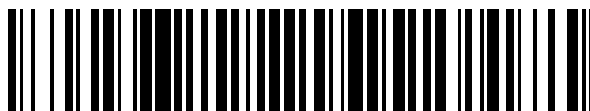


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 753 220**

51 Int. Cl.:

G06K 9/00 (2006.01)

G06F 16/00 (2009.01)

G06K 9/62 (2006.01)

G08G 1/01 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.02.2017** **E 17154099 (0)**

97 Fecha y número de publicación de la concesión europea: **17.07.2019** **EP 3358541**

54 Título: **Un procedimiento para predecir un comportamiento del tráfico en un sistema de carreteras**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.04.2020

73 Titular/es:

KAPSCH TRAFFICCOM AG (100.0%)
Am Europlatz 2
1120 Wien , AT

72 Inventor/es:

GIL AGUIRREBEITIA, CARLOS y
GARCÍA CASADO, RUBÉN

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 753 220 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

Un procedimiento para predecir un comportamiento del tráfico en un sistema de carreteras

La invención se refiere a un procedimiento para predecir un comportamiento de tráfico en un sistema de carreteras sobre la base de un conjunto de perfiles de tráfico, cada uno de los cuales representa una medida de tráfico de vehículos en el sistema de carreteras registrada a lo largo del tiempo, en el que una o más etiquetas predefinidas asociadas con un momento de registro del perfil de tráfico específico se atribuye a dicho perfil de tráfico.

La invención se refiere así al campo de medir y evaluar el comportamiento del tráfico de vehículos, por ejemplo, el comportamiento del tráfico en una calle principal de una ciudad. Una vez que se ha analizado el comportamiento del tráfico en esta calle durante un período de tiempo prolongado, en general durante un año o más, el objetivo es estimar el comportamiento del tráfico para un determinado día o ventana de tiempo en el futuro. Esto se puede usar para predecir qué carretera estará congestionada o bloqueada en un momento determinado, por ejemplo, para los cargueros que planifican el transporte de carga o las familias que planifican sus vacaciones. Dependiendo del comportamiento del tráfico previsto, se pueden elegir diferentes rutas u horarios en los que viajar en esta calle, por lo que en cualquier caso se puede reducir el tiempo de viaje de principio a fin para un vehículo.

Con este fin, los procedimientos conocidos de predicción de tráfico utilizan una agrupación de similitud de los perfiles grabados y una asignación de vectores característicos ("etiquetas") de etiquetas a los grupos, cuyas etiquetas se pueden consultar para identificar un grupo de perfiles de tráfico como comportamiento de tráfico previsto para un día determinado, por ejemplo, un día de semana específico, un día "lluvioso", etc.

Weijermars, van Berkomp: "Analyzing highway flow patterns using cluster analysis", Conferencia de IEEE sobre sistemas de transporte de inteligencia, del 13 al 16 de septiembre de 2005, muestra un procedimiento para analizar patrones de tráfico. En el presente documento, se han registrado 118 perfiles de tráfico para una carretera, cada uno en un lapso de tiempo de 24 horas. A partir de entonces, los perfiles de tráfico se han agrupado mediante un algoritmo de agrupación estándar. Después de etiquetar manualmente los grupos, se evaluaron las desviaciones dentro de los grupos.

El documento US2015/127245 A1 divulga un procedimiento para analizar patrones de tráfico.

Además, en el estado de la técnica se sabe que adopta un comportamiento de tráfico estándar y extrapola el comportamiento de tráfico para un pronóstico a corto plazo determinando la desviación actual del comportamiento de tráfico estándar. Este procedimiento se conoce como regresión paramétrica y produce un pronóstico de 15 minutos.

Por lo tanto, existe la necesidad de un procedimiento fiable de predicción de tráfico a largo plazo. El procedimiento debería dar resultados incluso con ordenadores con bajos recursos computacionales, de modo que el procedimiento se pueda usar en teléfonos móviles y similares en cualquier lugar.

Por lo tanto, la invención establece su objetivo de crear un procedimiento mejorado para predecir el comportamiento del tráfico.

Con este fin, la invención proporciona un procedimiento tal como se define en la reivindicación independiente 1.

El procedimiento según la invención procesa una petición que puede formularse fácilmente eligiendo etiquetas apropiadas. Por ejemplo, si un usuario desea predecir el comportamiento del tráfico para un lunes de julio, y sabe que este lunes será festivo, puede formular la petición mediante las etiquetas "lunes", "julio" y "vacaciones". A continuación, el procedimiento utiliza un algoritmo de eliminación que proporciona un número mínimo de pasos para encontrar el grupo de coincidencia más cercano para dicha petición.

El procedimiento de predicción de la invención se basa en la idea de que el grupo con el mayor número de perfiles de tráfico representa un flujo de tráfico promedio en el sistema de carreteras. Se prefieren las etiquetas en la petición que son muy peculiares, es decir, corresponden a grupos que son muy diferentes al grupo con el mayor número de perfiles de tráfico, y, por lo tanto, es menos probable que se eliminen. Basándose en este principio, el procedimiento comprende un bucle en el que las etiquetas menos interesantes se eliminan sucesivamente hasta que la petición corresponde a un vector característico de uno de los grupos, que luego se puede utilizar para predecir un comportamiento de tráfico a partir del mismo.

El procedimiento se puede refinar aún más teniendo en cuenta también las combinaciones de las etiquetas de la petición. Con este fin, al elegir qué etiqueta de petición se eliminará de la petición, también se puede comprobar si una combinación de etiquetas de petición corresponde a un vector característico, y si es así, y si este vector característico tiene la medida de disimilitud más baja entre las vectores característicos correspondientes a todas las etiquetas de petición o combinaciones de etiquetas de petición, dicha combinación de etiquetas se descarta hasta que se elimina una etiqueta de petición. Por medio de esto, la precisión de la predicción se puede mejorar mientras que el esfuerzo computacional no se incrementa sustancialmente. Si bien el ciclo puede repetirse con mayor frecuencia, no se produce una situación de ciclo en ciclo, por lo que los requisitos informáticos siguen siendo bajos.

La medida de disimilitud se puede obtener de varias maneras. Preferentemente, la medida de disimilitud de un grupo se obtiene comparando un perfil de tráfico promedio de dicho grupo con un perfil de tráfico promedio del grupo de referencia. Aquí, el término "promedio" también incluye promedios ponderados. De forma alternativa, en lugar de un promedio, también se puede usar un valor modal o mediano. De esta manera, el perfil de tráfico más significativo (desviado) de un grupo se comparará con el perfil de tráfico más significativo (desviado) del grupo de referencia. El paso de comparación también se puede realizar de varias maneras, por ejemplo, calculando las normas L1 o L2 entre los perfiles para comparar.

Hay dos ejemplos alternativos al determinar qué vector característico corresponde a una etiqueta o combinación de etiquetas de la petición. En el primer ejemplo, "equivalencia uno a uno", un vector característico corresponde a una etiqueta de petición o combinación de etiquetas de petición si el vector característico solo comprende esta etiqueta de petición o combinación de etiquetas de petición. Esto tiene la ventaja de que se encuentran menos vectores característicos como vectores característicos correspondientes, lo cual mejora la velocidad computacional.

En el segundo ejemplo de "equivalencia inclusiva", un vector característico corresponde a una etiqueta de petición o combinación de etiquetas de petición si el vector característico comprende esta combinación de etiquetas de petición o etiqueta de petición, ya sea sola o en combinación con al menos una etiqueta adicional. Esto tiene la ventaja de que hay más opciones disponibles, lo cual aumenta la fiabilidad del procedimiento y evita que las etiquetas "raras" e importantes se eliminen demasiado pronto cuando no tienen un vector característico equivalente estricto.

Para mejorar aún más la fiabilidad del procedimiento, se calcula un valor de significancia de cada vector característico de cada grupo y, cuando dos o más vectores característicos corresponden a una etiqueta de petición o combinación de etiquetas de petición de la petición, el vector característico con el valor de significancia más baja no se tiene en cuenta. Esto está disponible para los dos ejemplos alternativos mencionados anteriormente y evita posibles problemas de colisión, además de aumentar la precisión debido al valor de significancia adicional.

Para evitar aún más los problemas de colisión de dos o más vectores correspondientes, es decir, si dos vectores característicos corresponden a una etiqueta de petición o una combinación de etiquetas de petición, y ambos tienen el mismo valor de significancia, si corresponde, el vector característico del grupo con el número más bajo de los perfiles de tráfico no se tiene en cuenta. Esta es la opción más segura, pero también es posible utilizar el vector característico del conjunto con el menor número de perfiles de tráfico para no descartar etiquetas o combinaciones de etiquetas relevantes demasiado pronto.

