(19) DANMARK

(10) DK/EP 3948231 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *G 01 N 21/359 (2014.01)* *A 23 K 40/00 (2016.01)*

(45) Oversættelsen bekendtgjort den: **2022-10-24**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2022-08-03**

(86) Europæisk ansøgning nr.: **20734186.8**

(86) Europæisk indleveringsdag: **2020-06-23**

(87) Den europæiske ansøgnings publiceringsdag: **2022-02-09**

(86) International ansøgning nr.: **EP2020067422**

(87) Internationalt publikationsnr.: **WO2020260234**

(30) Prioritet: **2019-06-24 EP 19181948**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Evonik Operations GmbH, Rellinghauser Straße 1-11, 45128 Essen, Tyskland**

(72) Opfinder: **Reimann, Ingolf, Albert-Schweitzer-Str. 19, 64354 Reinheim, Tyskland**
REISING, Joachim, Julius-Echter-Str. 32, 63801 Kleinostheim, Tyskland
MÜLLER, Christoph, Mathildenstrasse 66, 63065 Offenbach am Main, Tyskland

(74) Fuldmægtig i Danmark: **RWS Group, Europa House, Chiltern Park, Chiltern Hill, Chalfont St Peter, Bucks SL9 9FG, Storbritannien**

(54) Benævnelse: **FREMGANGSMÅDE TIL FORUDSIGELSE AF ET FODERSTOF OG/ELLER ET FODERSTOF-RÅMATERIALE**

(56) Fremdragne publikationer:
WO-A1-2016/141198
REEVES J B ET AL: "Spectral library searching: mid-infrared versus near-infrared spectra for classification of powdered food ingredient", APPLIED SPECTROSCOPY, THE SOCIETY FOR APPLIED SPECTROSCOPY. BALTIMORE, US, vol. 53, no. 7, July 1999 (1999-07), pages 836-844, XP000835471, ISSN: 0003-7028, DOI: 10.1366/0003702991947414
CHU X ET AL: "Algorithms, Strategies and Application Progress of Spectral Searching Methods", CHINESE JOURNAL OF ANALYTICAL CHEMISTRY, vol. 42, no. 9, September 2014 (2014-09), pages 1379-1386, XP055495211, AMSTERDAM, NL ISSN: 1872-2040, DOI: 10.1016/S1872-2040(14)60768-4 cited in the application
LOUDERMILK J B ET AL: "Novel Search Algorithms for a Mid-Infrared Spectral Library of Cotton Contaminants", APPLIED SPECTROSCOPY., vol. 62, no. 6, June 2008 (2008-06), pages 661-670, XP055495443, US ISSN: 0003-7028, DOI: 10.1366/000370208784657968
OSBORNE B G: "Near-infrared spectroscopy in food analysis" In: MEYERS R A & McGorrin R J: "Encyclopedia of Analytical Chemistry", 2006, Wiley & Sons Ltd, XP008124298, DOI: 10.1002/9780470027318.a1018, page 3, left-hand column, line 47 - right-hand column, line 2 figures 1,2

Fortsættes ...

DESCRIPTION

[0001] The present invention relates to a method for predicting a feedstuff raw material and/or feedstuff of unknown type by means of near infrared spectroscopy and similarity analysis.

[0002] Animal diets typically contain a variety of different feedstuffs and/or feedstuff raw materials. It is therefore necessary to know the identity and type of a feedstuff and/or feedstuff raw material as precisely and as quickly as possible. This is relevant, in particular, when different feedstuffs and/or feedstuff raw materials shall be mixed to yield a diet with a specific composition for a specific species. The methods of qualitative analysis of feedstuffs and feedstuff raw materials in principle allow a precise identification of feedstuffs and/or feedstuff raw materials of unknown type, i.e. unknown identity, origin etc. However, these methods require cost- and maintenance-intensive lab equipment. Further disadvantages of these methods are their high standards for the time required and the expertise and experience of the operating staff. In principle, near infrared spectroscopy would be a suitable means for the identification and determination of feedstuffs and/or feedstuff raw materials. The article "Near-infrared Spectroscopy in Food Analysis" (Osborne B. G., Encyclopedia of Analytical Chemistry, Wiley & Sons, 2006, pages 1-14) gives an overview over the application of near infrared (NIR) spectroscopy in this field. According to this article the wide range of applications of NIR in food analysis is possible because of the different sample presentation techniques. There is a technique available for any type of liquid, slurry, powdered or solid sample. When used as a routine method, however, near infrared spectroscopy requires the knowledge of the identity and type of feedstuffs and/or feedstuff raw materials. However, human mistakes in the selection of feedstuffs and/or feedstuff raw materials can already lead to an incorrect classification of a feedstuff and/or feedstuff raw material regarding its identity and present form. Based on incorrect classification, the wrong calibration method would be chosen for the near infrared analysis of the ingredients and their specific amounts in the feedstuff and/or feedstuff raw material. Thus, the data obtained from the incorrectly calibrated NIR spectrometer would be erroneous. Consequently, these data would be misleading for any further operating steps, in which the respective feedstuff raw material and/or feedstuff is involved.

