Classification and identification of molecules through factor analysis method based on terahertz spectroscopy

Jianglou Huang [a], Jinsong Liu [a,*], Kejia Wang [a], Zhengang Yang [a], Xiaming Liu [b]

[a] Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China
[b] Department of Urology, Tongji Hospital, Tongli Medical College, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

ARTICLE INFO

ABSTRACT

By means of factor analysis approach, a method of molecule classification is built based on the measured terahertz absorption spectra of the molecules. A data matrix can be obtained by sampling the absorption spectra at different frequency points. The data matrix is then decomposed into the product of two matrices: a weight matrix and a characteristic matrix. By using the K-means clustering to deal with the weight matrix, these molecules can be classified. A group of samples (spirobenzopyran, indole, styrene derivatives and inorganic salts) has been prepared, and measured via a terahertz time-domain spectrometer. These samples are classified with 75% accuracy compared to that directly classified via their molecular formulas.

© 2018 Published by Elsevier B.V.

1. Introduction

During years of researches, terahertz (THz) wave has found itself an important role in chemical and biological fields [1–4] due to its non-destructive and non-invasive way of detecting samples. It has low sample damage or photoionization [5], with photon energy 1 million times weaker than X-ray. THz absorption spectrum has been measured for many types of molecules, including amino acids, saccharides, proteins, nucleotides and DNA [6–13]. However, most organic or biological molecules present very complex and obscure THz spectral lines, making very hard to resolve [14].

The classification of molecules is necessary and important in some fields, such as chemical, material, biological and medical treatment. In conventional infrared or visible light region, molecular spectra have very narrow spectral peaks and large peak separation, so the characteristic absorption peaks can be picked and assigned to distinguish molecular structures [15]. For THz spectra, on the contrary, molecular absorption bands overlap each other heavily, making hard to use the THz spectra to classify molecules. In order to solve this problem, some methods have been developed to perform the classification of molecules by analyzing the THz absorption spectrum of the molecules, such as support vector machine (SVM) [16], principal component analysis (PCA) [17,18] and multivariate data analysis (MDA) [18]. In these methods, a $n \times m$ order data matrix is obtained by sampling the measured THz absorption curves of a group of samples at different frequency points, in which n is the sample number in the group and m is

the number of the frequency points. The classification can be performed by treating the dada matrix. A quite satisfying accuracy rate can be obtained by SVM [19], but high algorithmic complexity and extensive memory requirement are unavoidable, although PCA and MDA can overcome these weaknesses in a certain extent.

In this paper, inspired by a multivariate analysis statistical method, named as factor analysis (FA) [20,21], we present a new method to classify molecules based on the measured THz absorption spectra of molecules. FA was firstly used in the field of psychology to study the underlying factor influencing human mental ability. With years of development, FA has also been widely used in physical and chemical sciences, such as ecology, molecular biology, astrophysics and cosmology [22–24]. Following the approach of FA, we can take an operation to write the data matrix as the product of two matrices: One is a weight matrix and the other is a characteristic matrix. Such an operation means to project the data matrix toward the characteristic matrix because each row of the characteristic matrix can be seen as a feature vector that relates to one mutual substructure of some kinds of molecules. The element of the weight matrix denotes the projection value of a molecule spectrum toward to a corresponding feature vector. By analyze the weight matrix via K-means clustering (KC) [25], the classification of molecules can be performed. Let us call this new method *THz factor analysis* (TFA). In an actual application, we need firstly to use a group of molecules to build and train a characteristic matrix, in which the categories of the molecules have been known. Once an effective and credible characteristic matrix is obtained, we can use it to classify a group of new molecules.

In order to check the validity of the TFA, sixteen kinds of molecules are selected and their THz absorption spectra are measured. These

---

* Corresponding author.
  E-mail address: jsliu4508@vip.sina.com (J. Liu).

molecules can be intuitively classified via their molecular formula (MF). Such a classification result via MF can be taken as a judgment standard to train a characteristic matrix as well as to examine the classification result by TFA. The sixteen kinds of molecules are divided into two groups. The first group includes twelve samples to be used to build and train a characteristic matrix, and the second group includes the other four samples to be used to check the classification result via TFA based on the built characteristic matrix. By using the measured THz absorption spectra, the data matrix of the twelve samples is obtained. The data matrix can be decomposed into the product of a weight matrix and a characteristic matrix, and there are several possible decompositions. By using the KC method to deal with one of these weight matrixes to obtain a classification result, and comparing the result with the standard one, we can determine which characteristic matrix is effective and credible. Finally, one characteristic matrix is selected based on the corresponding classification result with 75% accuracy compared to the standard result. On this basis, we can classify the second group of samples. To do so, we put their data matrix projecting on the finally selected characteristic matrix and get a weight matrix. By using the KC method to analyze the weight matrix, the four kinds of molecules are classified with 75% accuracy compared to that via MF. These results demonstrate that the TFA method is effective, which can succeed in multi-class classification combining the agility of PCA and the accuracy of SVM.

