datos de ejemplo a través de la etapa de nivelación de efecto, especialmente, mediante el componente 116 y/o los componentes 116a, 116b, 116n de software de nivelación de efecto. Los árboles de efecto generados en el proceso de nivelación de efecto, en este ejemplo, tienen la misma cantidad de niveles que los árboles de variable, como se ilustra en **308**. Debido a que hay cuatro efectos -- Intercepto, Género, Droga e Y-- en este ejemplo (ver **310**), se puede generar una matriz X'X, por ejemplo, mediante el componente 118 de software de ensamblaje de matriz y/o en los pasos 246 a partir de 10 submatrices: X'_IX_I, X'_GX_I, X'_DX_I, X'_YX_I, X'_GX_G, X'_DX_G, X'_YX_G, X'_DX_D, X'_YX_D, X'_YX_Y (como se ilustra en **310**), las cuales ocupan las ubicaciones en la Matriz X'X especificada en **310**.

Estas 10 submatrices se generan en memoria no contigua para permitir que crezcan según sea necesario. Cada una de las 10 submatrices se genera utilizando los árboles de nivelación de efecto y de variable y las 9 observaciones. En base a las 9 observaciones, en este ejemplo, las dimensiones de cada una de las submatrices son las siguientes X'_IX_I = [1x1], X'_DX_I = [3x1], X'_GX_I = [2x1], X'_YX_I = [1x1], X'_DX_D = [3x3], X'_GX_D = [2x3], X'_YX_D = [1x3], X'_GX_G = [2x2], X'_YX_G = [1x2] y X'_YX_Y = [1x1].

Para cada una de las 9 observaciones, se acumulan sus contribuciones a las diversas submatrices. Si se detectan niveles adicionales a partir de observaciones posteriores, las submatrices pueden expandirse porque se almacenan en memoria no contigua para permitir que se agreguen más filas y columnas según sea necesario cuando se detecten nuevos niveles de un efecto. Si las submatrices se "concatenaran" en una sola porción contigua de memoria, se obtendría la matriz **X'X** en el orden de datos mostrado en **312**.

Si están disponibles nuevos datos para el análisis, la **X'X** final se puede construir en memoria contigua, por ejemplo en el paso 246. Como se ilustra en la Fig. **16**, los niveles de efecto deben determinarse en base al orden especificado por la aplicación cliente (paso **244**). En este ejemplo, el orden correcto de los niveles de variable de clase se proporciona en la última columna de la tabla en **314**.

En este ejemplo, el modelo contiene sólo los efectos principales (sin interacciones) y el orden de nivel de efecto es el mismo que el orden de variable. Después de asignar un bloque de memoria suficiente para contener la matriz **X'X** (ahora se conoce el tamaño en base a todos los datos vistos anteriormente), los elementos de las submatrices se permutan a la ubicación correcta dentro de la matriz **X'X**, como se muestra en el ejemplo en **316**. Por lo tanto, se puede generar una matriz **X'X** como se ilustra en los ejemplos mencionados anteriormente.

Esta descripción escrita utiliza ejemplos para divulgar la invención, incluyendo el mejor modo y, también, para permitir que una persona experta en la técnica haga y utilice la invención. El alcance de la invención, como se define en las reivindicaciones, puede incluir otros ejemplos.

Por ejemplo, los métodos y sistemas descritos en el presente documento, pueden implementarse en muchos tipos diferentes de dispositivos de procesamiento mediante código de programa que comprende instrucciones de programa que son ejecutables por el subsistema de procesamiento del dispositivo. Las instrucciones del programa de software pueden incluir código fuente, código objeto, código máquina o cualquier otro dato almacenado que sea operable para hacer que un sistema de procesamiento realice los métodos y operaciones descritos en el presente documento. Sin embargo, también se pueden utilizar otras implementaciones, como firmware o incluso hardware diseñado adecuadamente, configurado para llevar a cabo los métodos y sistemas descritos en el presente documento.

Los datos de sistemas y de métodos (p. ej., las asociaciones, las asignaciones, la entrada de datos, la salida de datos, los resultados intermedios de datos, los resultados finales de datos, etc.) pueden almacenarse e implementarse en uno o más tipos diferentes de almacenes de datos implementados en computadora, tales como diferentes tipos de dispositivos de almacenamiento y construcciones de programación (p. ej., RAM, ROM, memoria Flash, archivos planos, bases de datos, estructuras de datos de programación, variables de programación, construcciones de declaraciones IF-THEN (o de tipo similar), etc.). Se observa que las estructuras de datos describen formatos para su uso en la organización y el almacenamiento de datos en bases de datos, programas, memoria u otros medios legibles por computadora para uso por un programa informático.

Los componentes de computadora, módulos de software, funciones, almacenes de datos y estructuras de datos descritas en el presente documento, pueden estar conectados directa o indirectamente el uno al otro con el fin de permitir el flujo de datos necesarios para sus operaciones. Los componentes del software y/o la funcionalidad pueden ubicarse en una sola computadora o distribuirse a través de múltiples computadoras, dependiendo de la situación actual.

Se debe entender que, tal como se utiliza en la descripción en el presente documento y en todas las reivindicaciones que siguen, el significado de “un”, “una” y “el” incluye la referencia plural a menos que el contexto dicte claramente lo contrario. Además, tal como se utiliza en la descripción en el presente documento y en todas las reivindicaciones que siguen, el significado de “en” incluye “en” y “sobre” a menos que el contexto indique claramente lo contrario.

REIVINDICACIONES

1. Un método implementado por computadora para generar una matriz X'X de productos cruzados, donde la matriz X es representativa de un modelo lineal que involucra variables y los efectos de esas variables, cada uno de los efectos está asociado con ciertas columnas de la matriz X, donde las columnas de la matriz X dependen de las variables de clasificación, en donde las variables de clasificación son variables cuyos valores sin procesar se asignan a una codificación de enteros llamada niveles, en una sola pasada a través de datos, en un sistema de procesamiento de datos con múltiples hilos ejecutables y memoria (102) de búfer, el método que comprende los pasos de:
recibir datos a ser analizados en la memoria (102) de búfer;
designar una porción única de la memoria (102) de búfer para ser leída por una pluralidad de hilos ejecutables;
generar árboles (122a - 122n) binarios específicos de hilo para una variable de clasificación de dichas variables de clasificación descritas por los datos en la porción de la memoria (102) leída por el hilo;
combinar los árboles (122a - 122n) específicos de hilo para la variable de clasificación en un árbol (124) general para la variable de clasificación;
generar submatrices (126a - 126m) parciales de la matriz de productos cruzados utilizando el árbol (124) general para la variable de clasificación, en donde las submatrices parciales son matrices de productos cruzados asociadas con los efectos en el modelo;
almacenar las submatrices parciales en espacio de memoria no contiguo; y
generar la matriz (100) de productos cruzados almacenando y ordenando los elementos de las submatrices (126a - 126m) almacenadas de forma no contigua en espacio de memoria contiguo.

2. El método de la reivindicación 1, en donde el paso de generar un árbol (122a - 122n) binario específico de hilo para una variable de clasificación comprende:
determinar si un conjunto de datos asignado debe utilizarse para el análisis;
derivar un valor formateado para cada uno de los valores sin procesar;
derivar un número de observación para cada uno de los valores sin procesar; y
contar la frecuencia con la que se produce cada uno de los valores sin procesar.

3. El método de la reivindicación 2, en donde el paso de generar un árbol (122a - 122n) binario específico de hilo para una variable de clasificación comprende además:
generar un árbol (122a - 122n) binario para la variable de clasificación utilizando cada uno de los valores sin procesar único para la variable de clasificación en el conjunto de datos asignado.

4. El método de la reivindicación 2 o 3, en donde el paso de generar un árbol (122a - 122n) binario específico de hilo para una variable de clasificación comprende además:
determinar si existe un conjunto de datos asignado adicional;
utilizar un valor único en el conjunto de datos asignado adicional para generar el árbol binario para la variable de clasificación.

5. El método de una de las reivindicaciones 2 a 4, en donde el paso de generar un árbol (122a -122c) binario específico de hilo para una variable de clasificación comprende, además, generar un árbol (122a -122c) binario específico de hilo para cada una de las variables de clasificación, que tenga valores sin procesar en el conjunto de datos asignado.

