

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 927 136**

51 Int. Cl.:

G01N 21/359 (2014.01)

A23K 40/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.06.2020** **PCT/EP2020/067422**

87 Fecha y número de publicación internacional: **30.12.2020** **WO20260234**

96 Fecha de presentación y número de la solicitud europea: **23.06.2020** **E 20734186 (8)**

97 Fecha y número de publicación de la concesión europea: **03.08.2022** **EP 3948231**

54 Título: **Método para predecir un pienso y/o la materia prima para piensos**

30 Prioridad:

24.06.2019 EP 19181948

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

02.11.2022

73 Titular/es:

EVONIK OPERATIONS GMBH (100.0%)

Rellinghauser Straße 1-11

45128 Essen, DE

72 Inventor/es:

REIMANN, INGOLF;

REISING, JOACHIM y

MÜLLER, CHRISTOPH

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 927 136 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

Método para predecir un pienso y/o la materia prima para piensos

La presente invención se refiere a un método para predecir la materia prima para piensos y/o un pienso de tipo desconocido por medio de espectroscopia de infrarrojo cercano y análisis de similitud.

- 5 Las dietas de los animales contienen típicamente una variedad de diferentes piensos y/o materias primas para piensos. Por lo tanto, es necesario conocer la identidad y el tipo de un pienso y/o la materia prima para piensos con la mayor precisión y rapidez posible. Esto es relevante, en particular, cuando se deben mezclar diferentes piensos y/o materias primas para piensos para obtener una dieta con una composición específica para una especie específica. Los métodos de análisis cualitativo de piensos y materias primas para piensos permiten en principio una identificación precisa de piensos
- 10 y/o materias primas para piensos de tipo desconocido, es decir, identidad, origen, etc. desconocidos. Sin embargo, estos métodos requieren equipos de laboratorio de alto coste y mantenimiento. Otras desventajas de estos métodos son sus altos estándares para el tiempo requerido y la pericia y experiencia del personal operativo. En principio, la espectroscopia de infrarrojo cercano sería un medio adecuado para la identificación y determinación de piensos y/o materias primas para piensos. El artículo "Near-infrared Spectroscopy in Food Analysis" ("Espectroscopia de infrarrojo cercano en el Análisis de Alimentos") (Osborne B. G., Encyclopedia of Analytical Chemistry, Wiley & Sons, 2006, páginas 1-14) ofrece una descripción general de la aplicación de la espectroscopia de infrarrojo cercano (NIR) en este campo. Según este artículo, la amplia gama de aplicaciones de NIR en el análisis de alimentos es posible gracias a las diferentes técnicas de presentación de muestras. Existe una técnica disponible para cualquier tipo de muestra líquida, en pasta, en polvo o sólida. Sin embargo, cuando se utiliza como un método de rutina, la espectroscopia de infrarrojo cercano requiere el conocimiento
- 20 de la identidad y el tipo de piensos y/o materias primas para piensos. Sin embargo, los errores humanos en la selección de piensos y/o materias primas para piensos ya pueden conducir a una clasificación incorrecta de un pienso y/o materia prima para piensos con respecto a su identidad y forma actual. En base a una clasificación incorrecta, se elegiría el método de calibración incorrecto para el análisis de infrarrojo cercano de los ingredientes y sus cantidades específicas en el pienso y/o la materia prima para piensos. Por lo tanto, los datos obtenidos del espectrómetro NIR mal calibrado serían erróneos.
- 25 En consecuencia, estos datos serían engañosos para cualquier etapa operativa adicional, en el que esté involucrada la materia prima para piensos respectiva y/o el pienso.

- Una opción para superar este problema es realizar un análisis de similitud del espectro de infrarrojo cercano de una sustancia de muestra con espectros de referencia para identificar coincidencias, determinando así cuál es la sustancia de muestra. El principio general de este enfoque se describe en el documento WO 2016/141 198 A1 y en el artículo "Algorithms, Strategies and Application Process of Spectral Searching Methods" ("Algoritmos, Estrategias y Proceso de Aplicación de Métodos de Búsqueda Espectral" (Chu X.-L., Li J.-Y., Chen P., Xu Y.-P., Chinese Journal of Analytical Chemistry, 2014, 42(9), 1379-1386). En detalle, un análisis de similitud implica el cálculo de una medida de similitud o una medida de distancia entre el espectro de la sustancia de muestra y los espectros de referencia. Un valor alto para la medida de similitud entre el espectro de muestra y un espectro de referencia indica una gran similitud de una sustancia de muestra con la sustancia de referencia del espectro de referencia respectivo. En comparación, cuando el análisis de similitud implica una medida de distancia, un valor de similitud bajo indica una similitud alta de una sustancia de muestra con la sustancia de referencia del espectro de referencia respectivo. Los valores de similitud obtenidos en un análisis de similitud generalmente se clasifican con las sustancias de referencia de las entradas mejor clasificadas que tienen la mayor similitud con la sustancia de muestra.
- 30
- 35

- 40 El artículo "Spectral Library Searching: Mid-Infrared Versus Near-Infrared Spectra or Classification of Powdered Food Ingredient" ("Búsqueda en la Biblioteca Espectral: Espectros de Infrarrojos Medios Frente a Espectros de Infrarrojos Cercanos para la Clasificación de Ingredientes Alimentarios en Polvo" (Reeves J. B. et al., Applied Spectroscopy, 1999, vol. 53, nº 7, páginas 836-844) describe un método para predecir un ingrediente alimentario que comprende las etapas de proporcionar espectros de infrarrojo cercano (NIR) de muestras de ingredientes alimentarios en polvo, se construyeron bibliotecas de espectros y se buscaron sobre la base de algoritmos de espectro completo. Los resultados de las diferentes búsquedas se presentan como distancias entre la muestra desconocida o de prueba y las muestras en la biblioteca de los algoritmos en cuestión. Una coincidencia se consideró exitosa cuando la muestra más cercana (distancia más pequeña) a una muestra de prueba era para una muestra del mismo grupo.
- 45