It is worth to point out that the previous methods hardly provide molecular physics and chemical information because they are mainly based on statistical and mathematical ways [16–19]. In contrast, the TFA method can reflect the molecular structure information due to each row of the characteristic matrix is essentially the THz absorption spectrum of a vibrational mode that is corresponding to a common substructure of molecules. Our work paves a new way for molecular classification based on THz spectrum, and should have unique potential for a wide range of applications from the field of chemical, material, biological to medical treatment.

2. THz Factor Analysis Method

The basic formula in the FA method is [26].

$$\underset{data}{\mathbf{D}} = \underset{loadings}{\mathbf{L}}\ \underset{factors}{\mathbf{F}}, \tag{1}$$

where $\mathbf{D}$, $\mathbf{L}$ and $\mathbf{F}$ stand for data, loading and factor matrixes, respectively. Specific to TFA method, let us consider a group of analogous molecules. To measure the THz absorption spectrum for each kind of the molecules and to take the absorption coefficient at many frequency points, we can get a data matrix, denoted by $\mathbf{A}_{mh}$, for which the subscripts m and h denote m kinds of molecules are taken values at h frequency points from the measured THz spectrum curve. In general, a huge number of the frequency points should be taken. We are interested in how many basic vibration modes contribute to the absorption spectra of these analogous molecules. Following Eq. (1), we can write $\mathbf{A}_{mh}$ as the product of two matrices,

$$\underset{molabs}{A_{mh}} = \underset{wei}{\mathbf{W}_{mn}}\ \underset{vibabs}{C_{nh}}, \tag{2}$$

where $\mathbf{W}_{mn}$ stands for the weight matrix of some possible vibration modes in a group of analogous molecules, and $\mathbf{C}_{nh}$ stands for the characteristic matrix that is essentially the absorption matrix of these vibration modes. The symbol n denotes the number of the modes. In terms of matrix elements, Eq. (2) can be expressed as

$$a_{ik} = \sum_{j=1}^{n} w_{ij}c_{jk}, \tag{3}$$

where a_{ik} denotes the measured THz absorption coefficient of molecule i at frequency k, w_{ij} denotes the weight of molecule i related to the

vibration mode j, and c_{jk} denotes the absorption coefficient of the vibration mode j at wavelength k. It should be pointed out that a_{ik} comes from measurement, while w_{ij} and c_{jk} come from calculation. Each row of $\mathbf{C}$ can be thought as a feature vector, which is, in fact, the THz absorption spectrum of a possible vibrational mode that relates to one mutual substructure of some kinds of molecules. Eq. (2) stands for an operation to project the data matrix $\mathbf{A}$ toward the characteristic matrix $\mathbf{C}$, thus creating a proportional coefficient matrix, i.e., the weight matrix $\mathbf{W}$ whose element w_{ij} is the projection value of the kind i of molecules to the vibration mode j.

At first, we need to select a group of samples whose classification has been known, and then use it to build and train an effective and credible characteristic matrix. The data matrix $\mathbf{A}_{mh}$ of these samples can be obtained from their THz absorption spectra. According to Eq. (2), such $\mathbf{A}_{mh}$ can be written as $\mathbf{A}_{mh} = \mathbf{W}_{mn}\mathbf{C}_{nh}$. In fact, we can take different n to obtain different sets of $\mathbf{W}_{mn}$ and $\mathbf{C}_{nh}$. By using the KC method to deal with $\mathbf{W}_{mn}$, the m molecules can be classified by clustering them to different vibration modes. By taking different values of n, we are able to obtain different classification results. By comparing these results with the previous known ones, we can determine which characteristic matrix is effective and credible. Once such a characteristic matrix has been selected, the corresponding value of n means there should exist n kinds of basic vibration modes in the m molecules.

In order to verify whether the basic idea of the TFA method described above is available, we should firstly prepare a group of samples for which the molecular formula of each sample is known, thus allowing us to intuitively and accurately classify these samples by observing their molecular formulas, shorthand for MFC (molecular formula classification). Then we need to measure the THz absorption spectrum to obtain a dada matrix, and calculate out the corresponding characteristic and weight matrixes. By analyzing the weight matrix, we can give a classification result for these samples. If the result is very closer to that of MFC, we have enough reason to believe that the characteristic matrix is effective, which could accurately reveal how many common substructures existing in these molecules. Once an effective and credible characteristic matrix is obtained, we can use it to classify a new group of samples.

