On the possibility of scientific knowledge formalization by fuzzy logic methods

To cite this article: Vladimir V Serov 2019 J. Phys.: Conf. Ser. 1399 033043

View the article online for updates and enhancements.
On the possibility of scientific knowledge formalization by fuzzy logic methods

Vladimir V Serov
Russian state social University, str. 4, p. 1, Wilhelm Peak, Moscow, 129226, Russia

E-mail: vvs030948@mail.ru

Abstract. The issues of formalizing scientific information by fuzzy logic methods are considered. The object of study is the generally accepted theory and knowledge; in our case it will be molecular spectroscopy. The method will be applied calculus of fuzzy first-order predicates. Examples of formalizing knowledge from the field of IR spectroscopy are given, as well as methods for solving problems of a qualitative nature (interpretation, prediction, identification) with an assessment of the truth of the results. The possibility of using these methods for the development of expert and intelligent systems in various subject areas is shown.

1. Introduction
Among the huge and rapidly growing flow of scientific information, one can single out areas of knowledge in which there are generally recognized, relatively stable theories, methods, and experimental data. Such knowledge is digitized, converted into monographs, algorithms, programs, databases and knowledge and is widely used in practical activities. Well-known methods can be used to formalize such knowledge, in particular, the applied logical calculus of fuzzy predicates proposed by the author [1-4]. Applied calculus differs from classical logic of first-order predicates in that instead of constants and variables it allows the use of fuzzy sets [5] with their characteristic membership functions. Knowledge is represented by logical expressions with their truth values, organized in some structure. All this allows us to take into account the fuzziness, uncertainty of knowledge and evaluate the truth of solutions to various problems of a qualitative nature (interpretation, forecasting, identification, etc.). Let us consider the application of this theory for molecular IR spectroscopy.

Molecular spectroscopy studies the optical spectra that arise when molecules emit or absorb radiation. The complexity of the molecular spectra is due to the fact that in the processes of emission, absorption and scattering of light by a molecule, along with electrons, nuclei participate, the motion of which manifests itself in the spectrum of the molecule. The molecule can oscillate around certain equilibrium positions and can rotate as a whole relative to the center of gravity. Both the vibrational and rotational motions of the molecule are quantized. Vibrational absorption spectra of the molecule are located in the infrared region (IR spectra). In addition, molecular vibrations also appear in Raman spectra in the visible region (Raman spectra). To obtain IR spectra, special equipment is used — spectrometers, which allow one to obtain a spectral curve (spectrogram) that gives the dependence of the spectrum intensity on the wavelength. In IR spectroscopy, the concept of characteristic features is generally accepted, the essence of which is that there are groups of atoms (fragments), the presence of which in the composition of the molecule leads to the appearance on the spectrogram of absorption bands in certain frequency ranges with known (but not always) intensity values and half widths. From...
the point of view of the representation of knowledge, information has a clearly expressed logical character in the form IF \rightarrow TO, i.e. production systems. At the same time, knowledge is fuzzy due to the complexity of the object of research, which is a polyatomic molecule. The characteristic frequencies are set by possible intervals of numerical values, the intensity and half-width of the strip are described linguistically (weak, strong, very strong, narrow, wide, etc.).

2. Results

Applied calculus of fuzzy predicates allows one to describe spectrochemical knowledge [3, 4] without significant loss of information and to solve qualitative problems of molecular spectroscopy, such as predicting and interpreting spectra, structural group analysis, establishing the structure of molecules with evaluating the truth of the results.

In the representation of knowledge, network, frame, and production models are used. Knowledge is structurally integrated into a hierarchical network in the form of a set of trees whose roots are the main structural fragments of molecules. Each network node contains a description of all facts and patterns related to it (frame). Each individual description is a product in the applied calculus of fuzzy predicates.

1. Predicates characterizing the IR spectrum.

I_1(a, b, x_1, y_1, z_1) - the spectrum contains a type of b vibration band group with x_1 frequency, y_1 intensity, z_1 half-width.

Intensity and half-width are specified by integers in the interval (0.5), frequency values by positive integers, a variable is defined on the set of possible types of vibrations ((\nu_s - symmetrical stretching vibrations, \delta_\alpha - plane deformation, etc.).

I_2(a, b, x_2, n) - the spectrum contains n split vibrational bands of a type of b group with x_2 frequency.

2. Predicates characterizing the structure of a molecule.

R_1(r_1) – the molecule contains r_1 fragment.

R_2(r_1, r_2, n) – the molecule contains r_1 fragment and r_2 fragment in n connections from it.

R_3(r_3, n_3) – the molecule contains n_3 structural of r_3 element.

R_5(r_5, j) – there is a cycle of j dimension containing r_5 fragment.

3. Predicates characterizing additional conditions.

As units for structuring knowledge, we take the main fragments that have characteristic features (C=C, C≡C, and C=O etc.). Each main fragment is written off as a whole, and then environment atoms are sequentially attached to it, forming fragments of the 1st, 2nd, 3rd, etc. levels, to the border of existing knowledge. The resulting hierarchy of knowledge is a graph in the form of a tree (a network of declarative knowledge) (figure 1).