- Sin embargo, los enfoques de la técnica anterior son muy susceptibles a determinaciones o predicciones falsas cuando las entradas positivas falsas están en la parte superior de la clasificación de valores de similitud. Las razones de las entradas de falsos positivos en dicha clasificación pueden ser la asignación errónea de un espectro a la sustancia de referencia incorrecta o a la clase incorrecta de sustancias de referencia, la heterogeneidad o el desorden de una clase de sustancias de referencia, cuyos espectros se registraron, o la similitud de algunas sustancias de referencia entre sí, lo que hace bastante difícil una coincidencia exacta. Cualquiera de estos casos dificulta o imposibilita la coincidencia precisa y fiable del espectro de muestra de una sustancia de muestra con el espectro de una sustancia de referencia.
- 50
- 55

El artículo "Novel Search Algorithms for a Mid-Infrared Spectral Library of Cotton Contaminants" ("Algoritmos de Búsqueda Novedosos para una Biblioteca Espectral de Infrarrojo Medio de Contaminantes de Algodón") (Loudermilk J. B. et al, Applied Spectroscopy, junio de 2008, vol. 62, nº 6, páginas 661-670) describe algoritmos de esquema de votación utilizados en la biblioteca MIR búsqueda de contaminantes de algodón. Se realiza un conteo en el llamado algoritmo de "grupo", p.

ej., el número de veces que la sustancia/categoría "capa de revestimiento de la semilla" aparece en una lista de aciertos. Se realiza una ponderación según el rango, seguida de una suma para dar como resultado una puntuación, con el algoritmo de "frecuencia ponderada", para espectros individuales y su aparición en listas de aciertos de múltiples algoritmos de búsqueda. Dicho algoritmo de "frecuencia ponderada" trabaja por lo tanto con múltiples algoritmos de búsqueda para un único espectro.

Según la presente invención, este problema se resuelve porque primero se realiza un análisis de similitud, seguido del recuento de las apariciones de una materia prima para piensos y/o pienso en la clasificación de valores de similitud. A continuación, el número así determinado de valores de similitud de la materia prima para piensos para animales y/o del pienso para animales se pondera según su posición de rango para dar una posición de rango ponderada, la suma se forma con las posiciones de rango ponderadas para dar puntuaciones de las materias primas para piensos y/o piensos, y la puntuación más alta indica la materia prima para piensos y/o el pienso con mayor similitud con la sustancia de la muestra.

Por lo tanto, el objeto de la presente invención es un método implementado por ordenador para predecir un pienso y/o materia prima para piensos según la reivindicación 1, que comprende las etapas de

- a) recibir un espectro de infrarrojo cercano (NIR) de una muestra de una materia prima para piensos y/o pienso desconocidos,
- b) transformar las intensidades de absorción de longitudes de onda o números de onda en el espectro de la etapa a) para dar un vector de consulta,
- c) proporcionar un conjunto de vectores de base de datos de una población de espectros de materias primas para piensos y/o piensos conocidos, en el que la población de espectros de piensos y/o materias primas para piensos conocidos de la etapa c) comprende al menos 50 espectros de muestras de cada pienso y/o materia prima para piensos de cada una de sus áreas de cultivo globales,
- d) analizar la similitud entre el vector de consulta de la etapa b) y el conjunto de vectores de base de datos de la etapa c) que comprende las etapas
 - d1) calcular una medida de similitud y/o una medida de distancia entre cada vector de base de datos de la etapa c) y el vector de consulta de la etapa b) para dar un valor de similitud para cada vector de base de datos con el vector de consulta,
 - d2) clasificar los valores de similitud obtenidos en la etapa d1) en orden descendente, cuando la medida de similitud se calcula en la etapa d1) o en orden ascendente, cuando la medida de distancia se calcula en la etapa d1), en donde el vector de base de datos mejor clasificado tiene la mayor similitud con el vector de consulta,
 - d3) contar el número de ocurrencias de cada una de las materias primas para piensos y/o piensos entre los vectores de base de datos mejor clasificados en la clasificación de la etapa d2), en donde dicho número de ocurrencias se indica mediante la variable N por cada materia prima para piensos y/o pienso,
 - d4) ponderar los primeros N valores de similitud de cada una de las materias primas para piensos y/o piensos según su posición en la clasificación de la etapa d2) para dar posiciones de rango ponderadas de cada una de las materias primas para piensos y/o piensos, en donde la ponderación de valores de similitud se realiza tomando la posición de rango recíproca de la materia prima para piensos y/o del pienso en la clasificación de la etapa d2),
 - d5) formar la suma de las posiciones de rango ponderadas de la etapa d4) para cada una de las materias primas para piensos y/o piensos para dar puntuaciones de cada una de las materias primas para piensos y/o piensos, y
- e) asignar la materia prima para piensos y/o pienso del vector de base de datos con mayor puntuación a la muestra de la etapa a).

En el contexto de la presente invención, el término materia prima para piensos y/o piensos desconocidos se refiere a cualquier tipo de pienso y/o materia prima para piensos cuya identidad, composición, origen y/o forma, es decir, si está molido o sin moler, es desconocido. En comparación, en el contexto de la presente invención, el término materia prima conocida para piensos y/o pienso se refiere a cualquier tipo de pienso y/o materia prima para piensos cuya identificación, composición, origen y/o forma, es decir, si es molido o sin moler, es conocido. Por consiguiente, una población de espectros de materias primas para piensos y/o piensos conocidos es un número o multitud de espectros, que se sabe que pertenecen a un pienso y/o materia prima específica para piensos de identidad, composición, origen y/o forma conocidos.