3. Sample Preparation and Measurement

Sixteen kinds of samples are prepared including aromatic diazoniums (spirobenzopyran and indole), styrene derivatives and inorganic salts, which represent four kinds of commonly seen solid chemicals: Organic macromolecules, organic salts, simple organic and inorganic salts. The names and categories as well as the molecular formulas of the samples are listed Tables 1 and 2.

The aromatic diazonium samples are synthesized following the method in Ref. [27]. Other samples are purchased from *Sinopharm Chemical Reagent Co., Ltd.* All the final chemicals are made in powder. All the samples are prepared by mixing with KBr(SP) powder, and then being pressed at 2.0 MPa for 5 min into tablets with diameter of 13 mm and thickness about 1 mm.

Table 1
The elements of the weight matrix $\mathbf{W}_{12, 5}^{(1)}$ and the corresponding KCC categories.

Number	Factor1	Factor2	Factor3	Factor4	Factor5	KCC	MFC
1	0.134	0.782	−0.007	0.069	0.223	C	A
2	0.898	0.022	0.078	−0.008	0.176	A	A
3	0.947	−0.033	−0.157	−0.024	0.018	A	A
4	0.931	−0.007	−0.149	−0.069	0.106	A	A
5	0.153	0.744	−0.124	−0.057	0.063	C	B
6	−0.114	0.018	−0.025	0.634	0.200	B	B
7	−0.011	−0.015	−0.125	0.907	−0.110	B	B
8	0.138	0.824	−0.078	0.003	−0.177	C	C
9	0.029	0.724	−0.072	0.190	0.653	C	C
10	−0.150	−0.153	0.477	−0.265	0.427	D	C
11	−0.079	0.024	0.692	0.069	−0.119	D	D
12	0.341	−0.004	0.821	0.013	0.008	D	D

Table 2
Elements in the weight matrix of samples #13–16 and the corresponding KCC and MFC categories.

Number	Factor1	Factor2	Factor3	Factor4	Factor5	KCC	MFC
13	−0.199	−0.120	0.408	−0.267	0.243	D	D
14	0.093	0.574	−0.232	0.031	0.074	C	A
15	0.088	0.056	0.003	0.408	0.068	B	B
16	−0.006	0.047	0.586	−0.025	−0.013	D	D

KBr is used as supporting material in the mixture, because organic samples are hard to be pressed into rigid pellets. We take about 0.05 g of each sample and 0.05 g of KBr to make a mixture, and to eliminate the influence of KBr, we make another pure KBr pellet with the same mass as in the mixture and set it as the reference. When KBr is treated as an independent sample, pure KBr pellet is used as sample and blank signal is set as reference.

We chose the samples #1–12 to build a characteristic matrix and leave the samples #13–16 waiting for classification. The molecular formulas of the samples #1–12 are listed in Fig. 1. Based on the MF method

they can be classified into four categories, i.e., types A to D. The samples #1–4 are classified as type A due to all the molecules are included in the indoline substructure, as shown in Fig. 1(a). The samples #5–7 are classified as type B due to they are all diphenyl-ethylene-based molecules, as shown in Fig. 1(b). The samples #8–10 are classified as type C due to they are all tetraphenylethylenes, only different in number of methyl substitute groups, as shown in Fig. 1(c). The samples #11–12 are classified as type D due to they are inorganic salts, as shown in Fig. 1(d). The categories based on the MF method are listed in Table 1. The structure molecular formulas of the samples #13-#16 are shown in Fig. 1(e), and their molecular formulas and their categories via the MF method are listed in Table 2.

By using a THz-TDS (Z3 Type, *Zomega, THz Corp., Troy, USA*), the THz absorption spectra of the samples are measured according to the traditional way [28,29] in N_2 atmosphere, in an ISO 6 Cleanroom with temperature controled between 22 °C–24 °C. Each sample is measured 10 times and average signal is calculated.