En ocasiones, pueden producirse situaciones en las que una etiqueta de petición no coincide exactamente con ninguno de los vectores característicos de los grupos. Esto podría significar que no se han acumulado suficientes datos para obtener un resultado exacto para esta petición. Para obtener la coincidencia más cercana con los datos disponibles, si en el paso de eliminar no se encuentra un vector característico para una etiqueta de petición, se supone la medida de disimilitud más baja para esta etiqueta. Sin embargo, esto también significa que es probable que esta etiqueta de petición se elimine temprano en el ciclo.

En todos los modos de realización de la invención, se prefiere que cada perfil de tráfico represente la medida del tráfico de vehículos en el sistema de carreteras durante un lapso de tiempo de 24 horas. De forma alternativa, también se pueden usar intervalos de tiempo de 6 horas, 12 horas o 18 horas, especialmente cuando el comportamiento del tráfico en ciertas horas del día, por ejemplo, horas nocturnas u horas punta, es sustancialmente el mismo para todos los perfiles de tráfico.

Preferentemente, al menos una etiqueta indica un día de la semana, un mes del año, una semana del mes, una indicación de vacaciones escolares o una indicación especial de vacaciones de dicho momento de registro. Dichas etiquetas pueden, por ejemplo, generarse automáticamente asociando un atributo de un calendario a un perfil de tráfico. El calendario se puede definir de antemano, en tiempo real o en retrospectiva, y también es posible agregar etiquetas posteriormente a los perfiles de tráfico cuando ocurren por primera vez en una petición.

De forma alternativa o adicional a esto, al menos una etiqueta indica una condición de la carretera o una condición climática en dicho momento de registro. Dichas etiquetas representan una medida física de los datos ambientales, que la etiqueta puede codificar por adelantado, en tiempo real o en retrospectiva.

La invención se explicará ahora con más detalle a continuación sobre la base de modos de realización a modo de ejemplo preferentes de la misma con referencia a los dibujos adjuntos, en los que:

La Figura 1 muestra una vista esquemática de un sistema de carreteras;

la Figura 2 muestra un diagrama de bloques de bases de datos y motores de procesamiento para realizar el procedimiento según la invención;

la Figura 3 muestra un ejemplo de grupos y etiquetas asociadas que se producen en las etapas de agrupación y etiquetado del procedimiento según la invención;

la Figura 4 muestra un árbol de decisión utilizado para determinar etiquetas para grupos durante el procedimiento

según la invención;

la Figura 5 muestra un diagrama de flujo de la etapa de predicción del procedimiento según la invención; y

la Figura 6 muestra un ejemplo de la eliminación de etiquetas de la petición durante la etapa de predicción de la Figura 5.

5 En la Figura 1, se representa un sistema de carreteras 1 con una pluralidad de carreteras 2 en el que los vehículos 3 se desplazan en las direcciones R. El sistema de carreteras 1 está equipado con al menos un dispositivo de medición de tráfico 4, que puede usarse, por ejemplo, para medir flujo de tráfico en una carretera 2 o en una intersección de carreteras 2. El dispositivo de medición de tráfico 4 puede montarse en un pórtico 5 que se extiende a lo ancho de la carretera 2 o, de forma alternativa, puede montarse al lado de la carretera. El dispositivo de medición de tráfico 4
10 puede realizarse como cualquier tipo conocido en el estado de la técnica, por ejemplo, como bucles inductivos, dispositivos de radar, barreras de luz, cámaras de procesamiento de vídeo o como lectores inalámbricos que leen una respuesta de radio de las llamadas unidades a bordo (OBU) transportadas por los vehículos 3.

El dispositivo de medición de tráfico 4 puede registrar el flujo de tráfico en la carretera 2 directamente, por ejemplo, como un número v de vehículos 3 por unidad de tiempo t , o puede registrar el número v/t de vehículos 3 que viajan en una sección dada de la carretera 2 en un instante dado en el tiempo, por ejemplo, contando los vehículos 3 en imágenes de cámara de dicho tramo de carretera tomadas a intervalos de tiempo regulares.
15

En lugar de medir el flujo de tráfico, también se pueden medir otras condiciones de tráfico en la carretera por medio de diferentes valores de medición, por ejemplo, registrando una velocidad promedio en la carretera 2 en un cierto punto de tiempo t o midiendo el tiempo de viaje que toman los vehículos recorrer un tramo de la carretera 2.

20 El dispositivo de medición de tráfico 4 registra el flujo de tráfico v/t , o valores de medición equivalentes como se describe anteriormente, por tiempo t durante un período de tiempo predefinido de, por ejemplo, 24 horas como un perfil de tráfico 6 (Figura 2). De forma alternativa, también se podrían usar períodos de tiempo más cortos, por ejemplo, solo los períodos de tiempo más interesantes de, por ejemplo, 6, 12 o 18 horas al día.

El dispositivo de medición de tráfico 4 registra dichos perfiles de tráfico 6 durante un período de tiempo prolongado, por ejemplo, día tras día durante todo un año o más. Un perfil de tráfico puede estar compuesto de muestras discretas, por ejemplo, una muestra cada 5 minutos, 15 minutos o cada hora, o puede ser una curva al menos parcialmente suave con el tiempo. De esta manera, cada dispositivo de medición de tráfico 4 registra un conjunto de perfiles de tráfico 6 para su carretera 2 o intersección de carreteras 2 y lo almacena en una base de datos 7 a través de una conexión de datos 8. Un procesador 9 está conectado a la base de datos 7 a través de una conexión 10 para el procesamiento adicional de los perfiles de tráfico 6, como se describirá a continuación.
25
30

La Figura 2 muestra un motor de agrupamiento 11, que puede ser parte del procesador 9 mencionado anteriormente. El motor de agrupamiento 11 obtiene perfiles de tráfico 6 de la base de datos 7 y agrupa estos perfiles de tráfico 6 en al menos uno o al menos dos grupos CL_1 , CL_2 ,..., en general CL_i . El algoritmo de agrupamiento utilizado en el motor de agrupamiento 11 puede ser de cualquier tipo conocido en el estado de la técnica, por ejemplo un algoritmo de agrupamiento aglomerativo jerárquico. Para este fin, los perfiles de tráfico 6 pueden necesitar ser preprocesados de manera que contengan, por ejemplo, 96 muestras de tiempo cada una a intervalos de 15 minutos. Además, se pueden emplear filtros para tratar las anomalías locales que se producen en los perfiles de tráfico 6.
35

Como resultado, el motor de agrupación 11 envía grupos CL_i , cada uno de las cuales comprende múltiples perfiles de tráfico 6 que son similares entre sí, a una base de datos 12. Esta base de datos 12 puede ser la misma que la base de datos 7 o un subconjunto de la misma. Además, el motor de agrupación 11 puede controlarse por medio de una entrada de configuración de control 13 para ajustar los parámetros de agrupación.
40

Después de que los perfiles de tráfico 6 se hayan agrupado en grupos CL_i , se emplea un motor de etiquetado 14 para asignar una etiqueta L_1 , L_2 ,..., en general L_i , a cada grupo CL_i , en el que cada etiqueta L_i consiste en al menos un vector característico CV_1 , CV_2 ,..., en general CV_j . Las etiquetas L_i deben reflejar el contenido de los grupos CL_i con la mayor precisión posible.
45

Los pasos del procedimiento realizados por el motor de etiquetado 14 se describirán ahora en detalle por medio de la Figura 3, que muestra cuatro grupos a modo de ejemplo CL_1 , CL_2 , CL_3 , CL_4 producidos por el motor de agrupamiento 11. El primer grupo CL_1 comprende $N = 100$ perfiles de tráfico similares 6, de los cuales solo cuatro se muestran en el diagrama 15 para facilitar la representación. El diagrama 15 asigna el tiempo t en el eje horizontal y el flujo de tráfico en vehículos por tiempo v/t en el eje vertical. Se puede ver además que cada uno de los cuatro perfiles de tráfico a modo de ejemplo 6 puede tener una o más, en el presente ejemplo dos a cuatro, etiquetas T predefinidas atribuidas a las mismas.
50