La ponderación de los valores de similitud en la etapa d4) típicamente se realiza tomando la posición de rango recíproco de la materia prima específica para piensos y/o del pienso en la clasificación de la etapa d2). Por ejemplo, si hay una ocurrencia de una materia prima específica para piensos en la posición 2 en la clasificación, su ponderación da el valor $\frac{1}{2}$

para la posición de rango ponderado. Los primeros N valores de similitud de cada una de las materias primas para piensos y/o piensos se ponderan según su posición en la clasificación de la etapa d2) para dar posiciones de rango ponderadas de cada una de las materias primas para piensos y/o piensos en la etapa d4). A continuación, en la etapa d5) se forma la suma de las posiciones de rango ponderadas de la etapa d4) para cada una de las materias primas para piensos y/o piensos para dar puntuaciones de cada una de las materias primas para piensos y/o piensos.

En su sentido más amplio, un vector es un objeto geométrico que tiene magnitud (o longitud) y dirección. En un sistema de coordenadas Cartesianas, un vector se puede representar identificando las coordenadas de su punto inicial y terminal. Por lo tanto, un vector es adecuado para representar una intensidad de absorción en una longitud de onda específica o número de onda en un espectro de infrarrojo cercano bidimensional. Además, un vector no se limita a la descripción de un sistema bidimensional. Más bien, un vector puede describir espacios multidimensionales, tal como un espectro de infrarrojo cercano con una multitud de intensidades de absorción en una multitud de longitudes de onda o números de onda diferentes. En este caso, cada dimensión de dicho vector corresponde a una única intensidad de absorción a una longitud de onda o número de onda específico. En el contexto de la presente invención, un vector de base de datos también contiene la identificación de su materia prima para piensos y/o pienso y, opcionalmente, su origen u otra información, tal como si es molido o sin moler, o la temporada en que se cosechó, en caso de que sea un maíz. Alternativamente, cada vector de base de datos tiene un número de identificación y la información mencionada anteriormente se almacena bajo el número de identificación en una unidad de procesamiento, tal como un ordenador, una nube o un servidor.

En una realización del método implementado por ordenador según la presente invención, el vector en las etapas b) y c) es un vector multidimensional, correspondiendo cada dimensión a una intensidad de absorción de una longitud de onda o número de onda específico.

Al igual que el vector de consulta del espectro de una materia prima para piensos y/o pienso desconocido, también se obtiene el conjunto de vectores de base de datos proporcionados en la etapa c) del método implementado por ordenador según la presente invención transformando cada espectro de una población de espectros de materias primas para piensos y/o piensos conocidos en el vector correspondiente. Si el conjunto de vectores de base de datos no está ya presente, la etapa c) también comprende la transformación de cada espectro de una población de espectros de materias primas para piensos y/o piensos conocidos en el vector correspondiente para dar el conjunto de vectores de base de datos. En ese caso, la etapa c) del método implementado por ordenador según la presente invención comprende las etapas de transformar cada espectro de una población de espectros de materias primas para piensos y/o piensos conocidos en el vector correspondiente para dar un conjunto de vectores de conjunto de datos y proporcionar el conjunto así obtenido de vectores de base de datos de una población de espectros de materias primas para piensos y/o piensos conocidos.

Según la presente invención, se recibe en la etapa a) un espectro de infrarrojo cercano de una muestra de una materia prima para piensos y/o pienso desconocidos. En el contexto de la presente invención, esto significa que el lugar donde se registra el espectro que se ha de recibir y el lugar donde se realiza el método implementado por ordenador según la presente invención, pueden ser diferentes o idénticos. Por ejemplo, es posible que un espectro de infrarrojo cercano de una muestra de una materia prima para piensos y/o un pienso desconocido se registre en un lugar y se envíe de alguna manera a un lugar remoto, donde se realiza el método implementado por ordenador según la presente invención. Alternativamente, tanto el registro del espectro como la predicción de la materia prima para piensos y/o del pienso en base a dicho espectro se pueden realizar en el mismo lugar.

En una realización, la etapa a) del método implementado por ordenador, la etapa puede comprender el registro de un espectro de infrarrojo cercano de una muestra de una materia prima para piensos y/o pienso desconocidos.

En principio, el espectro de infrarrojo cercano (NIR) de la etapa a) se puede registrar en longitudes de onda de 700 a 2.500 nm con cualquier espectrómetro de infrarrojo cercano adecuado que funcione con el monocromador o con el principio de transformada de Fourier. Sin embargo, se encontró que solo la parte de los espectros de infrarrojo cercano de 1.100 a 2.500 nm es necesaria para la predicción de la materia prima para piensos y/o del pienso en el método implementado por ordenador según la presente invención. Por lo tanto, se prefiere que el espectro cubra el registrado en el intervalo de 1.100 a 2.500 nm en la etapa a) del método implementado por ordenador según la presente invención. Por consiguiente, la población de espectros de materias primas para piensos y/o piensos conocidos de la etapa c) del método implementado por ordenador solo necesita cubrir el intervalo de 1.100 a 2.500 nm. Dado que las longitudes de onda se pueden convertir fácilmente en los números de onda respectivos, el espectro de infrarrojo cercano de la etapa a) y/o los espectros de la etapa c) también se pueden registrar en los números de onda correspondientes. Cuando la muestra de una materia prima para piensos y/o pienso en la etapa a) no es translúcida, se mide la reflectancia de la luz emitida por la muestra y la diferencia entre la luz emitida y la luz reflejada se da como absorción. Las absorciones así obtenidas, es decir, sus intensidades y sus longitudes de onda o números de onda, se utilizan para la generación de los vectores en la etapa b) y/o en la etapa c). Por consiguiente, un espectrómetro de infrarrojo cercano adecuado para su uso en el método según la presente invención puede funcionar en el modo de transmitancia o en el modo de reflectancia.

En una realización del método implementado por ordenador según la presente invención, el espectro en la etapa a) y/o en la etapa c) se registra en un intervalo de 1.100 a 2.500 nm.

