Using Rough Set Theory to Achieve Reliable Chemical Solvents Selection: A Multi-attribute Decision Case Study

Chaozhe Jiang [1] Pei Hu[2] Dun Liu[2] Fang Xu[3] Jixue Yuan[4]

[1]College of Traffic and Transportation, Southwest Jiaotong University, Chengdu 610031, P. R. China
[2]School of Economics & Management, Southwest Jiaotong University, Chengdu 610031, P. R. China
[3]Sichuan Tourism University, Chengdu 610031, P. R. China
[4]School of Physical Education, Yunnan Normal University, Kunming 650092, P. R. China

Abstract

The basic concepts of the rough set theory (RST) are introduced. An example of the rough set theory application to the chemical solvents selection problem (CSSP) is presented. Through the RST, we get a good multi-attributes decision-making effect in the chemical solvents selection. This work demonstrates how the use of RST in chemical process development by allowing efficient and reliable improvement of a given synthetic step. And other numerous earlier applications of rough set theory to the various scientific domains suggest that it also can be a useful tool for the analysis of inexact, uncertain, or vague chemical data.

Keywords: RST (Rough set theory), Chemical solvents selection problem (CSSP), Multi-attributes Decision Making (MADM), RIDAS (An Rough Set Based Intelligent Data Analysis System)

1. Introduction

The rough set theory, introduced by Pawlak in 1982[1], although popular in many other disciplines, is nearly unknown in chemistry. So this paper intends to propose a research communication of the rough set theory application to chemical field.

The rough set approach can be considered as a formal framework for discovering facts from imperfect data. The results of the rough set approach are presented in the form of classification or decision rules derived from a set of examples or cases.

The aim of this paper is to introduce the basic concepts of the Rough Set Theory (RST) and also to show its possible applications in the field of chemical industry.

2. Rough Set Theory

2.1 Basic concepts of the RST

2.1.1. Information system and indiscernibility relation

Formally, an information system, can be seen as a system IS=(U, A) Where U is the universe(a finite set of objects, U={x_1,x_2,…..,x_m}) and A is the set of attributes. Each attribute a∈A defines an information function f_a: $U \rightarrow V_a$, where V_a is the set of values of a, called the domain of attribute a.

For every set of attributes B⊂A, an indiscernibility relation Ind(B) is defined in the following way: two objects, x_i and x_j , are indiscernible by the set of attributes B in A, if $b(x_i) = b(x_j)$ for every b⊂B. The equivalence class of Ind(B) is called elementary set in B because it represents the smallest discernibile groups of objects. For any element x_i of U, the equivalence class of x_i in relation Ind(B) is represented as $[x_i]_{Ind(B)}$. The construction of elementary sets is the first step in classification with rough sets.

2.1.2. Approximations of sets

So called the lower and the upper approximations of a set (Fig.1), referring to:

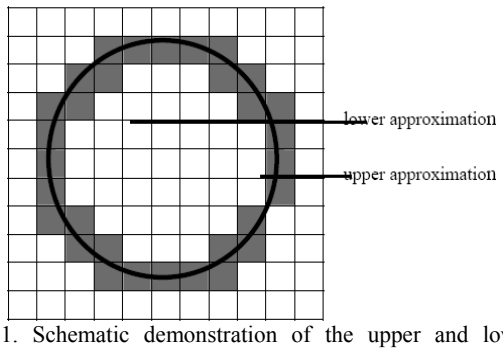


Fig.1. Schematic demonstration of the upper and lower approximation of set X.

Let X denotes the subset of elements of the universe U (X⊂U). The lower approximation of X in B (B⊆A), denoted as:

$$\underline{BX} = \{x_i \in U \mid [x_i]_{Ind(B)} \subset X\}$$

The upper approximation of the set X, denoted as: $\overline{BX} = \{x_i \in U \mid [x_i]_{Ind(B)} \cap X \neq \phi\}$

For any object x_i of the lower approximation of X(i.e., $x_i \in \underline{BX}$), it is certain that it belongs to X. For any object x_i of the upper approximation of X (i.e., $x_i \in \overline{BX}$), we can only say that x_i may belong to X. The difference: $BNX = \overline{BX} - \underline{BX}$ is called a boundary of X in U.

If the lower and upper approximation are identical(i.e., $\overline{BX} = \underline{BX}$), then set X is definable, otherwise, set X is indefinable in U. if $\underline{BX} \neq \phi$ and $\overline{BX} \neq U$, X is called roughly definable in U; where Φ denotes an empty set.

And $POS_B(X) = \underline{BX}$, called the B-positive region of X, $NEG_B(X) = U - \underline{BX}$, called the B-negative region of X.