[0003] An option to overcome this problem is to perform a similarity analysis of the near infrared spectrum of a sample substance with reference spectra to identify matches, thereby determining what the sample substance is. The general principle of this approach is described in WO 2016/141198 A1 and in the article "Algorithms, Strategies and Application Process of Spectral Searching Methods" (Chu X.-L., Li J.-Y., Chen P., Xu Y.-P., Chinese Journal of Analytical Chemistry, 2014, 42(9), 1379-1386). In detail, a similarity analysis involves the calculation of a similarity measure or a distance measure between the spectrum of the sample substance and the reference spectra. A high value for the similarity measure between the sample spectrum and a reference spectrum indicates a high similarity of a sample substance with the reference substance of the respective reference spectrum. By comparison, when the similarity analysis involves a distance measure, a low similarity value indicates a high similarity

of a sample substance to the reference substance of the respective reference spectrum. The similarity values obtained in a similarity analysis are typically ranked with the reference substances of the top-ranked entries having the greatest similarity with the sample substance.

[0004] The article "Spectral Library Searching: Mid-Infrared Versus Near-Infrared Spectra for Classification of Powdered Food Ingredient" (Reeves J. B. et al., Applied Spectroscopy, 1999, vol. 53, no. 7, pages 836-844) discloses a method for predicting a food ingredient comprising the steps of providing near infrared (NIR) spectra of samples of powdered food ingredients, libraries of spectra were constructed and searched on the basis of full spectrum algorithms. The results for the various searches are presented as distances between the unknown or test sample and samples in the library of the algorithms in question. A match was considered to be successful when the nearest sample (smallest distance) to a test sample was for a sample from the same group.

[0005] However, the approaches of the prior art are very susceptible to false determinations or predictions when false positive entries are at the top of the ranking of similarity values. Reasons for false positive entries in said ranking can be the erroneous assignment of a spectrum to the wrong reference substance or wrong class of references substances, the heterogeneity or messiness of a class of references substances, whose spectra were recorded, or the similarity of some reference substances to one another, which makes an exact matching rather difficult. Any of these cases make it difficult or impossible to precisely and reliably match the sample spectrum of a sample substance to the spectrum of a reference substance.

[0006] The article "Novel Search Algorithms for a Mid-Infrared Spectral Library of Cotton Contaminants" (Loudermilk J. B. et al, Applied Spectroscopy, June 2008, vol. 62, no. 6, pages 661-670) discloses voting scheme algorithms used in MIR library search of cotton contaminants. A counting is done in the so-called "group" algorithm, e.g. the number of times the substance/category "seed coat" is occurring in a hit list. A weighting according to rank, followed by summation to result in a score, is done the "weighted frequency" algorithm, for single spectra and their occurrence in hit lists of multiple search algorithms. Said "weighted frequency" algorithm therefore works with multiple search algorithms for a single spectrum.

[0007] According to the present invention this problem is solved in that first a similarity analysis is performed, followed by counting the occurrences of a feedstuff raw material and/or feedstuff in the ranking of similarity values. Next, the thus determined number of similarity values of the feedstuff raw material and/or feedstuff are weighted according to their rank position to give weighted rank position, the sum is formed of the weighted rank positions to give scores of the feedstuff raw materials and/or feedstuffs, and the highest score indicates the feedstuff raw material and/or feedstuff with the greatest similarity to the sample substance.

[0008] Object of the present invention is therefore a computer-implemented method for predicting a feedstuff and/or feedstuff raw material according to claim 1, comprising the steps of

1. a) receiving a near infrared (NIR) spectrum of a sample of an unknown feedstuff raw material and/or feedstuff,
2. b) transforming absorption intensities of wavelengths or wavenumbers in the spectrum of step a) to give a query vector,
3. c) providing a set of database vectors of a population of spectra of known feedstuff raw materials and/or feedstuffs, wherein the population of spectra of known feedstuffs and/or feedstuff raw materials of step c) comprises at least 50 spectra of samples of each feedstuff and/or feedstuff raw material from each of its global growing areas,
4. d) analyzing the similarity between the query vector of step b) and the set of database vectors of step c) comprising the steps

   d1) calculating a similarity measure and/or a distance measure between each database vector of step c) and the query vector of step b) to give a similarity value for each database vector with the query vector,

   d2) ranking the similarity values obtained in step d1) in descending order, when the similarity measure is calculated in step d1) or in ascending order, when the distance measure is calculated in step d1), wherein the top-ranked database vector has the greatest similarity with the query vector,

   d3) counting the number of occurrence of each of the feedstuff raw materials and/or feedstuffs among the top-ranked database vectors in the ranking of step d2), wherein said number of occurrences is indicated by the variable N per each feedstuff raw material and/or feedstuff,

   d4) weighting the first N similarity values of each of the feedstuff raw materials and/or feedstuffs according to their position in the ranking of step d2) to give weighted rank positions of each of the feedstuff raw materials and/or feedstuffs, wherein the weighting of similarity values is done by taking the reciprocal rank position of the feedstuff raw material and/or feedstuff in the ranking of step d2),

   d5) forming the sum of the weighted rank positions of step d4) for each of the feedstuff raw materials and/or feedstuffs to give scores of each of the feedstuff raw materials and/or feedstuffs, and

5. e) assigning the feedstuff raw material and/or feedstuff of the database vector with the highest score to the sample of step a).