6. El método de una de las reivindicaciones precedentes, en donde el paso de generar submatrices (126a - 126m) parciales de la matriz de productos cruzados utilizando el árbol (124) general para una variable de clasificación comprende:
determinar los niveles para cada uno de los efectos;
generar submatrices parciales de la matriz (100) de productos cruzados, en donde los niveles para cada uno de los efectos en las submatrices (126a - 126m) parciales están ordenados en el orden de datos; y
almacenar las submatrices (126a - 126m) parciales en memoria no contigua.

7. El método de la reivindicación 6, en donde el paso de generar submatrices (126a - 126m) parciales de la matriz (100) de productos cruzados utilizando el árbol (124) general para una variable de clasificación comprende, además:
determinar si hay nuevos datos disponibles para el análisis; y
agregar una fila y una columna adicional al final de una submatriz (126a - 126m), si se identifica un nivel adicional en los nuevos datos.

8. El método de la reivindicación 6 o 7, en donde el paso de determinar los niveles para cada uno de los efectos, comprende organizar los niveles para cada uno de los efectos en el orden en que aparecen los datos relativos a los niveles en la memoria (102).

9. El método de una de las reivindicaciones precedentes, en donde el paso de combinar los árboles (122a - 122n) específicos de hilo para una variable de clasificación en un árbol general para la variable de clasificación, comprende combinar los árboles (122a - 122n) específicos de hilo para una pluralidad de variables de clasificación, en árboles (124) generales para la pluralidad de árboles de clasificación.

10. El método de una de las reivindicaciones 6 a 9, en donde el paso de generar la matriz de productos cruzados comprende:
reordenar los niveles de los efectos en el orden especificado por el software cliente; y
generar la matriz de productos cruzados copiando los elementos de las submatrices (126a - 126m), almacenadas por separado en memoria contigua, en el orden especificado por el software cliente.

11. El método de una de las reivindicaciones precedentes, que comprende además verificar si hay datos adicionales después de generar la matriz (100) de productos cruzados y generar una matriz (100) de productos cruzados actualizada utilizando los datos adicionales.

12. El método de una de las reivindicaciones precedentes, en donde el paso de generar submatrices (126a - 126m) parciales de la matriz de productos cruzados utilizando el árbol (124) general para la variable de clasificación, se ejecuta mediante múltiples hilos.

13. Un sistema de procesamiento de datos que tiene múltiples hilos ejecutables para generar una matriz (100) de productos cruzados en una sola pasada a través de datos a ser analizados, el sistema que comprende:
memoria (102) para recibir los datos a ser analizados;
uno o más procesadores, configurados para realizar los pasos del método de cualquiera de las reivindicaciones 1 a 12.

14. Un programa informático que comprende instrucciones que, cuando el programa se ejecuta por un sistema de procesamiento de datos, hacen que el sistema de procesamiento de datos lleve a cabo los pasos del método de cualquiera de las reivindicaciones 1 a 12.

15. Un medio de almacenamiento legible por computadora que comprende instrucciones que, cuando se ejecutan mediante un sistema de procesamiento de datos, hacen que el sistema de procesamiento de datos lleve a cabo los pasos del método de cualquiera de las reivindicaciones 1 a 12.

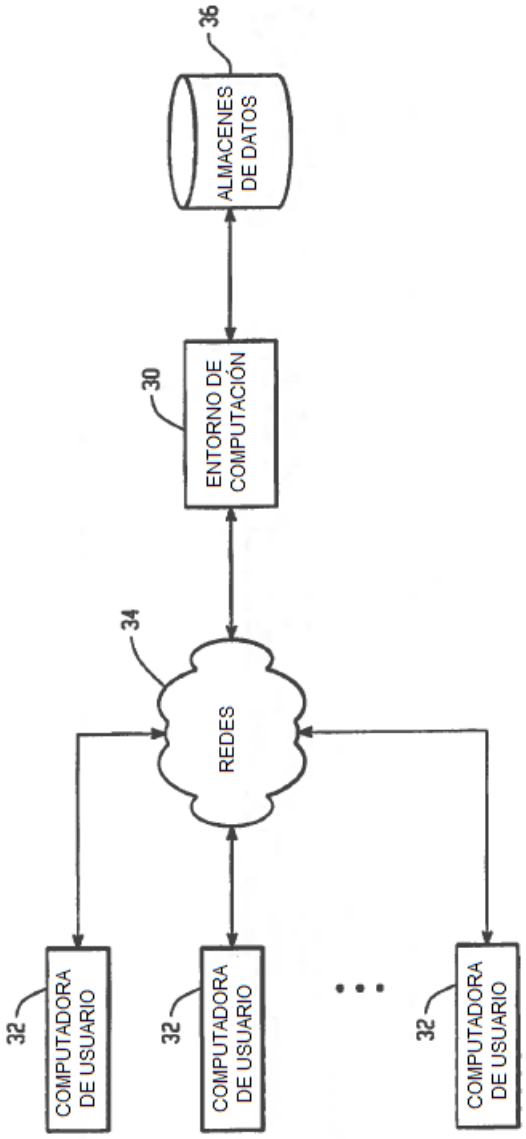
36
ALMACENES DE DATOS
30
ENTORNO DE COMPUTACIÓN
34
REDES
32
COMPUTADORA DE USUARIO
32
COMPUTADORA DE USUARIO
32
COMPUTADORA DE USUARIO

Fig. 1

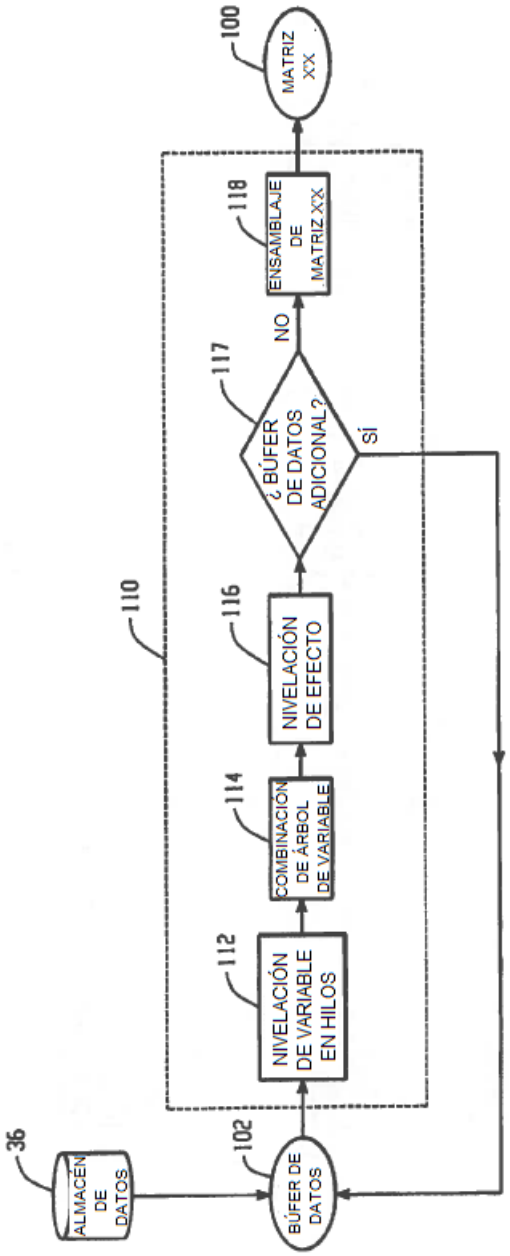
100
MATRIZ XX
118
ENSAMBLAJE DE MATRIZ XX
NO
117
¿BÚFER DE DATOS ADICIONAL?
Sí
110
116
NIVELACIÓN DE EFECTO
114
COMBINACIÓN DE ÁRBOL DE VARIABLE
112
NIVELACIÓN DE VARIABLE EN HILOS
36
ALMACÉN DE DATOS
102
BÚFER DE DATOS