Figure 1. Declarative knowledge network.
If some simplifications are allowed, then in the calculus of fuzzy predicates the main hypothesis that describes the concept of characteristic features in IR spectroscopy can be represented as the following set (conjunction) of expressions of the form:

$$\forall_{i} R_i(A_i) \rightarrow \mu_i \cdot I_i(a_i, b_i, x_i, y_i, z_i),$$

where \( n_i \) is the number of spectral regions in which characteristic features of \( A_i \) fragment appear, \( \mu_i \) is the external truth value of \( I_i(a_i, b_i, x_i, y_i, z_i) \) predicate, which is a numerical estimate of the manifestation of the characteristic features of \( A_i \) fragment in \( t \) region. The expression should be understood as follows: if the structure of the molecule contains \( A_i \) fragment, then in each region \( t \) of \( n_i \) regions of the spectrum with \( \mu_i \) truth \( a_i \in \{v, u, \delta, \delta_p, \ldots\} \) vibration bands of \( b_i \) type of \( A_i \) structure group bit fragment with frequency, intensity and half-width described by the values of \( x_i, y_i, z_i \) fuzzy variables.

The solution of the problem, as in classical logic, is carried out by the method of resolutions. The difference is that as a result we get an assessment of the truth of the solutions. Thus, knowledge in a generalized form can be written as \( \wedge(R^k \rightarrow I^k) \), the conditions for solving the problem as \( R^c \), the problem itself can be formulated as \( 'T' \), then to solve it, it is necessary to prove the impracticability of the expressions \( \wedge(R^k \rightarrow I^k) \wedge R^c \wedge 'T' \), or, after the equivalent transformations, \( \wedge(R^k \wedge R^c \wedge 'T') \), or \( \wedge(R^k \wedge R^c \wedge 'T' \lor I^k \rightarrow 'T') \). If fuzzy predicates \( R^k, I^k, R^c \) contain constants, and \( 'T' \) contains only variables, then the qualitative problem is formulated as a forecasting problem. If \( 'T' \) contains both constants and variables, then we are dealing with the task of interpretation.

Let us further consider the identification problem. For molecular spectroscopy, it is formulated as the task of establishing the structure of a molecule from its spectra. In this case, the conditions of the problem are written as \( 'F' \), the statement of the problem as \( R^k \), to solve it, it is necessary to prove the impracticability of the expressions \( \wedge(R^k \rightarrow I^k) \wedge R^c \wedge 'F' \) or the expressions \( \wedge(R^k \wedge 'F' \wedge R^c \wedge 'F') \wedge (R^c \wedge 'F') \). It is clear that in this formulation the identification problem has no solution, because it is impossible to get an empty clause due to the lack of cut-off type I letters. In order for a solution to be possible, it is necessary to change the form of knowledge representation to inverse: \( (R^k \rightarrow I^k) \Rightarrow (I^k \rightarrow R^k) \), where \( \Rightarrow \) is the transformation symbol. Knowledge in the original and transformed forms is nonequivalent, transformation is possible only if it allows us to make their internal structure with the use of additional hypotheses and conditions. In this case, we proceed to structural group analysis (SGA), as a result of which conjunct variants describing possible combinations of fragments can be obtained. If we return to the first stage, it is clear that for each solution there is a different interpretation of the correlation of absorption bands, fragments, and vibrating groups.

At the next stage of the formal solution of the problem of establishing the structure of polyatomic molecules, the possible variants of structural formulas are generated from a set of fragments established as a result of the SGA and the atoms remaining from the gross formula using graph theory methods [4]. Each of the obtained variants of the structural formula must be checked for compliance with the experimental spectrum and formalized knowledge. In the calculus of fuzzy predicates, this problem can be formulated as follows: for each \( A_i \) fragment from \( \{A_i\} \) set that is part of \( S \) checked structural formula, prove the impracticability of \( \neg R_i(A_i) \) formula. If this cannot be done for at least one of the fragments of \( \{A_i\} \) set, then \( S \) structural formula is excluded from consideration. The value of the truth function obtained as a result of the proof serves to rank the answers.

For the proof, the same form of knowledge representations is used as for solving the SGA problem at the first stage, and the response clauses for each \( A_i \) fragment contain information about the groups responsible for the appearance of a particular frequency and the types of vibrations of these groups, i.e. At the same time, the spectrum is interpreted.
3. Conclusions
The possibility of using the methods of applied calculus of fuzzy predicates to formalize scientific knowledge that has a logical structure and a certain degree of uncertainty or fuzziness by nature is shown. Methods for solving problems of a qualitative nature with an assessment of the truth of solutions and interpretation of the results are considered. Due to the fact that this approach allows solving problems in an automatic mode, it is applicable for creating various expert and intelligent systems [4, 6-9].

References
[1] Serov V V, Sokolov I V and Budnik A A 2019 IOP Conf. Ser.: Mater. Sci. Eng. 537(4) 04243
[2] Elyashberg M E, Serov V V and Gribov L A 1991 An expert system for molecular elucidation based on spectral data Computational and Theoretical Chemistry 230 191-203
[3] Serov V V, Elyashberg M E and Gribov L A 1988 System-related matters in the theory of solving qualitative problems of molecular spectroscopy Journal of Molecular Structure 178 1-21
[4] Elyashberg M E, Serov V V and Gribov L A 1987 Artificial intelligence systems for molecular spectral analysis Talanta 34(1) 21-30
[5] Zadeh L A 1965 Fuzzy sets Information and Control 8(3) 338-53
[6] Kałkowska J 2016 Information and communication technologies supporting fuzzy knowledge management (Springer: Cham) pp 363-73
[7] Liu H C et al 2017 Fuzzy Petri nets for knowledge representation and reasoning: A literature review Engineering Applications of Artificial Intelligence 60 45-56
[8] Wang W M et al 2014 Dynamic representation of fuzzy knowledge based on fuzzy petri net and genetic-particle swarm optimization Expert Systems with Applications 41(4) 1369-76
[9] Miroshnik M A et al 2015 Designing artificial intelligence systems using fuzzy logic Radio Engineering 182 42-50