The measured spectra for the samples #1–16 are shown in Fig. 2. To make the peaks in the figure easy to read, we process the spectra with base-line correction. However, in the establishment of characteristic
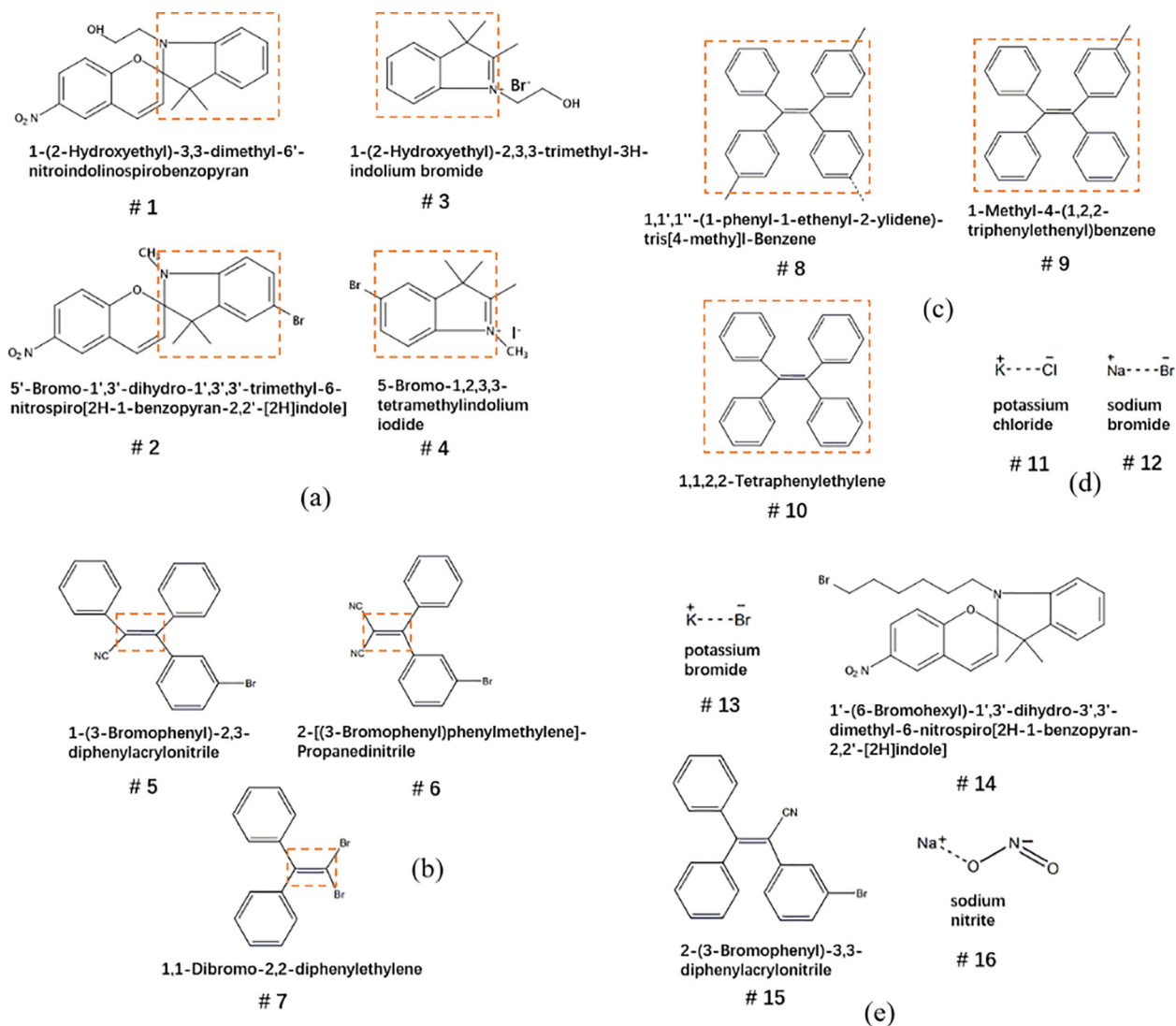

