DOI: 10.2428/ecea.2012.19(08)095 ECOL CHEM ENG A. 2012;19(8):973-981

Krystyna MACEK-KAMIŃSKA[1], Marek SUDOŁ[2]
and Sławomir STEMPLEWSKI[1*]

MIXTURES IDENTIFICATION OF CHEMICAL COMPOUNDS ON THE BASIS OF THEIR IR SPECTRA BY ARTIFICIAL INTELLIGENCE

IDENTYFIKACJA MIESZANIN ZWIĄZKÓW CHEMICZNYCH NA PODSTAWIE ICH WIDM W PODCZERWIENI Z WYKORZYSTANIEM SZTUCZNEJ INTELIGENCJI

Abstract: Infrared (IR) spectrometric identification of individual chemical compounds from their mixtures is still a challenging process. Therefore, we developed a method in which we use the IR "Fingerprint" spectra of a particular chemical substance followed by artificial intelligence (AI) – based analysis to correctly characterise components of relatively simple chemical mixtures. We describe here the assembly of tools developed especially for this purpose as well as the artificial neural network design together with the requirements that must be met for its proper functioning. To test our approach, we used a mixture of amphetamine and creatinine which are difficult to identify in mixtures by standard "Fingerprint" rules. The advantages of the artificial neural network approach include the generalisation and adaptation of knowledge by fitting parameter values to change the object characteristics. All this renders the effective identification of a mixture of two substances possible.

Keywords: chemical compounds, infrared spectroscopy, artificial intelligence, neural network

The traditional organic compounds identification method using IR absorption spectra (FTIR) is based on the identification of chemical functional groups contained within the molecules of tested compounds. We perform this by reading characteristic bands frequencies of a spectrum [1–4]. In the functional group identification we are using correlation tables which bind a frequency of a spectrum with a specific functional group. After determining the type of a functional group, compounds must be classified to appropriate chemical compounds class (to homologous series). The last stage of a

[1] Institute of Electromechanical Systems and Industrial Electronics, Opole University of Technology, ul. Prószkowska 76, 45–758 Opole, Poland, phone: +48 77 449 80 00, fax: +48 77 453 64 39, email stempell@gmail.com

[2] Department of Chemical Technology and Polymer Chemistry, University of Opole, ul. Oleska 48, 45–052 Opole, Poland, phone: +48 77 452 71 38, fax: +48 77 452 71 01, email: msudol@uni.opole.pl

[*] Corresponding author.

compound identification is to find a spectrum in the IR spectrum catalog (in the identification of compound class section) with an identical image as the tested spectrum in the full intermediate IR range (4000; 400) cm^{-1}. The spectral identity of a compound is determined by searching for an identical spectrum in the catalog. This method is called *Fingerprint* (FP). The method is quite time-consuming and never gives a full guarantee of the identification accuracy.

For this reason, today, in the FP compounds identification method a suitable computer software is used to search the right compound in the spectral library. The library is located in the database and automatically determines the identity of a tested compound, specifying the spectral compatibility (sample – pattern) in a percentage.

Unfortunately, identification methods based on the FP are not applicable in the case of mixtures of two or more substances.

This paper presents an innovative method of identifying chemical compounds on the basis of their IR spectra, using the *artificial neural networks* (ANN) in a spectral analysis and a method for a quantitative analysis of binary mixtures. Due to the current demand, the analysis of binary systems using the ANN was conducted on the example of a creatinine and amphetamine mixture.

Identification method based on artificial neural network uses one of the main advantages of ANN – the possibility to generalise knowledge. This involves the identification of a chemical compound most similar to the one searched for in the database. In comparison with the FP, the methods of monosubstance and mixture identification using ANN have several advantages:

– In the FP method, the entire spectral ranges are compared with each other and the extent to which the overlap is checked, while the ANN compares the position and height of the top 20 maximum points.

– ANN is more sensitive to small spectral differences, therefore, it identifies substances in a mixture of compounds better than the FP method does; it is shown in this article on the example of mixtures of creatinine and amphetamines.

– ANN verifies the set of chemical functional groups in the substance, which gives more reliable results of identification.

– ANN "learns" finding the spectral differences between compounds, which is later used in the identification.

– ANN, even damaged, may continue to operate properly.

Artificial Neural Networks

ANN is a system performing certain calculations on the principle of simultaneous work of many interconnected elements (neurons). A similar structure has been observed in biological nervous system (ex. *human cerebellum*). It is composed of many neurons, which constitute a very simplified version of the original, and are much more easily combined with each other.