El método según la presente invención no se limita a una medida específica de distancia o similitud para analizar la similitud

entre el vector de consulta de la etapa b) y los vectores de base de datos de la etapa c). Por lo tanto, cualquier medida de distancia o similitud que sea adecuada para determinar la similitud del vector de la etapa b) con los vectores de la etapa c) se puede utilizar en el método según la presente invención. En principio, un análisis de similitud se basa en una búsqueda del vecino más cercano. Se encontró que el coeficiente del coseno es una medida de similitud particularmente adecuada para una búsqueda del vecino más cercano en el método según la presente invención. Por ejemplo, el coeficiente del coseno, que permite el cálculo de la similitud entre dos vectores de forma extremadamente rápida con una alta precisión, es particularmente adecuado para el método según la presente invención. El coeficiente $S_{A,B}$ del coseno de dos vectores A y B se representa mediante la siguiente fórmula

$$S_{A,B} = \frac{[\sum_{j=1}^{j=n} x_{jA} x_{jB}]}{\left[\sqrt{\sum_{j=1}^{j=n} (x_{jA})^2} \sqrt{\sum_{j=1}^{j=n} (x_{jB})^2} \right]}$$

donde x_{jA} y x_{jB} son componentes de los vectores A y B, respectivamente, y n es el número de espacios, aquí el número de intensidades de absorción en longitudes de onda o números de onda específicos. Los valores para la similitud van desde -1, que significa exactamente lo opuesto entre sí, hasta 1, que significa identidad, con 0 que indica ortogonalidad (descorrelación) y valores intermedios que indican similitud o diferencia intermedia.

Alternativamente, la similitud entre los vectores también se puede calcular por medio de una medida de distancia. Por ejemplo, la distancia euclidiana, que también permite el cálculo de la similitud entre dos vectores de forma extremadamente rápida y precisa, es particularmente adecuada para el método según la presente invención. La distancia euclidiana $D_{A,B}$ de dos vectores A y B se representa mediante la siguiente fórmula

$$D_{A,B} = \sqrt{\sum_{j=1}^{j=n} (x_{jA} - x_{jB})^2}$$

donde x_{jA} y x_{jB} son componentes de los vectores A y B, respectivamente, y n es el número de espacios, aquí el número de intensidades de absorción en longitudes de onda o números de onda específicos.

Por tanto, se prefiere que en la etapa d1) del método implementado por ordenador según la presente invención, la medida de similitud sea el coeficiente del coseno y la medida de la distancia sea la distancia euclidiana.

Según la etapa b) de la presente invención, las intensidades de absorción de longitudes de onda o números de onda en un espectro se transforman para dar un vector de consulta. En principio, uno podría seleccionar las intensidades de absorción más fuertes y, por lo tanto, más significativas en un espectro y transferir solo dichas intensidades de absorción para dar un vector. Sin embargo, esto requeriría un análisis exhaustivo de cada espectro individual de una sustancia de muestra, lo que no solo requiere mucho tiempo sino que también requiere una muy buena comprensión de los espectros de infrarrojo cercano. Por lo tanto, este enfoque no sería adecuado para un análisis de rutina. Además, este enfoque tiene la desventaja de que pueden ignorarse intensidades de absorción significativas pero relativamente débiles en un espectro, de modo que la información se perdería. Esto podría dar lugar a una asignación incorrecta de la materia prima para piensos y/o del pienso desconocidos al final. Por lo tanto, es favorable considerar la mayor cantidad de información posible en el espectro sin un análisis previo en profundidad del espectro. Por lo tanto, se prefiere transformar las intensidades de absorción de longitudes de onda o números de onda equidistantes en un espectro, es decir, en la etapa b) y/o c), para dar un vector de dicho espectro. Para permitir el mejor análisis de similitud posible entre el vector de consulta y los vectores de la base de datos, se prefiere transformar las intensidades de absorción de longitudes de onda o números de onda equidistantes en un espectro de la etapa b) y la etapa c) del método según la presente invención para dar un vector de dicho espectro o espectros. Preferiblemente, las distancias de las intensidades de absorción que se transforman en vectores en la etapa b) son idénticas a las distancias de las intensidades de absorción que se transforman en vectores en la etapa c). Esto permite una mayor precisión en la predicción del método implementado por ordenador según la presente invención, incluso sin tener ningún conocimiento específico de la sustancia de muestra y sus espectros.

En una realización en la etapa b) y/o la etapa c) del método implementado por ordenador según la presente invención, las intensidades de absorción de longitudes de onda y/o números de onda equidistantes en un espectro se transforman para dar un vector de un espectro de la etapa a) y/o de la etapa b).

En otra realización del método implementado por ordenador según la presente invención, las distancias de las intensidades de absorción que se transforman en vectores en la etapa b) son idénticas a las distancias de las intensidades de absorción que se transforman en vectores en la etapa c).

Preferiblemente, las intensidades de absorción de longitudes de onda o números de onda en un espectro, que se transforman para dar un vector de dicho espectro, tienen pequeñas distancias entre sí. Esto tiene la ventaja de que la mayoría, si no todas, las intensidades de absorción relevantes, es decir, la información, de un espectro se transforman en

un vector de dicho espectro. Se cree que esto permite una transformación muy precisa de toda la información relevante de un espectro en vectores, incluso sin tener conocimiento del pienso y/o de la materia prima para piensos cuyo espectro se registró, en particular de su identidad, composición, origen y/o forma. Preferiblemente, la distancia entre las longitudes de onda en la etapa b) del método según la presente invención es de $0,1 \pm 10\%$ a $10 \pm 10\%$ nm, de $0,1 \pm 10\%$ a $5 \pm 10\%$ nm, o de $0,1 \pm 10\%$ a $2 \pm 10\%$ nm. Por consiguiente, la distancia entre los números de onda en la etapa b) del método según la presente invención es de $10^8 \pm 10\%$ a $10^6 \pm 10\%$, de $10^8 \pm 10\%$ a $5 \cdot 10^6 \pm 10\%$ nm, o de $10^8 \pm 10\%$ a $2 \cdot 10^6 \pm 10\%$ nm. En el contexto de la presente invención, el término $\pm 10\%$ se utiliza con respecto a los valores explícitamente mencionados para indicar que las desviaciones de dichos valores explícitamente mencionados todavía están dentro del alcance de la presente invención, siempre que conduzcan esencialmente a los efectos de la presente invención. Las distancias entre las longitudes de onda o los números de onda en la etapa c) del método según la presente invención son preferiblemente las mismas que en la etapa b), para proporcionar la mejor comparación posible entre el espectro registrado de un pienso y/o materia prima para piensos desconocidos y los espectros de piensos y/o materias primas para piensos conocidos.

