Knowledge extraction from the problem-oriented data warehouses on chemical kinetics and thermochemistry

VLADIMIR E. TUMANOV
Department of Computing and Information Resources
Institute of Problems of Chemical Physics RAS
142432, Chernogolovka, Semenov ave, 1
RUSSIAN FEDERATION
tve@icp.ac.ru

Abstract: This paper discusses the technology of extracting the chemical knowledge from the structured electronic sources, problem-oriented systems of science data analytics, and methods of science data analysis. Application of the feed-forward artificial neural network for predicting the reactivity of a compound (the classical potential barrier) in reactions of hydrocarbons with hydrogen atom in solution is presented. Empirical indexes of reactionary centers for groups of such reactions had identified. The artificial neural network is learned using a set of the experimental thermochemical and kinetic data. The artificial neural network has predicted classic potential barrier for hydrogen atom in reactions with hydrocarbons in solution at a temperature of 298 K with satisfactory accuracy. Special attention is placed the use of fuzzy knowledge base to predict the classical potential barrier and to calculate the rate constants of the bimolecular radical reactions of phenyl radical with hydrocarbons in the liquid phase on the experimental data. The hybrid algorithm of calculation of rate constants of bimolecular radical reactions on the experimental thermochemical data and an empirical index of the reactionary center is offered. This algorithm uses a fuzzy knowledge base or an artificial neural network for prediction of classical potential barrier of bimolecular radical reactions at some temperature, a database of experimental characteristics of reaction and Arrhenius's formula for calculation of rate constant. Results of prediction of the classical potential barrier are discussed.

Key-Words: chemical kinetic, thermochemistry**,** data warehouse, data science analytic, expert system, artificial neural network, fuzzy logic. radical reaction, activation energy, classical potential barrier, reactionary center.

1 Introduction

Chemical society watches closely the development of methods of artificial intelligence and applies them in the research [1-3]. In particular it studies the use of applied methods of artificial intelligence for extraction and production of new knowledge on the kinetic and thermochemical data from the electronic collections.

Creation of electronic data banks and databases on kinetic and thermochemical data [4-9] gave researchers new opportunities for the system analysis of the experimental data and data obtained from the mathematical modeling (data discovery and data mining from database) to derive new data on the rate constants of radical reactions and thermochemical properties of molecules.

Authors of the kinetic and thermochemical databases initially developed computer systems to obtain new knowledge – the reaction activation energies, enthalpies of formation and molecule bond strengths not known previously. The use of mathematical modeling or artificial intelligence elements (development of knowledge models, mathematical deduction, expert systems, artificial neural networks) and application of methods of fuzzy modeling was supposed in the following evolution of the created data banks and databases. Development of such approach was the core of the process of intellectual data analysis applied in organic synthesis, chemical kinetics and thermochemisty.

Thus the problem of extraction of new knowledge from the electronic sources of chemical kinetics and thermochemistry called for the development of use of the new information models for presentation, conservation, selection and production of new knowledge base of the intellectual data analysis.

The purpose of this article is to give a review of the most important open sources of kinetic and thermochemical information and to discuss the most promising approaches to extract new knowledge from them.

Particularly, one of the purposes of this article is to develop and discuss the use of the feed-forward artificial neural network, which have learned using a set of the experimental thermochemical and kinetic data, to predict the classical potential barrier for hydrogen atom in reactions with hydrocarbons in solution at a temperature of 298 K.

Other of the purposes of this paper is the use of the Mamdani method [10] to identify the empirical dependence of classical potential barrier of reactions between phenyl radicals and hydrocarbons in the liquid phase based on experimental kinetic data.

Section 2 contains short review on application of data warehouse technology for creating electronic collections of the kinetic and thermochemical data.

Software architecture of data warehouse for production of new knowledge in kinetics and thermochemistry (based on the example of storing rate constants of liquid-phase radical reactions and bond dissociation energies of organic compounds) is given.

Section 3.1 describes in short empirical models of radical bimolecular reactions of abstraction that give the expert a chance to classify such reactions based on the empirical index of the reactionary center obtained by statistical methods.

Section 3.2 describes how such classification opens a way to build hybrid production expert systems based on the decision tables. The decision table lets successfully use multi-agent technologies in realization of the expert systems built into the portal.

Section 3.3 shows an example of the use of artificial neural networks for approximation of a classical potential barrier in reactions of hydrogen atom with hydrocarbons in solutions.

Section 3.4 shows an example of the use of fuzzy knowledge base for approximation of a classical potential barrier in reactions of phenyl radical with hydrocarbons in solutions, using the Mamdani algorithm.

Algorithm is proposed that combines approaches of artificial neural networks and fuzzy databases for prediction of a classical potential barrier of radical bimolecular reactions of abstraction.

As conclusion, the advantages and disadvantages of the approaches examined above are discussed.

2 Problem Formulation

One of the approaches to representation of domain knowledge in electronic resources designed for data analysis is the use of data warehouse technology [11]. The eye of the problem is in integration of the problem-oriented databases on chemical kinetics and thermochemistry based on data warehouse and data mart technology.

In the NIST report of 2004 [12] Allison T.C. mentions the possibility to use data warehouse technology for the database [5]. In 2004 with the participation of NIST employees the data warehouse technology was used in the NIST-PrIMe Warehouse project [13], in the same place the question of integration of the kinetic and thermochemical data in the data warehouse for modeling of combustion processes was introduced.

In 2001 in the IPCP RAS in the context of development of data bank on rate constants of liquid-phase radical reactions and the database of bond dissociation energies of organic compounds a research of the use of data warehouse technology in physical chemistry of radical reactions was started. The use of data warehouse technology to represent data in electronic data collections on liquid-phase radical kinetics and thermochemistry was proposed in [14].

Data warehouse of this system contains empirical and calculated facts, production rules and procedures for calculating, and in conjunction with expert systems constitutes a virtual subsystem for production of new professional knowledge (the constants of rate and energies of radical reactions activation, energies of molecular bonds dissociation).

Data warehouse as a component of new professional knowledge production includes (Fig. 1):

- Exploration Data Warehouse, which contains experimental data on the reactivity of radical reactions in the liquid phase;
- integrated expert system to manage the evaluation of reactivity of reagents radical reactions, representing a combination of intellectual and reactive agents);
- intellectual agent to evaluate the rate constants and activation energies of reactions in the liquid and gas phase;
- web service, through which the call of trained artificial neural networks is performed to predict the values of rate constants and activation energies of liquid-phase radical reactions of certain classes;
- derivative data warehouse containing the calculated data on the reactivity of radical reactions in liquid and gas phases;
- integrated expert system to evaluate the bond dissociation energy of molecular by kinetic data of radical reactions;

- data warehouse of bonds dissociation energies of organic molecules, which can be supplemented with new data as a result of work of expert system for evaluation of bonds dissociation energy of molecular by kinetic data of radical reactions;
- thesaurus of key terms and concepts in the subject domain;
- thesaurus of descriptions of algorithms and procedures for predicting physicochemical characteristics of molecules;
- knowledge base containing production rules and facts, which are used by integrated expert systems.

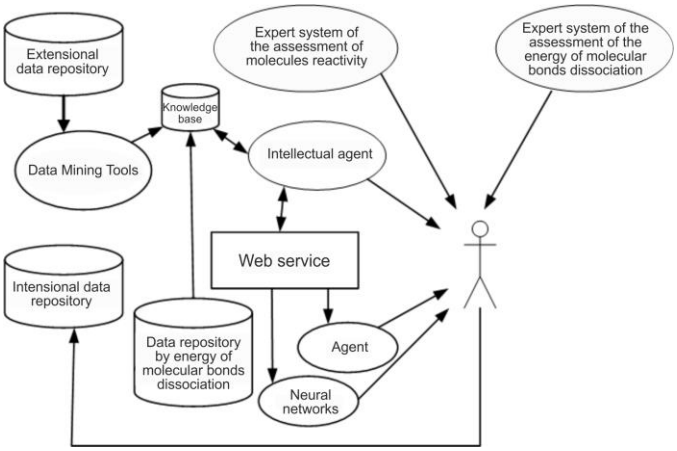

Fig. 1. Data warehouse of subject-oriented science intelligence system in physical chemistry of radical reactions.

As a result of the work of the system users the data warehouse is updated with new professional knowledge. As the mechanisms of production of new knowledge in this present system the expert systems integrated into the portal, trained artificial neural networks and intellectual agents are used.

3 Problem Solution

3.1 Empirical models of free radical reactions

Experimentally the activation energy (E) or a classical potential barrier (E_e) determines the reactivity of organic molecules in the bimolecular radical reaction:

$$E_e = E - 0.5(hL\nu_i - RT) \quad (1)$$

where ν_i is a frequency of the stretching vibrations for the bond being broken, R is the gas constant, h is the Planck constant, L is the Avogadro constant, and T is the reaction temperature (K).

In [15,16] proposed empirical models of elementary bimolecular radical reactions of abstraction, which allows to build non-linear correlations between the classical potential barrier radical bimolecular reaction and thermochemical properties of reactants (nonlinear correlation):

a. non-linear correlation [15]:

$$br_e = \alpha\sqrt{E_e - \Delta H_e} - \sqrt{E_e} \quad (2)$$

b. non-linear correlation [16]:

$$br_e = D_{ei}^{1/2}\ln\left(\frac{D_{ei}^{1/2}}{D_{ei}^{1/2} - E_e^{1/2}}\right) + \alpha D_{ef}^{1/2}\ln\left(\frac{D_{ef}^{1/2}}{D_{ef}^{1/2} - \left(E_e - \Delta H_e\right)^{1/2}}\right) \quad (3)$$

where $\Delta H_e = D_i - D_f + 0.5\,(hL\nu_i - hL\nu_f)$ is the enthalpy of reaction, including the difference between the zero-point energies to rupture or form the bonds, ν_i and ν_f are the frequencies of the stretching vibrations for the bond being broken and the bond being formed, respectively; D_i and D_f are the dissociation energies for the breaking and the forming bond, respectively; $D_{ei} = D_i + 0.5hL\nu_i$,, $D_{ef} = D_f + 0{,}5hL\nu_f$; μ_i and μ_f are the reduced masses of atoms for the breaking and the forming bond, respectively; the coefficients $b = \pi(2\mu_i)^{1/2}\nu_i$, $b_f = \pi(2\mu_f)^{1/2}\nu_f$ and $\alpha = b/b_f$, the distance r_e, which the abstracted atom is displaced in the course of the reaction.

We will designate br_e empirical index of the reactionary center of the radical reaction.

These models are the basis for the developed approaches to extraction of new knowledge from subject-oriented data warehouse of the radical reactions given on physical chemistry.

3.2 Expert Systems

Research review on development of expert systems in chemistry is given in the monograph [17].

The expert system for prediction of reactivity of organic molecules in liquid and gas phase radical reactions of abstraction based on the experimental kinetic and thermochemical data and the expert system for estimation of bond dissociation energies in liquid phase radical reactions based on the kinetic data were developed in the IPCP RAS [6]. As a result the estimation of more than 1000 bond

dissociation energies in complex organic compounds was made.

Fig. 2 schematically shows the multi-agent software architecture integrated into the portal of expert system.

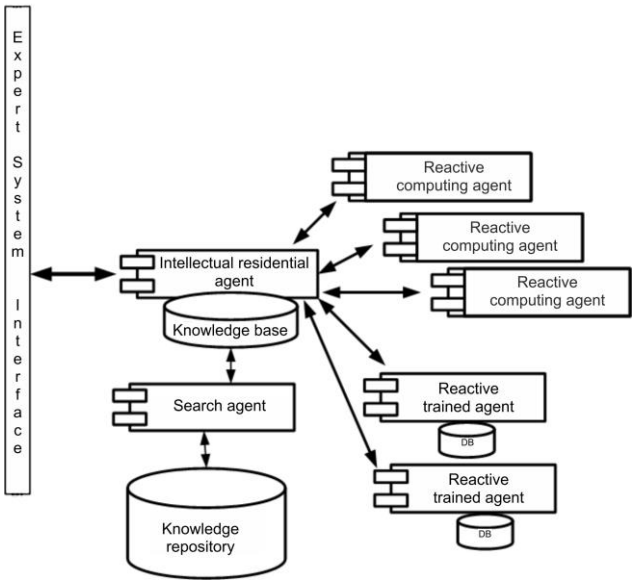

Fig. 2. The multi-agent software architecture integrated into the portal of expert system.

At Fig. 2 software agents operate in a simple model "Query-Answer-Agreement." After receiving the input data a survey of resident agents is performed. Based on the responses received the decision is made to what agent should be delegated with performing actions referred to in the expert system. After interviewing the agents the matrix of responses is formed and based on its analysis it is decided which agent shall be given the task to be solved. Under certain conditions, the solution of the task may be given to the two agents.

The approach that we use to form knowledge base of the expert system for prediction of reactivity is based on the statistical manipulation of the experimental kinetic data, which is partially typical for the EROS expert system too [18].

Only a small number of publications are devoted to problems of the use of expert systems for prediction of molecule reactivity in radical reactions. Among those the works made under the guidance of Prof. J. Gasteiger within the project of development of the EROS expert system should be marked [18]. It should be observed that this expert system is more directed towards the organic synthesis support, rather than the prediction of reactivity of organic molecules in radical reactions.

Molecule reactivity can be estimated in the SPARC expert system [19], which is available in the Internet. This program uses quantum chemical methods by calculation of the rate constants reaction, equilibrium constants and enthalpies of molecule formation.

Thus, the expert system technology was used for the prediction of the bond dissociation energies of molecules, classical potential barrier and enthalpies of free radical formation.

3.3 Artificial neural networks

Currently artificial neural networks are widely used by solution of the applied problems of automatic research data processing in interdisciplinary research in chemistry [20, 21]. One of such topical applied problems is the prediction of molecule reactivity in chemical reactions (activation energies and rate constants), bond dissociation energies and enthalpies of molecule formation.

A series of author's works [22-25] is devoted to research of the use of artificial neural network trained on the experimental selection from databases [6], for estimation of rate constants and activation energies of radical reactions.

A series of works is devoted to the use of artificial neural network N technology for prediction of bond dissociation energies [26] (based on the experimental data from databases [6]) and [27] (with the use of the set of chemical descriptors).

We can assume (2) the existence of a non-linear dependence of the classical potential barrier of the reactions of hydrocarbons with hydrogen atom in solution from the thermochemical characteristics of reagents and kinetic characteristics of this reaction:

$$E_e = f(D_{ei}, D_{ef}, br_e, \alpha) \tag{4}$$

and approximate of this dependence by artificial neural network.

The experimental data set consist of 53 radical reactions of the hydrogen atom with a variety of hydrocarbons in solution, of which 8 were in test set. The rate constants are taken from the database object-oriented system of scientific knowledge in the physical chemistry of radical reactions [6], the dissociation energies of the C-H bonds are used from [28,29].

The activation energy in the bimolecular reactions of hydrocarbons with hydrogen in solution was calculated by the formula:

$$E = -RT \ln(\frac{k}{nA_0}) \tag{5}$$

where A_0 is the pre-exponential factor per equireactive bond in the molecule and equal to 10^{11}

L mol^{-1} s^{-1}, n is the number of equireactive bonds in the molecule, k is the reaction rate constant, E is the activation energy from (1).

For a given set D_f = 436 kJ/mol [28] and α = 0.905 are permanent. Reactions were selected which occurred in solution. Therefore, the dependence (4) becomes:

$$E_e = \varphi(D_{ei}, br_{e\_ind}) \quad (6)$$

where br_{e_ind} is the empirical index of the reactionary center, calculated by the statistical method according to the formula (2) for groups of compounds with similar centers.

Empirical indexes of the reactionary centers for this set are shown in Table 1.

Table 1. Empirical indexes of the reactionary centers.

Classes of Organic Compounds	The radical reaction center	An empirical index $(kJ/mol)^{0.5}$
Alkanes	-C˚$(CH_3)_2$	15.88±0.14
	-C˚HCH_3	15.82±0.10
	cyclo-[C˚$H(CH_2)_k$]	16.17±0.16
Alkenes	=CHC˚$H(CH_2)_kCH_3$	17.31±0.04
	cyclo-[CH=CHC˚$H(CH_2)_k$]	17.96±0.06
Arenas	-C_6H_4C˚H_2	15.83±0.08
	C_6H_5C˚$H(CH_2)_kCH_3$	16.78±0.19
	-C_6H_4C˚$(CH_3)_2$	17.57
Alcohols	C˚H_2OH	14.38
	-C˚HOH	13.90
	-CH_2C˚OH	14.33
	=C(OH)C˚H_2	15.03
Ketones	-C(O)C˚H_2	15.00
	-C(O)C˚HCH_3	15.60±0.01
	-C(O)C˚$(CH_3)_2$	16.07
	cyclo-[C(O)C˚$H(CH_2)_k$]	15.40
Ethers	-OC˚$(CH_3)_2$	14.34
	cyclo-[OC˚$H(CH_2)_k$]	14.36
	cyclo-[OC˚$H(CH_2)_kO(CH_2)_m$]	14.52
Acids	C˚H_2C(O)OH	14.68
	-C˚HC(O)OH	15.29±0..02
	=C˚C(O)OH	15.58

To approximate the dependence (6) was used feed-forward artificial neural network with two inputs, two hidden inner layers with 7 and 5 neurons, respectively, and one output. In the neural network, each neuron has an activation function [30]:

$$\psi(x) = \frac{1}{1+e^{-x}} \quad (7)$$

An artificial neural network was trained for 52000 iterations using the error back propagation algorithm [30]. The results of training on the test set are shown in Table. 2.

Table 2. Training results of artificial neural network

Compound	E_e	E_{ANN}	E_e - E_{ANN}
	kJ/mol		
$CH_3(CH_2)_4CH_3$	54.74	54.05	0.69
cyclo-[$CH(CH_3)(CH_2)_4$]	49.37	50.40	-1.03
CH_2=$CH(CH_2)_3CH_3$	43.61	44.50	-0.89
$C_6H_5CH_2CH_3$	45.59	45.95	-0.36
CH_3CH_2OH	34.17	34.20	-0.03
$CH_3C(O)CH_2CH_3$	44.78	44.71	0.07
cyclo-[$O(CH_2)_4$]	34.59	34.52	0.07
$CH_3CH_2C(O)OH$	44.66	44.09	0.57

As seen from Table 2, developed artificial neural network with high accuracy predict the classical reaction barrier (activation energy) E_{ANN} in reactions of hydrocarbons with hydrogen in solution. The mean square error for the entire set is 0.80±0.75 kJ/mol. Maximum absolute error calculation classical potential barrier is ±2.83 kJ/mol, the minimum is ±0.02 kJ/mol, which is in good agreement with the experimental methods of determining the accuracy of the activation energy for such reactions ±4 kJ/mol.

Thus the possibility to use artificial neural network technology for prediction of bond dissociation energies of molecules, classical potential barriers and enthalpies of molecule formation based on the experimental selections was shown.

3.4 Fuzzy Knowledge Base

Recently high attention is emphasized on problems of the use of fuzzy logic methods and fuzzy knowledge bases in chemistry [31]. However there are currently only a few papers dedicated to the use of fuzzy logic in chemical kinetics and thermochemisty [see the review in 22].

A work on determination of bond dissociation energies with the use of fuzzy knowledge base and fuzzy neural networks [26] based on selections from databases [6] should be noted.

Thus, we can assume the existence of nonlinear dependence of the classical potential barrier bimolecular reaction of radical abstraction from the

thermochemical characteristics of reagents and kinetic characteristics of such a reaction:

$$E_e = f(D_{ei}, D_{ef}, br_e, \alpha) \quad (8)$$

and approximate of this dependence.

To determine the parameter space of identification (input data) of activation energy of radical reactions E_e we use the correlation ratio (2).

Here we consider experimental sample of the reactions:

$$Ph^{\cdot} + RH \rightarrow PhH + R^{\cdot}$$

The experimental sample of 97 reactions of the phenyl radical with various hydrocarbons is obtained from the database on rate constants of liquid phase radical reactions [6], which 12 were the control sample. The dissociation energy of the C-H bonds is taken from [28]. For this sample D_f = 474 kJ/mol [28] and α = 0.945 are constant. Therefore, the dependence (8) takes the form:

$$E_e = \varphi(D_{ei}, br_e) \quad (9)$$

Then the analysis of the kinetic parameter values br_e was performed for various reaction centers and the experimental index of the reaction center br_{e_ind} was calculated, as shown in Table 3. We used br_{e_ind} calculated for the certain reaction centers for prediction of the classical potential barrier.

Table 3. Empirical indexes of the reaction centers.

The class of compounds	The radical reaction center	An empirical index $(kJ/mol)^{0.5}$
Alkanes	$-CH_2C^{\cdot}H_2$	16.96
	$-CH_2C^{\cdot}HC(CH_3)$	16.88
	$(CH_3)_3C^{\cdot}$	17.48
	cyclo-$[C^{\cdot}H(CH_2)_k]$	17.10
	cyclo-$[C^{\cdot}(CH_3)(CH_2)_k]$	16.85
Alkenes	$CH_2=CHC^{\cdot}HCH_2-$	18.39
	$CH_2=C(CH_3)C^{\cdot}H-$	18.45
	$-C=C(CH_3)C^{\cdot}H_2$	18.53
	cyclo-$[CH=CHC^{\cdot}H(CH_2)_k]$	18.38
Alkynes	$-C\equiv CC^{\cdot}H_2$	18.51
	$-C\equiv CC^{\cdot}H-$	18.51
Arenes	$PhC^{\cdot}H_2$	17.83
	$PhC(CH_3)_2C^{\cdot}H_2$	16.85
	$X\text{-}C_6H_4C^{\cdot}H_2$	17.8
	$(CH_3)_5C_6C^{\cdot}H_2$	17.87
	$(PH)_2C^{\cdot}H$	18.36
	$(Ph)_3C^{\cdot}$	17.27
Alcohols	$C^{\cdot}H_2OH$	17.29
	$CH_3C^{\cdot}HOH$	15.82
	$C^{\cdot}H_2(CH_3)_2COH$	16.05
	$(CH_3)_2C^{\cdot}OH$	15.89
	$PhC^{\cdot}HOH$	16.01
Aldehydes	$CH_3C^{\cdot}(O)$	17.60
	$PhC(CH_3)_2C^{\cdot}(O)$	17.61
	$X\text{-}C_6H_4C^{\cdot}(O)$	17.87
Ketones	$-C(O)C^{\cdot}H_2$	17.38
	$PhC(O)C^{\cdot}H2$	16.80
	$-CH_2OC^{\cdot}H-$	17.27
	$=CHOC^{\cdot}(CH_3)_2$	17.44
Ethers	$-OC^{\cdot}H2$	17.21
	cyclo-$[OC^{\cdot}H_2(CH_2)_k]$	16.44
	$PhOC^{\cdot}H_2$	17.81
Esters	$-C(O)OC^{\cdot}H_2$	17.84
	$C^{\cdot}H[C(O)O-]_2$	17.92
	$PhC(O)OC^{\cdot}H_2$	17.01
Nitrites	$C^{\cdot}H2CN$	17.86
	$-C^{\cdot}HCN$	17.86
	$PhC(CH_3)(C^{\cdot}H_2)CN$	17.05
Nitro compounds	$C^{\cdot}H_2NO_2$	17.76
	$-C^{\cdot}HNO_2$	16.76
	$=C^{\cdot}NO_2$	17.76

In this case, the kinetic parameter br_e used as an experimental index of the reaction center of radical bimolecular reaction.

Fuzzy knowledge base was built by experts and includes 85 linguistic rules of the form:

$\boldsymbol{R}_i$ **IF** D_{ei} about 433.3257 **AND** br_e about 17.48 **THEN** E_e about 50,077

The membership function of fuzzy terms was:

$$\mu^G(x) = \frac{1}{1 + \left(\frac{x-b}{c}\right)^2} \quad (10)$$

To approximate the values of the classical potential barrier the Mamdani's fuzzy inference method was used based on the using the matching degree to which they belong to each of the fuzzy rule via membership functions, equal the real number α_i, characterizing the degree of membership of input $A_1', A_2', ..., A_n'$ to fuzzy sets $A_{i1}, A_{i2}, ..., A_{in}$ in the background of the i-th rule

$$\alpha_i = \min_{j=1}^{n} \left[\max_{X_i}(A_j^{'}(x_j) \wedge A_{ij}(x_j)) \right]$$

where X_j is the domain of the variables ($x_1 = D_{ei}$ и $x_2 = br_e$).

In our case, the algorithm output consists of the following steps:

- For each rule R_i, i = 1,2, ..., 85 is computed matching degree of each rules

$$\alpha_i = \min \left[\max_{D_{ei}}(A_1^{'}(x_1) \wedge A_{i1}(x_1)), \max_{be_e}(A_2^{'}(x_2) \wedge A_{i2}(x_2)) \right]$$

- For each rule defined individual outputs (Mamdani implication)

$$B_i^{'}(y) = \min(\alpha_i, B_i(y))$$

- Calculate the aggregation of individual outputs

$$B^{'}(y) = \max(B_1^{'}(y), B_2^{'}, \ldots, B_m^{'}(y))$$

- The resulting fuzzy set is converted to a clear value by defuzzification technique "center of gravity"

$$\bar{y} = \frac{\sum y_i B(y_i)}{\sum B(y_i)}$$

As a result, when D_{ei} = 432.326 and br_e = 17.48 the system produces E_e = 49.97 in well agreement with the experimental value of 50.07 kJ/mol.

Table 4 shows a comparison of the values of classical potential barrier activation of reactions for phenyl radicals with hydrocarbons E_e, obtained with the fuzzy knowledge base, and the calculated ones from the experimental values of the activation energy of these reactions E_{exp}.

Table 4. Comparison of the values of classical potential barrier activation of reactions for phenyl radicals with hydrocarbons.

Hydrocarbons	**E_e**	**E_{exp}**	**Δ**
	kJ/mol		
$(CH_3)_2CHCH(CH_3)_2$	37.28	37.24	0.04
$(CH_3)_2CH(CH_2)_2CH_3$	38.19	38.16	0.03
$CH_3CH_2CH(CH_3)(CH_2)_2CH_3$	37.58	37.54	0.04
$(CH_3)_3CCH_2CH_3$	42.93	42.90	0.03
$(CH_3)_4C$	50.08	50.06	0.02
$(CH_3)_2CHCH_2CH(CH_3)_2$	40.79	40.77	0.02
$(CH_3)_2CHCH_2CH_2CH(CH_3)_2$	39.02	38.98	0.04
$(CH_3)_3CCH_2CH(CH_3)_2$	40.75	40.71	0.04
$(CH_3)_3CCH_2CH_2CH(CH_3)_2$	38.35	38.32	0.03
$C_6H_5OCH_3$	40.98	40.94	0.04
$CH_3C(O)OH$	47.60	47.56	0.04
$C_6H_5CH(CH_3)_2$	31.59	31.55	0.04

As can be seen in Table 2, there is good agreement between the experimental and obtained by fuzzy inference values of classical potential barrier. The average error is 0.034±0.008 kJ/mol.

The values of classical potential barrier allow to calculate the activation energy (1) and the rate constant of radical reactions by the formula:

$$k = nA_0 \exp(-E/RT)$$

where: A_0 is the collision frequency per one equireactive bond, n is the number of equireactive bonds in a molecule.

Figure 3 show a basic scheme of the algorithm of calculation of rate constant of bimolecular radical reaction on the experimental data with using fuzzy knowledge base or artificial neural network.

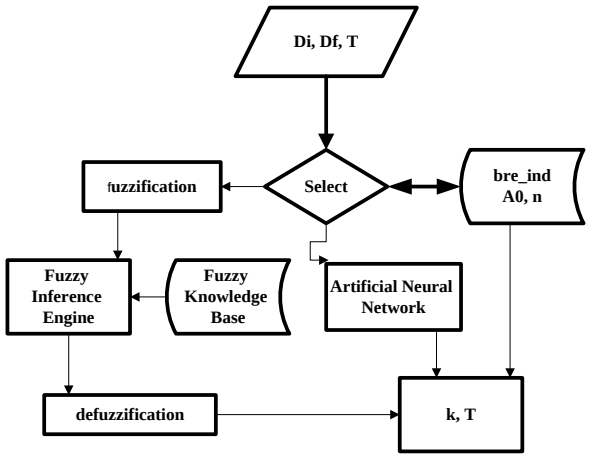

Fig. 3. The algorithm of calculation of rate constant of bimolecular radical reaction on the experimental data with using fuzzy knowledge base or artificial neural network.

3.5 Discussion

Using expert systems for prediction of a classical potential barrier of radical bimolecular reactions and bond dissociation energies based on the

experimental kinetic and thermochemical data has one important advantage. The obtained estimations of dissociation energy and classical potential barrier are in accordance with the kinetics of the elementary bimolecular reactions. Using the developed expert system allowed to increase by a third the number of known values of dissociation energies of C-H-bonds. Further it will be also possible to fill the data warehouses by experts and users through the Internet. The disadvantage of the approach is the lack of necessary set of the experimental kinetic and thermochemical data on radical reactions.

Using artificial neural networks for estimation of classical potential barrier gives advantage in approximation of the dependence of classical potential barrier from the enthalpy of reaction and properties of reagents, for such dependence may be specified implicitly (3).

Main disadvantages of using artificial neural networks for prediction of a classical potential barrier of radical bimolecular reactions of abstraction are:

1) the choice of the variable space (for example, if one examines the reaction of hydrogen atom with hydrocarbons in water, one has to take into account pH of the solution, the effect of which is non-linear),

2) the ambiguity in the training set (there may be several identical reactions with various activation energies obtained by different authors; this fact may cause the divergence of the network in training if the values of rate constants differ significantly).

Using fuzzy neural networks for estimation of classical potential barrier in radical bimolecular reactions is the most promising approach, although there is an ambiguity problems mentioned above. However it may be solved either by adding another linguistic variable, or by averaging if values of classical potential barrier don't differ much. I.e. it is practically possible to use fuzzy neural approximator for ambiguous functions.

4 Conclusion

In this paper a review of the most important open sources of kinetic and thermochemical information and of the main approaches to analysis and extraction of the new knowledge from them is given.

Developed artificial neural network predicts the value of the classical potential barrier for hydrogen atom in reactions with hydrocarbons in solution with an accuracy of about 1 - 3 kJ/mol, which agrees well with an accuracy of determining the activation energy of these reactions experimental methods.

On the example of reactions between substituted phenyl radicals and hydrocarbons an attempt was made to identify the dependence of classical potential barrier of radical reactions by the fuzzy knowledge base built on basis of quantitative and qualitative parameters.

Using a fuzzy knowledge base built by experts, and Mamdani's fuzzy inference method using membership functions (10) produces a good approximation of the values of the classical potential barrier for phenyl radical reactions with hydrocarbons.

5 Acknowledgement

The study was partially supported by RFBR, research project No. 15-07-08645.

References:

[1] Brereton, R. G., *Applied Chemometrics for Scientists*, John Wiley & Sons Ltd, 2007.

[2] Banville D.L., Chemical Information Mining, *Facilitating Literature-Based Discovery*, CRC Press, Taylor & Francis Group., 2009.

[3] Gasteiger, J., Engel, T., *Chemoinformatics. A Texbook*, Wiley-VCH GmbH & co.KGaA, 2003.

[4] Manion, J.A., Huie, R.E., Levin, R.D., Burgess, D.R., Jr., Orkin, V.L., Tsang, W., McGivern, W.S., Hudgens, J.W., Knyazev, V.D, Atkinson, D.B.,. Chai, E,. Tereza, A.M, Lin, C.-Y., Allison, T.C., Mallard, W.G., Westley, F., Herron, J.T., Hampson, R.F., Frizzell, D.H., NIST *Chemical Kinetics Database, NIST Standard Reference Database 17, Version 7.0 (Web Version), Release 1.4.3, Data version 2008.12*, National Institute of Standards and Technology, Gaithersburg, Maryland, 20899-8320. From: http://kinetics.nist.gov/

[5] X5. *NDRL/NIST Solution Kinetics Database on the Web*. From: http://kinetics.nist.gov/solution/

[6] Tumanov, V., Gaifullin, G., Subject-oriented Science Iintelligent System on Physical Chemistry of Radical Reactions, *Modern Advances in Intelligent Systems and Tools*, Vol. 431, 2012, pp. 121-126.

[7] Gurvich, L.V., Iorish, V.S., Chekhovskoi, D.V., Yungman, V.S. *IVTANTHERMO - A Thermodynamic Database and Software System for the Personal Computer*, User's Guide. Boca Raton, CRC Press, Inc., 1993.

[8] Teixeira, A.L, Santos, R.C., Leal, J.P., Simoes, J.A.M., Falcao A.O., ThermInfo: *Collecting, Retrieving, and Estimating Reliable Thermochemical Data*, From: http://arxiv.org/ftp/arxiv/papers/1302/1302.0710.pdf.

[9] *NIST Chemistry WebBook. NIST Standard Reference Database Number 69*. From: http://webbook.nist.gov/chemistry/

[10] Yager, R., Filev D.P., *Essentials of Fuzzy Modeling and Control*, USA: John Wiley & Sons, 1994.

[11] Thierauf, R.J., *Effective Business Intelligence Systems*, Westport. Quorum Books, 2001.

[12] Allison, T., *The NIST Chemical Kinetics Database*/ From: http://www.auburn.edu/~stanbdm/Alison27s20report/files/NISTKineticsDB-TomAllison.pdf.

[13] Frenklach M., *PrIMe Update: Workflow 2.0 and Automated UQ Predictive Modeling of Hydrogen Combustion*. From: http://kinetics.nist.gov/RealFuels/macccr/macccr2011/MACCCR_2011_Frenklach.pdf

[14] Tumanov, V.E., Data Warehousing and Data Mining in Thermochemistry of Free Radical Reactions, *Fourth Winter Symposium on Chemometrics "Modern Methods of Data Analysis"*,Chernogolovka, 2005, pp. 28-29.

[15] Denisov, E.T., New Empirical Models of Radical Abstraction Reactions, *Russian Chemical Reviews*, Vol. 66, No. 10, 1997, pp. 859–876.

[16] Denisov, E.T., Tumanov, V.E., Transition-State Model as the Result of 2 Morse Terms Crossing Applied to Atomic-Hydrogen Reactions, *Zurnal fiziceskoj himii*, Vol. 68, No. 4, 1994, pp. 719-725.

[17] Hemmer, M.C., *Expert System in Chemistry Research*, CRC Press, Taylor & Francis Group, 2008.

[18] Gasteinger, J., Jochum, C., EROS - A Compute Programm for Generating Sequences of Reactions, *Topics Curr. Chem.*, Vol. 74, 1978, pp. 93-126.

[19] Long, J. M., *SPARC: An Expert System for Estimating Physical and Chemical Reactivity: User Manual for Calculating Ionization pKa*, Ecosystems Research Division, National Exposure Laboratory, U.S. Environmental Protection Agency: Athens, GA, 1996.

[20] Himmelblau ,D.M., Applications of Artificial Neural Networks in Chemical Engineering, *Korean Z Chem. Eng*., Vol. 17, No. 4. 2000, pp. 373-392.

[21] Gasteiger, J., Zupan, J., Neural Networks in Chemistry, *Angev. Chem. Int. Ed. Engl*., Vol. 32, 1993, pp. 503-527.

[22] Tumanov, V.E., Application of the Artificial Neural Networks and Fuzzy Logic for the Prediction of Reactivity of Molecules in Radical Reactions, *Computational Problems in Engineering. Series: Lecture Notes in Electrical Engineering*, Switzerland: Springer International Publishing, Vol. 307, 2014, pp. 261-269.

[23] Tumanov, V.E., Hybrid Algorithm of Application of Artificial Neuronets for an Evaluation of Rate Constants of Radical Bimolecular Reactions, *Advances in Neural Networks, Fuzzy Systems and Artificial Intelligence. Recent Advances in Computer Engineering Series*, WSEAS Press. Gdansk, Poland, Vol. 21, 2014, pp. 58-61.

[24] Tumanov ,V.E., Gaifullin, B. N., Evaluation of the Rate Constants of Reactions of Phenyl Radicals with Hydrocarbons with the Use of Artificial Neural Network, *Current Approaches in Applied Artificial Intelligence*, Springer International Publishing Switzerland, Vol. 9101, 2015, pp. 394–403.

[25] Tumanov, V., Gaifullin, G., Application of the Artificial Neural Networks for the Prediction of Reactivity of Molecules in Radical Reactions, *Mathematical Modelling and Simulation in Applied Sciences. Proc. of the 3rd International Conference on Energy, Environment, Devices, Systems, Communications, Computers* (INEEE '12). Rovaniemi, Finland, 2012, pp. 62-65.

[26] Lazarev, D. Yu., Tumanov, V.E., Prediction of Bond Dissociation Energies of Organic Molecules by Kinetic Data of the Radical Reactions Using the Fuzzy Knowledge Base, *BIT,s 3rd Annual Conference and EXPO of AnalytiX-2014*, Dalian, China, 2014, p. 379.

[27] Dutot, A.-L., Rude, J., Aumont, B., Neural Network Method to Estimate the Aqueous Rate Constants for the OH Reactions with Organic Compounds, *Atmospheric Environment*, Vol. 37, 2003, pp. 269–276.

[28] Luo, Y.-R., *Comprehensive Handbook of Chemical Bond Energies*, London - New York: CRC Press, Boca. 2007.

[29] Denisov, E.T., Tumanov, V.E., Estimation of the Bond Dissociation Energies from the Kinetic Characteristics of Liquid-Phase Radical Reactions, *Russian Chemical Reviews*, Vol. 74, No. 9, 2005, pp. 825-858.

[30] Larose, D.T., *Data Mining Methods and Models*, John Wiley & Sons, 2006.
[31] Rouvray, D.H., *Fuzzy Logic in Chemistry*, Academic Press, 1997.