[0009] In the context of the present invention the term unknown feedstuff raw material and/or feedstuff refers to any kind of feedstuff and/or feedstuff raw material whose identity, composition, origin and/or form, i.e. whether it is ground or unground, is not known. By comparison, in the context of the present invention the term known feedstuff raw material and/or feedstuff refers to any kind of feedstuff and/or feedstuff raw material whose identify, composition, origin and/or form, i.e. whether it is ground or unground, is known. Accordingly, a

population of spectra of known feedstuff raw materials and/or feedstuffs is a number or multitude of spectra, which are known to belong to a specific feedstuff and/or feedstuff raw material of known identity, composition, origin and/or form.

[0010] The weighting of the similarity values in step d4) is typically done by taking the reciprocal rank position of the specific feedstuff raw material and/or feedstuff in the ranking of step d2). For example, if there is an occurrence of a specific feedstuff raw material at position 2 in the ranking, its weighing gives the value ½ for the weighted rank position. The first N similarity values of each of the feedstuff raw materials and/or feedstuffs are weighted according to their position in the ranking of step d2) to give weighted rank positions of each of the feedstuff raw materials and/or feedstuffs in step d4). Next, in step d5) the sum is formed of the weighted rank positions of step d4) for each of the feedstuff raw materials and/or feedstuffs to give scores of each of the feedstuff raw materials and/or feedstuffs.

[0011] In its broadest meaning a vector is a geometric object that has magnitude (or length) and direction. In a Cartesian coordinate system, a vector can be represented by identifying the coordinates of its initial and terminal point. Therefore, a vector is suited to represent an absorption intensity at a specific wavelength or wavenumber in a two-dimensional near infrared spectrum. In addition, a vector is not limited to the description of a two-dimensional system. Rather, a vector can describe multi-dimensional spaces, such as a near infrared spectrum with a multitude of absorption intensities at a multitude of different wavelengths or wavenumbers. In this case, each dimension of the said vector corresponds to a single absorption intensity at a specific wavelength or wavenumber. In the context of the present invention a database vector also contains the identification of its feedstuff raw material and/or feedstuff and optionally, its origin, or other information, such as whether it is ground or unground, or the season when it was harvested, in case it is a corn. Alternatively, each database vector has an identification number and the information mentioned before are stored under the identification number on a processing unit, such as a computer, cloud, or server.

[0012] In one embodiment of the computer-implemented method according to the present invention the vector in steps b) and c) is a multi-dimensional vector, with each dimension corresponding to an absorption intensity of a specific wavelength or wavenumber.

[0013] Like the query vector of the spectrum of an unknown feedstuff raw material and/or feedstuff, also the set of database vectors provided in step c) of the computer-implemented method according to the present invention is obtained by transforming each spectrum of a population of spectra of known feedstuff raw materials and/or feedstuffs into the corresponding vector. If the set of database vectors is not already present, the step c) also comprises the transformation of each spectrum of a population of spectra of known feedstuff raw materials and/or feedstuffs into the corresponding vector to give the set of database vectors. In that case the step c) of the computer-implemented method according to the present invention comprises the steps of transforming each spectrum of a population of spectra of known feedstuff raw materials and/or feedstuffs to the corresponding vector to give a set of data set vectors and providing the thus obtained set of database vectors of a population of spectra of known

feedstuff raw materials and/or feedstuffs.

[0014] According to the present invention a near infrared spectrum of a sample of an unknown feedstuff raw material and/or feedstuff is received in step a). In the context of the present invention this means that the place where the spectrum to be received is recorded and the place where the computer-implemented method according to the present invention is performed, can be different or identical. For example, it is possible that a near infrared spectrum of a sample of an unknown feedstuff raw material and/or feedstuff is recorded at one place, and sent in any way to a remote place, where the computer-implemented method according to the present invention is performed. Alternatively, both recording of the spectrum and the prediction of the feedstuff raw material and/or feedstuff based on said spectrum can be performed at the same place.

[0015] In an embodiment the step a) of the computer-implemented method the step may comprise the recording of a near infrared spectrum of a sample of an unknown feedstuff raw material and/or feedstuff.

[0016] In principle, the near infrared (NIR) spectrum of step a) can be recorded at wavelengths of from 700 to 2,500 nm with any suitable near infrared spectrometer working either on the monochromator or on the Fourier transform principle. However, it was found that only the part of near infrared spectra of from 1,100 to 2,500 nm is necessary for prediction of feedstuff raw material and/or feedstuff in the computer-implemented method according to the present invention. Therefore, it is preferred that the spectrum covers recorded in the range of from 1,100 to 2,500 nm in step a) of the computer-implemented method according to the present invention. Accordingly, the population of spectra of known feedstuff raw materials and/or feedstuffs of step c) of the computer-implemented only needs to cover the range of from 1,100 to 2,500 nm. Since wavelengths are easily convertible into the respective wavenumbers, the near infrared spectrum of step a) and/or the spectra of step c) can also be recorded at the corresponding wavenumbers. When the sample of a feedstuff raw material and/or feedstuff in step a) is not translucent, the reflectance of the emitted light from the sample is measured and the difference between the emitted light and the reflected light is given as absorption. The thus obtained absorptions, i.e. their intensities and their wavelengths or wavenumbers, are used for the generation of the vectors in step b) and/or in step c). Accordingly, a near infrared spectrometer suitable for use in the method according to the present invention can work either in the transmittance mode or in the reflectance mode.