En una realización del método implementado por ordenador según la presente invención, la distancia entre las longitudes de onda o los números de onda en la etapa b) y/o la etapa c) es de $0,1 \text{ nm} \pm 10\%$ a $10 \text{ nm} \pm 10\%$ o de $10^8 \text{ cm}^{-1} \pm 10\%$ a $10^6 \text{ cm}^{-1} \pm 10\%$.

En principio, el método implementado por ordenador según la presente invención no está sujeto a ninguna limitación con respecto al número de intensidades de absorción a transformar para dar un vector. Más bien, el número de información relevante en un espectro de una materia prima para piensos y/o un pienso depende en gran medida de la materia prima para piensos y/o del pienso individuales y, en particular, de su composición y componentes. Cuanto más complejo es un pienso y/o una materia prima para piensos, es decir, cuantos más componentes contiene un pienso y/o una materia prima para piensos, más información se requiere de un espectro de infrarrojo cercano para predecir un pienso y/o una materia prima para piensos desconocidos. Nuevamente, no sería práctico realizar un análisis en profundidad para averiguar las intensidades de absorción que necesariamente deben transferirse para dar un vector. Una opción adecuada para transformar el número de intensidades de absorción en un vector es correlacionarlas con la distancia entre las longitudes de onda o números de onda correspondientes, p. ej., de $0,1 \text{ nm} \pm 10\%$ a $10 \text{ nm} \pm 10\%$ o de $0,1 \pm 10\%$ a $2 \pm 10\%$ nm, y el intervalo de registro del espectro, p. ej., de 1.100 a 2.500 nm . Preferiblemente, el número de intensidades de absorción en cada espectro que se va a transformar en un vector es de al menos 100, en particular, dicho número oscila entre $150 \pm 10\%$ y $15.000 \pm 10\%$ o entre $700 \pm 10\%$ y $15.000 \pm 10\%$.

En principio, la etapa d3) del método según la presente invención no está sujeta a ninguna limitación en cuanto al número de vectores de mayor rango, aparte del número de entradas en la clasificación de la etapa d2). Sin embargo, la precisión de la predicción en el método según la presente invención no se ve afectada cuando la etapa d3) implica la utilización de un número relativamente bajo de vectores de base de datos ni se mejora la precisión de ninguna manera, cuando la etapa d3) implica la utilización de un número relativamente alto de vectores de base de datos. Más bien, cuando d3) implicó la utilización de una mayor cantidad de vectores de base de datos en la clasificación de la etapa d2), más allá de que los vectores de base de datos estuvieran realmente en la parte superior de la clasificación, esto daría como resultado un número creciente de vectores de base de datos menos significativos, lo que disminuiría la precisión de la predicción en el método según la presente invención. Por lo tanto, se prefiere que el número N de los primeros vectores de la base de datos mejor clasificados en la etapa d3) oscile de 10 a 100, de 10 a 90, de 10 a 80, de 10 a 70, de 10 a 60, de 10 a 50, de 10 a 40, 10 a 30, o 10 a 20. Esto conduce a predicciones muy precisas equivalentes a la utilización de una mayor cantidad de vectores de bases de datos, sin embargo, sin aumentar innecesariamente el esfuerzo de cálculo, como es el caso de más vectores de bases de datos.

En una realización del método según la presente invención, el número de vectores de base de datos mejor clasificados que se ha de considerar en la etapa d3) oscila de 10 a 100, preferiblemente de 10 a 50.

La población de espectros de materias primas para piensos y/o piensos conocidos utilizados en el método según la presente invención no se limita a materias primas para piensos y/o piensos específicos. Más bien, comprende preferiblemente espectros de todas las materias primas para piensos y/o piensos utilizados en nutrición animal, preferiblemente en la nutrición de aves, cerdos, cerdos y/o animales criados en acuicultura, tales como peces y/o crustáceos. Los espectros de las materias primas para piensos y/o de los piensos pueden diferir significativamente dependiendo de su forma o apariencia, p. ej., cuando están presentes en forma molida o sin moler. Por lo tanto, la población de espectros también comprende preferentemente espectros, que se registraron a partir de las materias primas para piensos antes mencionadas y/o de los piensos en forma molida y/o sin moler.

En otra realización del método según la presente invención, la población de espectros de piensos y/o de materias primas para piensos conocidos en la etapa c) de dicho método comprende espectros de todos los piensos y/o materias primas para piensos en forma molida y/o no molida utilizados en nutrición animal.

En principio, el método según la presente invención no está limitado en modo alguno en cuanto al número y tipos de piensos y/o materias primas para piensos, cuyos espectros, registrados en forma molida y/o sin moler, constituyen la población de espectros. No obstante, se prefiere que la población de espectros de piensos y/o materias primas para piensos conocidos en la etapa c) de dicho método comprenda espectros de todos los piensos y/o materias primas para