2.1.3. The weight ac Approximation of sets

An accuracy measure of the set X in B⊆A is defined as: $\mu_B(X) = card(\underline{BX}) / card(\overline{BX})$

Where *card(.)* means the cardinality of a set. As one can notice, $0 \le \mu_B(X) \le 1$. If X is definable in U then $\mu_B(X) = 1$; if X is undefinable in U ,then $\mu_B(X) < 1$.

2.1.4. Reduction and Independence of attributes

If Ind(A)=Ind(A-a_i), then the attribute a_i is called superfluous. Otherwise, the attribute a_i is indispensable in A.

If the set of attributes is dependent, one can be interested in finding all possible minimual subsets of attributes, which lead to the same number of elementary sets as the whole set of attributes(reducts) and in finding the set of all indispensable attributes (core).

The concepts of core and reduct are two fundamental concepts of the rough sets theory. The reduct is the essential part of an IS, which can discern all objects discernible by the original IS. The core is the common part of all reducts. To compute reducts and core, the discernibility matrix is used. The discernibility matrix has the dimension n×n, where n denotes the number of elementary sets and its elements are defined as the set of all attributesd which discern elementary sets $[x]_i$ and $[x]_j$:

Simplification of the IS can be achieved by dropping certain values of attributes, which are unnecessary for the system, i.e., by eliminating some of these values in such a way that we are still able to discern all elementary sets in the system. The procedure of finding core and reducts of the attribute values is similar to that of finding core and reducts of the attributes. All computations are performed based on the discernibility matrix, but the definition of the discernibility function is now slightly different. Instead of one discernibility function, we have to construct as many discernibility functions, as there are elementary sets in the IS.

2.1.5. Classification

Let F={X1,X2,…..,Xn},Xi⊂U be a family of subsets of the universe U. If the subsets in F do not overlap, i.e., Xi∩Xj=Φ. And the entity of them contains all elementary sets, i.e.,∪Xi=U for . i=1,…, n. Then, F is called a classification of U, whereas Xi are called classes.

The lower and upper approximations of F in B⊆A are defined as:

$$\underline{B}(F) = \{\underline{B}(X_1), \underline{B}(X_2), \cdots, \underline{B}(X_n)\}$$

$$\overline{B}(F) = \{\overline{B}(X_1), \overline{B}(X_2), \cdots, \overline{B}(X_n)\}$$

respectively, The quality of classification is defined as: $\eta_B F = \cup card\, \underline{B}(X_t) / card U$ and the accuracy of classification F in B can be calculated according to the following formula:

$$\beta_B F = \cup card\, \underline{B}(X_t) / \cup card\, \overline{B}(X_t)$$

2.2 Decision table

A knowledge representation system containing the set of attributes A(now called condition attributes) and the set of decision attributes S is called a decision table. As we will show further, decision tables are also useful for classification.

2.2.1. D-superfluous attributes

For the attibute a_i, belonging to the condition set of attributes B (where B⊆A), is D-superfluous if it exerts no influence on the lower approximation of D, i.e., if $POS_B(D)=POS_{(B-a_i)}(D)$, Otherwise, attribute a_i is D-indispensable in A.

2.2.2. Relative core and relative reducts of attributes

The set of all D-indispensable attributes in A is called the D-core of A, whereas, the minimal subsets of condition attributes that discern all equivalence classes of the relation Ind (D) discernable by the entire set of attributes are called D-reducts.

Relative reducts can be computed using a slightly

modified discernibility matrix. An element of the D-discernibility matrix of A is defined as the set of the relation Ind(D), i.e., to the same class. The D-core is the set of all single elements of the D-discernibility matrix of A.

2.3 Main steps of decision table analysis

- Construction of elementary sets in D-space,
- calculation of upper and lower approximations of the elementary sets in D.
- finding D-core and D-reducts of A attributes,
- finding D-core and D-reducts of A attribute values.

2.3.1. Decision rules

The above described decision table can also be regarded as a set of decision(classification) rules of the form: $a_{k_i} \Rightarrow d_j$, where a_{k_i} means that "attribute a_k has value i" and the symbol"$\Rightarrow$"denotes prepositional implication. In the decision rule $\theta \Rightarrow \phi$, formulas θ and ϕ are called condition and decision, respectively. Minimization of a set of attributes and values of attributes with respect to another set of attributes simply means a reduction of unnecessary conditions in the decision rules, which is also known as the generation of decision rules from the data.

2.3.2.New decisions

Logical rules derived from experimental data may be used to support new decisions. Matching their description to one of logical rules can support classification of new objects. The matching procedure may lead to one of four situations.

(a) the new object matches exactly one of the deterministic logical rules;
(b) the new object matches exactly one of the non-deterministic logical rules;
(c) the new object matches no logical rules;
(d) the new object matches more than one logical rule.