[0017] In an embodiment of the computer-implemented method according to the present invention the spectrum in step a) and/or in step c) is recorded in a range of from 1,100 to 2,500 nm.

[0018] The method according to the present invention is not limited to a specific distance or similarity measure for analyzing the similarity between the query vector of step b) and the database vectors of step c). Therefore, any distance or similarity measure, which is suited to determine the similarity of the vector of step b) with the vectors of step c) can be used in the

method according to the present invention. In principle, a similarity analysis is based on a nearest neighbor search. It was found that the Cosine coefficient is a particularly suitable similarity measure for a nearest neighbor search in the method according to the present invention. For example, the Cosine coefficient, which allows the calculation of the similarity between two vectors extremely rapidly with a high precision, is particularly suitable for the method according to the present invention. The Cosine coefficient $S_{A,B}$ of two vectors A and B is represented by the following formula

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

where x_{jA} and x_{jB} are components of the vectors A and B, respectively, and n is the number of spaces, here the number of absorption intensities at specific wavelengths or wavenumbers. The values for the similarity range from -1, meaning exactly the opposite to each other, to 1, meaning identity, with 0 indicating orthogonality (decorrelation), and in-between values indicating intermediate similarity or dissimilarity.

[0019] Alternatively, the similarity between the vectors can be also be calculated by means of a distance measure. For example, the Euclidian distance, which also allows the calculation of the similarity between two vectors extremely rapidly and precisely, is particularly suitable for the method according to the present invention. The Euclidian distance $D_{A,B}$ of two vectors A and B is represented by the following formula

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

where x_{jA} and x_{jB} are components of the vectors A and B, respectively, and n is the number of spaces, here the number of absorption intensities at specific wavelengths or wavenumbers.

[0020] It is therefore preferred that in step d1) of the computer-implemented method according to the present invention the similarity measure is the Cosine coefficient and the distance measure is the Euclidian distance.

[0021] According to step b) of the present invention absorption intensities of wavelengths or wavenumbers in a spectrum are transformed to give a query vector. In principle, one could select the strongest and therefore most meaningful absorption intensities in a spectrum and transfer only said absorption intensities to give a vector. However, this would require a thorough analysis of each individual spectrum of a sample substance, which is not only time-consuming but also requires a very good understanding of near infrared spectra. Hence, this approach would not be suitable for a routine analysis. Further, this approach has the disadvantage that meaningful but relatively weak absorption intensities in a spectrum may be ignored so that information would get lost. This could lead to a wrong assigning of the unknown feedstuff raw material and/or feedstuff in the end. It is therefore favorable to consider as many information as possible in the spectrum without a preceding in-depth analysis of the spectrum. It is therefore preferred to transform the absorption intensities of equidistant wavelengths or

wavenumbers in a spectrum, i.e. in step b) and/or c), to give a vector of said spectrum. In order to allow to the best possible similarity analysis between the query vector and the database vectors, it is preferred to transform the absorption intensities of equidistant wavelengths or wavenumbers in a spectrum of step b) and step c) of the method according to the present invention to give a vector of said spectrum or spectra. Preferably, the distances of the absorption intensities being transformed to vectors in step b) are identical with the distances of the absorptions intensities transformed to vectors in step c). This allows for a higher precision in the prediction of the computer-implemented method according to the present invention, even without having any specific knowledge of the sample substance and its spectra at all.

[0022] In an embodiment in step b) and/or step c) of the computer-implemented method according to the present invention the absorption intensities of equidistant wavelengths and/or wavenumbers in a spectrum are transformed to give a vector of a spectrum of step a) and/or of step b).

[0023] In a further embodiment of the computer-implemented method according to the present invention the distances of the absorption intensities being transformed to vectors in step b) are identical with the distances of the absorptions intensities transformed to vectors in step c).

[0024] Preferably, the absorption intensities of wavelengths or wavenumbers in a spectrum, which are transformed to give a vector of said spectrum, have small distances between each other. This has the advantage that most if not all relevant absorption intensities, i.e. information, of a spectrum are transformed to a vector of said spectrum. This is believed to allow a very precise transformation of all relevant information of a spectrum into vectors, even without having knowledge of the feedstuff and/or feedstuff raw material whose spectrum was recorded, in particular of its identity, composition, origin and/or form. Preferably, the distance between the wavelengths in step b) of the method according to the present invention is from 0.1 +/- 10% to 10 +/- 10% nm, from 0.1 +/- 10% to 5 +/-10% nm, or from 0.1 +/- 10% to 2 +/- 10% nm. Accordingly, the distance between the wavenumbers in step b) of the method according to the present invention is from 10^8 +/- 10% to 10^6 +/- 10%, from 10^8 +/- 10 to 5 * 10^6 +/-10% nm, or from 10^8 +/- 10% to 2 * 10^6 +/- 10% nm. In the context of the present invention, the term +/- 10% is used with respect to explicitly mentioned values to indicate that deviations from said explicitly mentioned values are still within the scope of the present invention, provided that they essentially lead to the effects of the present invention. The distances between the wavelengths or wavenumbers in step c) of the method according to the present invention are preferably the same as those in step b), in order to provide for the best possible comparison between the recorded spectrum of an unknown feedstuff and/or feedstuff raw material and the spectra of known feedstuffs and/or feedstuff raw materials.