Fig. 1. The structure formulas of molecules #1–16. (a) They all have the indoline structure and are classified as Type A. (b) They are all diphenyl-ethylene-based molecules and are classified as Type B. (c) They are all tetraphenylethylenes, only different in number of methyl substitute groups. They are classified as Type C. (d) They are inorganic salts and are classified as Type D. (e) They are chosen to test the classification.
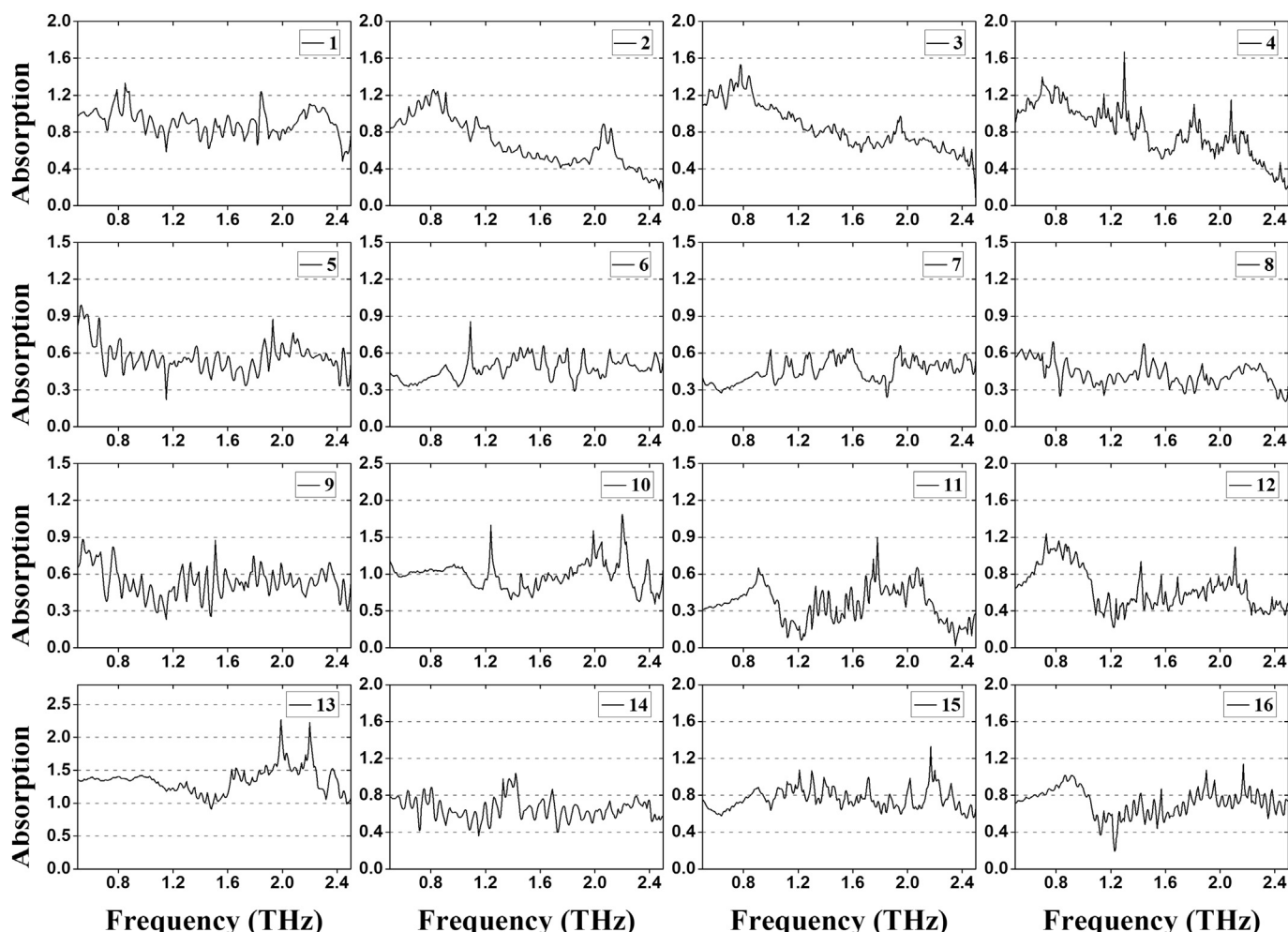
Fig. 2. The measured THz absorption spectra of the samples #1–16.

matrix, we do not correct the base-line but rescale the data matrix instead. As can be seen, these spectra do not present any comparability although some analogous substructures are included in these molecules.

4. Establishment of Characteristic Matrix

The samples #1–12 are chosen to build a characteristic matrix. Firstly, a data matric $\mathbf{A}_{12, 500}^{(1)}$ is made by taking the absorption coefficients at 500 frequency points (0.01–5.00 THz) from the corresponding spectral lines. Then we use software *Microsoft R Open 3.4.3* and its *maximum-likelihood factor analysis algorithm* to perform a calculation [26] to write $\mathbf{A}_{12, 500}^{(1)}$ as $A_{12, 500}^{(1)} = \mathbf{W}_{12, n}^{(1)} C_{n, 500}$. In doing so, we increase n from 2 to large. For each n, we use the KC method to analyze the corresponding weight matrix $\mathbf{W}_{12, n}^{(1)}$ and classify the 12 samples. The optimal n is determined if the KC classification (KCC) result is closest to MFC. Through a lot of calculation, we find $n = 5$ is the optimal value for which the KCC can achieve an 75% accuracy rate comparing to MFC. The finally determined weight matrix $\mathbf{W}_{12, 5}^{(1)}$ and the corresponding KCCs are listed in Table 1. $n = 5$ means five common factors are selected out. Each factor relates to a common substructure that is relative to a vibration or rotation mode whose THz absorption spectrum can be obtained by reading out each row of $C_{5, 500}$ as shown in Fig. 3, which cannot be obtained directly through experimental detection.

We use *OriginLab Pro 9.0* and its *Hierarchical Cluster Analysis* to calculate the cluster dendrogram of the 12 samples. Fig. 4 suggests how these 12 molecules are related to each other. For example, #3 and #4 are directly linked, this means they are very close to each other, in our case, it means their molecular structures are very alike. #2 joins the group

of #3 and #4 in a higher level, which tells us that #2 is similar to #3 but not as close as #4 is.