- piensos en forma molida y/o sin moler utilizados en nutrición animal, preferiblemente en la nutrición de aves, cerdos, cerdos y/o animales mantenidos en acuicultura, tales como pescados y/o crustáceos. La materia prima para piensos y/o el pienso puede ser de origen animal y/o vegetal. Los piensos y/o las materias primas para piensos particularmente preferidos son las materias primas para piensos y/o los piensos sin procesar y/o procesados. Son materias primas para piensos y/o
- 5 piensos procesados, aquellos que fueron sometidos a cualquier tipo de tratamiento térmico o a presión con el fin de eliminar o desintoxicar los factores antinutritivos. Los piensos y/o la materia prima para piensos preferidos son semillas oleaginosas, en particular harina de extracción y torta de presión de soja, soja entera, harina de colza y torta de presión, harina de extracción de algodón, harina de extracción de cacahuete, harina de extracción de girasol, harina de extracción de coco y/o harina de extracción de almendra de palma; legumbres, en particular harina de guar tostada; subproductos de
- 10 cervecía y destilación, en particular granos de destilería secos con solubles (DDGS), subproductos del procesamiento de granos y producción de piensos, en particular gluten de maíz, harina de semilla de maíz y/o subproductos de panadería; subproductos animal, en particular harina de pescado, harina de carne, harina de aves, harina de sangre y/o harina de huesos; y también cualquier tipo de granos. En particular, la materia prima para piensos es soja, semillas de soja o un producto de soja.
- 15 Dependiendo de factores tales como el clima, el suelo y la genética de las plantas, las materias primas para piensos y/o los piensos de diferentes áreas de cultivo globales pueden diferir en sus ingredientes y en el contenido de dichos ingredientes. Para permitir predicciones fiables y reproducibles de un pienso y/o materias primas para piensos, por lo tanto, la materia prima para piensos y/o el pienso, cuyos espectros son parte de la población de espectros, proviene de todas sus áreas de cultivo globales.
- 20 El número de espectros de las materias primas para piensos y/o de los piensos de la población de espectros utilizado en el método según la presente invención debería ser representativo para permitir una predicción fiable y reproducible de la materia prima para piensos y/o del pienso en cuestión. Por lo tanto, la población de espectros comprende al menos 50 espectros de muestras de cada pienso y/o materia prima para piensos, es decir, cada pienso y/o materia prima para piensos en forma molida y/o sin moler utilizado en la alimentación animal, preferentemente en la nutrición de aves, cerdos, cerdos
- 25 y/o animales mantenidos en acuicultura, tales como pescado y/o crustáceos, de cada una de sus áreas de cultivo globales. El método según la presente invención no está sujeto a ninguna limitación con respecto al número de espectros de muestras de cualquier pienso y/o materia prima para piensos de cualquiera de sus áreas de cultivo globales. Por lo tanto, el número de espectros de muestras de cualquier pienso y/o materia prima para piensos de cualquiera de sus áreas de cultivo globales puede oscilar entre 50 y 10.000, entre 50 y 5.000, entre 50 y 2.500, entre 50 y 2.000, entre 50 y 1.500, de 50 a 1.000, de 100 a 1.000, de 50 a 500, de 100 a 500, de 50 a 250, de 100 a 250, o de 50 a 100.
- 30 Cuando la población de espectros de piensos y/o materias primas para piensos de tipo conocido considera cada área de cultivo global de un pienso y/o materia prima para piensos y el número de espectros de cada área de cultivo global es representativo, el método según la presente invención permite no solo una predicción fiable y reproducible de la materia prima para piensos y/o del pienso en cuestión, sino también una predicción del origen de la materia prima para piensos y/o del pienso en cuestión.
- 35 Por lo tanto, dicha población de espectros de piensos y/o materia prima para piensos conocidos de la etapa c) comprende preferiblemente al menos 50 espectros de muestras de cada pienso y/o materia prima para piensos de cada una de sus áreas de cultivo globales. El número de espectros de muestras de un pienso y/o materia prima para piensos de cada área de cultivo global no está sujeto a ninguna limitación. Por lo tanto, el número de espectros de muestras de cualquier pienso y/o materia prima para piensos de cada área de cultivo global puede oscilar de 50 a 10.000, de 50 a 5.000, de 50 a 2.500, de 50 a 2.000, de 50 a 1.500, de 50 a 1.000, de 100 a 1.000, de 50 a 500, de 100 a 500, de 50 a 250, de 100 a 250, o de 50 a 100.
- 40 La población de espectros de piensos y/o materia prima para piensos conocidos puede ser una base de datos ya existente, preferiblemente una base de datos constantemente actualizada, y/o puede ser una nueva base de datos, que se actualiza constantemente o se desarrolla más. Una base de datos de espectros de piensos y/o materia prima para piensos de tipo conocido adecuada para utilizar en el método según la presente invención es, por ejemplo, AMINODat® 5.0 de Evonik.
- 45 Pueden surgir casos en los que la posición de un pico de señal en un espectro, ya sea en la etapa b) y/o en la etapa c) del método según la presente invención no se puede ubicar porque los máximos y mínimos de los picos individuales no se pueden identificar claramente en tal espectro. Es posible una localización más fácil de picos individuales en el espectro, cuando los mínimos y máximos de los picos son más fáciles de identificar. Tomar la primera derivada de un espectro facilita la identificación de los picos en el espectro porque da un cruce por cero de picos máximos o picos mínimos. Tomar la segunda derivada da un pico mínimo exactamente en esa posición, donde había un pico máximo en el espectro original y viceversa. Tomar la primera o la segunda derivada de un espectro también facilita la identificación de un valor atípico en la población de espectros de materias primas para piensos y/o piensos conocidos.
- 50 En otra realización del método según la presente invención, se forma una derivada del espectro de la materia prima para piensos y/o piensos desconocidos de la etapa a) y/o de los espectros de las materias primas para piensos y/o piensos de la etapa c).
- 55 Preferiblemente, se forma la primera derivada del espectro de un pienso y/o materia prima para piensos de tipo

desconocido de la etapa a) y/o de los espectros de piensos y/o materias primas para piensos conocidos de la etapa c).