[0025] In an embodiment of the computer-implemented method according to the present invention, the distance between the wavelengths or wavenumbers in step b) and/or step c) is from 0.1 nm +/- 10% to 10 nm +/- 10% or from 10^8 cm^{-1} +/- 10% to 10^6 cm^{-1} +/- 10%.

[0026] In principle, the computer-implemented method according to the present invention is not subject to any limitation regarding the number of absorption intensities to be transformed to give a vector. Rather, the number of relevant information in a spectrum of a feedstuff raw material and/or feedstuff strongly depends on the individual feedstuff raw material and/or feedstuff, and in particular, on its composition and components. The more complex a feedstuff and/or feedstuff raw material is, i.e. the more components a feedstuff and/or feedstuff raw material contains, the more information are required from a near infrared spectrum for predicting an unknown feedstuff and/or feedstuff raw material. Again, it would not be practical to perform an in-depth analysis in order to find out the absorption intensities which necessarily must be transferred to give a vector. A suitable option for the number of absorption intensities to be transformed to give a vector is to correlate them with the distance between the corresponding wavelengths or wavenumbers, e.g. from 0.1 nm +/- 10% to 10 nm +/- 10% or 0.1 +/- 10% to 2 +/- 10% nm, and the recording range of the spectrum, e.g. from 1,100 to 2,500 nm. Preferably, the number of absorption intensities in each spectrum to be transformed into a vector is at least 100, in particular, said number ranges from 150 +/- 10% to 15,000 +/- 10% or from 700 +/- 10% to 15,000 +/- 10%.

[0027] In principle, the step d3) of the method according to the present invention is not subject to any limitations regarding the number of top-ranked vectors, apart from the number of entries in the ranking of step d2). However, the preciseness of the prediction in the method according to the present invention is neither affected when the step d3) involves the use of a relatively low number of database vectors nor is the preciseness improved in any way, when the step d3) involves the use of a relatively high number of database vectors. Rather, when d3) involved the use of a larger number of database vectors in the ranking of step d2), beyond the database vectors being really top-ranked, this would result in an increasing number of less meaningful database vectors, which would decrease the preciseness of prediction in the method according to the present invention. It is therefore preferred that the number N of the first top-ranked database vectors in step d3) ranges from 10 to 100, 10 to 90, 10 to 80, 10 to 70, 10 to 60, 10 to 50, 10 to 40, 10 to 30, or 10 to 20. This leads to very precise predictions equivalent to the use of a larger number of database vectors, however, without unnecessarily increasing the computing effort, as is the case for more database vectors.

[0028] In one embodiment of the method according to the present invention the number of the top-ranked database vectors to be considered in step d3) ranges from 10 to 100, preferably from 10 to 50.

[0029] The population of spectra of known feedstuff raw materials and/or feedstuffs used in the method according to the present invention is not limited to specific feedstuff raw materials and/or feedstuff. Rather, it preferably comprises spectra of all feedstuff raw materials and/or feedstuffs used in animal nutrition, preferably in the nutrition of poultry, pigs, pigs and/or animals kept in aquacultures, such as fish and/or crustacea. The spectra of the feedstuff raw materials and/or feedstuffs may significantly differ depending on their form or appearance, e.g. when they are present in ground or unground form. Therefore, the spectra population preferably also comprises spectra, which were recorded from the aforementioned feedstuff raw

materials and/or feedstuff in ground and/or unground form.

[0030] In another embodiment of the method according to the present invention the population of spectra of known feedstuffs and/or feedstuff raw materials in step c) of said method comprises spectra of all feedstuffs and/or feedstuff raw materials in ground and/or unground form used in animal nutrition.

[0031] In principle, the method according to the present invention is not limited in any way regarding the number and types of feedstuffs and/or feedstuff raw materials, whose spectra, recorded in ground and/or unground form, make up the spectra population. Notwithstanding, it is preferred that the population of spectra of known feedstuffs and/or feedstuff raw materials in step c) of said method comprises spectra of all feedstuffs and/or feedstuff raw material in ground and/or unground form used in animal nutrition, preferably in the nutrition of poultry, pigs, pigs and/or animals kept in aquacultures, such as fish and/or crustacea The feedstuff raw material and/or feedstuff can be of animal and/or vegetable origin. Particularly preferred feedstuffs and/or feedstuff raw materials are unprocessed and/or processed feedstuff raw materials and/or feedstuff. Processed feedstuff raw materials and/or feedstuff are those, which were subjected to any type of heat or pressure treatment in order to remove or detoxify anti-nutritive factors. Preferred feedstuffs and/or feedstuff raw material are oilseeds, in particular soy extraction meal and expeller, full-fat soybeans, rapeseed meal and expeller, cotton extraction meal, peanut extraction meal, sunflower extraction meal, coconut extraction meal, and/or palm kernel extraction meal; legumes, in particular toasted guar flour; brewery and distillation by-products, in particular dried distiller's grain with solubles (DDGS), by-products of grain-processing and feedstuff production, in particular corn gluten, maize seed meal and/or bakery by-products; anima by-products, in particular fish meal, meat meal, poultry meal, blood meal, and/or bone meal; and also any type of grains. In particular, the feedstuff raw material is soy, soybeans or a soybean product.