5. Classify the Unknown Samples

Based on the samples #1–12, we obtain a characteristic matrix $C_{5, 500}$. Relying on the 75% accuracy rate of KCC relating to MFC, we can believe that the characteristic matrix is credible and effective. We can see each row of $C_{5, 500}$ as a feature vector and can use them to classify the unknown samples #13–16. To do so, we add the measurement data of the samples #13–16 into $\mathbf{A}_{12, 500}^{(1)}$ to get a new data matrix $\mathbf{A}_{16, 500}^{(2)}$, and then project $\mathbf{A}_{16, 500}^{(2)}$ to $C_{5, 500}$ making $A_{16, 500}^{(2)} = \mathbf{W}_{16, 5}^{(2)} C_{5, 500}$. As a result, we obtain a new weight matrix $\mathbf{W}_{16, 5}^{(2)}$ whose matrix elements are listed in Tables 1 and 2. The sixteen samples can be classified by using the KC method to analyze $\mathbf{W}_{16, 5}^{(2)}$. For the samples #1–12, their KCCs are just the same as those listed in Table 1. For the samples #13–16, their KCCs show that the samples #13 and #16 belong to D category, the sample #14 belongs to C category and the sample #15 belongs to B category. The accuracy of such classification results can be checked via MFC. For three of the four samples, their KCCs are consistent with their MFCs, achieving 75% accuracy rate. This result indicates that the proposed TFA method is available and has a considerable accuracy.

Furthermore, we use F_1 score to validate the performance of this method on classifying each kind of molecules.

F_1 score is calculated through
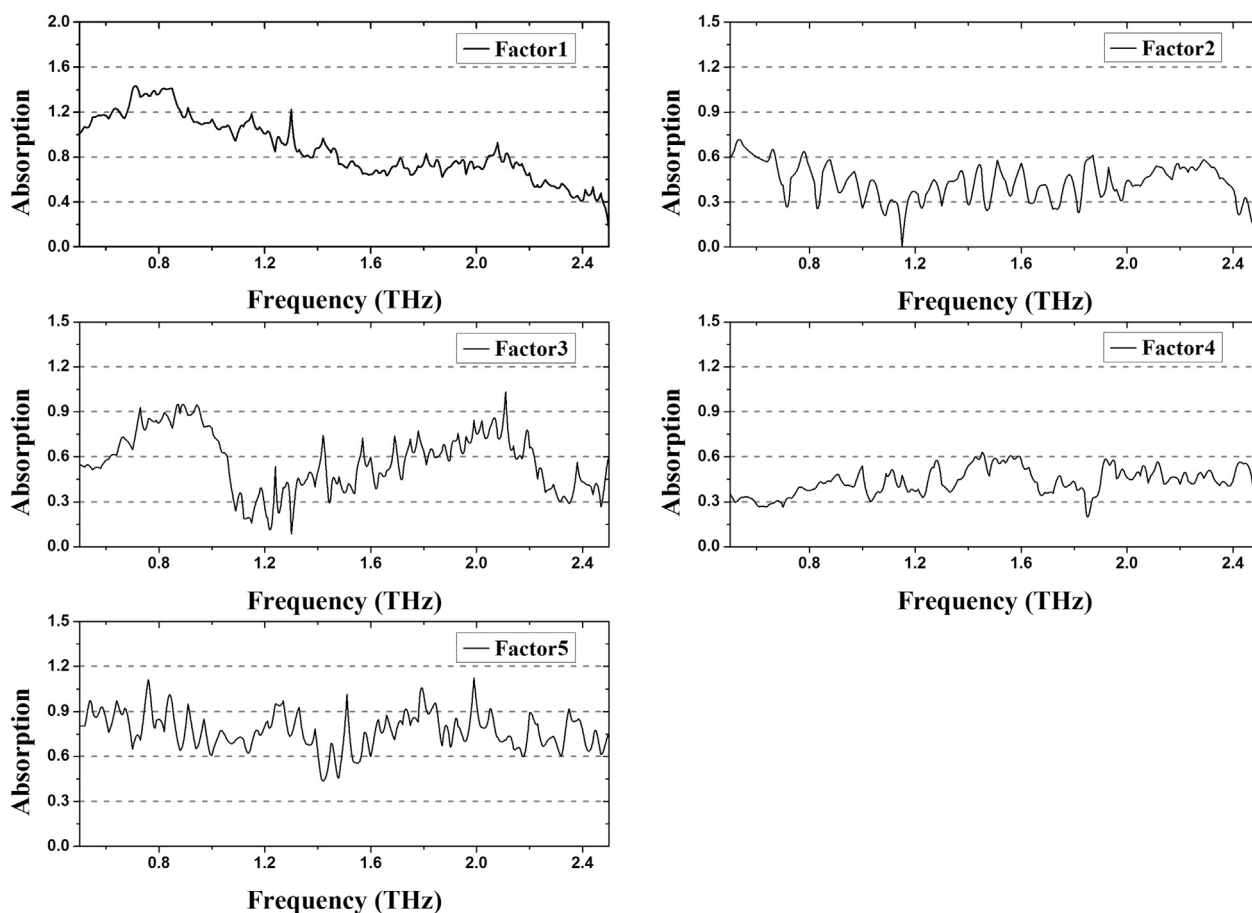
$$F_1 = \frac{2 \times P \times R}{P + R} \tag{4}$$

Fig. 3. Common factors spectra (base-line corrected).

where, P is precision, R is recall.