Fig. 2

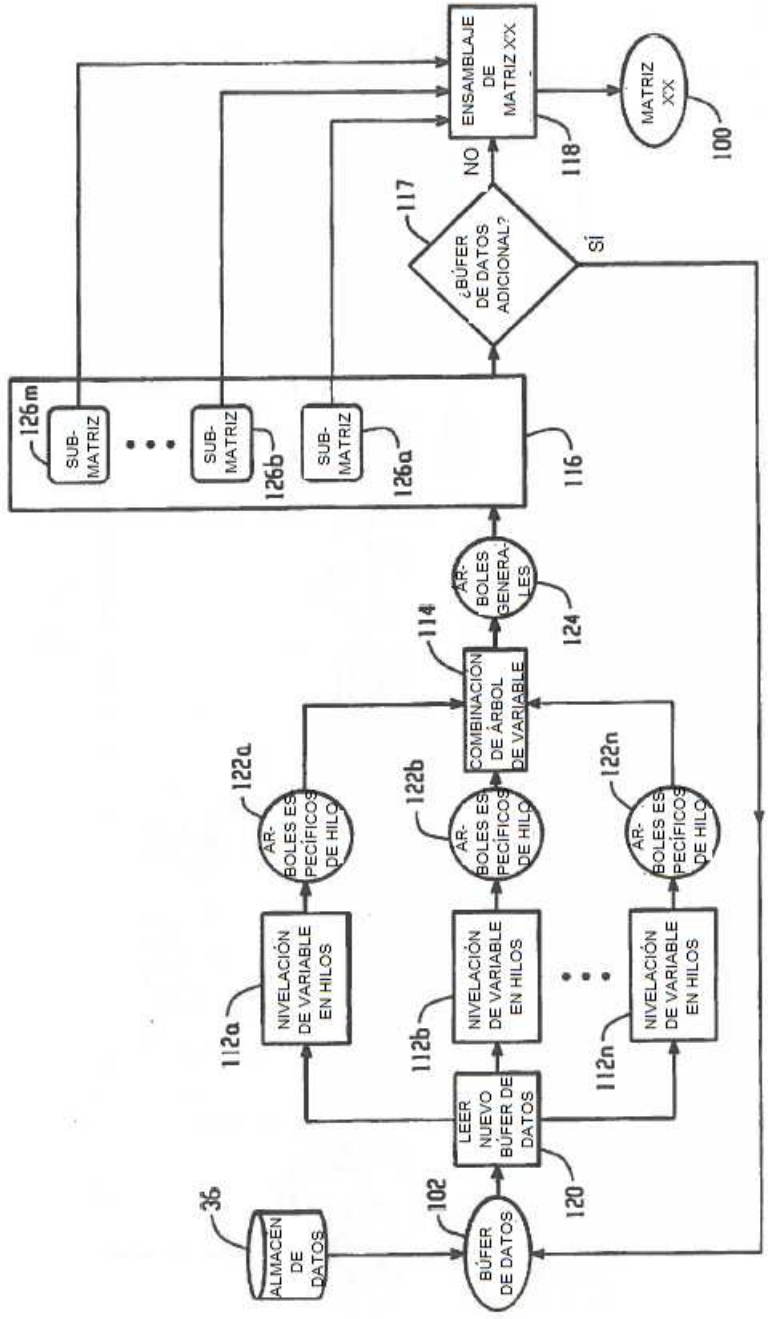
ALMACÉN DE DATOS
36
BÚFER DE DATOS
102
LEER NUEVO BÚFER DE DATOS
120
NIVELACIÓN DE VARIABLE EN HILOS
112a
NIVELACIÓN DE VARIABLE EN HILOS
112b
NIVELACIÓN DE VARIABLE EN HILOS
112n
AR-BOLES ESPECÍFICOS DE HILO
122a
AR-BOLES ESPECÍFICOS DE HILO
122b
AR-BOLES ESPECÍFICOS DE HILO
122n
COMBINACIÓN DE ÁRBOL DE VARIABLE
114
AR-BOLES GENERALES
124
SUB-MATRIZ
126m
SUB-MATRIZ
126b
SUB-MATRIZ
126a
116
¿BÚFER DE DATOS ADICIONAL?
117
NO
SÍ
ENSAMBLAJE DE MATRIZ XX
118
MATRIZ XX
100

Fig. 3

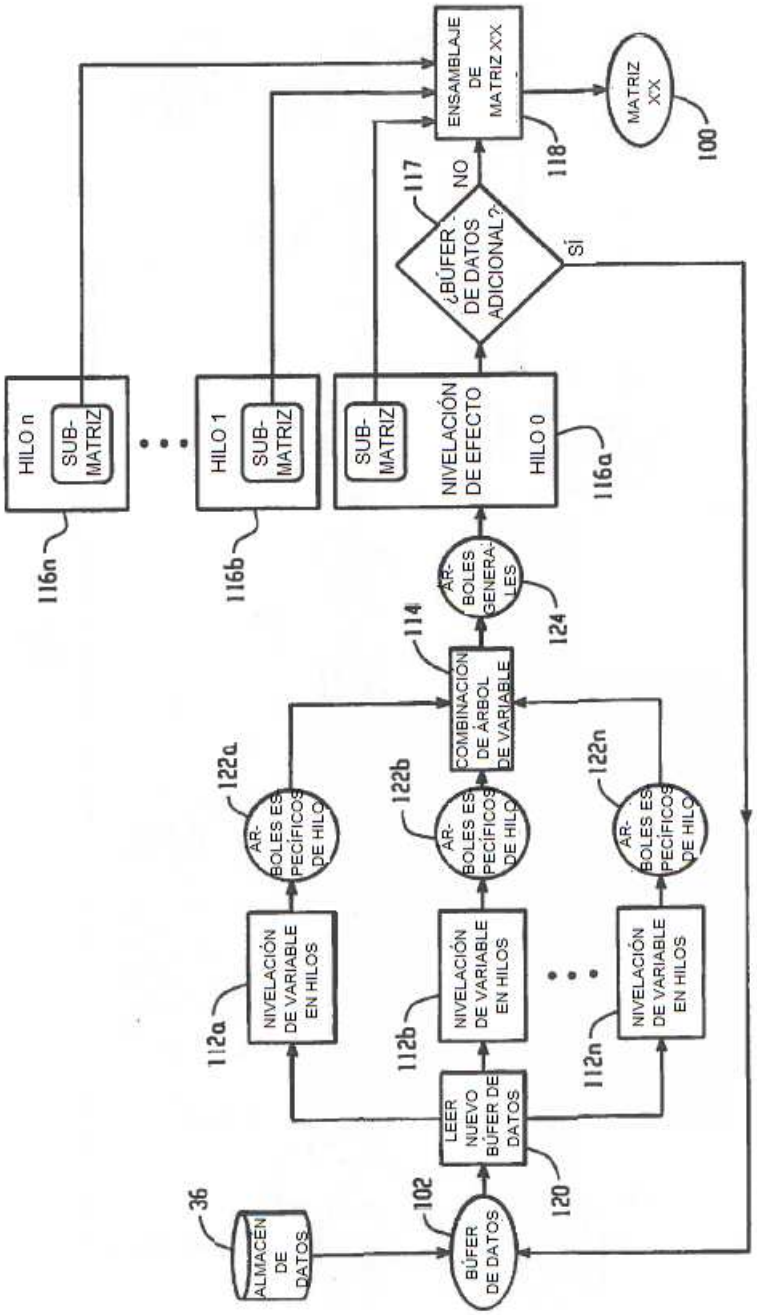
ALMACÉN DE DATOS
36
BÚFER DE DATOS
102
LEER NUEVO BÚFER DE DATOS
120
NIVELACIÓN DE VARIABLE EN HILOS
112a
NIVELACIÓN DE VARIABLE EN HILOS
112b
NIVELACIÓN DE VARIABLE EN HILOS
112n
ÁRBOLES ESPECÍFICOS DE HILO
122a
ÁRBOLES ESPECÍFICOS DE HILO
122b
ÁRBOLES ESPECÍFICOS DE HILO
122n
COMBINACIÓN DE ÁRBOL DE VARIABLE
114
ÁRBOLES GENERALES
124
NIVELACIÓN DE EFECTO
HILO 0
116a
SUB-MATRIZ
HILO 1
116b
SUB-MATRIZ
HILO n
116n
SUB-MATRIZ
¿BÚFER DE DATOS ADICIONAL?
117
NO
SÍ
ENSAMBLAJE DE MATRIZ XX
118
MATRIZ XX
100