2.4 Types of attributes

There are different types of attributes.

Quantitative attributes represent measurable properties of objects. Their values are ordered by definition. Examples: temperature, pH, concentration.

Qualitative attributes are expressed in linguistic terms. They can be divided into two classes. (1) ordered qualitative attributes. The values of these attributes can be ordered along an axis of significance. The order of the linguistic values can be represented by a sequence of increasing or decreasing numbers encoding them. Example: polarity=(low, medium, high)can be coded as low=1; medium=2 and high=3. (2) unordered qualitative attributes (nominal attributes. The linguistic values of these attributes cannot be ordered; in other words, it is impossible to arrange them along any axis of significance.

Application of RST to qualitative attributes is straightforward. For nominal attributes, RST offers evident advantages when compared with other classifiers. To use this type of attribute as an input to classical classifiers, one has to code it in a special way. Each linguistic value is represented by separate input(variable, feature). This encoding, called one-from-k, creates a binary vector the elements of which correspond to new inputs. When an attribute takes a particular value, the corresponding vector element is set to 1, while others are set to 0. this type of coding causes a drastic extension of data dimensionality.

Continuous condition attributes present a problem, as in this case, a discretezation is required. Both the number of subranges and their intervals have to be optimized. The number of subranges decides about the number of logical rules considered. The number of rules is not given in advance, but is limited by the general requirement that the learning objects should confirm the rules.

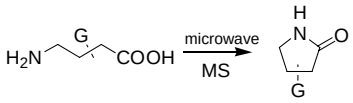
There are two possible approaches to the discretization problem. One can optimize coding taking into account only the similarities of the objects in the attributes' space or one can maximize the predictive properties of the information system in the stage of coding. Although the second approach seems to be more interesting it has its limitations. As pointed out by Ziarko et al. in ref. a low roughness setting, i.e., many small subranges, will lead to weak rules, i.e., rules not supported by many examples, which may be in contradiction to the expert's experience. When the roughness parameter is set high, generalized rules are produced with many supporting cases which lead to strong rules.

3. RST application to the chemical solvents selection- a case study

In order to illustrate RST, we have chosen to apply it to a chemical solvents selection problem. A small data set taken from the experimental data record of Asta Company is used (see Table 3). The example is chosen to be simple for tutorial reasons but it illustrates most aspects of the RST approach.

γ-lactam (pyrrolidinone) derivatives is a very important class of bioactive compounds that widely exist in natural products, and also serves as a versatile building block for the synthesis of drug material and natural products1. In our search for practical process to manufacture optical pure γ-lactam, we investigated all factors that affect the formation of chiral γ-lactam. In this communication, we like to report an interesting approach to obtain high optical purity of γ-lactam by

using molecular sieves (MS) as catalyst and also by microvave irradiation. Most interestingly, Rough Set theory (RST), a soft computation method introduced previously, was successfully used in solvent selection for this reaction.



Scheme 1

3.1. Molecular Sieves Promoting formation of γ-lactam

γ-lactams can be synthesized from amino acids3 (Scheme 1) under traditional condition by using water as solvent. However, this traditional method not only requires long reaction time, but also gives a certain percentage of racemized products for chiral amino acids. For example, producing (D) or (L)-pyroglutamic acid from (D) or (L)-glutamic acid give 15% racemized product.

In order to avoid racemization, we tried various reaction conditions and found that reaction time is key to optical selectivity. To speed up reaction, we added 4 Å molecular sieves (10 wt%) to above reaction mixture. As expected, chemical yield increased from 45% to 58% (Table 1). Meanwhile, racemized product decreased from 15% to 10%.

Production	Yield	
	Without MS	With MS
(D)-pyroglutamic acid	45	58
(L)-pyroglutamic acid	45	59

Table 1: The yield affected by Molecular sieves in reaction of (D) and (L)-pyroglutamic acid formation.

3.2. Microvave Irradiation in formation of γ-lactam

Although the yield increased 14% and optical purity increased 5% with molecular sieves, it is still not suited for scale-up.

Microwave technology using in chemical reaction becomes more and more popular in recent years4. Microwave can promote reaction. With this idea, we did a set of comparison experiments under 900 W microwave irradiation. The result showed that the reaction time of pyroglutamic acid formation in water decreased from 48 hrs to 2 hrs and the yield also was improved. Combining MS and microwave irradiation, the reaction time further decreased 30 minutes and the yield was improved. Only 2 to 3% racemized product was detected (Table 2).

Production	Yield(%)		Time (min)	
	Without MS	With MS	Without MS	With MS
(D)-pyroglutamic acid	65	78	120	30
(L)-pyroglutamic acid	64	77	120	30

Table 2: The microwave effection in the formation (D) and (L)-pyroglutamic acid.