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN} \qquad (5)$$

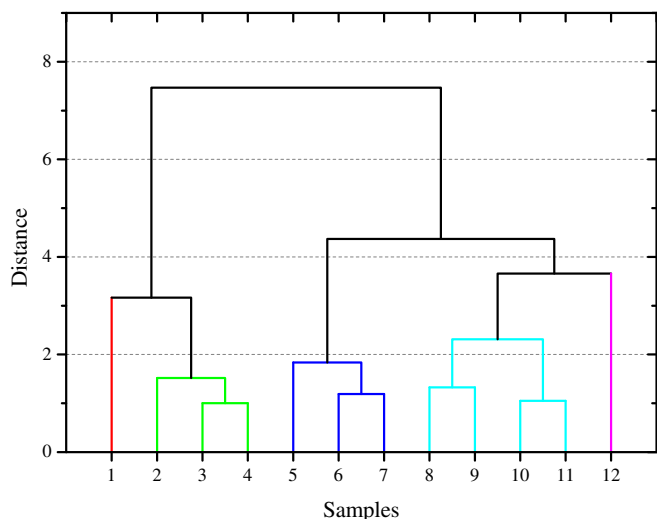where, TP is true positives, FP is false positives, FN is false negatives.

Fig. 4. Cluster dendrogram of the samples #1–12.

According to Eqs. (4)–(5), each F_1 score for classifying A–D molecules is listed in Table 3.

For A, B and D type molecules, their F_1 scores are 0.8–0.9, while only C type molecules fail in this validation, we speculate that characteristic vector of C type molecules is not distinct enough, so the method would occasionally classify other types of molecules into C type.

Overall, our method is sufficient to classify A, B and D type molecules.

6. Conclusion

In this paper, a kind of molecular classification method, namely THz factor analysis, has been built by developing the previous factor analysis approach to deal with the measured THz absorption spectra of the molecules. In an actual application, we need to select a group of sample whose categories have been known and need to measure their THz spectra. Finally, we can obtain a characteristic matrix that can reflect the molecular structure information due to its each row essentially being the THz absorption spectrum of a possible vibrational mode corresponding to a common substructure of molecules. Once the characteristic matrix is determined we can use it to classify a group of new samples. Our work provides a new way to make molecular classification via using THz spectra, being expected to be widely used.

Table 3
Validation for classifying A–D molecules.

	A	B	C	D
Precision	1.00	1.00	0.40	0.80
Recall	0.60	0.75	0.67	1.00
F1 score	0.75	0.86	0.50	0.89

Acknowledgments

References

[1] S.C. Shen, L. Santo, L. Genzel, THz spectra for some bio-molecules, Int. J. Infrared Millimeter Waves 28 (8) (2007) 595–610.

[2] T.R. Globus, D.L. Woolard, T. Khromova, T.W. Crowe, M. Bykhovskaia, B.L. Gelmont, J. Hesler, A.C. Samuels, THz-spectroscopy of biological molecules, J. Biol. Phys. 29 (2–3) (2003) 89–100.

[3] K. Yamamoto, K. Tominaga, H. Sasakawa, A. Tamura, H. Murakami, H. Ohtake, N. Sarukura, Terahertz time-domain spectroscopy of amino acids and polypeptides, Biophys. J. 89 (3) (2005) L22–L24.

[4] D.F. Plusquellic, K. Siegrist, E.J. Heilweil, O. Esenturk, Applications of terahertz spectroscopy in biosystems, ChemPhysChem 8 (17) (2007) 2412–2431.

[5] J.V. Rudd, J.L. Johnson, D.M. Mittleman, Quadrupole radiation from terahertz dipole antennas, Opt. Lett. 25 (20) (2000) 1556.

[6] R.M. Woodward, B.E. Cole, V.P. Wallace, R.J. Pye, D.D. Arnone, E.H. Linfield, M. Pepper, Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue, Phys. Med. Biol. 47 (21) (2002) 3853.

[7] M. Yamaguchi, F. Miyamaru, K. Yamamoto, M. Tani, M. Hangyo, Terahertz absorption spectra of L-, D-, and DL-alanine and their application to determination of enantiometric composition, Appl. Phys. Lett. 86 (2005), 053903. .