[0032] Depending on factors such as climate, soil and genetics of the plants, feedstuff raw materials and/or feedstuff from different global growing areas may differ in their ingredients and the contents of said ingredients. In order to allow for reliable and reproducible predictions of a feedstuff and/or feedstuff raw materials, therefore the feedstuff raw material and/or feedstuff, whose spectra are part of the spectra population, is from all of its global growing areas.

[0033] The number of spectra of feedstuff raw materials and/or feedstuffs of the spectra population used in the method according to the present invention should be representative to allow for a reliable and reproducible prediction of the feedstuff raw material and/or feedstuff in question. Therefore, the population of spectra comprises at least 50 spectra of samples of each feedstuff and/or feedstuff raw material, i.e. each feedstuff and/or feedstuff raw material in ground and/or unground form used in animal nutrition, preferably in the nutrition of poultry, pigs, pigs and/or animals kept in aquacultures, such as fish and/or crustacea, from each of its global growing areas. The method according to the present invention is not subject to any limitations regarding the number of spectra of samples of any feedstuff and/or feedstuff raw

material from any of its global growing areas. Hence, the number of spectra of samples of any feedstuff and/or feedstuff raw material from any of its global growing areas may range from 50 to 10,000, from 50 to 5,000, from 50 to 2,500, from 50 to 2,000, from 50 to 1,500, from 50 to 1,000, from 100 to 1,000, from 50 to 500, from 100 to 500, from 50 to 250, from 100 to 250, or from 50 to 100.

[0034] When the population of spectra of feedstuffs and/or feedstuff raw material of known type considers each global growing area of a feedstuff and/or feedstuff raw material and the number of spectra from each global growing area is representative, the method according to the present invention allows not only a reliable and reproducible prediction of the feedstuff raw material and/or feedstuff in question but also a prediction of the origin of the feedstuff raw material and/or feedstuff in question.

[0035] Therefore, said population of spectra of known feedstuffs and/or feedstuff raw material of step c) preferably comprises at least 50 spectra of samples of each feedstuff and/or feedstuff raw material from each of its global growing areas. The number of spectra of samples of a feedstuff and/or feedstuff raw material from each global growing area is not subject to any limitations. Hence, the number of spectra of samples of any feedstuff and/or feedstuff raw material from each global growing area may range from 50 to 10,000, from 50 to 5,000, from 50 to 2,500, from 50 to 2,000, from 50 to 1,500, from 50 to 1,000, from 100 to 1,000, from 50 to 500, from 100 to 500, from 50 to 250, from 100 to 250, or from 50 to 100.

[0036] The population of spectra of known feedstuffs and/or feedstuff raw material may be an already existing database, preferably a constantly updated database, and/or it may be a new database, which is constantly updated or further developed. A database of spectra of feedstuffs and/or feedstuff raw material of known type suited for use in the method according to the present invention is for example Evonik's AMINODat® 5.0.

[0037] Cases may arise, where the position of a signal peak in a spectrum, either in step b) and/or in step c) of the method according to the present invention cannot be located because the maxima and minima of the individual peaks cannot be clearly identified in such a spectrum. An easier locating of individual peaks in the spectrum is possible, when the minima and maxima of the peaks are easier identifiable. Taking the first derivative of a spectrum facilitates the identification of the peaks in the spectrum because it gives a zero crossing of peak maxima or peak minima. Taking the second derivative gives a peak minimum at exactly that position, where a peak maximum was in the original spectrum and vice versa. Taking the first or second derivative of a spectrum also facilitates the identification of an outlier in the population of spectra of known feedstuff raw materials and/or feedstuffs.

[0038] In another embodiment of the method according to the present invention a derivative of the spectrum of the unknown feedstuff raw material and/or feedstuff of step a) and/or of the spectra of known feedstuff raw materials and/or feedstuffs of step c) is/are formed.

[0039] Preferably, the first derivative of the spectrum of a feedstuff and/or feedstuff raw

material of unknown type of step a) and/or of the spectra of known feedstuffs and/or feedstuff raw materials of step c) is/are formed.

[0040] It is preferred to provide the set of database vectors of a population of spectra of known feedstuff raw materials and/or feedstuffs directly in step c) of the computer-implemented method according to the present invention. Notwithstanding, it is also possible that first only a population of spectra of known feedstuff raw materials and/or feedstuffs is provided, which is transformed in the next step to the set of database vector for the similarity analysis in step d). In this case, the step c) of the computer-implemented method according to the present invention also comprises the step of transforming the absorption intensities of wavelengths or wavenumbers in each spectrum of a population of spectra of known feedstuff raw materials and/or feedstuffs into vectors. The thus obtained multitude of vectors of the population of spectra of known feedstuff raw materials and/or feedstuffs is then the set of database vectors, as mentioned above. In any case, it is preferred to store the population of spectra of known feedstuff raw materials and/or feedstuffs or the set of database vectors of said population of spectra of known feedstuff raw materials and/or feedstuffs on a processing unit, such as a computer or a cloud. The processing unit on which the population of spectra or the database vectors are stored can be identical with or different from the processing unit, which carries out the computer-implemented method according to the present invention. In the second case, the first processing unit, which carries out the computer-implemented method according to the present invention, and the second processing unit, on which the population of spectra or the database vectors are stored, form a network. For example, it is also possible that the population of spectra or the database vectors are stored on a cloud. In that case, the first processing unit, e.g. a computer, which carries out the computer-implemented method according to the present invention, and the second processing unit, e.g. a cloud, on which the population of spectra or the database vectors are stored, form a network.