3.3. Rough Set Theory in Solvent Selection

Rough Set Theory may predict the suitable solvents. According to RST, we created a Rough decision table (Table 3)

Solvents	ε	Cp	μ	Oa
acetic acid	6.15	29.84	1.68	…
acetone	20.71	17.73	2.69	…
acetonitrile	37.52	12.48	3.44	…
benzene	2.28	19.52	0.00	…
chlorobenzene	7.33	0.39	1.54	…
chloroform	4.92	0.23	1.15	…
1,2- chloride	7.35	0.31	1.86	…
ether	4.19	0.55	1.12	…
DMF	38.25	37.45	3.86	…
H_2O	78.30	3.54	6.47	…
ethanol	23.80	0.58	1.68	…
dioxane	2.21	1.72	0.00	…
Dimethyl sulfoxide	48.91	36.40	4.30	…
…	…	…	…	…

Table 3: The physical constant of the selected solvents.

We only choose 8 attributes to describe the solvents and only for 20 objectives. Based on the rough set method introduced previously, or the software RIDAS to compute the final decision rules.

At last we got some rules, under the real environment we chose one and explained as follows: if $\varepsilon \geq 35 \wedge Cp \geq 35 \wedge \mu \geq 3$, their corresponding solvents' microwave effect should be strong.

We assumed: the stronger the microwave effect is, the higher the yield is. The microwave effect of thirteen solvents was compared. Based on RST calculation, DMSO and DMF should display high microwave effect. Following order was obtained for selected solvents: DMSO > DMF > CH_3CN> H_2O >toluene > dioxane.

To prove the calculation, we repeated the cyclization of glutamic acid with microwave irradiation and catalytic amount of molecular sieves with DMSO and DMF as solvents. The result is

encouraging: over 90% yield was obtained and no racemized product was detected in DMSO (Table 4).

Production	Yield\Time (in DMF)	Yield\Time(in DMSO)
(D)-pyroglutamic acid	86%\30min	95%\5min
(L)-pyroglutamic acid	87%\30min	95%\5min

Table 4: The solvent effected in the formation (D) and (L)-pyroglutamic acid.

Synthesis of serial of functional γ-lactams from studied. (The experiments results were shown in Table 5.).Substituted 4-aminobutyric acids in DMSO as solvent was studied. (The experiments results were shown in Table 5.).

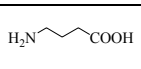
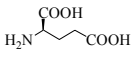
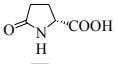
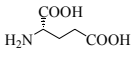
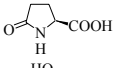
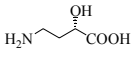
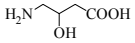
entry	substrate	Product	Yield(%)	ee%
1			96	racemic
2			95	99
3			95	98
4			87	99
5			85	racemic

Table 5: Substituted 4-aminiobutyric acid to γ-lactam under microwave radiation in DMSO promoted by molecular sieves

The yield for above cyclization under combining microwave irradiation and molecular sieves is very good6. Most importantly, no racemization was observed for entry 2, 3 and 4.

In conclusion, γ-lactams were synthesized directly from substituted 4-aminiobutyric acids under microwave radiation with high yield and without racimazition. It has advantage of short reaction time, more economic and environmental friendly than traditional methods. Two solvents were selected by RST and showed good activity in synthesis of γ-lactams.

4. Conclusions

RST is a methodology which has demonstrated its usefulness in the multi-attribute decision making, and the above case study has shown that solvents selection based on rough set theory can optimize the reaction process, reduce the cost in chemical research and production.

Acknowledgement

This work is partially supported by the SWJTU Doctoral Innovation Fund(2005JCZ), the Sichuan Province Education Dept. Nature Science Youth Fund(2006B090), and the MEPRC Doctoral Science Research Fund(20060613019).

References

[1] Z. Pawlak, Rough sets, *Int. J. Inf. Comput. Sci.* 11: 341-356,1982

[2] Z. Pawlak, Rough sets. *Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publisher, ,1991.

[3] G. M. B. Anthony, H. John, L. S. Marie, Nicholas S S, White A J P, and. Williams D J, *J Org Chem*, 64: 600-605, 1999.

[4] Jiang Chaozhe, Hupei, Xufang. Survey of Rough Set theory and application, *Computer Science and Practice*. 180:160-170,2004

[5] Turui, Jiang Chaozhe, Wurui, The application of Rough set theory in chemical synthesis1, *Computer Science*. 31: 124-127,2004.

[6] M. Mary and H. Paul, *Tetrahedron Letter*, 29: 3049-3056,1998.

[7]. P. T. Anastas, J. C. Warner, Green Chemistry, *Theory and Practice*, Oxford University Press, 1998.