Las etiquetas T atribuidas a un perfil de tráfico 6 están asociadas de alguna manera con el momento de registrar dicho perfil de tráfico 6. Por ejemplo, una etiqueta T puede indicar un día de la semana, un mes del año, una semana del mes, una indicación de vacaciones escolares, una indicación especial de vacaciones, una condición de la carretera o una condición climática en el lapso de tiempo que su perfil de tráfico 6 había sido registrado.
55

Las etiquetas T pueden atribuirse al perfil de tráfico 6 directamente en el momento de registro por medio del dispositivo de medición de tráfico 4, en la base de datos 7, o incluso después de la agrupación en la base de datos 12. Para este fin, cada uno de los dispositivos correspondientes 4, 7, 12 puede comprender un reloj, calendario, estación meteorológica, o cualquier medio para recibir datos medioambientales o del estado de la carretera. De forma alternativa, las etiquetas T pueden atribuirse a los perfiles de tráfico 6 en retrospectiva, por ejemplo, mediante calendarios de las bases de datos 7, 12. La información del calendario, el clima, el medio ambiente o el estado de la carretera se puede atribuir en retrospectiva, por ejemplo, si una semana o un mes más tarde surge información de que la carretera en la que se registró el flujo de tráfico se dañó o estaba en construcción. Las etiquetas T incluso pueden atribuirse a los perfiles de tráfico 6 durante el paso de etiquetado por medio del motor de etiquetado 14, que puede tener su propio calendario o entradas de datos para este propósito.

Volviendo al ejemplo específico de la Figura 3, se puede ver que un primer perfil de tráfico 6 del primer grupo CL_1 tiene las etiquetas "lunes", "lluvia", "día normal (sin vacaciones)" y "junio". Un segundo perfil de tráfico 6 de este grupo tiene diferentes etiquetas T, aquí "lunes", "soleado", "día normal (sin vacaciones)" y "julio", y así sucesivamente.

Para extraer una etiqueta L_i que describa con mayor precisión un grupo CL_i de todas las etiquetas T de este grupo CL_i , el motor de etiquetado 14 calcula en primer lugar un valor de significancia SV para al menos algunas de las etiquetas T dentro de este grupo CL_i . En algunos ejemplos, se calculan los valores de significancia SV de todas las etiquetas T de todos los perfiles de tráfico 6 de un grupo CL_i , mientras que en otros ejemplos los valores de significancia SV se calculan solo para las etiquetas más comunes T de los perfiles de tráfico 6 de un grupo CL_i .

El valor de significancia SV se puede calcular mediante una estadística Chi-cuadrado. Por ejemplo, la estadística Chi Cuadrado de Pearson se puede usar si hay 5 o más perfiles de tráfico 6, y un procedimiento combinatorio exacto como la prueba exacta de Fisher se puede usar si hay menos de 5 perfiles de tráfico 6. Esto produce una descripción más precisa de la importancia de una etiqueta individual T.

Para evaluar si una etiqueta T es de hecho significativa o no, el motor de etiquetado 14 obtiene un valor umbral TH a partir de una entrada umbral 15. La entrada de umbral 15 se puede ajustar de forma manual o digital si está conectada al ordenador. Si el valor de significancia SV de una etiqueta T dentro de un grupo CL_i se encuentra por encima del umbral TH, un vector característico CV_j que comprende esta etiqueta T se asigna a este grupo CL_i ; ver Figura 3 a la derecha.

Para el primer grupo CL_1 de la Figura 3 se puede ver que cuatro etiquetas T dentro de este grupo tienen un valor de significancia SV que se encuentra por encima del valor umbral TH, de modo que se han almacenado como vectores característicos CV_1 , CV_2 , CV_3 , CV_4 . Las etiquetas correspondientes T fueron "lunes", "martes", "miércoles" y "jueves". La etiqueta L_1 de este grupo CL_1 consta de cuatro vectores característicos CV_j , cada uno de los cuales contiene una sola etiqueta T.

La Figura 4 muestra un ejemplo de cómo generar la etiqueta L_i correspondiente al segundo grupo CL_2 y la segunda etiqueta L_2 de la Figura 3. Aquí, en primer lugar, se han calculado los valores de significancia SV de tres etiquetas "vacaciones", "domingo" y "julio". Suponiendo un valor umbral TH del 90%, se puede ver que solo la etiqueta "vacaciones" con un valor de significancia SV del 92% califica como un vector característico CV_j .

Además, los valores de significancia SV para *combinaciones*, es decir, combinaciones "y", de etiquetas T pueden calcularse, ya sea para todas las etiquetas T cuyo valor de significancia SV se encuentra por debajo del valor umbral TH o para etiquetas T cuyo valor de significancia SV se encuentra por debajo del valor umbral TH pero por encima de un valor umbral adicional, por ejemplo, 25%. Para esas etiquetas T, el valor de significancia SV se calcula en *combinación* con otra etiqueta T, y si el valor de significancia SV de dicha combinación de etiquetas TC se encuentra por encima del umbral TH (o un umbral diferente), esta combinación de etiquetas TC se asigna como un vector característico CV_j para este grupo CL_i .

Dependiendo del procedimiento para calcular el valor de significancia SV, podría ser beneficioso considerar también combinaciones "o" en la medida en que no se hayan encontrado en forma de múltiples etiquetas individuales T cuyos valores de significancia SV estén por encima del umbral dentro del mismo grupo CL_i . Con este fin, el valor de significancia SV de una combinación "o" de dos etiquetas cuyo valor de significancia SV se encuentra por debajo del nivel umbral TH se calcula para ver si se deben incluir cada una como un vector característico CV_j para un grupo CL_i .

En general, este procedimiento se puede realizar también para combinaciones de etiquetas TC que comprenden tres o más etiquetas T, por ejemplo cuando el valor de significancia SV de una combinación de dos etiquetas TC nuevamente se encuentra en un rango específico de valores de umbral. Por medio de esto, se pueden asignar combinaciones de etiquetas de cualquier longitud a un grupo CL_i como vectores característicos CV_j .

Un algoritmo a modo de ejemplo de identificación de las etiquetas significativas T de un grupo CL_i se da de la forma siguiente. Para empezar, las etiquetas T son valores específicos de las llamadas variables V, por ejemplo, las variables V pueden ser "día de la semana", "día especial", "día de vacaciones escolares", "mes", "mes semana", y/o "condición climática". Las etiquetas T se asignan a su vez a cada variable V, por ejemplo, etiquetas "Lun", "Mar", "Mié",... a la variable "día de la semana", etiquetas "vacaciones", "pre-vacaciones",... a la variable "día especial", "Sí", "No" a la variable "día de vacaciones escolares" y así sucesivamente.

En un primer paso del algoritmo a modo de ejemplo, se determina mediante una de dichas pruebas estadísticas (prueba Chi cuadrado de Pearson, prueba exacta de Fisher,...) si una variable V es significativa para un grupo CL_i , por ejemplo, si la distribución del día de la semana es *a priori* significativa en este grupo CL_i .

5 En un segundo paso, para la variable V más significativa se determinan las etiquetas T que aparecen más de lo esperado.

10 En un tercer paso, se determina el valor de significancia SV para cada etiqueta T con respecto al resto de las etiquetas T para esa misma variable V y se compara con un umbral TH predefinido como se detalla anteriormente. De esta manera, se puede determinar, por ejemplo, si la distribución "sábado" es significativa con respecto al resto de la semana en este grupo. Además, se genera un SCL_i "subgrupo" adicional, que solo contiene el resto de los perfiles de tráfico 6 que no corresponden a las etiquetas significativas T determinadas anteriormente.

En un cuarto paso, se repiten los pasos dos y tres para el subgrupo SCL_i hasta que no haya más variables V o perfiles de tráfico 6 para analizar.

15 En un quinto paso, la prueba de significancia para las etiquetas T se realiza comenzando con la variable V menos significativa de acuerdo con el paso uno, pero solo para las muestras dentro de este subgrupo SCL_i y no con el tamaño completo del grupo. Si el valor de significancia SV está por encima de dicho umbral TH , esas etiquetas T se agregarán a la etiqueta del grupo.

20 En un sexto paso, el quinto paso se repite para el siguiente subgrupo SCL_i en orden de importancia creciente. Esta vez, no solo se consideran sus propios perfiles de tráfico 6 sino también los elementos de los subgrupos SCL_i analizados previamente que no eran significativos. De esta manera, se pueden calcular todos los subgrupos "significativos" SCL_i para el primer nivel, y las etiquetas T correspondientes se agregan como una etiqueta al grupo inicial CL_i .