[8] M. Walther, B.M. Fischer, P. Uhd Jepsen, Noncovalent intermolecular forces in polycrystalline and amorphous saccharides in the far infrared, Chem. Phys. 288 (2–3) (2003) 261–268.

[9] B.M. Fischer, H. Helm, P.U. Jepsen, Chemical recognition with broadband THz spectroscopy, Proc. IEEE 95 (8) (2007) 1592–1604.

[10] P.C. Upadhya, Y.C. Shen, A.G. Davies, E.H. Linfield, Far-infrared vibrational modes of polycrystalline saccharides, Vib. Spectrosc. 35 (1–2) (2004) 139–143.

[11] P.C. Upadhya, K.L. Nguyen, Y.C. Shen, J. Obradovic, K. Fukushige, R. Griffiths, L.F. Gladden, A.G. Davies, E.H. Linfield, Characterization of crystalline phase-transformations in theophylline by time-domain terahertz spectroscopy, Spectrosc. Lett. 39 (3) (2006) 215–224.

[12] T.M. Korter, R. Balu, M.B. Campbell, M.C. Beard, S.K. Gregurick, E.J. Heilweil, Terahertz spectroscopy of solid serine and cysteine, Chem. Phys. Lett. 418 (1–3) (2006) 65–70.

[13] P.C. Upadhya, Y.C. Shen, A.G. Davies, E.H. Linfield, Terahertz time-domain spectroscopy of glucose and uric acid, J. Biol. Phys. 29 (2–3) (2003) 117–121.

[14] B.C. Smith, Infrared Spectral Interpretation: A Systematic Approach, CRC Press, 1998.

[15] B.H. Stuart, Infrared Spectroscopy: Fundamentals and Applications, John Wiley & Sons, 2004.

[16] X. Yin, B.W.-H. Ng, B. Fischer, B. Ferguson, S.P. Mickan, D. Abbott, Feature Extraction From Terahertz Pulses for Classification of RNA Data via Support Vector Machines, 2006.

[17] W. LI, B. HU, M. WANG, Discrimination of varieties of borneol using terahertz spectra based on principal component analysis and support vector machine, Spectrosc. Spectr. Anal. 34 (12) (2014) 3235–3240.

[18] A. Pohl, N. Deßmann, K. Dutzi, H.-W. Hübers, Identification of unknown substances by terahertz spectroscopy and multivariate data analysis, J. Infrared Millimeter Terahertz Waves 37 (2) (2016) 175–188.

[19] C. Xi-Ai, Z. Guang-Xin, H. Ping-Jie, H. Di-Bo, K. Xu-Sheng, Z. Ze-Kui, Classification of the green tea varieties based on support vector machines using Terahertz spectroscopy, 2011 IEEE Instrumentation and Measurement Technology Conference (I2MTC), 2011.

[20] C. Spearman, 'General intelligence,' objectively determined and measured, Am. J. Psychol. 15 (2) (1904) 201–292.

[21] R.B. Cattell, Theory of fluid and crystallized intelligence: a critical experiment, J. Educ. Psychol. 54 (1) (1963) 1–22.

[22] K.H. Riitters, R.V. O'Neill, C.T. Hunsaker, J.D. Wickham, D.H. Yankee, S.P. Timmins, K. B. Jones, B.L. Jackson, A factor analysis of landscape pattern and structure metrics, Landsc. Ecol. 10 (1) (1995) 23–39.

[23] D. Risso, J. Ngai, T.P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples, Nat. Biotechnol. 32 (9) (2014) 896–902.

[24] Z. Bagoly, L. Borgonovo, A. Mészáros, L.G. Balázs, I. Horváth, Factor analysis of the long gamma-ray bursts, Astron. Astrophys. 493 (1) (2009) 51–54.

[25] J. MacQueen, Some methods for classification and analysis of multivariate observations, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, The Regents of the University of California, 1967.

[26] E.R. Malinowski, Factor Analysis in Chemistry, 3rd edition Wiley, New York, 2002.

[27] D.A. Davis, A. Hamilton, J. Yang, L.D. Cremar, D. Van Gough, S.L. Potisek, M.T. Ong, P. V. Braun, T.J. Martínez, S.R. White, J.S. Moore, N.R. Sottos, Force-induced activation of covalent bonds in mechanoresponsive polymeric materials, Nature 459 (7243) (2009) 68–72.

[28] Y.-S. Lee, Principles of Terahertz Science and Technology, Springer Publishing Company, Incorporated, 2008.

[29] L. Duvillaret, F. Garet, J.-L. Coutaz, A reliable method for extraction of material parameters in terahertz time-domain spectroscopy, IEEE J. Sel. Top. Quantum Electron. 2 (3) (1996) 739–746.