Fig. 4

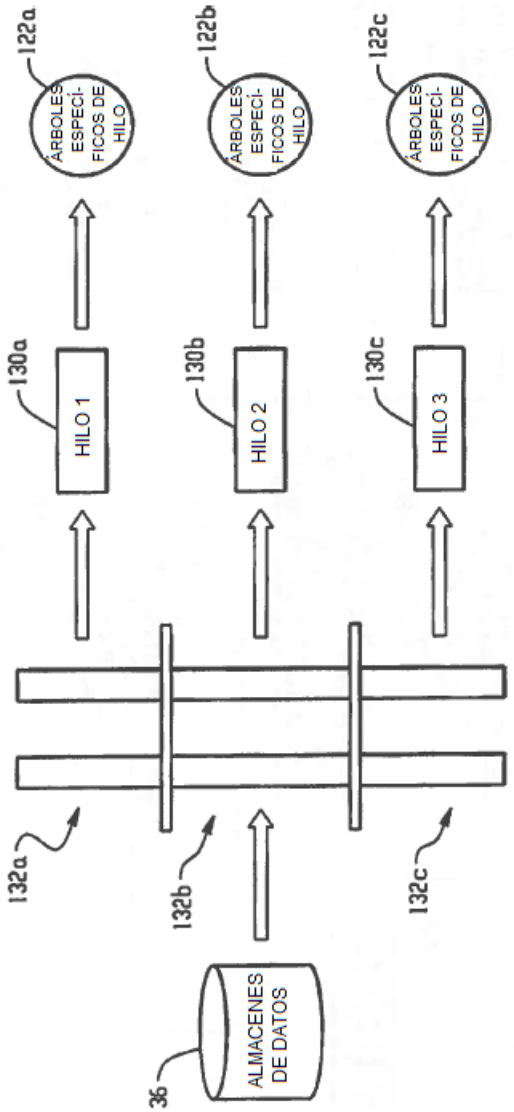
122a
ÁRBOLES ESPECÍFICOS DE HILO
122b
ÁRBOLES ESPECÍFICOS DE HILO
122c
ÁRBOLES ESPECÍFICOS DE HILO
130a
HILO 1
130b
HILO 2
130c
HILO 3
132a
132b
132c
BÚFER
100
36
ALMACENES DE DATOS

Fig. 5

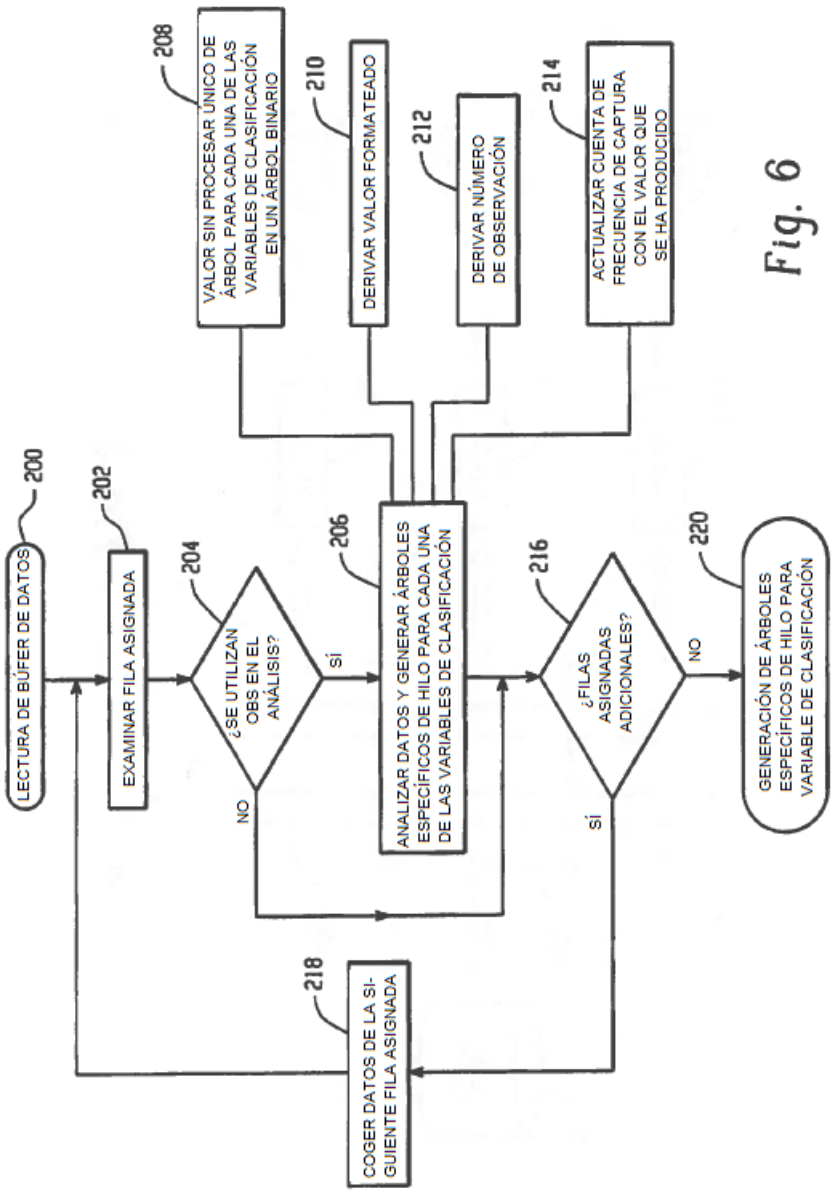
200
LECTURA DE BÚFER DE DATOS
202
EXAMINAR FILA ASIGNADA
204
¿SE UTILIZAN OBS EN EL ANÁLISIS?
SÍ
NO
206
ANALIZAR DATOS Y GENERAR ÁRBOLES ESPECÍFICOS DE HILO PARA CADA UNA DE LAS VARIABLES DE CLASIFICACIÓN
208
VALOR SIN PROCESAR ÚNICO DE ÁRBOL PARA CADA UNA DE LAS VARIABLES DE CLASIFICACIÓN EN UN ÁRBOL BINARIO
210
DERIVAR VALOR FORMATEADO
212
DERIVAR NÚMERO DE OBSERVACIÓN
214
ACTUALIZAR CUENTA DE FRECUENCIA DE CAPTURA CON EL VALOR QUE SE HA PRODUCIDO
216
¿FILAS ASIGNADAS ADICIONALES?
SÍ
NO
218
COGER DATOS DE LA SIGUIENTE FILA ASIGNADA
220
GENERACIÓN DE ÁRBOLES ESPECÍFICOS DE HILO PARA VARIABLE DE CLASIFICACIÓN

Fig. 6

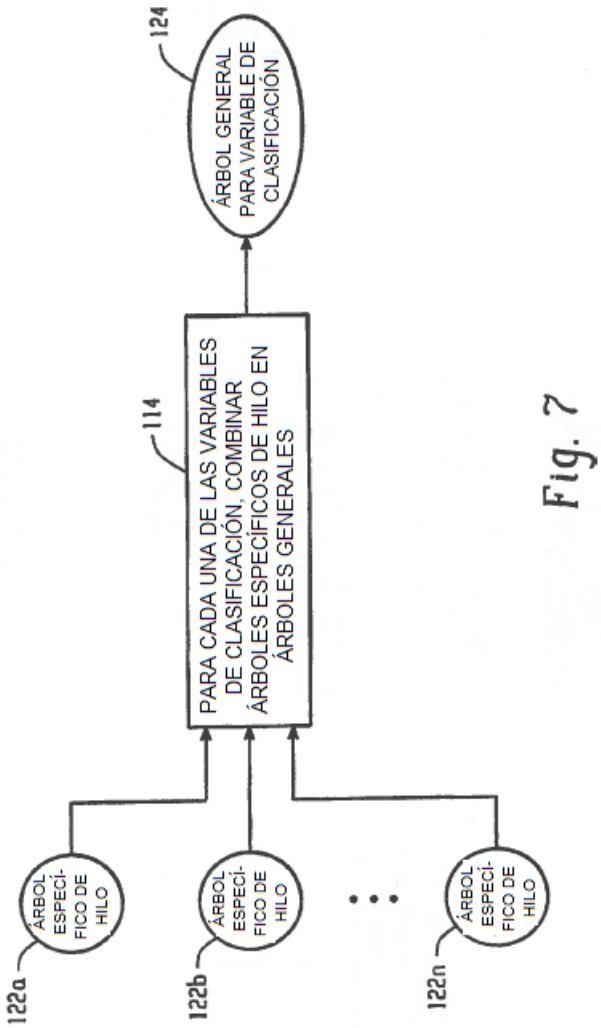
124
ÁRBOL GENERAL PARA VARIABLE DE CLASIFICACIÓN
114
PARA CADA UNA DE LAS VARIABLES DE CLASIFICACIÓN, COMBINAR ÁRBOLES ESPECÍFICOS DE HILO EN ÁRBOLES GENERALES
ÁRBOL ESPECÍFICO DE HILO
122a
ÁRBOL ESPECÍFICO DE HILO
122b
ÁRBOL ESPECÍFICO DE HILO
122n