Si hay uno o más subgrupos importantes SCL_i en el primer nivel, en un séptimo paso, el proceso se repite para cada subgrupo SCL_i en el siguiente nivel para determinar nuevos "subgrupos secundarios", considerando solo las características o variables que no han sido analizadas antes.

25 Estos pasos se repiten hasta que no haya nada más para analizar en cualquier rama y todas las etiquetas T "combinadas" se puedan agregar a la etiqueta para los subgrupos secundarios SCL_i .

Además, si un subconjunto SCL_i incluye más de una etiqueta T significativa para una variable V , entonces cada etiqueta T se analizará en una rama separada. Solo si después de analizar todas estas ramas no tienen subramas, todas las etiquetas T se agrupan.

30 Además, para calcular el valor de significancia SV de una etiqueta T en cada rama, solo se consideran los perfiles de tráfico 6 correspondientes a las variables de todos los niveles por encima de la rama analizada. Por ejemplo, cuando se analiza la etiqueta "lluvioso" de un subgrupo "viernes", el valor esperado para "lluvioso" se calculará considerando solo las condiciones climáticas para todos los "viernes" de la población total.

35 Volviendo al segundo grupo CL_2 de la Figura 3 detallado en la Figura 4, se puede ver que las combinaciones de etiquetas "domingo y lluvia" y "domingo y julio" tienen valores de significancia SV por encima del valor umbral TH . Por lo tanto, esas dos combinaciones de etiquetas TC se asignan como vectores característicos CV_1 , CV_2 al segundo grupo CL_2 .

En este ejemplo, se puede ver que si bien el segundo grupo CL_2 no es representativo de luns, días lluviosos o días solo en julio, este grupo CL_2 es representativo de domingos y domingos lluviosos en julio.

40 Además, si a dos grupos se les asignan los mismos vectores característicos CV_j , el vector característico CV_j cuya etiqueta T o combinación de etiquetas TC tiene el valor de significancia más bajo SV se puede eliminar de la etiqueta L_i de este grupo CL_i .

45 Por medio de esto, a todos los grupos CL_i se les puede asignar una etiqueta L_i que comprenda cualquier número de vectores característicos CV_j que representen este grupo CL_i . El motor de etiquetado 14 a su vez almacena estos grupos etiquetados CL_i en una base de datos 16 que puede incorporarse nuevamente como la base de datos 7 o un subconjunto de la misma, como la base de datos 12 o un subconjunto de la misma, o incorporarse como una base de datos independiente.

50 Del primer grupo CL_1 de la Figura 3 también se puede ver que, dado que cada uno de los vectores característicos CV_1 , CV_2 , CV_3 , CV_4 comprende una sola etiqueta T , pueden agruparse opcionalmente, por ejemplo, como días laborables. Esto se puede hacer para reducir el número de vectores característicos CV_j , lo cual a su vez reduce los esfuerzos computacionales cuando se trata de encontrar un determinado grupo.

Los grupos etiquetados CL_i en la base de datos 16 ahora se pueden usar para predecir un comportamiento de tráfico pasado, presente o futuro PTB (Figura 2) en el sistema de carreteras 1. Para este fin, una petición Req se puede alimentar a un motor de predicción 17, que para este propósito obtiene los grupos etiquetados CL_i de la base de datos

16. La petición Req comprende una o más etiquetas de petición RT, que pueden ser formuladas por un usuario. El motor de predicción 17 hace coincidir las etiquetas de petición RT con los vectores característicos CV_j de los grupos etiquetados CL_i para determinar un grupo "coincidente" CL_i . Este grupo coincidente CL_i , o cualquier dato derivado del mismo, es luego enviado por el motor de predicción 17 como el comportamiento de tráfico previsto PTB.

- 5 Una configuración a modo de ejemplo de los motores 11, 14, 17 puede ser la siguiente: Mientras que el dispositivo de medición de tráfico 4 adquiere datos en tiempo real en el campo, después de un período prolongado de, por ejemplo, un año, el motor de agrupamiento 11 y el motor de etiquetado 14 pueden producir los agrupamientos etiquetados CL_i en el procesador 9 de una estación central. Por otro lado, el motor de predicción 17 puede incorporarse como una entidad físicamente distinta de los motores de agrupamiento y etiquetado 11, 14. De forma alternativa, la misma máquina puede implementar el motor de agrupamiento 11, el motor de etiquetado 14 y el motor de predicción 17.

El funcionamiento del motor de predicción 17 se describe ahora con más detalle por medio de las Figs. 5 y 6. Después de comenzar el proceso de predicción, en el paso 18 se determina el grupo CL_i que comprende el mayor número N de perfiles de tráfico 6 y se establece como un grupo de referencia. En el ejemplo de la Figura 3, el primer grupo CL_1 contiene la mayoría de los perfiles de tráfico ($N = 100$) 6.

- 15 En el paso 19, se calcula una medida de disimilitud D para cada uno de los grupos CL_i . La medida de disimilitud D es una medida de cuánto difieren los perfiles de tráfico 6 de un grupo específico CL_i de los perfiles de tráfico 6 del grupo de referencia (aquí: CL_1). Una vez que se calcula una medida de disimilitud D para un grupo CL_i , se puede asignar a todos los vectores característicos CV_j de este grupo CL_i . Por definición, el grupo de referencia no difiere de sí mismo y, por lo tanto, tiene la medida de disimilitud más baja, por ejemplo, $D = 0$, asignada.
- 20 Para calcular una medida de disimilitud D para los otros grupos, por ejemplo, un perfil de tráfico promedio de un grupo CL_i se compara con un perfil de tráfico promedio del grupo de referencia. En algunos ejemplos, esto se puede hacer usando la misma métrica de distancia que se usó para generar los grupos. En el ejemplo de la Figura 3, el tercer grupo CL_3 se desvía más del grupo de referencia CL_1 y tiene una medida de disparidad $D = 12$ asignada al mismo. El cuarto grupo CL_4 solo se desvía un poco del grupo de referencia y tiene asignada una medida de disimilitud $D = 3$, mientras que el segundo grupo CL_2 tiene una medida de disimilitud $D = 7$ asignada. Se enfatiza que la medida de disimilitud D es independiente del número N de perfiles de tráfico 6 en un grupo CL_i y solo representa una diferencia del valor promedio, modal o mediano, o similar, de los perfiles de tráfico 6 entre los grupos CL_i . Cualquier medida de (di)similitud entre las curvas conocidas en el estado de la técnica puede usarse como la medida de disimilitud D tal como, por ejemplo, una norma L2, una norma L1 o normas híbridas. Los pasos 19 y 20 pueden, por ejemplo, realizarse en el procesador 9 o en el motor de etiquetado 14.

- Después de que una petición Req con etiquetas de petición RT se introduce en la máquina de predicción 17 en el paso 20, se comprueba en el paso 21 si todas las etiquetas de petición RT de la petición Req coinciden con uno de los vectores característicos CV_j de los grupos CL_i . Para este propósito, las etiquetas de petición RT se pueden combinar en la forma de un vector característico CV_j para ver si alguno de los vectores característicos CV_j de los grupos coincide con el mismo. Si este es el caso (rama "Y"), el grupo CL_i que contiene el vector característico coincidente CV_j se determina como el grupo CL_i coincidente en el paso 22.

- Al final del paso 22, el motor de predicción 17 puede generar todo el grupo coincidente CL_i o un perfil de tráfico promedio o cualquier otro dato derivado de este grupo coincidente CL_i como el comportamiento de tráfico pronosticado PTB para la petición Req. De forma alternativa, si en el paso 21 no se puede encontrar el vector característico CV_j entre los grupos CL_i (rama "N"), una de las etiquetas de petición RT se elimina de la petición Req en un paso 23, y el paso 21 se repite.

- El paso 23 de eliminar una etiqueta de petición RT de la petición Req se describe en detalle por medio de un ejemplo que se muestra en la Figura 6, que nuevamente usa los grupos a modo de ejemplo de la Figura 3. Aquí, al principio, la petición Req comprende las etiquetas de petición "miércoles", "vacaciones" y "lluvia". Cuando se realiza el paso 21 por primera vez, se comprueba si la combinación de estas tres etiquetas de petición RT aparece en cualquiera de las etiquetas L_i de los cuatro grupos CL_i ; vea las cuatro etiquetas L_1, L_2, L_3, L_4 en el lado derecho de la Figura 3. En este ejemplo, no se puede encontrar el vector característico CV_j que comprenda las tres etiquetas de petición RT.