ANN was designed in such a way as to be convenient to track and easy to implement, which makes ANN have a flat and a set regular structure. Moreover, ANN should usually contain many neurons and should be able to work only as a whole.

The program described here uses ANN, whose neurons introduce corrections to their state of knowledge from the known values of the error committed. Output layer error is calculated from the difference of the expected and obtained value at the output of the network, while the error of neurons below the output layer is calculated by the back error propagation. A significant feature of this method is that its effect is independent of any theoretical assumptions. This means that the backward error propagation will always work [5–11].

Using the ANN for identification opens up new possibilities. Not only does it speed up the process, but also gives it the ability to adapt: to adjust the parameters to the changes of the characteristics of the object and to generalizes which means generating an appropriate response to the data not used in the learning process [12]. Also, ANN's ability to ignore redundant data and the data whose impact on the phenomenon is negligible has been used. Therefore, ANN focuses on the input data having a major impact for the modelled process. These are the crucial features which let ANN find application in many branches of science and technology [12–15].

ANN used in the program consists of:

– 20 entries,
– One hidden layer (consisting of 40 neurons),
– 10 outputs.

In order to identify the object, two neural networks were used: one to identify the number of wave bands, and the other with a maximum absorption of this bands.

A computer program for the diagnosis of chemical compounds based on the "fingerprint" of the infrared range using artificial neural networks

The program for the diagnosis of chemical compounds using ANN is written in C# and does not use any other tools to create an ANN such as Matlab. Figure 1 presents the

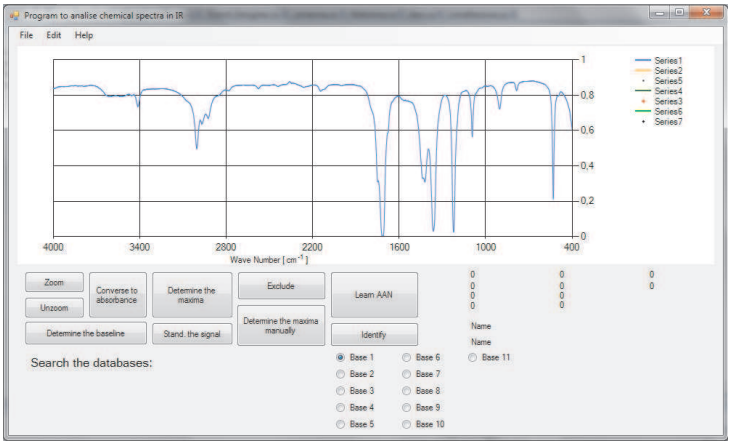

Fig. 1. General view of a computer program for the diagnosis of chemical compounds

main panel where the visible buttons are triggering processes responsible for the initial analysis of the test spectrum and for its further identification.

Before the spectrum is about to be subjected to a diagnosis, it should be examined first. The analysis involves determining transmission spectra and possibly replacing it with the absorption spectrum as well as designing the so-called base line (Fig. 2). The baseline is important because the values of the maximum individual bands are calculated from it and not from zero.

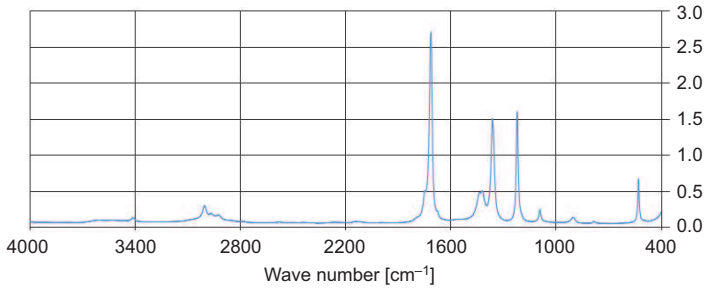

Fig. 2. Absorption spectra of acetone with a designated base line

After this process, the program determines the maximum points in each band (Fig. 3). This is done automatically using an algorithm written specially for this purpose and built into the program identifier. After determining the points, there are two possibilities of adjustment. The first concerns the selection of places where the program ignored the key points, while the second relates to the points that should not have been selected.

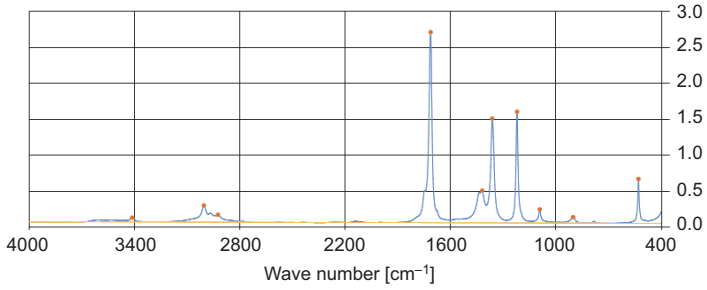

Fig. 3. Absorption spectra of acetone with designated points of maximum

The next process is the standardisation of the signal which will be then sent to the input ANN. The highest determined value of the absorption spectrum is reduced to unity, while lower values according to the principles of proportion to the respective values (Fig. 4). This solution increases the effectiveness of training an ANN, and thus the identification of chemical compounds. The substances may be present in different concentrations and peak heights of individual absorption bands may differ, even though

New Max	Old Max
1	2,652698
0,5837261	1,548449
0,547359	1,451978
0,2332636	0,6187778
0,1693595	0,4492596
0,08729737	0,2315736
0,07154877	0,1897973
0,03888361	0,1031465
0,03149752	0,0835534
0,02213695	0,05872265

Fig. 4. Standardisation of the input signal

they belong to the same compound. The introduction of standardised signal solves this problem, because all values are imported by the proportion of one value, and as the identification method theory says the substances have a fixed relationship of the individual absorption bands and this operation is therefore possible.

Before the program proceeds to the final stage of identification, it still has to designate of specific chemical functional groups, which the program executes automatically, too.

This is a part of the verification because the program checks whether the spectrum may belong to a given substance by checking the chemical functional groups (it may designate more than one group) with the group stored in the database. Figure 5 shows the window for selecting the chemical functional groups at the time of creating a database of compounds.

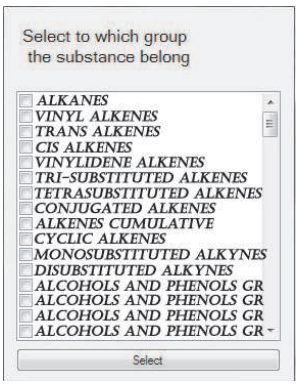

Fig. 5. The choice of chemical functional groups

After determining the chemical functional groups, in the main panel there appears the information suggesting which database should be searched to find the chemical compound (Fig. 6). The displayed information is merely a hint and does not block the possibility of searching other databases.

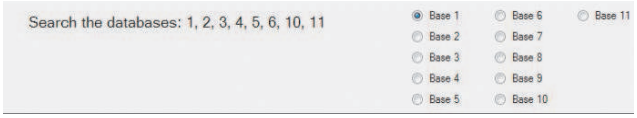

Fig. 6. Designation database

The last stage of the program is the identification by ANN. It takes ANN quite a lot of time to learn, but the identification process itself takes a fraction of a second. ANN outputs are shown in Fig. 7; they are saved bitwise, so it is possible to write up to 1023 chemical compounds on them. Activation function that was used in the program is a bipolar function, so the program receives the output value of –1 and 1. The following figure has value of 1 in only the first position, which is the information to the program that the spectrum points are the first link in the database.

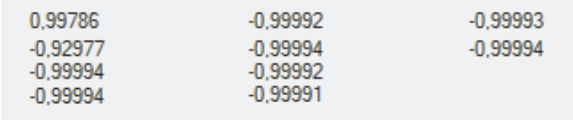

Fig. 7. ANN outputs

Results and discussion

In order to confirm the effectiveness for the identification of chemical compounds methods based on their infrared spectra using the ANN, a study of mixture of two chemical compounds – creatinine and amphetamine – was performed. Figure 8 shows absorption spectrum of amphetamine, while the absorption spectrum of creatinine is in Fig. 9.

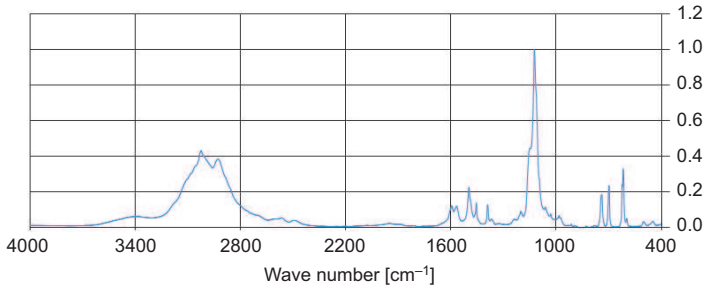

Fig. 8. The absorption spectrum of amphetamine

Because the samples were made using FTIR (Fourier Transform Infrared Spectroscopy), it is possible to mix them by the computer program. In this way, fourteen samples were created containing amphetamines from 5 % to 95 % in steps of 5 %.

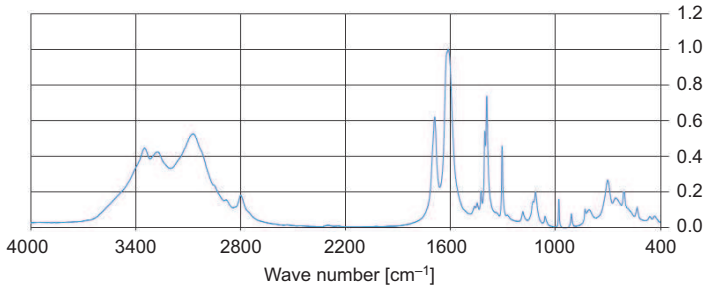

Fig. 9. The absorption spectrum of creatinine

Both spectra have characteristic bands in similar positions, and the additional difficulties is the fact that the band at the time of mixing the two substances overlap to themselves blurring the differences and make the mixture very difficult to identify later. Figure 10 shows the spectrum of the mixture of creatinine (40 %) and amphetamines (60 %).

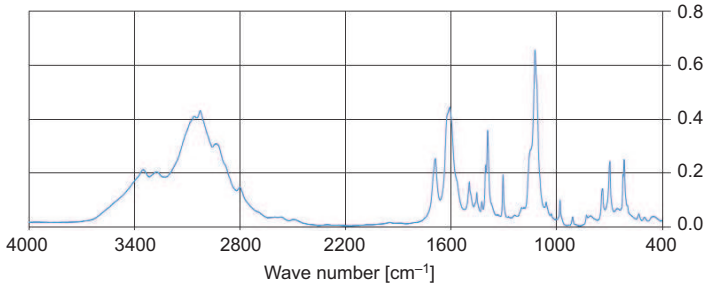

Fig. 10. Absorption spectrum of a mixture of creatinine (40 %) and amphetamines (60 %)

Figure 11 shows the same mixture, but in inverted proportions (creatinine – 60 %, amphetamines – 40 %). While in this case it is easy to discern the difference between the two spectra, it is far more difficult to notice the differences between the spectra of

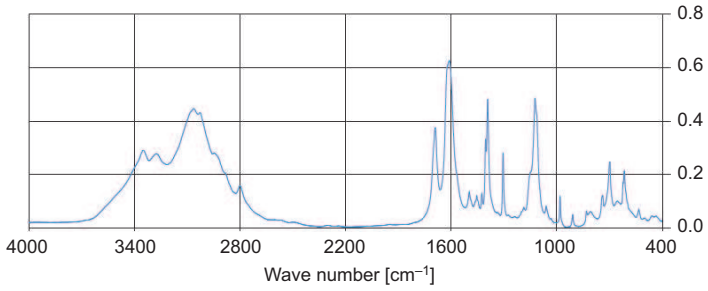

Fig. 11. Absorption spectrum of a mixture of creatinine (60 %) and amphetamines (40 %)

pure creatinine and the mixture having only 5 % of amphetamine, as it is shown in Fig. 12.

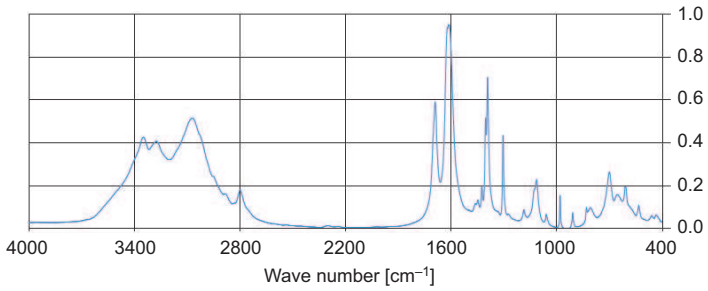

Fig. 12. Absorption spectrum of a mixture of creatinine (95 %) and amphetamines (5 %)

The database of 16 spectra was used by ANN to learn how to identify mixtures and compounds. ANN managed to find the minimum differences separating different spectra, and used this knowledge in their subsequent identification. After the learning process completed, ANN was able to recognise all of the 16 samples with a very high efficiency. In this case, ANN managed much better than the identification schemes used before, encountering problems only when there was less amphetamine than creatinine in a mixture.

Conclusions

Although the ANN has been known worldwide for nearly 50 years, its potential has still not been fully exploited. There are many areas of science and technology, where it will certainly find wider use, as it is the case with the identification of chemical compounds based on their infrared spectra. Software developed for this purpose can be a useful tool for rapid diagnosis of substance, for example in forensic laboratories. The presented method certainly needs improvement, but the results show that it is an effective and reliable method of identification.

Acknowledgement

Work co-financed by European Social Fund.

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

POLITECHNIKA OPOLSKA

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

References

[1] Kęcki Z. Podstawy spektroskopii molekularnej. Warszawa: Wyd. PWN; 1998.
[2] Silverstein R, Webster F, Kiemle D. Spektroskopowe metody identyfikacji związków organicznych. Warszawa: Wyd. Naukowe PWN; 2007.
[3] Zielinski W, Rajca A. Metody spektroskopowe i ich zastosowanie do identyfikacji związków organicznych. Warszawa: WNT; 1995.

[4] Kazicyna LA, Kupletska NB. Metody spektroskopowe wyznaczania struktury związków organicznych. Warszawa: PWN; 1976.
[5] Tadeusiewicz R. Odkrywanie właściwości sieci neuronowych przy użyciu programów w języku C#. Kraków: Polska Akademia Umiejętności; 2007.
[6] Tadeusiewicz R. Sieci neuronowe. Warszawa: Akademicka Oficyna Wydawnicza RM; 1993.
[7] Tadeusiewicz R. Elementarne wprowadzenie do techniki sieci neuronowych z przykładowymi programami. Warszawa: Akademicka Oficyna Wydawnicza PLJ; 1998.
[8] Korbicz J, Obuchowicz A, Uciński D. Sztuczne sieci neuronowe. Warszawa: Akademicka Oficyna Wydawnicza PLJ; 1994.
[9] Rutkowski L. Metody i techniki sztucznej inteligencji. Warszawa: PWN; 2006.
[10] Żurada J, Barski M, Jędruch W. Sztuczne sieci neuronowe – podstawy i zastosowania. Warszawa: PWN; 1996.
[11] Ossowski S. Sieci neuronowe do przetwarzania informacji. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej; 2000.
[12] Rojek R, Bartecki K, Korniak J. Metody sztucznej inteligencji w zastosowaniach automatyki. PAK. 2006;10:29-34.
[13] Chudzik S, Gryś S, Minkina W. Wykorzystanie sztucznych sieci neuronowych w zagadnieniu odwrotnym dyfuzji ciepła. PAK. 2009;02:83-88.
[14] Bartecki K, Czorny M. Implementacja sztucznej sieci neuronowej w architekturze równoległej z wykorzystaniem protokołu MPI. PAK. 2011;06:638-640.
[15] Giergiel M, Małka P. Sztuczne sieci neuronowe w sterowaniu mini robota kołowego. PAK. 2004;05:20-24.

IDENTYFIKACJA MIESZANIN ZWIĄZKÓW CHEMICZNYCH NA PODSTAWIE ICH WIDM W PODCZERWIENI Z WYKORZYSTANIEM SZTUCZNEJ INTELIGENCJI

[1] Instytut Układów Elektromechanicznych i Elektroniki Przemysłowej
Politechnika Opolska
[2] Katedra Technologii Chemicznej i Chemii Polimerów
Uniwersytet Opolski

Abstrakt: W artykule przedstawiono możliwości zastosowania w identyfikacji związków chemicznych metody tzw. odcisku palca oraz sztucznej inteligencji na podstawie widm w podczerwieni. Opisano budowę opracowanego specjalnie do tego celu narzędzia i sztuczną sieć neuronową oraz wymogi, jakie muszą być spełnione do jej poprawnego funkcjonowania. Obecnie stosowane programy użytkowe do identyfikacji związków chemicznych na podstawie ich widm w podczerwieni natrafiają na trudności z poprawną identyfikacją w przypadku mieszanin substancji. W przeprowadzonych badaniach testowych wykorzystano mieszaninę kreatyniny oraz amfetaminy – substancje z którymi obecnie wykorzystywane oprogramowania działające wg zasady Finger-print mają duże trudności. Dlatego też zastosowano sztuczną sieć neuronową, której zalety, takie jak uogólnianie zdobytej wiedzy oraz adaptacja, czyli dopasowania wartości parametrów do zmian charakterystyk obiektu, pozwalają na skuteczną identyfikację w mieszaninie dwóch substancji.

Słowa kluczowe: związki chemiczne, podczerwień, spektroskopia, sztuczna inteligencja, sieć neuronowa.