Se prefiere proporcionar el conjunto de vectores de base de datos de una población de espectros de materias primas para piensos y/o piensos conocidos directamente en la etapa c) del método implementado por ordenador según la presente invención. No obstante, también es posible que primero se proporcione solo una población de espectros de materias primas para piensos y/o piensos conocidos, que se transforma en la etapa siguiente al conjunto de vectores de base de datos para el análisis de similitud en la etapa d). En este caso, la etapa c) del método implementado por ordenador según la presente invención también comprende la etapa de transformar las intensidades de absorción de longitudes de onda o números de onda en cada espectro de una población de espectros de materias primas para piensos y/o piensos conocidos en vectores. La multitud de vectores así obtenida de la población de espectros de materias primas para piensos y/o piensos conocidos es entonces el conjunto de vectores de base de datos, como se mencionó anteriormente. En cualquier caso, se prefiere almacenar la población de espectros de materias primas para piensos y/o piensos conocidos o el conjunto de vectores de base de datos de dicha población de espectros de materias primas para piensos y/o piensos conocidos en una unidad de procesamiento, tal como un ordenador o una nube. La unidad de procesamiento en la que se almacena la población de espectros o los vectores de base de datos puede ser idéntica o diferente de la unidad de procesamiento, que lleva a cabo el método implementado por ordenador según la presente invención. En el segundo caso, la primera unidad de procesamiento, que lleva a cabo el método implementado por ordenador según la presente invención, y la segunda unidad de procesamiento, en la que se almacena la población de espectros o los vectores de base de datos, forman una red. Por ejemplo, también es posible que la población de espectros o los vectores de base de datos estén almacenados en una nube. En ese caso, la primera unidad de procesamiento, p. ej., un ordenador, que lleva a cabo el método implementado por ordenador según la presente invención, y la segunda unidad de procesamiento, p. ej., una nube, en la que se almacena la población de espectros o los vectores de base de datos, forman una red.

Otro objeto de la presente invención es por lo tanto un sistema de predicción de una materia prima para piensos y/o pienso según la reivindicación 11.

En el caso de que la población de espectros de materias primas para piensos conocidas se almacene en un ordenador, se prefiere que dicha población se almacene en un segundo ordenador, es decir, un ordenador, que no lleve a cabo ya el método implementado por ordenador según la presente invención. Como consecuencia, la carga de trabajo se distribuye uniformemente entre los ordenadores, lo que puede conducir a una ejecución más rápida del método implementado por ordenador según la presente invención. Esto también permite una comunicación entre el usuario y el proveedor de una base de datos con el vector de base de datos para utilizar en el método según la presente invención, por ejemplo, una actualización de los vectores de base de datos.

Descripción de la figura:

La figura 1 es una representación del método implementado por ordenador según la presente invención. Comenzando con un vector de consulta para un espectro de una materia prima para piensos y/o pienso desconocidos, los valores de similitud se calculan para cada vector de base de datos de una población de espectros de materias primas para piensos y/o piensos conocidos. Los valores de similitud así obtenidos se ordenan de forma ascendente, con el mayor en la parte superior, si los valores de similitud se calcularon con una medida de similitud, o descendente, con el menor en la parte superior, si los valores de similitud se calcularon con una medida de distancia, para dar una clasificación. Los n vectores mejor clasificados, que representan los vecinos más cercanos a los vectores de consulta en la búsqueda de similitud, se seleccionan a partir de esta clasificación y se determinan las ocurrencias de las clases correspondientes, es decir, de las materias primas para piensos y/o piensos. Dichas clases se ponderan según el rango de sus vectores y se suman para dar puntuaciones de clase. Las puntuaciones de clase se clasifican según los valores con el valor más alto en la parte superior. La materia prima para piensos y/o el pienso de la clase mejor clasificada tiene la mayor similitud con el vector de consulta y, por lo tanto, se asigna a la muestra del vector de consulta.

Ejemplo:

El siguiente es un ejemplo para ilustrar el método implementado por ordenador según la presente invención en comparación con un método de predicción simple de la técnica anterior y un método de predicción que implica una votación mayoritaria.

En la primera etapa, se recibió el espectro NIR de una muestra de la materia prima para piensos FRM3. La información relevante de dicho espectro, es decir, las intensidades de absorción de las longitudes de onda, se transformaron para dar un vector de consulta de dicho espectro. En la siguiente etapa, los valores de similitud de todos los vectores de base de datos con el vector de consulta se calcularon utilizando una medida de distancia. Los valores de similitud así obtenidos se clasificaron según sus valores, incluida la indicación de su materia prima para piensos y/o pienso correspondientes en orden descendente para dar una clasificación del vector de base de datos.

En el modelo de predicción simple según el estado de la técnica (en lo sucesivo, también denominado Método de predicción 1: 1-NN), los n vectores mejor clasificados representaban los vecinos más cercanos en el análisis de similitud entre el vector de consulta de la muestra de una materia prima para piensos y/o pienso desconocidos y los vectores de base de datos de una población de espectros de materias primas para piensos y/o piensos conocidos.

El método con votación por mayoría también partió de la clasificación de la base de datos como se mencionó anteriormente, pero también implica un recuento de las ocurrencias de los casos materiales, es decir, la materia prima para piensos y/o el pienso correspondientes a los vectores de la base de datos.

5 Finalmente, se utilizó el método según la presente invención, que implica una votación mayoritaria y una ponderación de los resultados así obtenidos.

Tabla 1: Clasificación de los vectores de base de datos según su clasificación e indicando su tipo correspondiente materia prima para piensos y/o pienso, es decir, FRM1, FRM2 o FRM3.

Rango	Valor de Similitud	Nombre del vector	Materia prima para piensos y/o pienso
1	0,93	V1	FRM1
2	0,91	V3	FRM3
3	0,74	V5	FRM3
4	0,69	V4	FRM3
5	0,42	V7	FRM2
6	0,33	V2	FRM3
7	0,27	V12	FRM2
8	0,15	V6	FRM2
9	0,04	V8	FRM1
10	0,03	V9	FRM1
11	0,02	V10	FRM2
12	0,01	V11	FRM2

10 En el método de predicción 1 (1-NN), la materia prima para piensos con el mayor valor de similitud según la clasificación de la Tabla 1, es decir, FRM1, se indica como la materia prima para piensos con mayor similitud con el material del vector de consulta.

En comparación, en el método de predicción 2 (votación mayoritaria), la materia prima para piensos con el mayor número de ocurrencias en la clasificación de la Tabla 1, es decir, FRM2 con 5 votos sobre FRM1 con 3 votos y FRM3 con 4 votos, se indica como la materia prima para piensos con mayor similitud con el material del vector de consulta.