- Para eliminar una de las etiquetas de petición RT de la petición Req, en un ejemplo del presente procedimiento para cada etiqueta de petición RT, se comprueba individualmente si existe un vector característico CV_j que coincida con esta etiqueta de petición RT, es decir, no se comprueba si esta etiqueta de petición RT simplemente ocurre en una combinación de etiquetas TC de un vector característico CV_j , sino solo si hay un vector característico CV_j que coincida exactamente con la etiqueta de petición RT. Por medio de las Figs. 3 y 6, se puede ver que la etiqueta de petición "miércoles" de la petición Req aparece como el vector característico CV_3 en la etiqueta L_1 del primer grupo CL_1 , la etiqueta de petición "vacaciones" de la petición Req aparece como el vector característico CV_1 en la etiqueta L_2 del segundo grupo CL_2 , y la etiqueta de petición "lluvia" de la petición Req aparece como el vector característico CV_1 en la etiqueta L_4 del cuarto grupo CL_4 .

Ahora se comprueba cuál de estos vectores característicos correspondientes CV_j tiene la medida de disimilitud más baja D. Dado que el vector característico CV_3 está en el primer grupo CL_1 , que tiene la medida de disimilitud más baja

$D = 0$, y los otros dos grupos CL_2 , CL_4 tienen medidas de disimilitud más altas $D = 7$, 3 , respectivamente, la etiqueta de petición "miércoles" se elimina de la petición Req porque esta etiqueta de petición RT corresponde a un vector característico CV_j del grupo CL_i con la medida de disimilitud D más baja, es decir, al grupo menos interesante CL_i , aquí: grupo CL_1 .

5 Posteriormente, se repite el paso 21 y se comprueba si existe un vector característico CV_j que contiene la combinación de etiquetas de las dos etiquetas de petición restantes "vacaciones" y "lluvia". Nuevamente, este no es el caso y, por lo tanto, se comprueba nuevamente cuál de las dos etiquetas de petición restantes RT corresponde a un grupo CL_i con la medida D de disimilitud más baja. En este caso, la etiqueta de petición "lluvia" se elimina de la petición Req porque tiene la menor medida de disimilitud D .

10 Posteriormente, se comprueba nuevamente si la petición Req, que ahora solo comprende la etiqueta "vacaciones", coincide con alguno de los vectores característicos CV_j de los grupos CL_i . Este es el caso del segundo grupo CL_2 , que tiene asignado el vector característico CV_1 . Por lo tanto, el segundo grupo CL_2 es el grupo coincidente, a partir del cual se obtiene el comportamiento de tráfico previsto PTB.

15 A veces, las etiquetas de petición RT no se producen en un vector característico CV_j que contiene solo esta etiqueta de petición RT en cualquiera de los grupos CL_i porque esta etiqueta de petición no está presente en ninguna de las etiquetas T de los perfiles de tráfico 6 que se usan para crear los grupos CL_i o porque esta etiqueta T solo se usa en un vector característico CV_j en combinación con una etiqueta adicional T. En este caso, la medida de disimilitud más baja, por ejemplo, $D = 0$, puede asignarse a esta etiqueta de petición RT.

20 En un ejemplo alternativo, cuando se comprueba cuál de las etiquetas de petición RT debe eliminarse, no solo se comprueba si hay un vector característico CV_j que solo comprende esta etiqueta de petición RT, sino que también se comprueba si hay un vector característico CV_j que comprende esta etiqueta de petición RT en combinación con una etiqueta adicional T. En este caso, se elige este vector característico CV_j y a su vez el grupo subyacente CL_i y su medida de disimilitud D , para la cual, por ejemplo, el valor de significancia SV del vector característico CV_j sea mayor o que tenga un número N mayor o menor de perfiles de tráfico 6. Por ejemplo, en la Figura 3 para la etiqueta "vacaciones", el vector característico CV_1 del segundo grupo CL_2 y el vector característico CV_2 del tercer grupo CL_3 , que tiene la combinación de etiquetas "vacaciones y sol", están disponibles. Debido a que el valor de significancia SV del vector característico CV_2 de la combinación de etiquetas "vacaciones y domingo", el tercer grupo CL_3 es más alto (98%) que el valor de significancia SV (92%) del vector característico CV_1 en el segundo grupo CL_2 , la medida de disimilitud $D = 12$ del tercer grupo CL_3 (y no $D = 7$ del segundo grupo CL_2) se elige para la etiqueta de petición "vacaciones" y se utiliza en los pasos de comparación de la Figura 6. Por lo tanto, en algunos ejemplos, en los que los valores de significancia SV de los vectores característicos CV_j se han utilizado para etiquetar los grupos CL_i , estos valores de significancia SV también se pueden utilizar al determinar el grupo CL_i que coincide con la petición Req.

35 En todos los ejemplos, cuando se comprueba qué etiqueta de petición RT corresponde al vector característico CV_j de un grupo CL_i y dos grupos tienen los mismos vectores característicos CV_j , se puede elegir el vector característico CV_j con el valor de significancia más alto SV para esta etiqueta de petición RT.

Además, cuando en el paso 21 coincidente se encuentran dos grupos CL_i coincidentes, el que tiene el alto número N de perfiles de tráfico 6 se puede determinar como el grupo coincidente y el otro se descarta.

40 Para mejorar aún más la precisión del procedimiento, no solo se puede comprobar si las etiquetas de petición única RT se producen en los vectores característicos CV_j de los grupos CL_i , sino también si las combinaciones de etiquetas de petición RT corresponden a un vector característico CV_j de un grupo CL_i . En este caso, si este vector característico CV_j correspondiente a una combinación de etiquetas de petición RT tiene la medida de disimilitud D más baja entre los vectores característicos CV_j correspondientes a todas las etiquetas de petición RT o combinaciones de etiquetas de petición, dicha combinación de etiquetas de petición no se tendrá en cuenta hasta que una etiqueta de petición RT se elimine de la petición Req.

45

REIVINDICACIONES

1. Un procedimiento para predecir un comportamiento de tráfico en un sistema de carreteras, que comprende los siguientes pasos realizados por al menos un procesador (9, 11, 14, 17) conectado a una base de datos (7, 12, 16), que contiene un conjunto de tráfico perfiles (6), con cada perfil de tráfico (6) que representa una medida de tráfico de vehículos (3) en el sistema de carreteras (1) registrado a lo largo del tiempo (t), con una o más etiquetas predefinidas (T) asociadas con un tiempo de registro del perfil de tráfico respectivo (6) atribuido a dicho perfil de tráfico (6):
 - agrupar los perfiles de tráfico (6) con sus etiquetas (T) en al menos dos grupos (CL_i) en cuanto a la similitud de los perfiles de tráfico (6);
 - asignar al menos un vector característico (CV_j) a cada grupo (CL_i) basándose en las etiquetas (T) en este grupo (CL_i), en el que cada vector característico (CV_j) contiene una etiqueta (T) o una combinación de etiquetas (TC) de dicho grupo (CL_i);
 - determinar qué grupo (CL_i) comprende la mayoría de los perfiles de tráfico (6) y establecer este grupo (CL_i) como un grupo de referencia;
 - para cada grupo (CL_i), determinar una medida de disimilitud (D) de este grupo (CL_i) con respecto al grupo de referencia y asignar la medida de disimilitud (D) a todos los vectores característicos (CV_j) de este grupo (CL_i), en el que la medida de disimilitud (D) es una medida de cuánto difieren los perfiles de tráfico (6) de un grupo (CL_i) de los perfiles de tráfico (6) del grupo de referencia;
 - recibir una petición (Req) que tiene una o más etiquetas de petición (RT);
 - comprobar si todas las etiquetas de petición (RT) coinciden con uno de los vectores característicos (CV_j) de los grupos (CL_i) y,
 - de ser así, generar un comportamiento de tráfico previsto (PTB) basado en el grupo (CL_i) con dicho vector característico coincidente (CV_j);
 - de lo contrario, eliminar una de las etiquetas de petición (RT) de la petición (Req) y volver al comienzo de dicho paso de comprobación;
 - en el que la etiqueta de petición (RT) a eliminar se elige determinando cuál de los vectores característicos (CV_j) corresponde a cada etiqueta de petición (RT) y eligiendo la etiqueta de petición (RT) cuyo vector característico correspondiente (CV_j) tiene la medida de disimilitud (D) más baja.
2. El procedimiento según la reivindicación 1, en el que, al elegir qué etiqueta de petición (RT) se eliminará de la petición (Req), también se comprueba si una combinación de etiquetas de petición (RT) corresponde a un vector característico (CV_j), y si es así, y si este vector característico (CV_j) tiene la medida de disimilitud más baja (D) entre los vectores característicos (CV_j) correspondientes a todas las etiquetas de petición (RT) o combinaciones de etiquetas de petición, dicha combinación de etiquetas de petición se descarta hasta que una etiqueta de petición (RT) se elimina de la petición (Req).
3. El procedimiento según la reivindicación 1 o 2, en el que la medida de disimilitud (D) de un grupo (CL_i) se obtiene comparando un perfil de tráfico promedio de dicho grupo (CL_i) con un perfil de tráfico promedio del grupo de referencia.
4. El procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que un vector característico corresponde a una etiqueta de petición (RT) o combinación de etiquetas de petición si el vector característico (CV_j) solo comprende esta etiqueta de petición (RT) o combinación de etiquetas de petición.
5. El procedimiento según cualquiera de las reivindicaciones 1 a 3, en el que un vector característico (CV_j) corresponde a una etiqueta de petición (RT) o combinación de etiquetas de petición si el vector característico (CV_j) comprende esta etiqueta de petición (RT) o combinación de etiquetas de petición sola o en combinación con al menos una etiqueta adicional (T).
6. El procedimiento según una cualquiera de las reivindicaciones 1 a 5, en el que se calcula un valor de significancia (SV) de cada vector característico (CV_j) de cada grupo (CL_i) y, cuando dos o más vectores característicos (CV_j) corresponden a una etiqueta de petición (RT) o combinación de etiquetas de petición de la petición (Req), no se tiene en cuenta el vector característico (CV_j) con el valor de significancia más baja (SV).
7. El procedimiento según una cualquiera de las reivindicaciones 1 a 6, en el que, si dos vectores característicos (CV_j) corresponden a una etiqueta de petición (RT) o una combinación de etiquetas de petición, y ambos tienen el mismo valor de significancia (SV), si corresponde, no se tiene en cuenta el vector característico (CV_j) del grupo (CL_i) con el número más bajo (N) de perfiles de tráfico (6).
8. El procedimiento según una cualquiera de las reivindicaciones 1 a 7, en el que cada perfil de tráfico (6) representa la medida del tráfico de vehículos (3) en el sistema de carreteras (1) durante un lapso de tiempo de 24 horas.