[0041] Another object of the present invention is therefore a system for predicting a feedstuff raw material and/or feedstuff according to claim 11.

[0042] In case the population of spectra of known feedstuff raw materials is stored on a computer, it is preferred that said population is stored on a second computer, i.e. a computer, which does not already carry out the computer-implemented method according to the present invention. As consequence, the working load is evenly distributed between the computers, which may lead to a quicker performance of the computer-implemented method according to the present invention. This also allows for a communication between the user and the provider of a database with the database vector for use in the method according to the present invention, for example, an update of the database vectors.

Description of the figure:

[0043] Fig. 1 is a representation of the computer-implemented method according to the present invention. Starting with a query vector for a spectrum of an unknown feedstuff raw

material and/or feedstuff, the similarity values are calculated for each database vector of a population of spectra of known feedstuff raw materials and/or feedstuffs. The thus obtained similarity values are ordered either ascending, with the highest at the top, if similarity values were calculated with a similarity measure, or descending, with the lowest at the top, if similarity values were calculated with a distance measure, to give a ranking. The top-n ranked vectors, which represent the nearest neighbors to the query vectors in the similarity search, are selected from this ranking and the occurrences of the corresponding classes, i.e. of the feedstuff raw materials and/or feedstuffs, are determined. Said classes are weighted according to the rank of their vectors and added up to give class scores. The class scores are ranked according to the values with the highest value at the top. The feedstuff raw material and/or feedstuff of the top-ranked class has the highest similarity with the query vector and therefore, it is assigned to the sample of the query vector.

Example:

[0044] The following is an example to illustrate the computer-implemented method according to the present invention in comparison to a simple prediction method of the prior art, and a prediction method involving a majority voting.

[0045] In the first step, the NIR spectrum of a sample of the feedstuff raw material FRM3 was received. The relevant information of said spectrum, i.e. absorption intensities of wavelengths, were transformed to give a query vector of said spectrum. In the next step, the similarity values of all database vectors with the query vector were calculated using a distance measure. The thus obtained similarity values were ranked according to their values including the indication of their corresponding feedstuff raw material and/or feedstuff in descending order to give a ranking of the database vector.

[0046] In the simple prediction model according to the prior art (in the following also referred to as Prediction method 1: 1-NN), the top-n vectors in this ranking represented the nearest neighbors in the similarity analysis between the query vector of the sample of an unknown feedstuff raw material and/or unknown feedstuff and the database vectors of a population of spectra of known feedstuff raw materials and/or feedstuffs.

[0047] The method with majority voting also started from the ranking of the database as mentioned above but also involves a counting of the occurrences of the material cases, i.e. the feedstuff raw material and/or feedstuff corresponding to the database vectors.

[0048] Finally, the method according to the present invention was used, which involves a majority-voting and a weighting of the thus obtained results.

Table 1: Ranking of database vectors according to their ranking and stating their corresponding type feedstuff raw material and/or feedstuff, i.e. FRM1. FRM2 or FRM3.

Rank	Similarity value	Vector name	Feedstuff raw material/ feedstuff
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

[0049] In the prediction method 1 (1-NN) the feedstuff raw material with the highest similarity value according to the rankling of Table 1, i.e. FRM1, is indicated as the feedstuff raw material with the highest similarity to the material of the query vector.

[0050] By comparison, in the prediction method 2 (majority voting), the feedstuff raw material with the greatest number of occurrences in the ranking of Table 1, i.e. FRM2 with 5 votings over FRM1 with 3 votings and FRM2 with 4 votings, is indicated as the feedstuff raw material with the highest similarity to the material of the query vector.

[0051] In the prediction method 3. i.e. the method according to the present invention, the feedstuff raw material FRM3 is indicated as the feedstuff raw material with the highest similarity to the material of the query vector.

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

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

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

[0052] The results of the three methods are extremely different. More importantly, two of the tested methods also give false positive results. Only the method according to the present invention gave the correct result.

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO2016141198A1 [0003]

Non-patent literature cited in the description

- Near-infrared Spectroscopy in Food Analysis**OSBORNE B. G.**Encyclopedia of Analytical ChemistryWiley & Sons200600001-14 [0002]
- **CHU X.-L.LI J.-Y.CHEN P.XU Y.-P.**Algorithms, Strategies and Application Process of Spectral Searching MethodsChinese Journal of Analytical Chemistry, 2014, vol. 42, 91379-1386 [0003]
- **REEVES J. B. et al.**Spectral Library Searching: Mid-Infrared Versus Near-Infrared Spectra for Classification of Powdered Food IngredientApplied Spectroscopy, 1999, vol. 53, 7836-844 [0004]
- **LOUDERMILK J. B. et al.**Novel Search Algorithms for a Mid-Infrared Spectral Library of Cotton ContaminantsApplied Spectroscopy, 2008, vol. 62, 6661-670 [0006]