Fig. 7

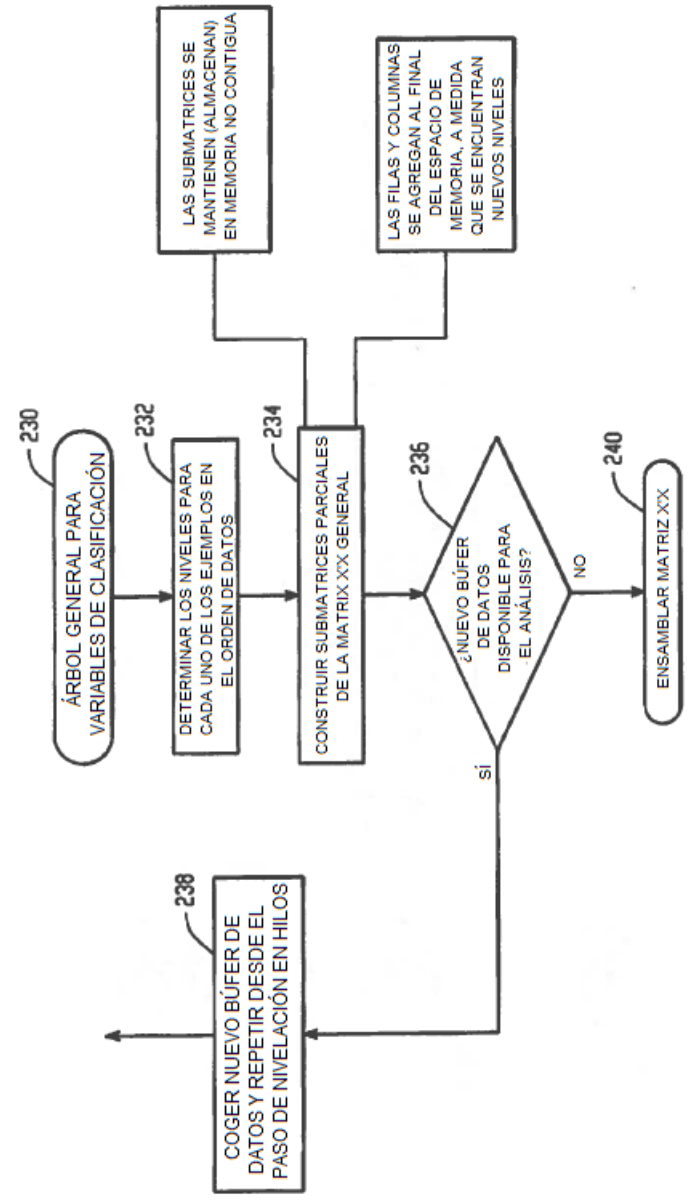
230
ÁRBOL GENERAL PARA VARIABLES DE CLASIFICACIÓN
232
DETERMINAR LOS NIVELES PARA CADA UNO DE LOS EJEMPLOS EN EL ORDEN DE DATOS
234
CONSTRUIR SUBMATRICES PARCIALES DE LA MATRIX XX GENERAL
LAS SUBMATRICES SE MANTIENEN (ALMACENAN) EN MEMORIA NO CONTIGUA
LAS FILAS Y COLUMNAS SE AGREGAN AL FINAL DEL ESPACIO DE MEMORIA, A MEDIDA QUE SE ENCUENTRAN NUEVOS NIVELES
236
¿NUEVO BÚFER DE DATOS DISPONIBLE PARA EL ANÁLISIS?
NO
240
ENSAMBLAR MATRIZ XX
SÍ
238
COGER NUEVO BÚFER DE DATOS Y REPETIR DESDE EL PASO DE NIVELACIÓN EN HILOS

Fig. 8

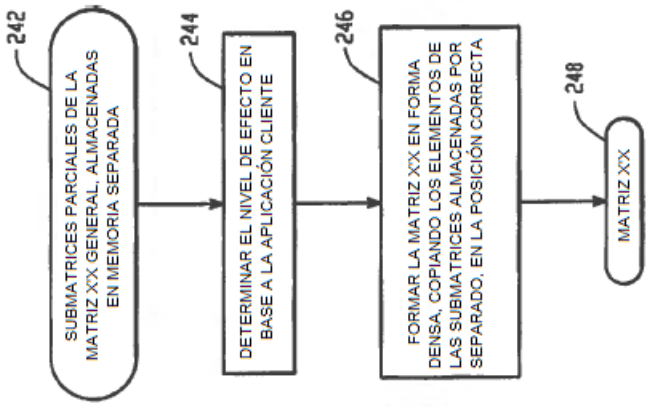
242
SUBMATRICES PARCIALES DE LA MATRIZ XX GENERAL, ALMACENADAS EN MEMORIA SEPARADA
244
DETERMINAR EL NIVEL DE EFECTO EN BASE A LA APLICACIÓN CLIENTE
246
FORMAR LA MATRIZ XX EN FORMA DENSA, COPIANDO LOS ELEMENTOS DE LAS SUBMATRICES ALMACENADAS POR SEPARADO, EN LA POSICIÓN CORRECTA
248
MATRIZ XX

Fig. 9

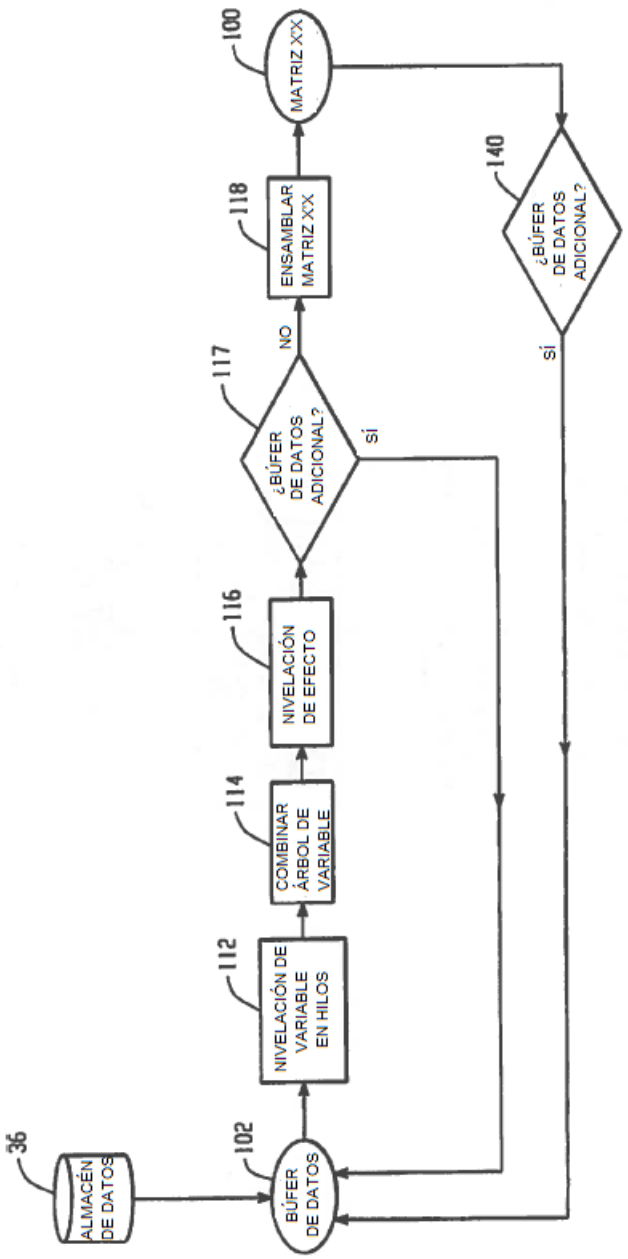
36
ALMACÉN DE DATOS
102
BÚFER DE DATOS
112
NIVELACIÓN DE VARIABLE EN HILOS
114
COMBINAR ÁRBOL DE VARIABLE
116
NIVELACIÓN DE EFECTO
117
¿BÚFER DE DATOS ADICIONAL?
SÍ
NO
118
ENSAMBLAR MATRIZ XX
100
MATRIZ XX
140
¿BÚFER DE DATOS ADICIONAL?
SÍ

Fig. 10

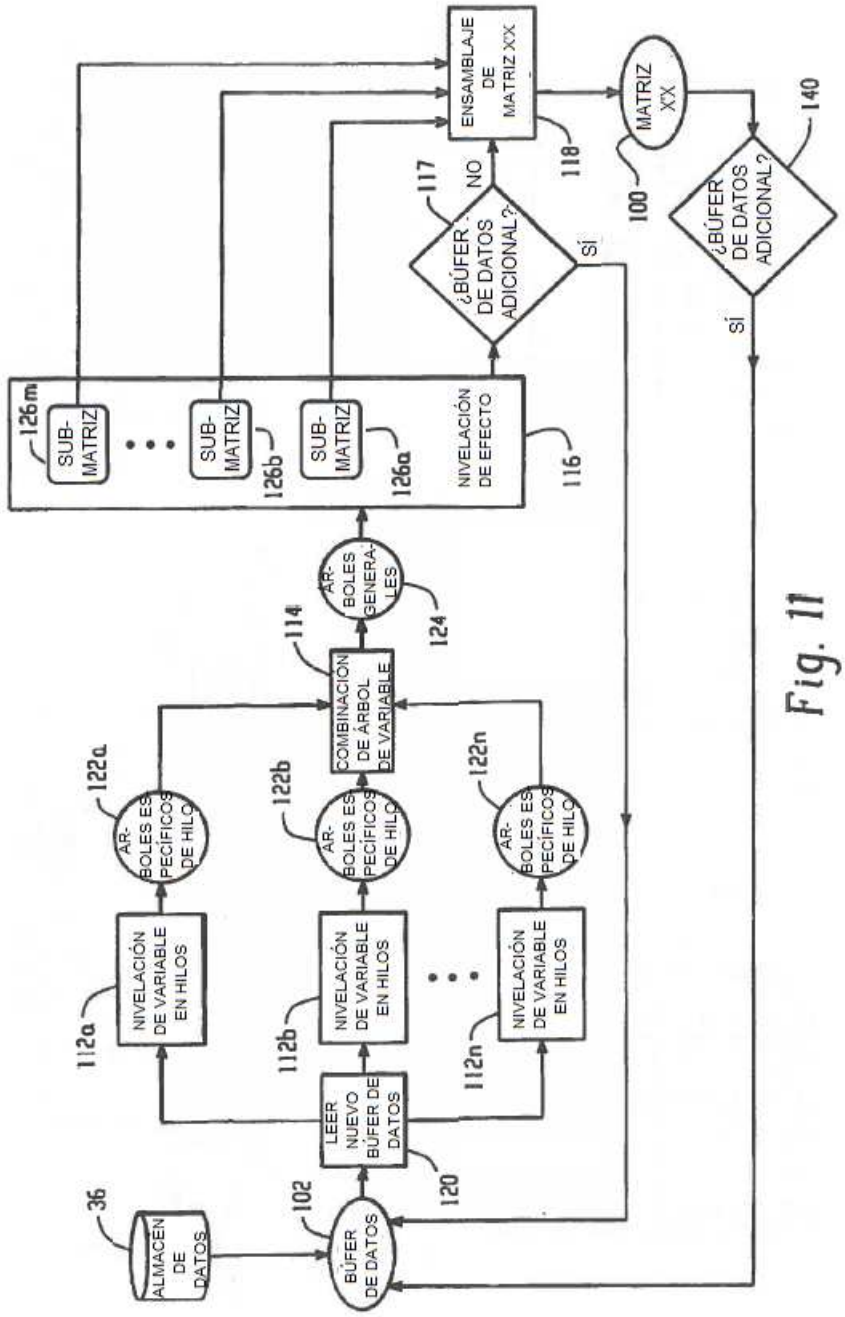
ALMACÉN DE DATOS
36
BÚFER DE DATOS
102
LEER NUEVO BÚFER DE DATOS
120
NIVELACIÓN DE VARIABLE EN HILOS
112a
NIVELACIÓN DE VARIABLE EN HILOS
112b
NIVELACIÓN DE VARIABLE EN HILOS
112n
ÁRBOLES ESPECÍFICOS DE HILO
122a
ÁRBOLES ESPECÍFICOS DE HILO
122b
ÁRBOLES ESPECÍFICOS DE HILO
122n
COMBINACIÓN DE ÁRBOL DE VARIABLE
114
ÁRBOLES GENERALES
124
NIVELACIÓN DE EFECTO
116
SUB-MATRIZ
126a
SUB-MATRIZ
126b
SUB-MATRIZ
126m
¿BÚFER DE DATOS ADICIONAL?
117
NO
SÍ
ENSAMBLAJE DE MATRIZ XX
118
MATRIZ XX
100
¿BÚFER DE DATOS ADICIONAL?
140
SÍ

Fig. 11

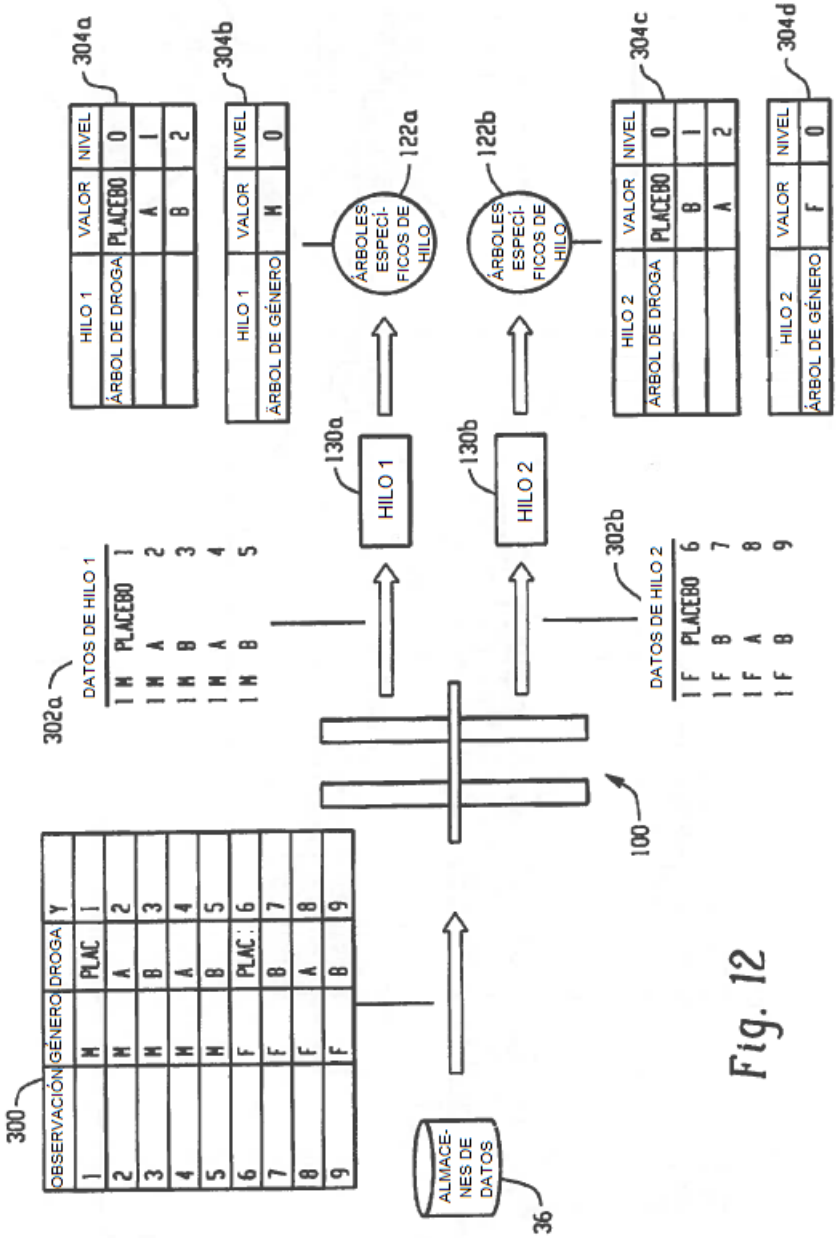
300
OBSERVACIÓN
GÉNERO
DROGA
Y
302a
DATOS DE HILO 1
302b
DATOS DE HILO 2
100
130a
HILO 1
130b
HILO 2
122a
ÁRBOLES ESPECÍFICOS DE HILO
122b
ÁRBOLES ESPECÍFICOS DE HILO
304a
HILO 1
ÁRBOL DE DROGA
VALOR
NIVEL
304b
HILO 1
ÁRBOL DE GÉNERO
VALOR
NIVEL
304c
HILO 2
ÁRBOL DE DROGA
VALOR
NIVEL
304d
HILO 2
ÁRBOL DE GÉNERO
VALOR
NIVEL
ALMACENES DE DATOS
36

Fig. 12

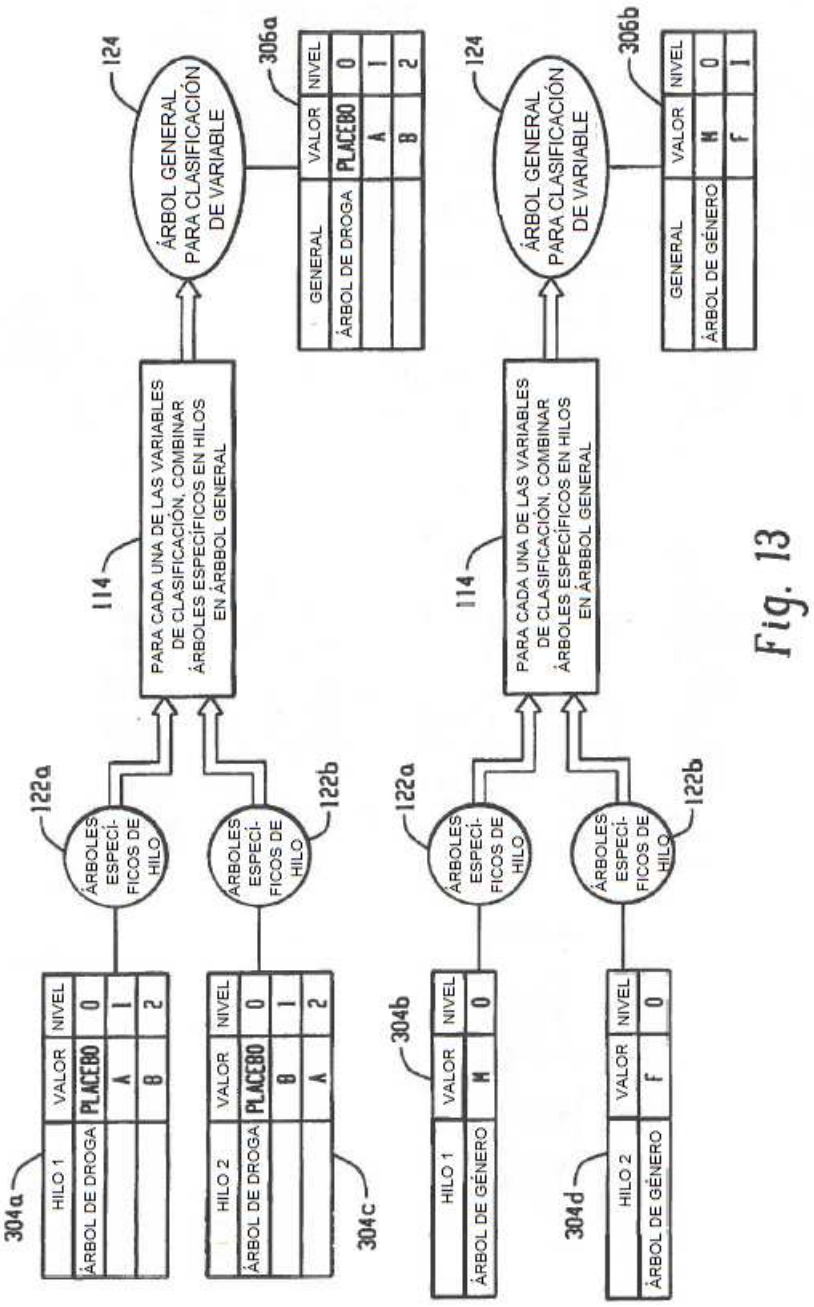
304a
HILO 1
ÁRBOL DE DROGA
VALOR
NIVEL
PLACEBO
0
A
1
B
2
304c
HILO 2
ÁRBOL DE DROGA
VALOR
NIVEL
PLACEBO
0
B
1
A
2
304b
HILO 1
ÁRBOL DE GÉNERO
VALOR
NIVEL
M
0
304d
HILO 2
ÁRBOL DE GÉNERO
VALOR
NIVEL
F
0
122a
ÁRBOLES ESPECÍFICOS DE HILO
122b
ÁRBOLES ESPECÍFICOS DE HILO
114
PARA CADA UNA DE LAS VARIABLES DE CLASIFICACIÓN, COMBINAR ÁRBOLES ESPECÍFICOS EN HILOS EN ÁRBBOL GENERAL
124
ÁRBOL GENERAL PARA CLASIFICACIÓN DE VARIABLE
306a
GENERAL
ÁRBOL DE DROGA
VALOR
NIVEL
PLACEBO
0
A
1
B
2
306b
GENERAL
ÁRBOL DE GÉNERO
VALOR
NIVEL
M
0
F
1

Fig. 13

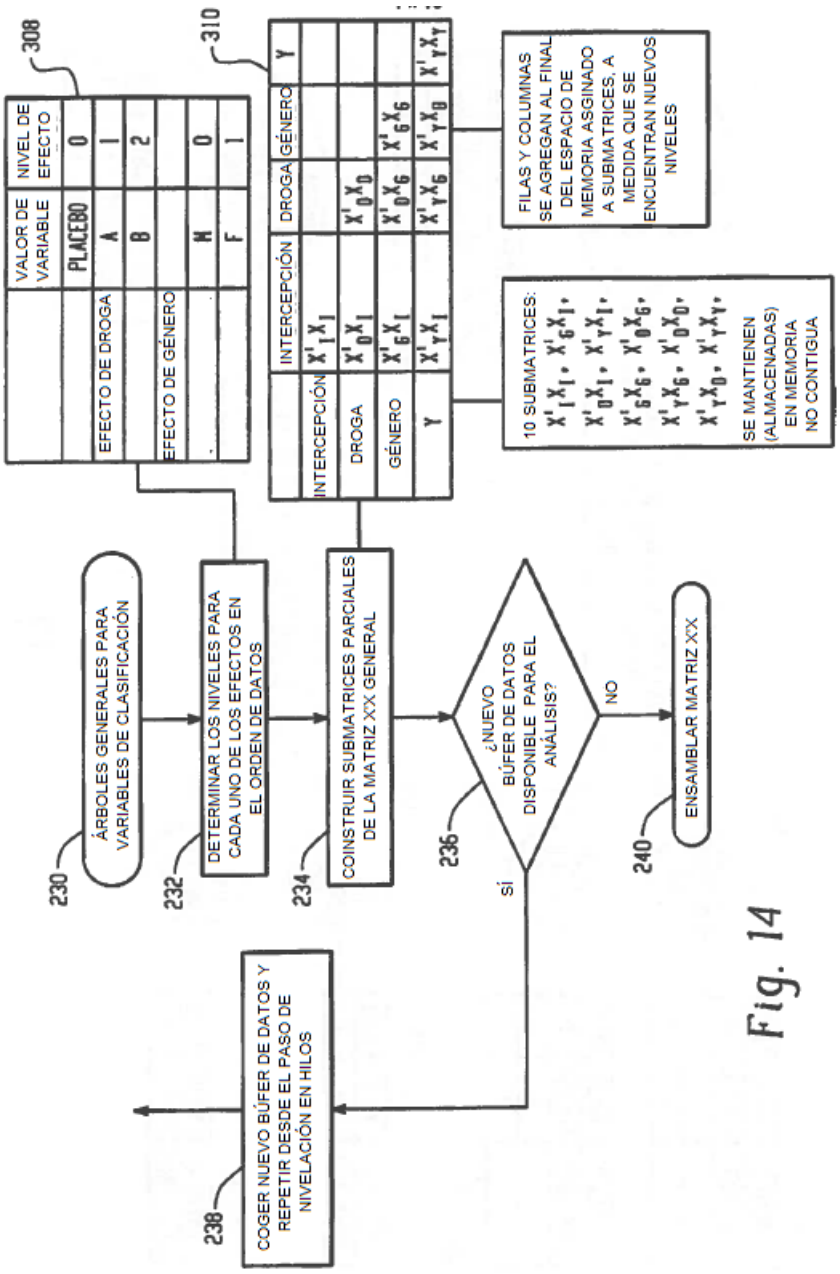
308
VALOR DE VARIABLE
NIVEL DE EFECTO
EFECTO DE DROGA
PLACEBO
0
A
1
B
2
EFECTO DE GÉNERO
M
0
F
1
310
INTERCEPCIÓN
DROGA
GÉNERO
Y
INTERCEPCIÓN
DROGA
GÉNERO
Y
FILAS Y COLUMNAS SE AGREGAN AL FINAL DEL ESPACIO DE MEMORIA ASGINADO A SUBMATRICES, A MEDIDA QUE SE ENCUENTRAN NUEVOS NIVELES
10 SUBMATRICES:
SE MANTIENEN (ALMACENADAS) EN MEMORIA NO CONTIGUA
ÁRBOLES GENERALES PARA VARIABLES DE CLASIFICACIÓN
230
DETERMINAR LOS NIVELES PARA CADA UNO DE LOS EFECTOS EN EL ORDEN DE DATOS
232
COINSTRUIR SUBMATRICES PARCIALES DE LA MATRIZ XX GENERAL
234
¿NUEVO BÚFER DE DATOS DISPONIBLE PARA EL ANÁLISIS?
236
NO
ENSAMBLAR MATRIZ XX
240
SÍ
COGER NUEVO BÚFER DE DATOS Y REPETIR DESDE EL PASO DE NIVELACIÓN EN HILOS
238

Fig. 14

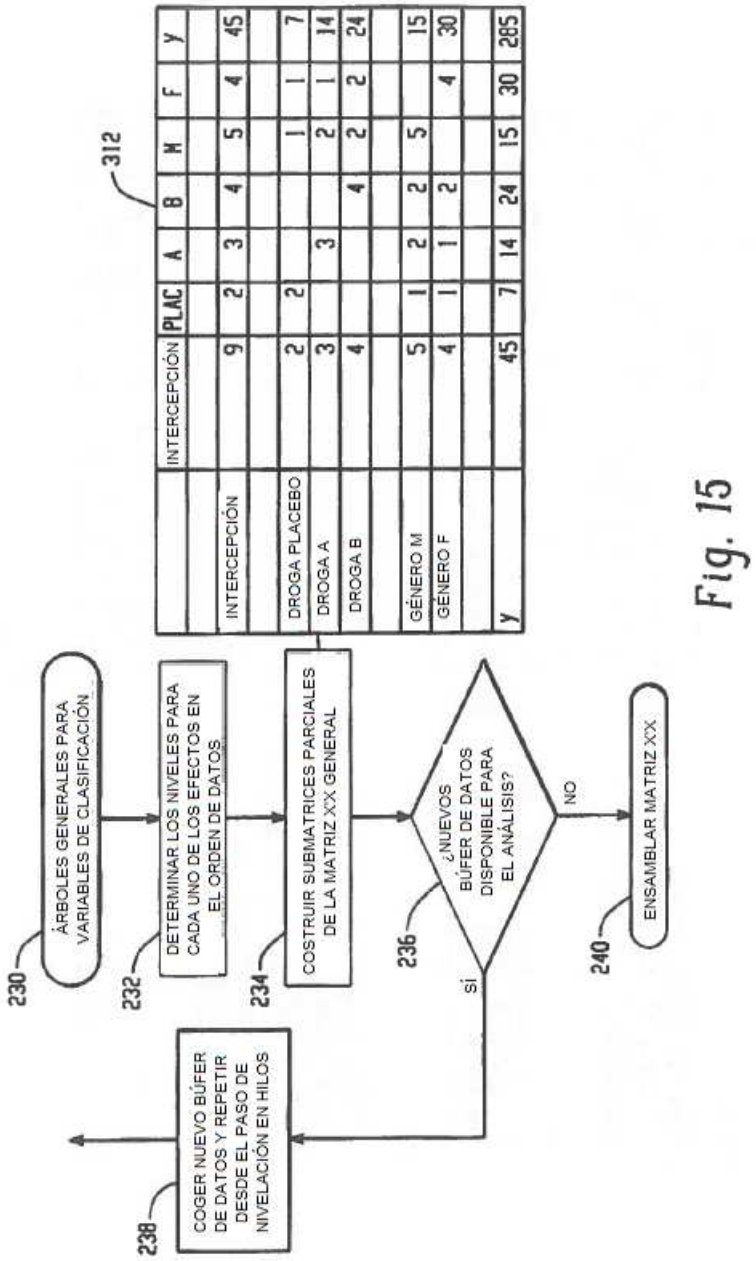
312
INTERCEPCIÓN
PLAC
A
B
M
F
y
INTERCEPCIÓN 9 2 3 4 5 4 45
DROGA PLACEBO 2 2 1 1 7
DROGA A 3 3 2 1 14
DROGA B 4 4 2 2 24
GÉNERO M 5 1 2 2 5 15
GÉNERO F 4 1 1 2 4 30
y 45 7 14 24 15 30 285
230
ÁRBOLES GENERALES PARA VARIABLES DE CLASIFICACIÓN
232
DETERMINAR LOS NIVELES PARA CADA UNO DE LOS EFECTOS EN EL ORDEN DE DATOS
234
COSTRUIR SUBMATRICES PARCIALES DE LA MATRIZ XX GENERAL
236
¿NUEVOS BÚFER DE DATOS DISPONIBLE PARA EL ANÁLISIS?
SÍ
NO
238
COGER NUEVO BÚFER DE DATOS Y REPETIR DESDE EL PASO DE NIVELACIÓN EN HILOS
240
ENSAMBLAR MATRIZ XX

Fig. 15

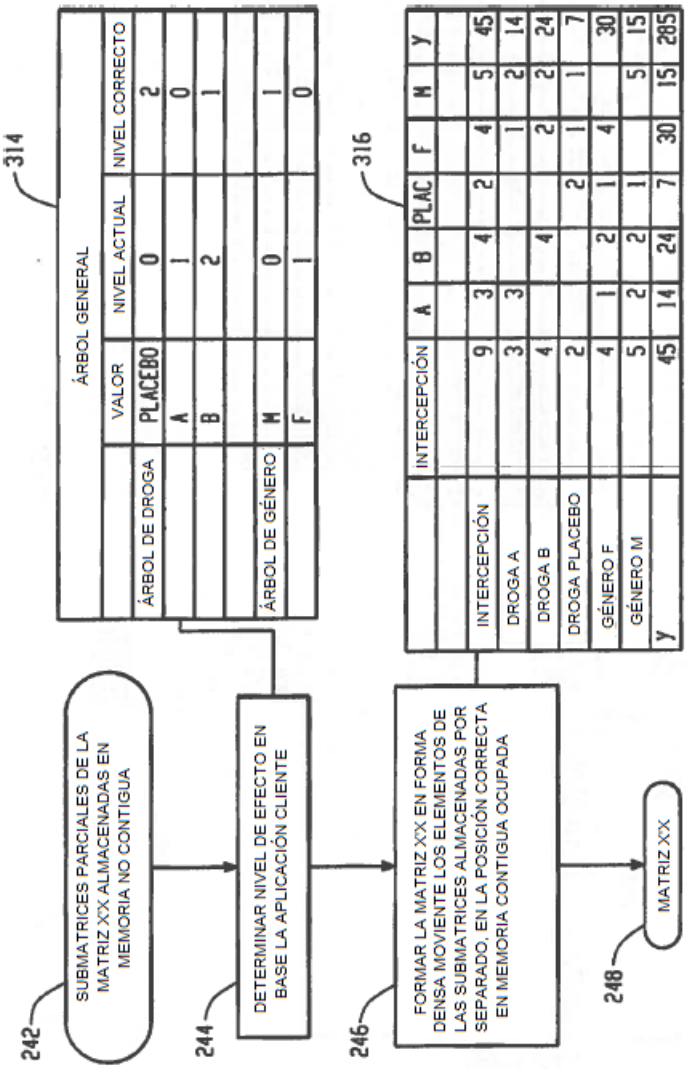
314
ÁRBOL GENERAL
ÁRBOL DE DROGA
VALOR
NIVEL ACTUAL
NIVEL CORRECTO
PLACEBO
0
2
A
1
0
B
2
1
ÁRBOL DE GÉNERO
M
0
1
F
1
0
316
INTERCEPCIÓN
A
B
PLAC
F
M
y
INTERCEPCIÓN
9
3
4
2
4
5
45
DROGA A
3
3
1
2
14
DROGA B
4
4
2
2
24
DROGA PLACEBO
2
2
1
1
7
GÉNERO F
4
1
2
1
4
30
GÉNERO M
5
2
2
1
5
15
y
45
14
24
7
30
15
285
242
SUBMATRICES PARCIALES DE LA MATRIZ X'X ALMACENADAS EN MEMORIA NO CONTIGUA
244
DETERMINAR NIVEL DE EFECTO EN BASE LA APLICACIÓN CLIENTE
246
FORMAR LA MATRIZ X'X EN FORMA DENSA MOVIENTE LOS ELEMENTOS DE LAS SUBMATRICES ALMACENADAS POR SEPARADO, EN LA POSICIÓN CORRECTA EN MEMORIA CONTIGUA OCUPADA
248
MATRIZ X'X

Fig. 16