9. El procedimiento según una cualquiera de las reivindicaciones 1 a 8, en el que al menos una etiqueta (T) indica un día de la semana, un mes del año, una semana del mes, una indicación de vacaciones escolares o una indicación especial de vacaciones de dicho momento de registro.

5 10. El procedimiento según una cualquiera de las reivindicaciones 1 a 9, en el que al menos una etiqueta (T) indica una condición de la carretera o una condición climática en dicho momento de registro.

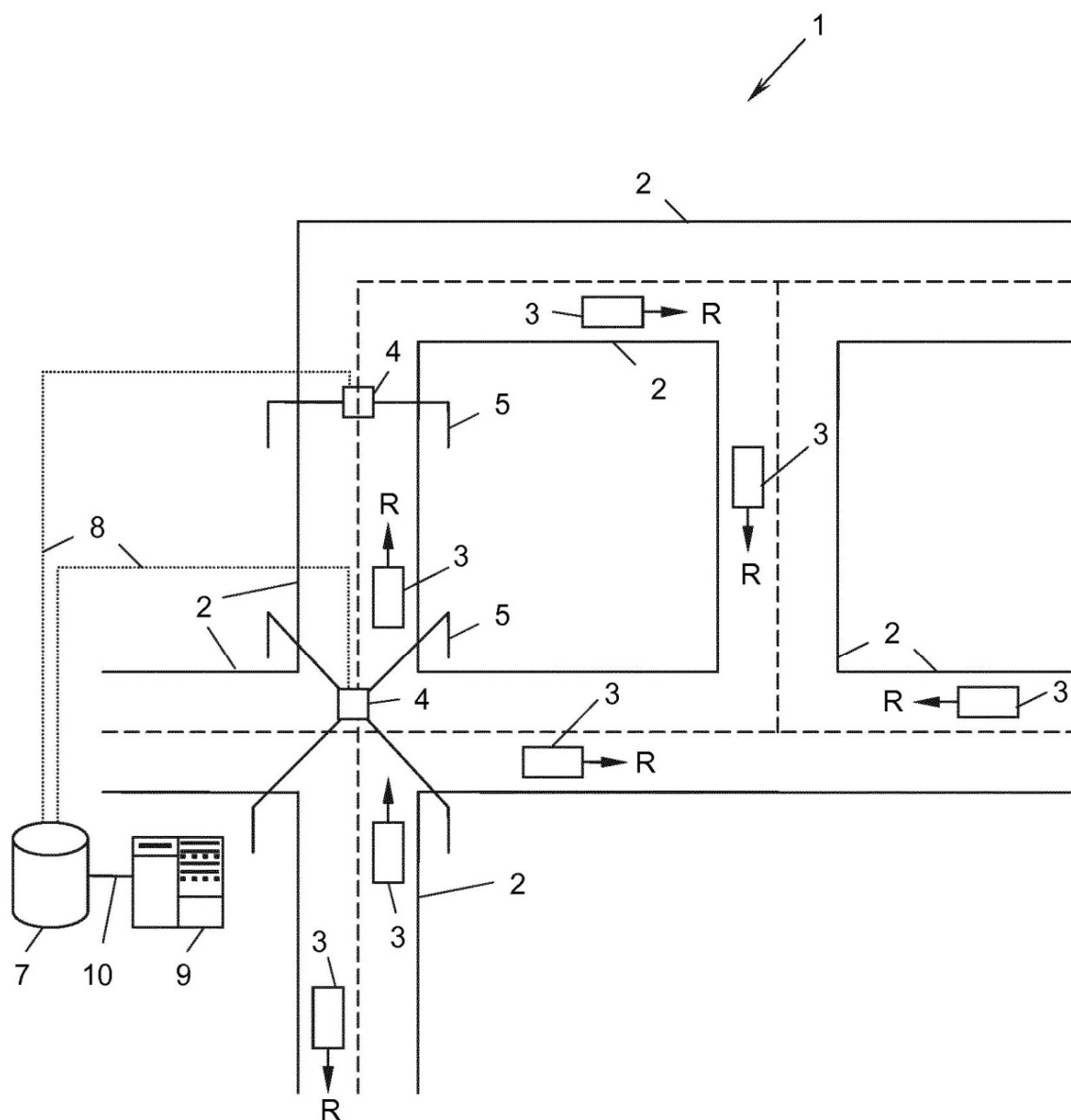


Fig. 1

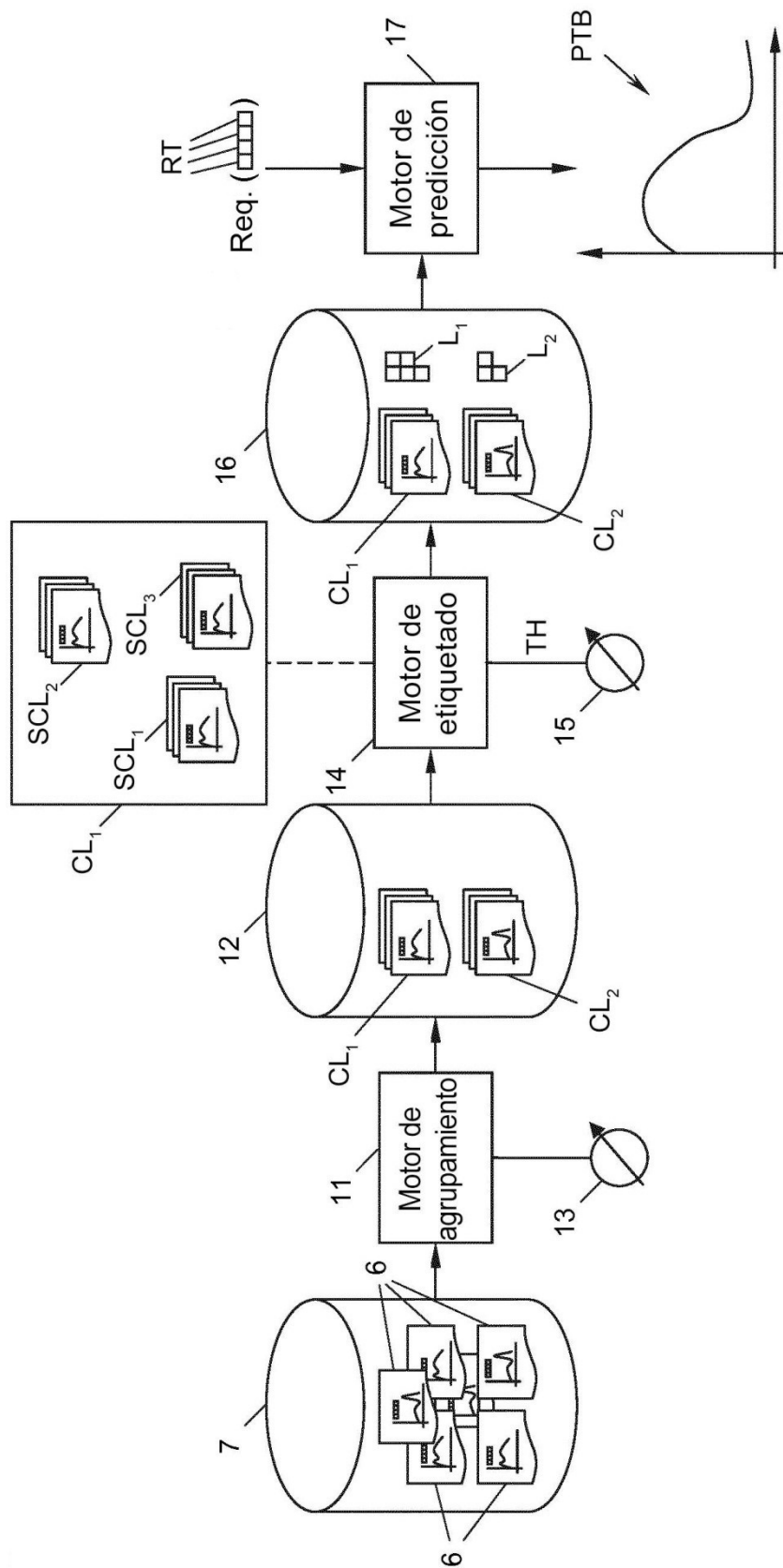


Fig. 2

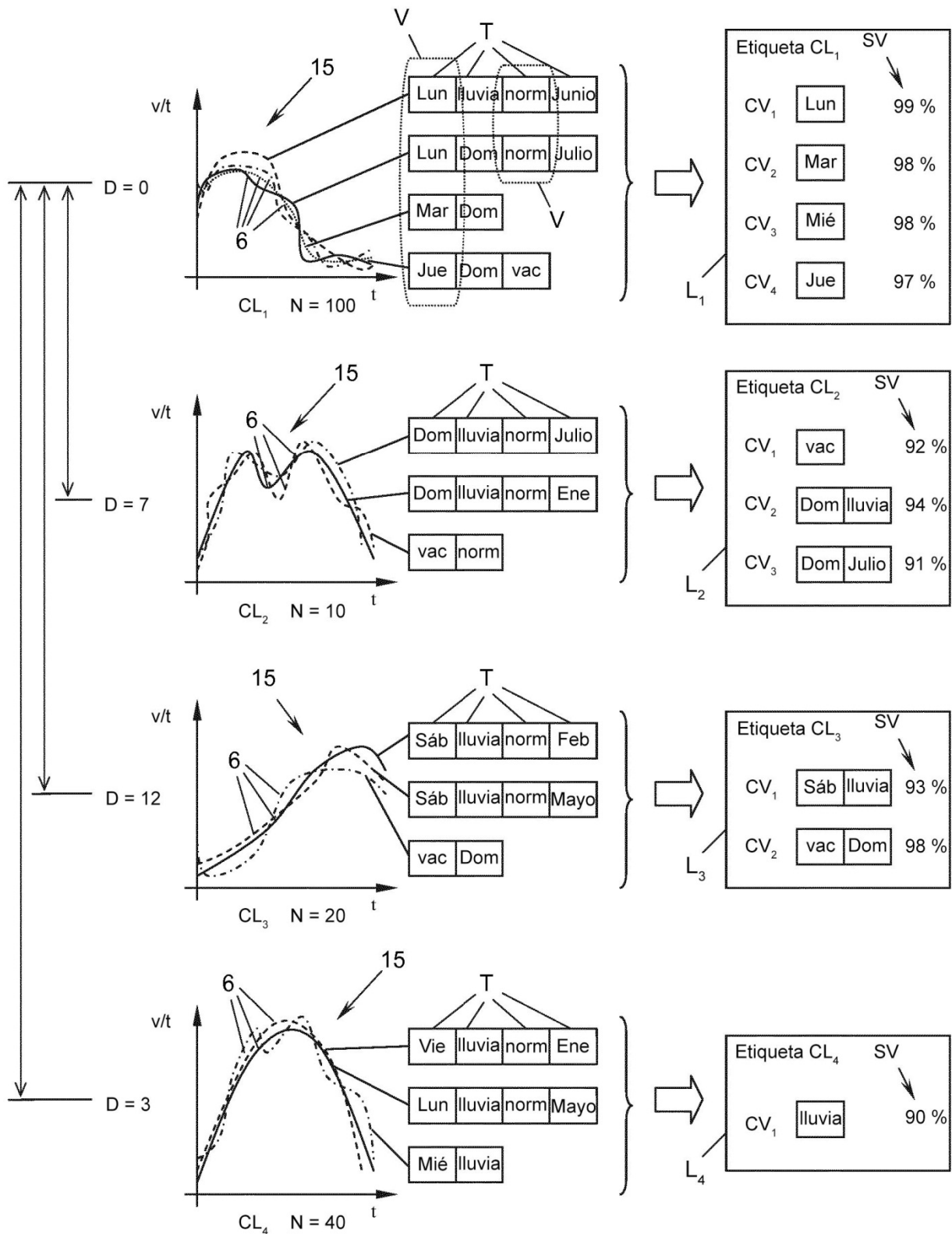


Fig. 3

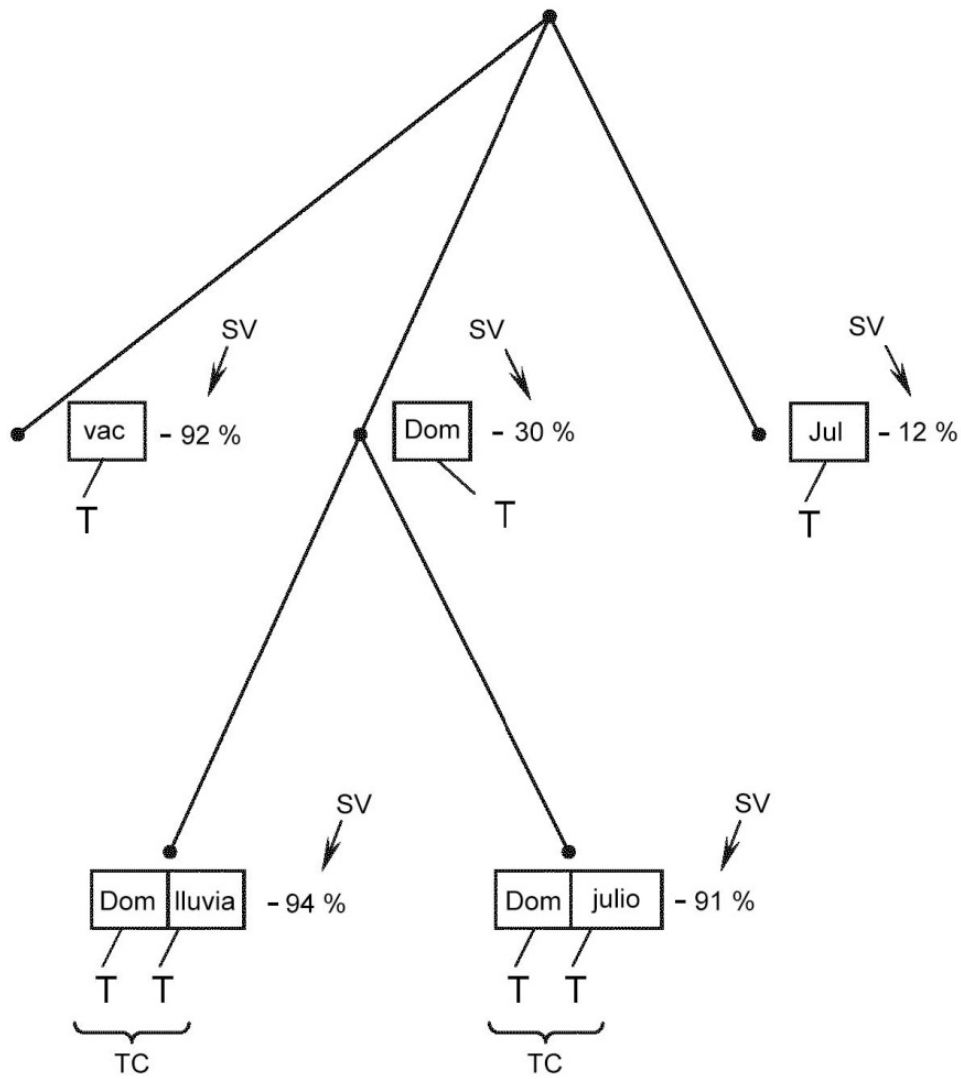
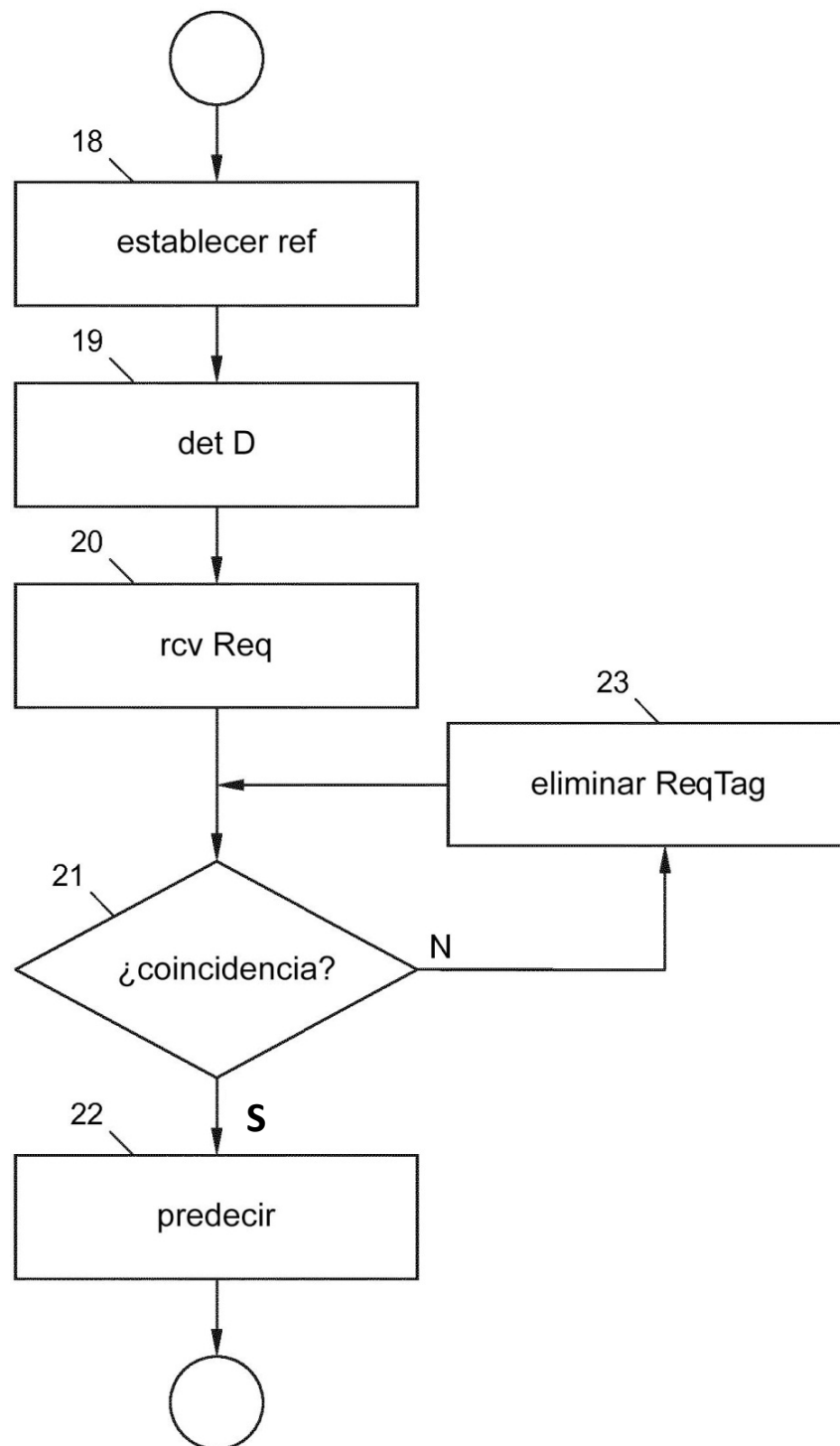


Fig. 4

**Fig. 5**

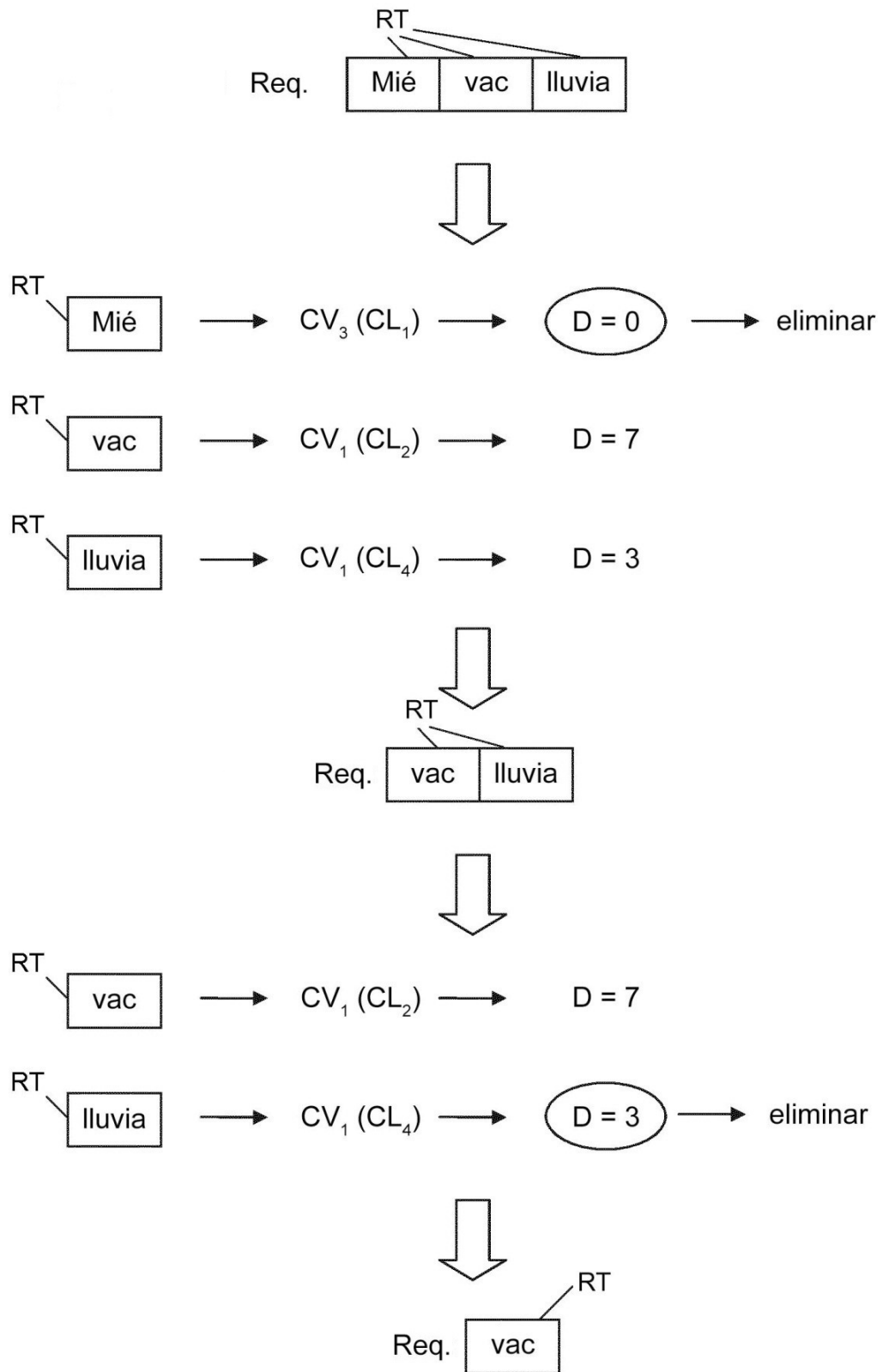


Fig. 6