Patentkrav

1. Computerimplementeret fremgangsmåde til forudsigelse af et foderstof og/eller et foderstof-råmateriale, hvilken fremgangsmåde omfatter de følgende trin:

a) modtagelse af et nærinfrarødt spektrum (NIR-spektrum) af en prøve af et ukendt foderstof-råmateriale og/eller foderstof

b) transformation af absorptionsstyrker for bølgelængder eller bølgetal i spektret fra trin a) for at give en forespørgselsvektor

c) tilvejebringelse af et sæt databasevektorer for en population af spektre af kendte foderstof-råmaterialer og/eller foderstoffer, hvor populationen af spektre af kendte foderstoffer og/eller foderstof-råmaterialer fra trin c) omfatter mindst 50 spektre af prøver af hvert foderstof og/eller foderstof-råmateriale fra hvert af dets globale vækstområder

d) analysering af ligheden mellem forespørgselsvektoren fra trin b) og sættet af databasevektorer fra trin c), omfattende de følgende trin:

d1) beregning et lighedsmål og/eller et afstandsmål mellem hver databasevektor fra trin c) og forespørgselsvektoren fra trin b) for at give en lighedsværdi for hver databasevektor med forespørgselsvektoren

d2) rangering af de i trin d1) opnåede lighedsværdier i faldende rækkefølge, når lighedsmålet beregnes i trin d1), eller i stigende rækkefølge, når afstandsmålet beregnes i trin d1), hvor den toprangerede databasevektor har største lighed med forespørgselsvektoren

d3) optælling af antallet af forekomster for hvert af foderstof-råmaterialerne og/eller foderstofferne blandt de toprangerede databasevektorer i rangeringen i trin d2), hvor antallet af forekomster angives ved hjælp af variablen N pr. hvert foderstof-råmateriale og/eller foderstof

d4) vægtning af de første N lighedsværdier for hvert af foderråmaterialerne og/eller foderstofferne i overensstemmelse med deres placering i rangordningen i trin d2) for at give vægtede rangpositioner for hvert af foderstof-råmaterialerne og/eller foderstofferne, hvor vægtningen af lighedsværdier

udføres ved at tage den reciprokke rangposition af foderstof-råmaterialet og/eller foderstoffet i rangordningen fra trin d2)
d5) dannelse af summen af de vægtede rangpositioner i trin d4) for hvert af foderstof-råmaterialerne og/eller foderstofferne for at give point for hvert af foderstof-råmaterialerne og/eller foderstofferne, og
e) tildeling af foderstof-råmaterialet og/eller foderstoffet i databasevektoren med det højeste point til prøven fra trin a).

2. Computerimplementeret fremgangsmåde ifølge krav 1, hvorved vektoren i trin b) og c) er en multidimensional vektor, hvor hver dimension svarer til en absorptionsstyrke for en specifik bølgelængde eller et specifikt bølgetal.

3. Computerimplementeret fremgangsmåde ifølge krav 1 eller 2, hvorved spektret fra trin
a) og/eller fra trin c) dækker et område fra 1100 til 2500 nm.

4. Computerimplementeret fremgangsmåde ifølge et af kravene 1 til 3, hvorved absorptionsstyrkerne for ækvidistante bølgelængder og/eller bølgetal i et spektrum transformeres til at give en vektor for et spektrum fra trin a) og/eller trin b).

5. Computerimplementeret fremgangsmåde ifølge et af kravene 1 til 4, hvorved afstandene for de absorptionsstyrker, der transformeres til vektorer i trin b), er identiske med afstandene for de absorptionsstyrker, der transformeres til vektorer i trin c).

6. Computerimplementeret fremgangsmåde ifølge et af kravene 1 til 5, hvorved afstanden mellem bølgelængderne eller bølgetallene i trin b) og/eller trin c) er fra 0,1 nm +/- 10 % til 10 nm +/- 10 % eller fra 10^8 cm^{-1} +/- 10 % til 10^6 cm^{-1} +/- 10 %.

7. Computerimplementeret fremgangsmåde ifølge et af kravene 1 til 6, hvorved antallet af de toprangerede databasevektorer, der skal tages i betragtning i trin d3), ligger i området fra 10 til

100.

8. Computerimplementeret fremgangsmåde ifølge et af kravene 1 til 7, hvorved antallet af de toprangerede databasevektorer, der skal tages i betragtning i trin d3), ligger i området fra 10 til 50.

9. Computerimplementeret fremgangsmåde ifølge et af kravene 1 til 8, hvorved populationen af spektre af kendte foderstoffer og/eller foderstof-råmaterialer fra trin c) i fremgangsmåden omfatter spektre af alle foderstoffer og/eller foderstof-råmaterialer i formalet og/eller ikke-formalet form, der anvendes som dyrefoder.

10. Computerimplementeret fremgangsmåde ifølge et af kravene 1 til 9, hvorved et derivat af spektret af det ukendte foderstof-råmateriale og/eller foderstof fra trin a) og/eller spektrene af kendte foderstof-råmaterialer og/eller foderstoffer fra trin c) dannes.

11. System til forudsigelse af et foderstof-råmateriale og/eller foderstof, hvilket system omfatter en forarbejdningsenhed eller en forarbejdningsenhed og en yderligere forarbejdningsenhed, der er tilpasset til at udføre den computerimplementerede fremgangsmåde ifølge et af kravene 1 til 10, hvor databasevektorerne i første tilfælde er lagret på behandlingsenheden, eller hvor databasevektorerne i det andet tilfælde er lagret på den yderligere behandlingsenhed, og behandlingsenheden og den yderligere behandlingsenhed danner et netværk.

DRAWINGS

Fig. 1:

