Available online at www.sciencedirect.com
ScienceDirect
Thermochimica Acta 452 (2007) 140–148
thermochimica acta
www.elsevier.com/locate/tca

An approach for interpreting thermogravimetric profiles using artificial intelligence

G.P. Voga, J.C. Belchior*

Departamento de Química-ICEx, Universidade Federal de Minas Gerais, Pampulha, 31.270-901 Belo Horizonte, MG, Brazil

Received 5 September 2006; received in revised form 16 October 2006; accepted 26 October 2006
Available online 10 November 2006

Abstract

Expert systems are proposed in order to analyze thermogravimetric profiles. The methodology showed to be an efficient tool as a complementary procedure for experimentalists in the area of themogravimetric analysis. Several simple and complex molecules were used to validate the developed methodology. The advantage of the actual approach can be attributed to its efficient error minimization for each weigh loss and global process. In addition, the proposed expert system presents a low computational demanding. The validation analysis using compounds such as salicylato (amine) Co (III) complexes, norfloxacin complexes of manganese (II) and cobalt (II), and calcium borate produced an average confidence interval of about 2%.

Keywords: Thermogravimetric analysis; Expert system; TG curve; Artificial intelligence

1. Introduction

The thermogravimetric (TG) analysis correlates the weight loss as a function of temperature. The experimental measurements allow evaluation of thermal stability, reaction rate, reaction mechanisms and the determination of sample composition [1,2]. In general, the thermogravimetric device is constituted by: (i) balance (thermo balance), (ii) furnace, (iii) furnace inert or reactive atmosphere and (iv) microcomputer for instrumental control, data acquisition and presentation [2,3].

The data interpretation is performed by an analyst and therefore it is required experimental expert analysts for thermogravimetric analyses. The human expert has an objective of attributing the weight losses for specified sample components (or compounds) released during the analysis. In the interpretation process, the analyst should provide all decompositions for each weight loss. However, such a task may be considered exhaustive for a human expert due to the number of possible decompositions for a specified system with particular attention to complex molecules. Usually, the amount of data (possible different compounds) is large enough to produce incorrect interpretations. Nevertheless, simple compounds can be analyzed, although uncorrected interpretations can be found. For example, the decomposition of acetylsalicylic acid was previously interpreted [4] as $H_2O + CO + CO_2$. However, the correct analysis determined in Ref. [5] showed to be a loss of acetic acid molecule.

Artificial intelligence techniques such as neural networks are widely applied in different branches of sciences. Particularly in chemistry we have recently proposed an approach to deal with drug controlled release [6] and blood plasma simulations [7]. Artificial intelligent systems based on mathematical techniques [8] may be also useful for analyzing thermogravimetric curves such as the determination of thermal decomposition models, searching patterns in TG curves, acquisition of thermal properties and the determination of instrumental parameters. For example, artificial neural networks have been applied for analyzing the influence of several models for thermal decompositions [9]. Similarly, multivariate statistical methods are also frequently applied in order to analyze thermogravimetric data [10–13]. Therefore, these approaches have shown to be important tools in order to contribute to the TG analysis. Particularly, the experimental parameters such as heating rate, final temperature and gas pressure can produce a great influence in the quantitative and qualitative results of thermal analysis. Mainly, these parameters are important for high-resolution analysis [14].

* Corresponding author. Tel.: +55 31 3499 5775; fax: +55 31 3499 5700.
E-mail address: jadson@ufmg.br (J.C. Belchior).

0040-6031/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.tca.2006.10.017

The determination of experimental conditions is able to accomplish computational systems, mainly based on expert system techniques [15]. The fuzzy logic approach, for instance, has been applied to determine the relationship between molecular fragments and spectral parameters using database of molecular spectra [16–19].

This work will deal with an algorithm that composes an expert system. The approach is proposed to be considered as an alternative or complementary tool for analyzing thermogravimetric data. The system is based on heuristic rules used for interpreting thermogravimetric data and acquisition of new knowledge, similar to human experts. For the latter case, the new acquisition may be added as new data or new rules in the proposed expert system. As will be discussed, the application of this expert system can contribute to mostly reduce errors and also to provide standard procedures for the interpretation process, particularly dealing with TG data.

2. Methodology—expert system

An expert system can be defined considering a system based on knowledge that is capable of solving problems in which human experience is generally necessary. Expert systems can be created based on rules that manipulate the knowledge and allow efficient search, based on symbolically or numerically reasoning [20–23]. The structure of expert systems can be, in principle, constructed considering three main parts: knowledge base, inference machine and user interface [24–26].

Examples of expert system development through knowledge base can be found elsewhere [24–32] and, in general, the knowledge of human expert can transform information into logic formulations for solving problems [27]. The knowledge base may be therefore represented by facts, search trees; semantics net and heuristic rules [28]. The Semantic net is generally defined as a direct graph that may be applied in several contents, for example, psychological models of human memory, physical structures or causal bonds and inter-relationship of elements into economical models [24]. Similarly, heuristic rules are employed in several problems such as spectra interpretations [19] or medical diagnose [29,30]. As shown in the latter works one can verify that expert systems provide a more efficient analysis than exhaustive standard procedures analyzed through space of possible solutions [31,32].

The machine inference can be considered as an elaborate protocol to search in heuristic rules and knowledge representation data in order to solve problems. In general, the inference selects and applies the more appropriate rules and it determines the application of special rules in each execution step of an expert system [20]. Particularly, the user interface makes the relationship between users and expert system [24]. Moreover, the validation is performed in order to evaluate the system capability of solving problems correctly and the corresponding computational performance [24].

For the particular case of thermo analysis, an expert system can be considered a useful tool in order to help the experimentalist for interpreting chemical thermal decompositions. In this case the thermogravimetric analysis may be carried out by considering the mass variation as a function of temperature and/or time. The application of this procedure can be used to determine mass changes for several sequential reactions. In general, the molecular decomposition route for typical thermogravimetric analyses has well defined steps and each of them has a specific molecular fragment [33]. Therefore, the analysis is performed through considering each weight loss for one or more molecular fragments.

In the present work the expert system interpretation process was constructed considering two steps. The first step (Fig. 1: Part I) was developed based on the fragment set of the molecule. The latter algorithm was then applied for preparing data to be used in the second step (Fig. 1: Part II). This second step combines

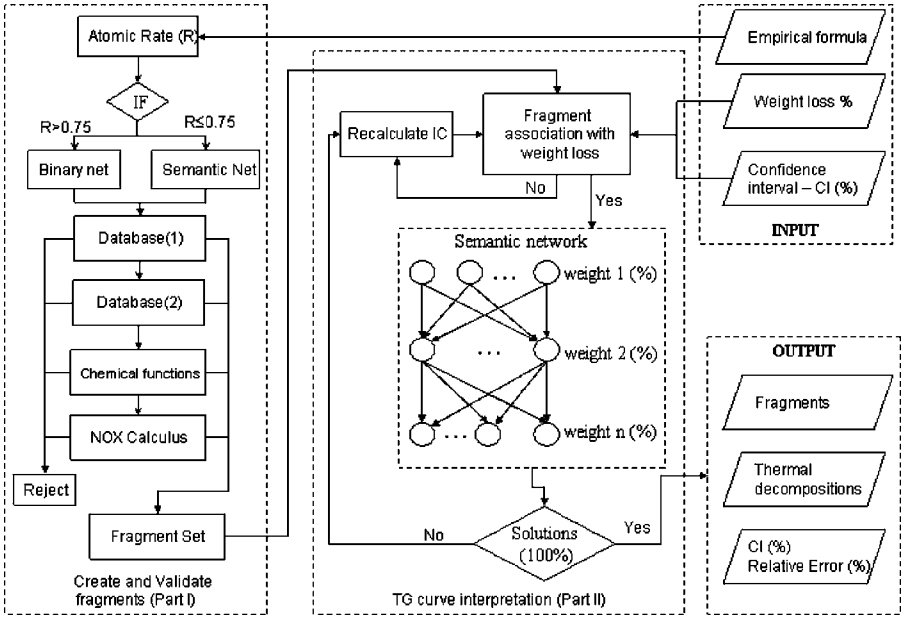

Fig. 1. Algorithm to analyze processes of thermogravimetric data. (Part I) fragment set (NOX = 0) and (Part II) interpretation TG curves.

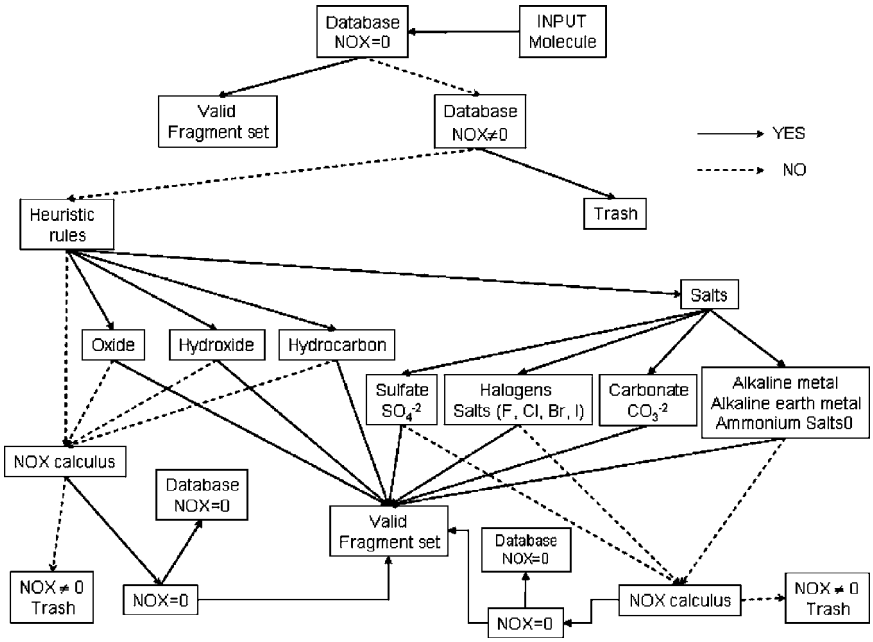

Fig. 2. Search tree for molecular fragment validation (NOX = 0).

the fragments in agreement with inference machine in order to find better interpretation of TG curves.

The fragmentation process has two algorithms namely semantic and binary, and each one is specialized to different kinds of molecules. The semantic net algorithm is more appropriated for molecules that present a large number of repeated atoms. This net is able to eliminate repeated fragments and thereby simplifies the structure of the search tree.

The present thermogravimetric analysis only considers neutral molecular fragments. Therefore, all analyzed molecules have an oxidation number (NOX) equal to zero. The expert system may employ a set of heuristic rules to classify molecular fragments into valid (NOX = 0) or not valid (NOX ≠ 0) NOX condition. Fig. 2 shows the search tree used to evaluate all classifications. The validation process is initiated by searching fragments in the database (database 1) that contains only structures with NOX equal to zero. For example, ammonium and carbonic gas frequently occur in thermogravimetric analyses [33,35]. If a fragment search is not found in the first database then it is searched in the second database (database 2). The latter contains structures with electrical charges (ions), for example, CO_3^{-2} and NH_4^{+} used in order to classify not valid structures.

The next test is based on a set of heuristic rules in which chemical functions are defined. At the present analysis, the system has seven heuristic rules, in which three are chemical functions: oxide, hydroxide and hydrocarbons, and four are salt definitions: sulfates, halogens salts, carbonates, ammonium and alkaline elements salts. If any fragment does not belong to any of these chemical functions, the system starts for the first analysis by evaluating the NOX calculation. For the latter case one has two output types: (i) if the calculated NOX is equal to zero, the molecular fragment is assign as valid and is added to the database (NOX = 0); (ii) if the NOX value is not zero, the fragment is eliminated. Therefore, those fragments already classified as not valid will not be taken into account in any new fragment calculations. Accordingly, this provides lower computational demanding and hence lower computational time is achieved.

The new inclusions in the search tree (Fig. 2) may be supervised by a human expert and modifications can be done to avoid increment of wrong information. Thus, the inclusion of any new information is added after occurrence of incorrect interpretation. The latter procedure provides an improvement of the whole expert system.

Fig. 3 shows a semantic net of molecular fragmentation taking $CaCO_3$ as an example. The validation process can illustrate the search tree algorithm. In this particular case, 31 molecular fragments (all possible allow combinatory components) should be found; however, the semantic net determines only 15 different fragments. All determined fragments are marked with different symbols to differentiate allow and forbidden molecular

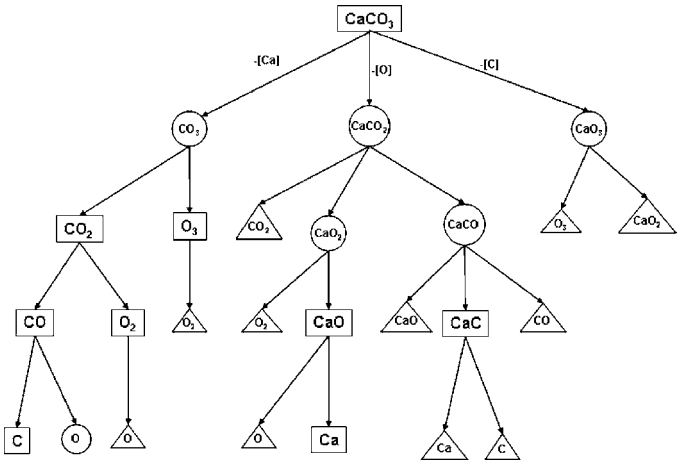

Fig. 3. Semantic tree for $CaCO_3$ fragmentation. Triangles repeated fragments, squares valid fragments and forbidden fragments (circles).

fragments. However, due to the combinatorial analysis these fragments do not exist and therefore their charges are not relevant for the final analysis. For example, for the particular case of $CaCO_3$ one observes in Fig. 3 these types of compounds such as CaO_3, CaCO, $CaCO_2$. Although some fragments are chemically charged, for simplify, these charges are not shown in Fig. 3. For example, CO_3 (circle) should be $CO_3{}^{-2}$. All these fragments are classified as not valid. The semantic net eliminates every repeated molecular fragment, thereby; the CO_2 is shown twice in Fig. 3. However, only one fragment is allowed and the other which is forbidden (triangle), is not generated as any fragment for the next layer. The valid fragment set is marked as square, in which has NOX value equal to zero, for example CaO, CO, CO_2, etc.

The second fragmentation algorithm, binary net, is more appropriated to molecules with few repeated atoms. The binary net makes all possible fragments, that is to say, $2^n - 1$, where n is the number of atoms in the molecule. Therefore, these two classification algorithms can be considered the most time consuming for the present expert system. Fig. 4 compares the competition of each classification based on these two algorithms.

The second step of the actual expert system (Fig. 1: Part II) is started with an association of molecular fragments with the correspondent weight loss, but this attribution is only used for absolute errors smaller or equal than confidence interval (I_C). This quantity can be defined by the expert analyst or evaluated by the proposed expert system. For that, one can calculate the I_C according to

$$I_C = \min\{|W^1_{exp} - W^1_{cal}|, |W^2_{exp} - W^2_{cal}|, \ldots, |W^N_{exp} - W^N_{cal}|\} \quad (1)$$

where N is the total number of weight loss, W^i_{exp} the experimental weight loss and W^i_{cal} is the calculated weight loss for fragment attribution. Therefore, the expert system evaluates the confidence analysis by itself or through the analyst definition. This means that each weight loss has a subset of valid fragment set. Each subset of weight loss presents at least one molecular fragment. In the case of empty subset, Eq. (1) is applied to restart the process. The system also evaluates a relative error, which is calculated as given by

$$\text{error} = \frac{|W^i_{exp} - W^i_{cal}|}{W^i_{exp}} \times 100 \quad (2)$$

where i corresponds to i-esim weight loss, W^i_{exp} the experimental weight loss and W^i_{cal} is the calculated weight loss for each fragment attribution. In the next step (Fig. 1), the subsets of all fragments are organized in a semantic network. This network is constructed in layers and each layer represents one weight loss and each vertex in one layer that corresponds to one associated molecular fragment. The edges characterize the sum operation between one vertex and an adjacent vertex in the next layer. The final result for the thermogravimetric analysis using a semantic network interaction is obtained only for absolute errors smaller or equal to the confidence interval. Particularly, if the solution is not determined, the I_C is modified and a new search is performed (Fig. 1: Part II). In the present study the analysis showed that the expert system can find more than one solution and the system provides all possible fragments. In this particular case, the expert analyst should select the correct interpretation solution that is more appropriated for the correct thermogravimetric data.

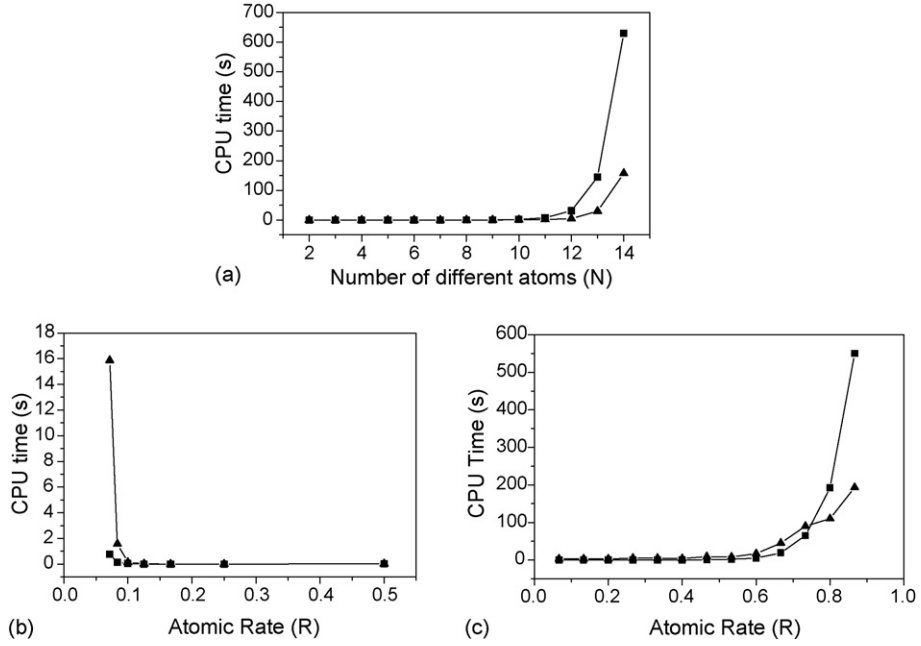

Fig. 4. Computational time for the fragmentation algorithm: binary net (triangle) and semantic net (square) using Intel®, PC Pentium 4®, 3.0 GHz, 1.0 GB of RAM. (a) CPU time for different number of atoms; (b) CPU time for homology series; (c) CPU time for molecules between homology series and compounds with all different atoms.

3. Mathematical background

The present expert system has been proposed based on the strategies of heuristic rules [24]. These rules can be constructed through the mathematical logic (ML) formalism [36–38]. Therefore, the flowchart shown in Fig. 2 defines the algorithm for these rules that produce the molecular fragment validation (NOX = 0). In order to describe the actual approach lets assume that a specified molecular fragment F will be analyzed using the ML procedure. The classification can be done according to valid fragments (F_v) and not valid ones (F_n). All fragments are produced considering a combinatorial decomposition analysis and we will return to this point later.

The fragment search in the database for a valid fragment is evaluated according to the following premise: "If there is F in valid fragment set then the fragment (F) is classified as valid fragment, otherwise F will be submitted to the next heuristic rule." This first rule can be written as[1]

$$[(\exists(F \in S_v)) \rightarrow (F = F_v)] \vee [(\neg\exists(F \in S_v)) \rightarrow N_r] \tag{3}$$

where S_v is a valid set of a specified chemical function, for example, ammonium and carbonic gas, F_v the valid molecular fragment and N_r means the next heuristic rule. A similar reasoning is used to set not valid fragments. The search in the database is performed according to

$$[(\exists(F \in S_{nv})) \rightarrow (F = F_n)] \vee [(\neg\exists(F \in S_{nv})) \rightarrow N_r] \tag{4}$$

where S_{nv} is a set of molecular structures with charge (ions), for example, NH_4^+ and CO_3^{2-}, and F_n is a not valid molecular fragment. However, if a fragment cannot be classified in database, the expert system activates a set of heuristic rules with chemical functions definitions. The first rules define inorganic compounds, with oxides and halogenated compounds, and these rules have the following premises:

- Every compound must have a functional atom, oxygen or halogen, and any other chemical element.
- There is a sum of NOX values for each atom that must be equal to zero. These premises can be mathematically formulated as

$$\left[\begin{array}{c}(F = F_b) \leftrightarrow \exists(F_{gru} \subset F) \wedge \exists(X_{oe}^i \subset F) \wedge \\ \exists\left(\left(n(F_{gru}) \times \mathrm{NOx}(F_{gru}) + \sum_{i=1}^{m} \mathrm{NOx}(X_{oe}^i)\right) = 0\right)\end{array}\right] \rightarrow (F = F_v) \tag{5}$$

where F_b is a binary chemical function (oxide or halogenated compound), F_{gru} an atom of functional group (O, F, Cl, Br or I), X_{oe}^i an other chemical atom, $n(X_{oe})$ the quantity of functional group, NOX(F_{gru}) and NOX(X_{oe}) are functions that return the NOX value of an i-esim atom or functional group and m is the total number of different atoms of each functional group composition. The same construction is used for other chemical functions and some salts, as hydroxide, sulfate, carbonate, salts of alkaline metals and ammonium salts, can be evaluated as

$$\left[(F = F_{hscaa}) \leftrightarrow \exists(F_{gru} \subset F) \wedge \exists\left(n(F_{gru}) \times \mathrm{NOx}(F_{gru}) + \sum_{i=1}^{m} \mathrm{NOx}(X_{oe}^i)\right) = 0\right] \rightarrow (F = F_v) \tag{6}$$

where F_{hscaa} is a representation for chemical functions and compounds classes (hydroxide, sulfate, carbonate, salts of alkaline metals and ammonium salts).

Organic functions need more specific rules. The heuristic for hydrocarbon function can be defined according to the following premises:

- Hydrocarbons must have only carbon and hydrogen atoms in their compositions.
- Each carbon atom must have four chemical bonds and hydrogen's have only one chemical bond.
- Each carbon atom is defined in order to have the following bonds: four single, two single and one double bonds; two double, one single and one triple bonds.

The following equation can summarize these rules:

$$\left[(F = F_h) \leftrightarrow \left(\begin{array}{c}\forall(C \subset F)\exists(4\,\mathrm{bonds}) \wedge \forall(H \subset F)\exists(1\,\mathrm{bond}) \wedge \\ \forall(C \subset F)\exists(B_n(\mathrm{single} \vee (\mathrm{double} \wedge \mathrm{single}) \vee (\mathrm{double} \wedge \mathrm{double}) \vee (\mathrm{triple} \wedge \mathrm{single})))\end{array}\right)\right] \rightarrow F = F_v \tag{7}$$

where F_h is a hydrocarbon function, C the carbon atom, H the hydrogen atom, B_n defines a combinatorial function for all types of carbon bonds. If the molecular fragment is not classified, the last rule for validation and the NOX value is calculated for all atoms according to

$$\left((F = F_v) \leftrightarrow \exists\left(\sum_{i=1}^{m} \mathrm{NOx}(X_{oe}^i) = 0\right)\right) \vee \left(F = F_n \leftrightarrow \neg\exists\left(\sum_{i=1}^{m} \mathrm{NOx}(X_{oe}^i) = 0\right)\right) \tag{8}$$

where m is the total number of atoms. After validation of a fragment set, the system starts to search the solution for thermogravimetric data through the semantic net. Each interaction for the semantic net is classified as solution (R) or not solution (L). This classification depends on the difference between evaluated weight loss and experimental weight loss. If this difference is smaller than the confidence interval (I_C) value then the interaction is classified as valid, otherwise the interaction is classified

[1] Details of the mathematical means of logic operators can be found in Refs. [24,27].

as a not valid solution. This premise can be written as

$$\left(R \leftrightarrow \sum_{i=1}^{h}(\mathrm{abs}(W_{\mathrm{cal}}(F^i) - W^i_{\mathrm{exp}}) \leq I_{\mathrm{C}})\right) \vee \left(L \leftrightarrow \sum_{i=1}^{h}(\mathrm{abs}(W_{\mathrm{cal}}(F^i) - W^i_{\mathrm{exp}}) > I_{\mathrm{C}})\right) \quad (9)$$

where R is a set of fragments and the corresponding calculated weight loss for a valid solution, L a not valid answer with fragments and the corresponding calculated weight loss, $W_{\mathrm{cal}}(F^i)$ a function that calculates the percentage of one fragment or a set of fragments associated with i-esim weight loss, h the number of weight lost and I_{C} denotes the confidence interval that corresponds to the maximum absolute error.

4. Discussion and results

The expert system can be validated using accurate tests and evaluating its performance. These tests have to be applied to simple and complex chemical systems. In the present study several features were tested: fragmentation capability, computational time of fragmentation algorithms; interpretation capability, total computational time of interpretation algorithm and interpretation of complexes TG curves.

The proposed expert system depends on the definition of parameters that are used for the classification rules of each algorithm (two) as shown in Fig. 1 (Part I) and Fig. 2. The first parameter, which is reported in Fig. 4a, is related to the molecular size. If the molecule is formed with a number of different atoms smaller than 12, the computational time for binary and semantic nets are almost the same. For molecules with more than 12 different atoms, the semantic net has a greater computational cost and both have almost an exponential behavior as can be seen in Fig. 4a. Another parameter is based on the atomic rate. This parameter is defined as the ration between the number of different atoms and total number of atoms. It can be considered as the contribution of each atom to generate repeated molecular fragments in the search tree with respect to the whole molecule. The dependence of computational time for this parameter is shown in Fig. 4(b and c). Fig. 4(b) presents the results for homology series (alkanes, alkenes, etc.), that is to say, more extensive chemical structures that produce atomic rates smaller than 0.1. The semantic set algorithm is more appropriated to be applied in compounds with large number of repeated atoms, in which the search tree can simplify the procedure by eliminating repeated fragments. The behavior between homology series and molecules with all different atoms is shown in Fig. 4(c). If the atomic rate is smaller than 0.75, one expects that the semantic net can produce better performance. However, if this ratio is greater than 0.75 one verifies an inversion of the time demanding and the binary net presents lower computational cost than the semantic net.

The fragmentation test is presented in Table 1. Only molecules with NOX = 0 were analyzed and used in order to verify simple fragmentation cases. In general, valid fragment sets produce fragments that may not exist for real analysis. For example, H_2, or O_3 cannot be detected in a thermogravimetric analysis. However, these structures are found due to the combinatorial behavior of the algorithms developed for the present study.

Table 1
Molecules and valid fragments sets analyzed with fragmentation algorithms: binary and semantic nets

Molecule	Fragments (NOX = 0)
$CaCO_3$	CO, CO_2, CaO, O_2, Ca, C
$CaC_2O_4{\cdot}H_2O$	H_2O, $Ca(OH)_2$, $CaCO_3$, H_2, CO, CO_2, CaO, CaC_2, O_2, Ca, C
$Mg(CHO_2)_2{\cdot}2H_2O$	H_2O, $Mg(OH)_2$, $MgCO_3$, MgO, H_2, CO, CO_2, O_2, Mg, C
$[Cu(NH_3)_4]SO_4{\cdot}H_2O$	CuO, Cu_2O, O_2, S, SO_2, SO_3, NH_3, CuS, N_2, H_2, $H_2S{\cdot}H_2O$, Cu, $CuSO_4$

The proposed expert system (Fig. 2) has three classes of validation fragments: search in database, heuristic rules to chemical functions and NOX evaluation. In the present case the heuristic rules provide the smallest computational time. The computational cost may be reduced with the addition of new heuristic rules, but very particular fragments do not need any new rules since these type of fragments may be added in the database. Every valid molecular fragment is found by calculating NOX and is added to the database and hence, the computational cost may be reduced of about 80%. The appropriate fragmentation algorithms for larger molecules, such as organometallic complexes, produce the valid fragment sets between 50 and 100 s.[2]

The fragmentation algorithms are based on a combinatorial behavior; therefore, the approach can eventually find uncommon molecular fragments. For example, the fragment set of $CaC_2O_4{\cdot}H_2O$ has CaC_2. However, this compound may not exist in its thermo decomposition [33]. The fragmentation set for larger molecules can produce larger numbers of uncommon molecular fragments. For example, in the case of $[Cu(NH_3)_4]SO_4{\cdot}H_2O$ compound, the system found 14 valid fragments but only four possible decompositions appear in the thermogravimetric analysis [39].

A detailed interpretation of those compounds shown in Table 1 is reported in Table 2. The relative error is calculated for each weight loss (Eq. (2)). In contrast, the I_{C} parameter (Eq. (1)) is evaluated for each analysis. The confidence interval shown in Table 2 can be changed to higher values and hence the expert system can find more solutions than the previous analysis. This condition allows the system to find more solutions. In this situation the selected fragments can be determined by the analyst. Therefore, better solutions for an adequate interpretation are found for the thermogravimetric data. All interpretation results reported in Table 2 quantitatively agree with the human expert interpretation according to Refs. [33–35]. For example, the calcium oxalate shown in Table 2 is used as standard calibration for TG devices [35]. For this particular compound three weight losses appear in the TG curves [35] and there is not simultaneous decompositions. In this particular case the system looses

[2] Intel®, PC Pentium 4®, 3.0 GHz, 1.0 GB of RAM.

Table 2
Thermogravimetric data interpretation evaluated by the present approach for different compounds

Molecule	Degradation step	Experimental weight lost (%)	Calculated weight lost (%)	Correspond to[a]	Relative error (%)[a]	I_C (%)[a]	Time (s)
$CaCO_3$	1	44.03	43.97	CO_2	0.14		
	Residue	55.94	56.03	CaO	0.16	0.09	0.015
$CaC_2O_4H_2O$	1	12.32	12.32	H_2O	0		
	2	18.88	19.17	CO	1.5		
	3	29.04	30.12	CO_2	3.71		
	Residue	39.77	38.38	CaO	3.49	1.39	0.015
$Mg(CHO_2)_2{\cdot}2H_2O$	1	23.67	23.96	$2H_2O$	1.22		
	2	49.10	49.23	$H_2O + 2CO_2$	0.26		
	Residue	27.23	26.80	MgO	1.58	0.42	0.200
$[Cu(NH_3)_4]SO_4{\cdot}H_2O$	1	20.83	21.19	$H_2O + 2NH_3$	1.72		
	2	6.03	6.93	NH_3	14.92		
	3	6.97	6.93	NH_3	0.57		
	4	32.56	32.57	SO_3	0.03		
	Residue	33.61	32.36	CuO	3.71	1.30	4.15

[a] Expert system results.

H_2O, starting at 76 °C and finishing at 190 °C [35]. The next step which is characterized by CO molecule occurs between 400 and 523 °C [35]. In contrast, the first and second decomposition steps for $[Cu(NH_3)_4]SO_4{\cdot}H_2O$ produce a shoulder in the TG curve. However, this fact does not have influence in the expert system interpretation and the attribution loss of NH_3 is correct as experimentally determined in Refs. [39,40]. The residue for the decomposition of $Mg(CHO_2)_2{\cdot}2H_2O$ described in Ref. [40] is also correctly determined by the expert system and it is attributed as MgO. However, this compound may contain a small amount of carbon that is considered as contamination and hence, this attribution may be not fully corrected. For the particular case of $[Cu(NH_3)_4]SO_4{\cdot}H_2O$ [39] at 892 °C, the expert system determined the residue in which was attributed to CuO with a higher relative error equal to 3.71% and 1.3% for the confidence interval. This result was also experimentally determined with the same relative error [39].

All these tests (Table 2) show that the proposed expert system can be, in principle, efficient to deal with themogravimetric analysis. However, all molecules described in Table 2 can be considered simple cases due to their chain size. Therefore, larger systems need to be tested in order to provide a more accurate analysis of the actual approach. Considering complex data from the literature in which the corresponding thermogravimetric curves usually show several steps decompositions can complete the actual analysis. For that, seven compounds namely salicylato (amine) Co (III) complexes, norfloxacin complexes of manganese (II) and cobalt (II), and calcium borate were taken as validation tests. Table 3 compares the literature results and those fragments obtained by using the present approach. In general, the expert system has a high efficiency since each attribution for decomposition steps is correct, according to a quantitative agreement with the literature results [41–43]. The determination of fractional stequiometric index is considered as a limitation of the expert system, also reported in Table 3.

For the particular example of compound 4 one observes a more difficult analysis. In this case, there is a partial overlap between steps II and III. In Ref. [41], these steps were reported as 0.6trien and 0.4trien + Cl + sal, respectively. The present expert system found only two weight losses for all decomposition process. The first loss was attributed as one water molecule and the second weight loss, which is equivalent to 80%; was defined as trien + Cl + sal. Therefore, if thermogravimetric analyses produce overlap steps and fractional stequimetric index, the expert human analyst needs to infer to obtain a better solution. In this case, the analyst needs to combine both weight losses (0.6trien and 0.4trien + Cl + sal) in order to determine the correct TG interpretation. Accordingly, the results of interpretation process provide an indicative for the decomposition attribution.

As described in Ref. [42], the Co(II) (5) and Mn(II) (6) carried out under a N_2 flown. These norfloxacin complexes have more than 100 atoms and this can be considered an efficient test for the proposed expert system. The cobalt complex (5) has two decomposition steps. The first was attributed as four hydration water molecules. And the second step was associated to others four water molecules and norfloxacin molecular fragments, that defines a larger step, with 71.21% of weight loss. The same behavior is observed for manganese complex (6), but the thermogravimetric analysis produces three steps. The first and second steps are attributed as four hydration water molecules each. The third step corresponds to norfloxacin and acetic ligant degradation. The residue composition has a common component, which is associated to carbon atoms, in both norfloxacin complexes. As observed for compound (5) there are four carbon atoms as a residue and this is considered as a common residue for this particular case of cobalt and manganese complexes.

In general, hydration water molecules are well defined during the weight loss and this always presents an absolute error smaller that 1% [33]. This condition is then used by the expert system to evaluate the number of hydration water molecules in the system. The substance (7) is a calcium borate mono hydrated [38]. The interpretation of this compound has a confidence interval equal to 0.6% and it is attributed to only one water molecule. However, if the compound has two water molecules then the

Table 3
Interpretation of TG curve of complex chemical systems with large structures

Complex	Molecule	Degradation step	Experimental weight loss (%)[a]	Calculated weight loss (%)[a]	Relative error (%)[a]	I_C (%)[a]	Correspond to[a]
1, Ref. [41]	$[Co(en)_2sal]Cl{\cdot}2H_2O$[b]	I	9.2	8.9	3.2		$2H_2O$
		II	30.2	29.7	5.7		2en
		III	43.9	42.9	2.3		Sal + Cl
		Residue	16.1	18.5	14.9	2.5	CoO[c]
2, Ref. [41]	$[Co(en)_2(C_2H_5NH_2)salH](NO_3)_2{\cdot}H_2O$[b]	I	3.60	3.57	0.8		H_2O
		II	31.00	31.62	2.0		$2NO_2 + O_2 + CH_3NH_2$
		III	24.20	23.83	1.5		2en
		IV	25.00	27.20	8.8		Sal
		Residue	16.20	14.85	8.3	2.5	CoO[c]
3, Ref. [41]	$[Co(NH_3)_5salH](NO_3)_2{\cdot}H_2O$[b]	I	4.20	4.24	0.9		H_2O
		II	4.60	4.01	12.8		NH_3
		III	75,00	77.84	3.8		$4NH_3 + NO_2 + O_2$ + salH
		Residue	16.20	17.66	9.0	3.0	CoO[c]
4, Ref. [41]	β-*cis*-$[Co(trien)sal]Cl{\cdot}H_2O$[b]	I	4.33	4.36	0.7		H_2O
		II	23.43	80	–		Trine + Cl + sal
		III	57.20		–		
		Residue	15.0	18.15	21	3.15	CoO[c]
5, Ref. [42]	$[Co(nor)_2]SO_4{\cdot}8H_2O$[b]	I	7.62	7.70	1.0		$4H_2O$
		II	71.21	70.67	0.8		$11C_2H_2 + 3C_2H_4 + 6NO + 2HF + 4H_2O$
		Residue	21.17	21.65	2.3	0.7	$CoSO_4 + 4C$
6, Ref. [42]	$Mn(nor)_2(CH_3CO_2)_2{\cdot}8H_2O$[b]	I	7.49	7.55	0.8		$4H_2O$
		II	7.51	7.55	0.5		$4H_2O$
		III	72.02	72.42	0.6		$11C_2H_2 + 4C_2H_4 + 6NO + H_2O + 2CO + 2HF$
		Residue	12.98	12.45	4.0	2	MnO + 4C
7, Ref. [43]	$(CaO)_2B_2O_3{\cdot}H_2O$	I	9.62	9.01	6.34		H_2O
			90.38	90.98	0.7	0.6	$Ca_2O_2B_2O_3$

[a] Expert system results.
[b] Here, salH is $C_6H_4(OH)CO_2^-$, en is ethylenediamine, trien is triethynetetramine and nor is norfloxacin.
[c] The correct attribution is Co_3O_4, that is reported in Ref. [41].

confidence interval increases to 1.9%. A more accurate analysis is performed by adding up to 10 water molecules. In this case the confidence interval is greater than 2%. Due to the smallest I_C (0.6%) the methodology attributes for this particular compound (7) only one hydration water molecule.

5. Conclusion

The present work dealt with an approach to analyze and carries out interpretation of thermogravimetric data using an expert system methodology. Efficient fragmentation analysis is done through two algorithms, namely binary and semantic nets. Interpretation process was performed via association between valid molecular fragments and each weight loss. The procedure was carried out using semantic net algorithm as a function of temperature. The automation for interpreting chemical processes was performed in order to create the inference method to determine thermogravimetric analysis.

The actual expert system can, in principle, generate a large number of decomposition routes in which may be not adequately analyzed by humans due to their inherent limitations. The advantage of the present approach can be attributed to its efficient error minimization for each weigh loss and global process. However, there are disadvantages. For example, for the particular case of multiple step TGA in which there are poor resolved steps the actual expert system can use the human inference or keeping evaluating the decomposition but considering the analysis through the differential thermal analysis (DTG) curve. In this case each weight loss in TG curve is determined by the difference between two local maximum in the DTG curve. Similarly, the analysis of polymers cannot be performed with the actual expert system in its present form. One of the reasons is attributed to the unresolved empirical formula for the polymer (number of monomers). The system needs empirical formula definitions to obtain a set of valid molecular fragments. In this case, it is necessary a new implementation to increase a distribution of the composition to be used for polymers analysis. The expert system may be also applied to the case of parallel decomposition reactions through the combination of valid fragments in only one weight loss. However, in its present form such analyses are not implemented. In addition, another limitation is concerned with the determination of fractional stequiometric indexes.

Finally, the validation tests showed the viability of using the proposed methodology for interpreting thermogravimetric data. In addition, the computational cost is compatible with microcomputers. Accordingly, the system may be considered as a

useful tool for the experimentalist in many thermogravimetric analyses.

Acknowledgements

The authors would like to thank the research agencies CNPq, FAPEMIG, and CAPES (GPV) for financial support.

References

[1] R.C. Mackenzie, Thermochim. Acta 28 (1979) 1–6.
[2] D.A. Skoog, F.J. Holler, T.A. Nieman, Principles of Instrumental Analysis, 4th ed., Saunders, Philadelphia, 1992.
[3] N. Morel, A. Faist, Build. Environ. 28 (1993) 465–473.
[4] G. Gupchup, K. Alexander, D. Dollimore, Thermochim. Acta 196 (1992) 267–278.
[5] A.S. Ribeiro, A.C.F. Caires, N. Borrale, M. Lonashiro, Thermochim. Acta 279 (1996) 177–181.
[6] M.A.A. Reis, J.C. Belchior, R.D. Sinisterra, J. Pharm. Sci. 92 (2004) 418–430.
[7] A. Liparini, S. Carvalho, J.C. Belchior, Clin. Chem. Lab. Med. 43 (2005) 939–946.
[8] D. Holzhauer, I. Grosse, Eng. Comput. 15 (1999) 315–325.
[9] R.C.O. Sebastião, J.P. Braga, M.I. Yoshida, Thermochim. Acta 412 (2004) 107–111.
[10] G.C. Xu, L. Zhang, L. Liu, G.F. Liu, D.Z. Jia, Thermochim. Acta 429 (2005) 31–42.
[11] M. Eloma, C.H. Lochmuller, M. Kudrjashova, M. Kaljurand, Thermochim. Acta 362 (2000) 137–144.
[12] M. Wesolowski, M. Czerwonka, Thermochim. Acta 398 (2003) 175–183.
[13] M. Wesolowski, P. Konieczynski, Thermochim. Acta 397 (2003) 171–180.
[14] C. Bernal, A.B. Couto, S.T. Trazzi, E.T.G. Cavalheiro, Quím. Nova 25 (2002) 849–855.
[15] T. Nakayama, K. Tanaka, J. Chem. Inform. Comput. Sci. 39 (1999) 819–832.
[16] H.F. Pop, C. Sarbu, O. Horowitz, D. Dumitrescu, J. Chem. Inform. Comput. Sci. 36 (1996) 465–482.
[17] B. Debska, B. Guzowska-Swider, J. Chem. Inform. Comput. Sci. 40 (2000) 325–329.
[18] B. Debska, B. Guzowska-Swider, D. Cabrol-Brass, J. Chem. Inform. Comput. Sci. 40 (2000) 330–338.
[19] S.G. Molodtov, M.E. Elyashberg, K.A. Blinov, A.J. Willians, E.E. Martirosian, G.E. Martin, B. Lefebvre, J. Chem. Inform. Comput. Sci. 44 (2004) 1737–1751.
[20] F.X. Rius, Anal. Chim. Acta 283 (1993) 518–527.
[21] J.-S.R. Jang, C.-T. Sun, E. Mizutani, Neuro-Fuzzy and Soft Computing, Prentice Hall, London, 1997.
[22] J. Vladmir, Collect. Czech. Chem. Commun. 57 (1992) 184–197.
[23] A. Ting, R. McGuire, A.P. Johnson, S. Green, J. Chem. Inform. Comput. Sci. 40 (2000) 347–353.
[24] E.A. Bender, Mathematics Methods of Artificial Intelligence, E.E. Comp. So. Press, Washington, 2000.
[25] G. Klopman, S.K. Chakravarti, H. Zhu, J.M. Ivanov, R.D. Saiakhov, J. Chem. Inform. Comput. Sci. 44 (2004) 704–715.
[26] T. Laidboeur, D. CabrolBass, O. Ivanciuc, J. Chem. Inform. Comput. Sci. 36 (1996) 811–821.
[27] W.F. Clocksin, C.S. Mellish, Progamming in Prolog, 4th ed., Springer, New York, 1994.
[28] I. Bratko, Prolog Programming for Artificial Intelligence, Addison–Wesley, Boston, MA, 1990.
[29] E.T. Keravnou, F. Dams, J. Washbrook, C.M. Hall, R.M. Dawood, D. Shaw, Comput. Meth. Programs Biomed. 45 (1994) 239–260.
[30] E.Y.K. Ng, N.M. Sudharsan, J. Eng. Med. 215 (2001) 25–37.
[31] N. Greene, J. Chem. Inform. Comput. Sci. 37 (1997) 148–150.
[32] P.N. Judson, C.A. Marchant, J.D. Vessey, J. Chem. Inform. Comput. Sci. 43 (2003) 1364–1370.
[33] W.W. Wendlandt, Thermal Analysis, John Wiley, New York, 1986.
[34] D.E. Brown, M.J. Hardy, Thermochim. Acta 85 (1985) 521–524.
[35] A. Arenillas, F. Rubiera, J.J. Pis, J. Anal. Appl. Pyrol. 50 (1999) 31–46.
[36] M. Pollet, V. Sorge, M. Kerber, Lect. Notes Comput. Sci. 3119 (2004) 317–331.
[37] R. Wille, Lect. Notes Comput. Sci. 3874 (2006) 1–29.
[38] F. Dau, Lect. Notes Artif. Intell. 2961 (2004) 156–171.
[39] R.F. Speyer, Thermal Analysis of Materials, Marcel Dekker, New York, 1994.
[40] J.P. Gupta, D.V. Nowell, Thermochim. Acta 30 (1979) 339–350.
[41] N. Das, Thermochim. Acta 257 (1995) 163–171.
[42] S.A. Sadeek, J. Mol. Struct. 753 (2005) 1–12.
[43] Z.H. Liu, C.F. Zuo, S.Y. Li, Thermochim. Acta 424 (2004) 59–62.