15 En el método de predicción 3, es decir, el método según la presente invención, la materia prima para piensos FRM3 se indica como la materia prima para piensos con la mayor similitud con el material del vector de consulta.

$$\text{FRM3} \rightarrow \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{6} = 1.249$$

$$\text{FRM2} \rightarrow \frac{1}{5} + \frac{1}{7} + \frac{1}{8} + \frac{1}{11} + \frac{1}{12} = 0.640$$

$$\text{FRM1} \rightarrow \frac{1}{1} + \frac{1}{9} + \frac{1}{10} = 1.211$$

Los resultados de los tres métodos son extremadamente diferentes. Más importante aún, dos de los métodos probados

también dan resultados falsos positivos. Sólo el método según la presente invención dio el resultado correcto.

REIVINDICACIONES

1. Un método implementado por ordenador para predecir un pienso y/o materia prima para piensos que comprende las etapas de
 - a) recibir un espectro de infrarrojo cercano (NIR) de una muestra de una materia prima para piensos y/o pienso desconocidos,
 - b) transformar las intensidades de absorción de longitudes de onda o números de onda en el espectro de la etapa a) para dar un vector de consulta,
 - c) proporcionar un conjunto de vectores de base de datos de una población de espectros de materias primas para piensos y/o piensos conocidos, en el que la población de espectros de piensos y/o materias primas para piensos conocidos de la etapa c) comprende al menos 50 espectros de muestras de cada pienso y/o materia prima para piensos de cada una de sus áreas de cultivo globales,
 - d) analizar la similitud entre el vector de consulta de la etapa b) y el conjunto de vectores de base de datos de la etapa c) que comprende las etapas
 - d1) calcular una medida de similitud y/o una medida de distancia entre cada vector de base de datos de la etapa c) y el vector de consulta de la etapa b) para dar un valor de similitud para cada vector de base de datos con el vector de consulta,
 - d2) clasificar los valores de similitud obtenidos en la etapa d1) en orden descendente, cuando la medida de similitud se calcula en la etapa d1) o en orden ascendente, cuando la medida de distancia se calcula en la etapa d1), en donde el vector de base de datos mejor clasificado tiene la mayor similitud con el vector de consulta,
 - d3) contar el número de apariciones de cada una de las materias primas para piensos y/o piensos entre los vectores de base de datos mejor clasificados en la clasificación de la etapa d2), en donde dicho número de apariciones se indica mediante la variable N por cada materia prima para piensos y/o pienso,
 - d4) ponderar los primeros N valores de similitud de cada una de las materias primas para piensos y/o piensos según su posición en la clasificación de la etapa d2) para dar posiciones de rango ponderadas de cada una de las materias primas para piensos y/o piensos, en donde la ponderación de valores de similitud se realiza tomando la posición de rango recíproco de la materia prima para piensos y/o pienso en la clasificación de la etapa d2),
 - d5) formar la suma de las posiciones de rango ponderadas de la etapa d4) para cada una de las materias primas para piensos y/o piensos para dar puntuaciones de cada una de las materias primas para piensos y/o piensos, y
 - e) asignar la materia prima para piensos y/o el pienso del vector de base de datos con mayor puntuación a la muestra de la etapa a).
2. El método implementado por ordenador según la reivindicación 1, en el que el vector en las etapas b) y c) es un vector multidimensional, correspondiendo cada dimensión a una intensidad de absorción de una longitud de onda o número de onda específico.
3. El método implementado por ordenador según la reivindicación 1 o 2, en el que el espectro en la etapa
 - a) y/o en la etapa c) cubre un intervalo de 1.100 a 2.500 nm.
4. El método implementado por ordenador según cualquiera de las reivindicaciones 1 a 3, en el que las intensidades de absorción de longitudes de onda y/o números de onda equidistantes en un espectro se transforman para dar un vector de un espectro de la etapa a) y/o de la etapa b).
5. El método implementado por ordenador según cualquiera de las reivindicaciones 1 a 4, en el que las distancias de las intensidades de absorción que se transforman en vectores en la etapa b) son idénticas a las distancias de las intensidades de absorción que se transforman en vectores en la etapa c).
6. El método implementado por ordenador según cualquiera de las reivindicaciones 1 a 5, en el que la distancia entre las longitudes de onda o los números de onda en la etapa b) y/o la etapa c) es de 0,1 nm +/- 10% a 10 nm +/- 10% o de 10^8 cm^{-1} +/- 10% a 10^6 cm^{-1} +/- 10%.
7. El método implementado por ordenador según cualquiera de las reivindicaciones 1 a 6, en el que el número de vectores de base de datos mejor clasificados a considerar en la etapa d3) oscila de 10 a 100.

8. El método implementado por ordenador según cualquiera de las reivindicaciones 1 a 7, en el que el número de los vectores de base de datos mejor clasificados a considerar en el paso d3) oscila entre 10 y 50.
- 5 9. El método implementado por ordenador según cualquiera de las reivindicaciones 1 a 8, en el que la población de espectros de piensos y/o materias primas para piensos conocidos en la etapa c) de dicho método comprende espectros de todos los piensos y/o materias primas para piensos molidos y/o o sin moler utilizada en la alimentación animal.
10. El método implementado por ordenador según cualquiera de las reivindicaciones 1 a 9, en el que se forma una derivada del espectro de la materia prima para piensos y/o pienso desconocidos de la etapa a) y/o de los espectros de las materias primas para piensos de la etapa c).
- 10 11. Un sistema para predecir una materia prima para piensos y/o un pienso, que comprende una unidad de procesamiento, o una unidad de procesamiento y una unidad de procesamiento adicional, adaptado para llevar a cabo el método implementado por ordenador según cualquiera de las reivindicaciones 1 a 10, en el que, en el primer caso, los vectores de base de datos se almacenan en la unidad de procesamiento, o en el que, en el segundo caso, los vectores de base de datos se almacenan en la unidad de procesamiento adicional y la unidad de procesamiento y la unidad de procesamiento adicional forman una red.

15

Fig. 1:

