Association Rule Granulation using Rough Sets on Intuitionistic Fuzzy Approximation Spaces and Granular Computing

B. K. Tripathy and D. P. Acharjya
School of Computing Sciences and Engineering
VIT University, Vellore, India

ABSTRACT: Association rule granulation is a common data mining technique used to extract knowledge from the universe. To characterize the elements of the universe and to extract knowledge about the universe we classify the elements of the universe based on indiscernibility relation. However, in many information systems we find numerical attribute values that are almost similar instead of full identical. To handle such type of information system, we use (α, β) - indiscernibility due to rough set on intuitionistic fuzzy approximation space with ordering rules. So, the classification results in a set of classes called granules and are basic building blocks of the knowledge about the universe. Granular computing processes these granules and produces possible patterns and associations. In this paper we study the concept of almost indiscernibility and knowledge granulation along with the association rules using granular computing method. We process the granules obtained as the result of classification and find the association rules between them.
KEYWORDS: Association rules, indiscernibility, rough sets, ordering rules, granule, granular computing.

Introduction

Data are being collected and accumulated at a dramatic pace, especially at the age of internet. It is very hard for human being to find knowledge about the universe that is hidden in the accumulated voluminous data. Thus finding knowledge about the universe is one of the most popular areas of recent research. Researchers proposed many methods to mine knowledge from the rapidly growing voluminous digital data. Most of our traditional tools to mine knowledge are crisp, deterministic, and precise in character. Real life situations are quite opposite to that. For a complete description of a real life system often one should require by far more detailed data than a human being could recognize simultaneously, process and understand. It leads to the extension of the concept of crisp sets so as to model imprecise data that can enhance their modeling power.

The first notion in this direction of information granulation, by L. A. Zadeh [Zad94, Zad96a, Zad96b] is very important for logic, computer science, philosophy and others. The basic idea behind this is to process vagueness and incompleteness.

Another approach in this direction to capture impreciseness is the notion of rough sets [Paw91, PS07a, PS07b, PS07c] introduced by Z. Pawlak. This is an effective tool for uncertainty management, and it has wide applications in artificial intelligence [DG98, SC02, JKJ06, Reb06, ZPZ06, JQ07]. In this notion it is assumed that information is associated with every object and objects can be seen through information only. Therefore, object with same information cannot be discerned and appear as same. This leads to indiscernibility of objects of the universe that form clusters of indistinguishable objects. Hence, according to rough set the information granulation is due to indiscernibility of objects caused by lack of information about them. Equivalence relation is the simplest representation of the indiscernibility and is sufficient for many real life applications [Paw98]. However, in many applications transitivity of equivalence relation is not being satisfied. Therefore, indiscernibility relation is formalized as a tolerance relation [SS96].

Universe can be considered as a large collection of objects. There is some information associated with each object. To find knowledge about the universe we need to extract some information about these objects. We can not uniquely identify some objects due to the lack of sufficient information about the objects of the universe. Therefore, we require classification of these objects into similarity classes to characterize these objects and to extract knowledge about the universe. Each class is called a granule which is a basic building block of the knowledge about the universe. The classification of the objects of the universe is done based on indiscernibility relation among the objects. It indicates that objects of a class can not discerned from one another based on available set of attributes of the objects. Creating granules from the objects of the universe as the result of the classification is called knowledge granulation (granularity of the knowledge). Therefore, the concept of indiscernibility is prior to the concept of granularity [Zad94, Zad96a, Zad96b, Paw98].

It is also seen that in real life data, two different objects x and y may have attribute values that are almost identical, if not exactly identical. It indicates that attribute values could be identical up to certain extent. We consider an information granulation where attribute values are not always quantitative, rather subjective having imprecise meanings. This indiscernibility relation on different domains of attributes is formalized as a fuzzy tolerance relation or fuzzy proximity relation [De99]. Thus from rough set on fuzzy approximation space point of view information granulation is due to almost indiscernibility relation of objects caused by insufficient information about them.

Processing the granules to find knowledge about the universe is called granular computing. Granular computing processes these granules and produces possible patterns and associations among the attribute values of the objects of the universe. Association rule [LLT05] is a type of relationship among attribute values of the objects of the universe.

In this paper we propose an association rule granulation model that use rough set on intuitionistic fuzzy approximation spaces, ordering rules and granular computing to establish associations between the attribute values of the objects of the universe of discourse.

1. Rough sets on intuitionistic fuzzy approximation spaces

Rough set on intuitionistic fuzzy approximation space is founded on the assumption that in real life applications with every object we associate some information (data, knowledge). This information is not exactly identical but almost identical. This is because objects characterized by the almost same information are almost indiscernible in the view of available information. For example, if objects are patients suffering from certain disease, symptoms of the disease form information about patients. These symptoms are almost identical rather full identical. At this point we generalize Pawlak's approach of indiscernibility [Paw98]. Keeping view to this the almost indiscernibility relation generated in this way is the mathematical basis of rough set on intuitionistic fuzzy approximation space.

Any set of all almost indiscernible objects is called an elementary concept, and forms a granule (atom) of knowledge about the universe. To decide amount of identity between two attribute values, we use intuitionistic fuzzy proximity relation on each domain of attributes. This domain of attributes helps the universal set to create an intuitionistic fuzzy approximation space on the universe.

The motivation behind the study is that the notion of almost indiscernibility relation is a generalization of indiscernibility relation with the use of intuitionistic fuzzy proximity relation and is mainly applicable on information systems where the attribute values are subjective having vague or imprecise meaning rather than quantitative. Also, the notion of intuitionistic fuzzy approximation space defined with the help of intuitionistic fuzzy proximity relation is a generalization of Pawlak's [Paw91] approximation space. Therefore, the rough set on intuitionistic fuzzy approximation space [Tri06] generalizes the Pawlak's [Paw91] rough set and provides better result in real life problems under study.

However, for completeness of the paper we provide the basic notions of rough sets on intuitionistic fuzzy approximation spaces. We use standard notation μ for membership and ν for non-membership functions associated with an intuitionistic fuzzy set.

Definition 1.1 An intuitionistic fuzzy relation R on a universal set U is an intuitionistic fuzzy set defined on $U \times U$.

Definition 1.2 An intuitionistic fuzzy relation R on U is said to be an intuitionistic fuzzy proximity relation if the following properties hold.

$$\mu_R(x, x) = 1 \text{ and } \nu_R(x, x) = 0 \quad \forall\, x \in U$$

$$\mu_R(x, y) = \mu_R(y, x), \nu_R(x, y) = \nu_R(y, x) \quad \forall\, x, y \in U$$

Definition 1.3 Let R be an intuitionistic fuzzy (IF) proximity relation on U. Then for any $(\alpha, \beta) \in J$, where $J = \{(\alpha, \beta) \mid \alpha, \beta \in [0,1]$ and $0 \le \alpha + \beta \le 1\}$, the (α, β)-cut '$R_{\alpha,\beta}$' of R is given by

$$R_{\alpha,\beta} = \{(x, y) \mid \mu_R(x, y) \ge \alpha \text{ and } \nu_R(x, y) \le \beta\}$$

Definition 1.4 Let R be an IF-proximity relation on U. We say that two elements x and y are (α, β)-similar with respect to R if $(x, y) \in R_{\alpha,\beta}$ and we write $x R_{\alpha,\beta}\, y$.

Definition 1.5 Let R is an IF-proximity relation on U. We say that two elements x and y are (α, β)-identical with respect to R for $(\alpha, \beta) \in J$, written as $xR(\alpha, \beta)y$ if and only if $x R_{\alpha,\beta}\, y$ or there exists a sequence of elements $u_1, u_2, u_3, \cdots, u_n$ in U such that $x R_{\alpha,\beta} u_1, u_1 R_{\alpha,\beta} u_2, u_2 R_{\alpha,\beta} u_3, \cdots, u_n R_{\alpha,\beta} y$. In the last case, we say that x is transitively (α, β)-similar to y with respect to R.

It is also easy to see that for any $(\alpha, \beta) \in J$, $R(\alpha, \beta)$ is an equivalence relation on U. We denote $R^*_{\alpha,\beta}$ the set of equivalence classes generated by the equivalence relation $R(\alpha, \beta)$ for each fixed $(\alpha, \beta) \in J$.

Definition 1.6 Let U be a universal set and R be an intuitionistic fuzzy proximity relation on U. The pair (U, R) is an intuitionistic fuzzy approximation space (IF-approximation space). An IF-approximation space (U, R) generates usual approximation space $(U, R(\alpha, \beta))$ of Pawlak for every $(\alpha, \beta) \in J$.

Definition 1.7 The rough set on X in the generalized approximation space $(U, R(\alpha, \beta))$ is denoted by $(\underline{X}_{\alpha,\beta}, \overline{X}_{\alpha,\beta})$, where

$$\overline{X}_{\alpha,\beta} = \bigcup\{Y \mid Y \in R^*_{\alpha,\beta} \text{ and } Y \subseteq X\}$$

and $$\overline{X}_{\alpha,\beta} = \bigcup\{Y \mid Y \in R^*_{\alpha,\beta} \text{ and } Y \cap X \neq \phi\}$$

Definition 1.8 Let X be a rough set in the generalized approximation space $(U, R(\alpha, \beta))$. Then we define the (α, β)-boundary of X with respect to R denoted by $BNR_{\alpha,\beta}(X)$ as $BNR_{\alpha,\beta}(X) = \overline{X}_{\alpha,\beta} - \underline{X}_{\alpha,\beta}$.

Definition 1.9 Let X be a rough set in the generalized approximation space $(U, R(\alpha, \beta))$. Then X is (α, β)*-discernible* with respect to R if and only if $\overline{X}_{\alpha,\beta} = \underline{X}_{\alpha,\beta}$ and X is (α, β)*-rough* with respect to R if and only if $\overline{X}_{\alpha,\beta} \neq \underline{X}_{\alpha,\beta}$.

2. Almost indiscernibility and knowledge granulation

From the previous section, it is clear that rough set on intuitionistic fuzzy approximation space is due to almost indiscernibility relation generated by information about the objects of interest. This is because we are unable to discern some objects up to certain extent. Therefore, we have to consider clusters of almost indiscernible objects as fundamental concepts of knowledge.

Let U be the universe and A be a set of attributes. With each attribute $x \in A$ we associate a set of its values V_x, called the domain of x. The pair $S = (U, A)$ will be called an information system. Let $B \subseteq A$. For a chosen $\alpha, \beta \in J$ we denote a binary relation $I_B(\alpha, \beta)$ on U defined by $x\, I_B(\alpha, \beta)\, y$ if and only if $f_x(a)\, I(\alpha, \beta)\, f_y(a)$ for all $a \in B$, where $f_x : U \to V_x$ defined by $f_x(a) = f(x, a)$ denotes the value of attribute a for element x. Obviously, it can be proved that the relation $I_B(\alpha,\beta)$ is an equivalence relation on U. Also, we notice that $I_B(\alpha,\beta)$ is not exactly the indiscernibility relation defined by Pawlak [Paw98]; rather it can be viewed as an *almost indiscernibility relation* on U. For $\alpha = 1, \beta = 0$ the almost indiscernibility relation, $I_B(\alpha,\beta)$ reduces to the indiscernibility relation. Thus, it generalizes the Pawlak's indiscernibility relation. The family of all equivalence classes of $I_B(\alpha,\beta)$, *i.e.*, the partition generated by B for $\alpha,\beta \in J$, will be denoted by $U/I_B(\alpha,\beta)$; an equivalence class of $I_B(\alpha,\beta)$, *i.e.,* the block of the partition $U/I_B(\alpha,\beta)$, containing x will denoted by $B_{\alpha,\beta}(x)$.

If $(x, y) \in I_B(\alpha,\beta)$, then we will say that x and y are $B_{\alpha,\beta}$-indiscernible. Blocks of the partition $U/I_B(\alpha,\beta)$ (or the equivalence classes of the relation $I_B(\alpha,\beta)$) are referred as $B_{\alpha,\beta}$ *– granules* or $B_{\alpha,\beta}$-*elementary concepts*. These are the basic building concepts of our knowledge in the rough set on intuitionistic fuzzy approximation space.

3. Ordered information table

Let $I = (U, A, \{V_x : x \in A\}, \{f_x : x \in A\})$ be an information system, where U is a finite non-empty set of objects called the universe and A is a non-empty finite set of attributes. For every $x \in A$, V_x is the set of values that attribute x may take and $f_x : U \to V_x$ is an information function. In practical applications, there are various possible interpretations of objects such as cases, states, patients, processes, and observations. Attributes can be interpreted as features, variables, and characteristics. A special case of information systems called information table or attribute value table where the columns are labeled by attributes and rows are by objects. For example: consider the information Table 1, in which we have $V_{R\&D} = \{\text{Yes, No}\}$. Similarly, $V_{\text{SoA}} = \{\text{Yes, No}\}$, $V_{\text{Mkt}} = \{\text{High, Very high, Average}\}$, and $V_{\text{Profit}} = \{200, 300, 250\}$.

An information table represents all available information and knowledge about the objects under consideration. Objects are only perceived or measured by using a finite number of properties. At the same time, it does not consider any semantic relationships between distinct values of a particular attribute [Yao00]. Different values of the same attribute are considered as distinct symbols without any connections, and therefore on simple pattern matching we consider horizontal analyses to a large extent. Hence, in general one uses the trivial equality relation on values of an attribute as discussed in standard rough set theory [Paw91].

Table 1. Information table

Object	**Availability of R&D facility** (a_1)	**Adoption of state of art facility** (a_2)	**Marketing expenses** (a_3)	**Profits in million of rupees** (a_4)
o_1	No	Yes	High	200
o_2	Yes	No	High	300
o_3	yes	Yes	Average	200
o_4	No	Yes	Very high	250
o_5	Yes	No	High	300
o_6	No	Yes	Very high	250

However, in many real life applications it is observed that the attribute values are not exactly identical but almost identical. This is because objects characterized by the almost same information are almost indiscernible in the view of available information. For example, if objects are patients suffering from certain disease, symptoms of the disease form information about patients. These symptoms are almost identical rather full identical. At this point we generalize Pawlak's approach of indiscernibility. Keeping view to this, the almost indiscernibility relation generated in this way is the mathematical basis of rough set on intuitionistic fuzzy approximation space as discussed in the previous section. Generalized information table may be viewed as information tables with added semantics. For the problem of knowledge mining, we introduce order relations on attribute values [YY01]. However, it is not appropriate in case of attribute values that are almost indiscernible. In this paper we use rough sets on intuitionistic fuzzy approximation space to find the attribute values that are (α, β)-identical before introducing the order relation. This is because exactly ordering is not possible when the attribute values are almost identical. For $\alpha = 1$, $\beta = 0$ the almost indiscernibility relation, reduces to the indiscernibility relation. Therefore, it generalizes the Pawlak's indiscernibility relation.

Generalized information tables may be viewed as information tables with added semantics. For association rule granulation, when attribute values are almost identical, we introduce order relation on attribute values. An ordered information table is defined as $\text{OIT} = \{\text{IT}, \{\prec_x : x \in A\}\}$ where, IT is a standard information table and $\prec_x$ is an order relation on attribute *x*. An ordering of values of a particular attribute *x* naturally induces an ordering of objects:

$$a \prec_{\{x\}} b \Leftrightarrow f_x(a) \prec_x f_x(b)$$

where, $\prec_{\{x\}}$ denotes an order relation on *U* induced by the attribute *x*. An object o_i is ranked ahead of object o_j if and only if the value of o_i on the attribute x is ranked ahead of the value of o_j on the attribute x. For example, information Table 1 becomes order information table on introduction of the following ordering relations.

$\prec_{a_1}$: Yes $\prec$ No

$\prec_{a_2}$: Yes $\prec$ No

$\prec_{a_3}$: Very high $\prec$ High $\prec$ Average

$\prec_{a_4}$: 300 $\prec$ 250 $\prec$ 200

For a subset of attributes $P \subseteq A$, we define:

$$a \prec_P b \Leftrightarrow f_x(a) \prec_x f_x(b) \quad \forall\, x \in P$$
$$\Leftrightarrow \bigwedge_{x \in P} f_x(a) \prec_x f_x(b)$$
$$\Leftrightarrow \bigcap_{x \in P} \prec_{\{x\}}$$

It indicates that a is ranked ahead of b if and only if a is ranked ahead of b according to all attributes in P. The above definition is a straightforward generalization of the standard definition of equivalence relations in rough set theory [Paw82, Paw91], where the equality relation is used. Knowledge mining based on order relations is a concrete example of applications on generalized rough set model with non equivalence relations [Yao98, YL96].

4. Association rule granulation model

In this section, we propose our association rule granulation model in figure 1 and show how granular computing is used in finding association rules from the information table or information system containing almost indiscernible attribute values. In granular computing, a set of attribute values is called an association rule if they satisfy certain criteria. If a set of attribute values is an association rule, all the attribute values in the set are said to be associated with one another. To find association rules we find granules based on each attribute. To check whether a set of attribute values is an association rule or not, we perform the AND operation among the bit representation of the attribute values of this set. If number of 1's in the result of AND operation is greater than or equal to minimum support, then it is an association rule, otherwise it is not an association rule.

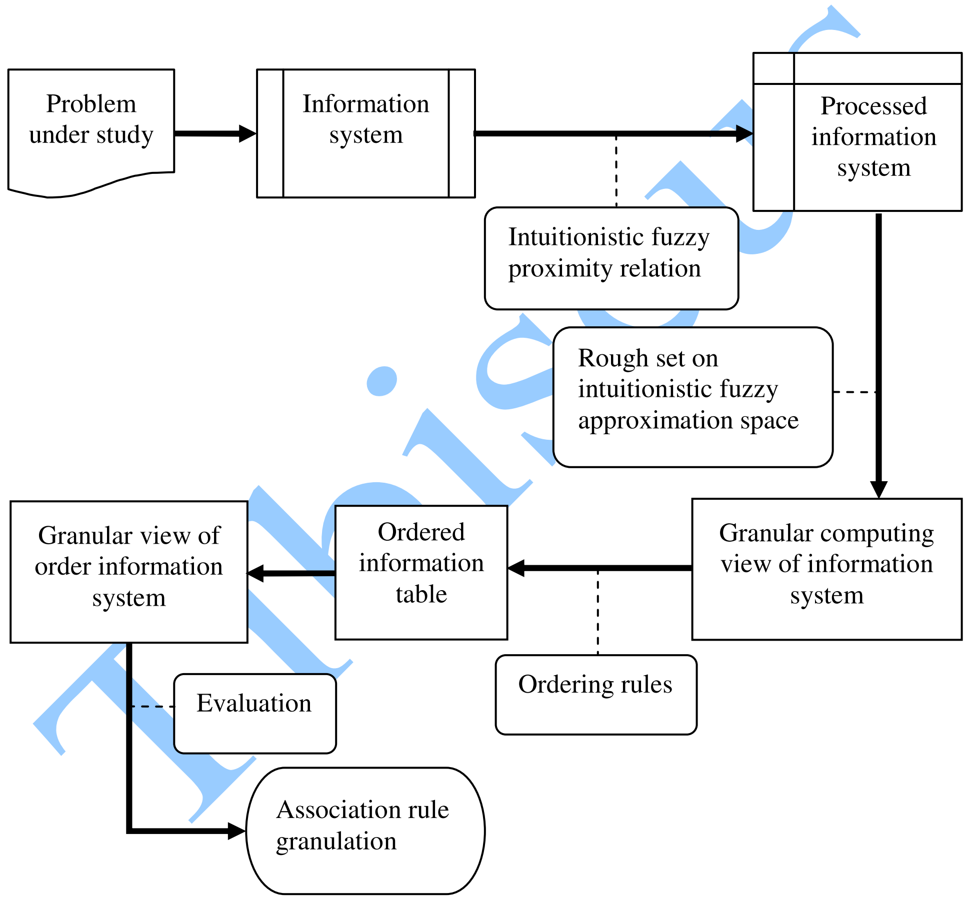

Figure 1. Association rule granulation model

The above proposed association rule granulation model consists of problem definition, information system, processed information system, granular view of the processed information system, ordered information table, and association rule granulation. Problem definition and incorporation of prior knowledge are the fundamental steps of any model in which we identify the right problem. The potential validity or usefulness of an individual data element or pattern of data elements may change dramatically from individual to individual, organization to organization, or task to task because of the acquisition of knowledge and reasoning may involve in vagueness and incompleteness. It is very difficult for common human to obtain useful information that is hidden in the accumulated voluminous data. Finding association between attribute values generally understandable is also highly problematic. There is much need for dealing with the incomplete and vague information in classification, concept formulation, and data analysis. To this end here we use intuitionistic fuzzy proximity relation for processing data and data classification after removal of noise and missing data. This step takes the most time needed for computing the association rule extracting process. This processed information system is further processed by using rough set on intuitionistic fuzzy approximation space to give a granular computing view of the information syatem. The resultant information system is further processed by ordering rules to produce ordered information table. Finally, we extract the association rules between attribute values with the help of evaluation and interpretation.

5. Association rule granulation using granular computing

In this section, to demonstrate our model we take into consideration an information system of a group of companies where we study the business strategies of different cosmetic companies in a country. In the Table 2 given below, we consider few parameters for business strategies to get maximum sales; their possible range of values and an intuitionistic fuzzy proximity relation which characterized the relationship between parameters. The companies with more expenditure in marketing, more expenditure in advertisement, more expenditure in distribution, more expenditure in miscellaneous, and more expenditure on research and development is being an ideal case. But such type of cases is rare in practice. So, it is essential to study the association rules between the attribute values of a company to get the maximum sales.

The membership and non-membership functions have been adjusted such that the sum of their values should lie in [0, 1] and also these functions must be symmetric. The first requirement necessitates a major of 2 in the denominators of the non-membership functions.

The companies can be judged by the sales outputs that are produced. The amount of sales can be judged by the different parameters of the companies. These parameters forms the attribute set for our analysis. Here the marketing expenditure means, all expenditure incurred for corporate promotion, which includes event marketing, sales promotion, direct marketing etc. which comes to around 6%. The advertising expenditure includes promotional activities using various medium like television, newspaper, internet etc. which comes around 36%. The miscellaneous expenditure is mainly incurred through activities like corporate social responsibility and it leads to maximum of 28%. The distribution cost includes expenses on logistic, supply chain etc. and it comes around 24%. The investment made on new product development and other research activities are taken on research and development activities and it takes around 6%. However we have not considered many other parameters that do not influence the sales of a company and to make our analysis simple. The average of the data collected is considered to be the representative figure and tabulated below. The notations and abbreviations used in the following analysis are presented in Table 2.

Table 2. Attribute description table

Attribute	Notation	Possible range	Membership function	Non-membership function
Expenditure on marketing	EM (a_1)	1 – 250	$1-\frac{\lvert x-y\rvert}{250}$	$\frac{\lvert x-y\rvert}{2(x+y)}$
Expenditure on distribution	ED (a_2)	1 – 200	$1-\frac{\lvert x-y\rvert}{200}$	$\frac{\lvert x-y\rvert}{2(x+y)}$
Expenditure on advertisement	EA (a_3)	1 – 385	$1-\frac{\lvert x-y\rvert}{385}$	$\frac{\lvert x-y\rvert}{2(x+y)}$
Expenditure on miscellaneous	EMi (a_4)	1 – 200	$1-\frac{\lvert x-y\rvert}{200}$	$\frac{\lvert x-y\rvert}{2(x+y)}$
Sales unit	SU(a_5)	1 – 60	$1-\frac{\lvert x-y\rvert}{60}$	$\frac{\lvert x-y\rvert}{2(x+y)}$
Expenditure on research and development	ERD (a_6)	1 – 80	$1-\frac{\lvert x-y\rvert}{80}$	$\frac{\lvert x-y\rvert}{2(x+y)}$

In Table 3 we present the data obtained from ten different companies. However, we keep confidential their identity due to various official reasons. Here, we use the notation C_i, $i=1,2,3,\cdots,10$ for different companies for the purpose of the study to demonstrate the method and not to probe the performance of individual company. It is to be noted that, in the information table all non-ratio figures shown in the table are ten million INR.

Table 3. Small universe of information system

Name of the company	EM (a_1)	ED (a_2)	EA (a_3)	EMi (a_4)	SU (a_5)	ERD (a_6)
C_1	229	151	304	169	56	49
C_2	227	143	298	169	53	79
C_3	226	145	266	167	54	63
C_4	191	110	316	163	41	64
C_5	179	117	247	160	53	53
C_6	148	102	180	147	43	27
C_7	131	78	138	145	46	25
C_8	124	61	130	142	38	9
C_9	88	58	121	143	40	34
C_{10}	92	48	100	137	32	2

Intuitionistic fuzzy proximity relation R_1 corresponding to attribute 'EM' is given in Table 4.

Table 4. Intuitionistic fuzzy proximity relation for attribute EM

R_1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
C_1	1, 0	0.992 0.002	0.989 0.003	0.848 0.045	0.802 0.061	0.676 0.108	0.610 0.135	0.581 0.149	0.437 0.222	0.451 0.214
C_2	0.992 0.002	1, 0	0.997 0.001	0.856 0.043	0.810 0.058	0.684 0.106	0.618 0.133	0.589 0.147	0.445 0.22	0.460 0.212
C_3	0.989 0.003	0.997 0.001	1, 0	0.859 0.042	0.813 0.058	0.686 0.105	0.621 0.132	0.591 0.146	0.448 0.219	0.462 0.211
C_4	0.848 0.045	0.856 0.043	0.859 0.042	1, 0	0.954 0.016	0.827 0.064	0.762 0.092	0.732 0.106	0.589 0.184	0.603 0.175
C_5	0.802 0.061	0.810 0.058	0.813 0.058	0.954 0.016	1, 0	0.873 0.048	0.808 0.077	0.778 0.091	0.635 0.170	0.649 0.162
C_6	0.676 0.108	0.684 0.106	0.686 0.105	0.827 0.064	0.873 0.048	1, 0	0.935 0.029	0.905 0.044	0.762 0.126	0.776 0.117
C_7	0.610 0.135	0.618 0.133	0.621 0.132	0.762 0.092	0.808 0.077	0.935 0.029	1, 0	0.970 0.014	0.827 0.098	0.841 0.089
C_8	0.581 0.149	0.589 0.147	0.591 0.146	0.732 0.106	0.778 0.091	0.905 0.044	0.970 0.014	1, 0	0.857 0.084	0.871 0.075
C_9	0.437 0.222	0.445 0.22	0.448 0.219	0.589 0.184	0.635 0.170	0.762 0.126	0.827 0.098	0.857 0.084	1, 0	0.986 0.01
C_{10}	0.451 0.214	0.460 0.212	0.462 0.211	0.603 0.175	0.649 0.162	0.776 0.117	0.841 0.089	0.871 0.075	0.986 0.01	1, 0

Intuitionistic fuzzy proximity relation R_2 corresponding to attribute 'ED' is given in Table 5. Intuitionistic fuzzy proximity relation R_3 corresponding to attribute 'EA' is given in Table 6.

Table 5. Intuitionistic fuzzy proximity relation for attribute ED

R_2	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
C_1	1, 0	0.956 0.015	0.966 0.011	0.792 0.080	0.827 0.065	0.751 0.098	0.632 0.161	0.549 0.212	0.532 0.224	0.485 0.258
C_2	0.956 0.015	1, 0	0.990 0.004	0.836 0.065	0.871 0.050	0.795 0.084	0.676 0.147	0.593 0.2	0.576 0.212	0.529 0.246
C_3	0.966 0.011	0.990 0.004	1, 0	0.825 0.069	0.86 0.054	0.785 0.087	0.666 0.150	0.583 0.203	0.565 0.215	0.519 0.249
C_4	0.792 0.080	0.836 0.065	0.825 0.069	1, 0	0.965 0.015	0.96 0.019	0.84 0.085	0.757 0.142	0.74 0.155	0.693 0.194
C_5	0.827 0.065	0.871 0.050	0.86 0.054	0.965 0.015	1, 0	0.925 0.034	0.806 0.1	0.723 0.156	0.705 0.169	0.659 0.207
C_6	0.751 0.098	0.795 0.084	0.785 0.087	0.96 0.019	0.925 0.034	1, 0	0.881 0.066	0.798 0.124	0.78 0.138	0.734 0.177
C_7	0.632 0.161	0.676 0.147	0.666 0.150	0.84 0.085	0.806 0.1	0.881 0.066	1, 0	0.917 0.060	0.9 0.074	0.853 0.116
C_8	0.549 0.212	0.593 0.2	0.583 0.203	0.757 0.142	0.723 0.156	0.798 0.124	0.917 0.060	1, 0	0.983 0.015	0.936 0.058
C_9	0.532 0.224	0.576 0.212	0.565 0.215	0.74 0.155	0.705 0.169	0.78 0.138	0.9 0.074	0.983 0.015	1, 0	0.954 0.044
C_{10}	0.485 0.258	0.529 0.246	0.519 0.249	0.693 0.194	0.659 0.207	0.734 0.177	0.853 0.116	0.936 0.058	0.954 0.044	1, 0

Table 6. Intuitionistic fuzzy proximity relation for attribute EA

R_3	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
C_1	1, 0	0.987 0.004	0.904 0.033	0.967 0.01	0.853 0.052	0.678 0.128	0.57 0.187	0.55 0.2	0.525 0.215	0.472 0.252
C_2	0.987 0.004	1, 0	0.917 0.028	0.954 0.015	0.866 0.047	0.691 0.124	0.583 0.184	0.563 0.196	0.539 0.212	0.485 0.249
C_3	0.904 0.033	0.917 0.028	1, 0	0.87 0.043	0.949 0.019	0.774 0.097	0.666 0.159	0.646 0.172	0.622 0.188	0.568 0.227
C_4	0.967 0.01	0.954 0.015	0.87 0.043	1, 0	0.819 0.062	0.645 0.138	0.537 0.196	0.517 0.208	0.492 0.224	0.439 0.259
C_5	0.853 0.052	0.866 0.047	0.949 0.019	0.819 0.062	1, 0	0.825 0.079	0.717 0.141	0.697 0.154	0.673 0.171	0.619 0.211
C_6	0.678 0.128	0.691 0.124	0.774 0.097	0.645 0.138	0.825 0.079	1, 0	0.892 0.065	0.872 0.079	0.847 0.098	0.794 0.142
C_7	0.57 0.187	0.583 0.184	0.666 0.159	0.537 0.196	0.717 0.141	0.892 0.065	1, 0	0.98 0.014	0.955 0.033	0.902 0.079
C_8	0.55 0.2	0.563 0.196	0.646 0.172	0.517 0.208	0.697 0.154	0.872 0.079	0.98 0.014	1, 0	0.975 0.019	0.922 0.065
C_9	0.525 0.215	0.539 0.212	0.622 0.188	0.492 0.224	0.673 0.171	0.847 0.098	0.955 0.033	0.975 0.019	1, 0	0.947 0.046
C_{10}	0.472 0.252	0.485 0.249	0.568 0.227	0.439 0.259	0.619 0.211	0.794 0.142	0.902 0.079	0.922 0.065	0.947 0.046	1, 0

Intuitionistic fuzzy proximity relation R_4 corresponding to attribute 'EMi' is given in Table 7. Intuitionistic fuzzy proximity relation R_5 corresponding to attribute 'SU' is given in Table 8.

Table 7. Intuitionistic fuzzy proximity relation for attribute EMi

R_4	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
C_1	1, 0	1, 0	0.99 0.003	0.97 0.009	0.955 0.014	0.89 0.035	0.88 0.038	0.865 0.043	0.87 0.042	0.84 0.052
C_2	1, 0	1, 0	0.99 0.003	0.97 0.009	0.955 0.014	0.89 0.035	0.88 0.038	0.865 0.043	0.87 0.042	0.84 0.052
C_3	0.99 0.003	0.99 0.003	1, 0	0.98 0.006	0.965 0.011	0.9 0.032	0.89 0.035	0.875 0.04	0.88 0.039	0.85 0.049
C_4	0.97 0.009	0.97 0.009	0.98 0.006	1, 0	0.985 0.005	0.92 0.026	0.91 0.029	0.895 0.034	0.9 0.033	0.87 0.043
C_5	0.955 0.014	0.955 0.014	0.965 0.011	0.985 0.005	1, 0	0.935 0.021	0.925 0.025	0.91 0.03	0.915 0.028	0.885 0.039
C_6	0.89 0.035	0.89 0.035	0.9 0.032	0.92 0.026	0.935 0.021	1, 0	0.99 0.003	0.975 0.009	0.98 0.007	0.95 0.018
C_7	0.88 0.038	0.88 0.038	0.89 0.035	0.91 0.029	0.925 0.025	0.99 0.003	1, 0	0.985 0.005	0.99 0.003	0.96 0.014
C_8	0.865 0.043	0.865 0.043	0.875 0.04	0.895 0.034	0.91 0.03	0.975 0.009	0.985 0.005	1, 0	0.995 0.002	0.975 0.009
C_9	0.87 0.042	0.87 0.042	0.88 0.039	0.9 0.033	0.915 0.028	0.98 0.007	0.99 0.003	0.995 0.002	1, 0	0.97 0.011
C_{10}	0.84 0.052	0.84 0.052	0.85 0.049	0.87 0.043	0.885 0.039	0.95 0.018	0.96 0.014	0.975 0.009	0.97 0.011	1, 0

Table 8. Intuitionistic fuzzy proximity relation for attribute ERD

R_5	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
C_1	1, 0	0.95 0.014	0.967 0.009	0.75 0.077	0.95 0.014	0.783 0.066	0.833 0.049	0.7 0.096	0.733 0.083	0.6 0.136
C_2	0.95 0.014	1, 0	0.983 0.005	0.8 0.064	1, 0	0.833 0.052	0.883 0.035	0.75 0.082	0.783 0.07	0.65 0.124
C_3	0.967 0.009	0.983 0.005	1, 0	0.783 0.068	0.983 0.005	0.817 0.057	0.867 0.04	0.733 0.087	0.767 0.074	0.633 0.128
C_4	0.75 0.077	0.8 0.064	0.783 0.068	1, 0	0.8 0.064	0.967 0.012	0.917 0.029	0.95 0.019	0.983 0.006	0.85 0.062
C_5	0.95 0.014	1, 0	0.983 0.005	0.8 0.064	1, 0	0.833 0.052	0.883 0.035	0.75 0.082	0.783 0.07	0.65 0.124
C_6	0.783 0.066	0.833 0.052	0.817 0.057	0.967 0.012	0.833 0.052	1, 0	0.95 0.017	0.917 0.031	0.95 0.018	0.817 0.073
C_7	0.833 0.049	0.883 0.035	0.867 0.04	0.917 0.029	0.883 0.035	0.95 0.017	1, 0	0.867 0.048	0.9 0.035	0.767 0.09
C_8	0.7 0.096	0.75 0.082	0.733 0.087	0.95 0.019	0.75 0.082	0.917 0.031	0.867 0.048	1, 0	0.967 0.013	0.9 0.043
C_9	0.733 0.083	0.783 0.07	0.767 0.074	0.983 0.006	0.783 0.07	0.95 0.018	0.9 0.035	0.967 0.013	1, 0	0.867 0.056
C_{10}	0.6 0.136	0.65 0.124	0.633 0.128	0.85 0.062	0.65 0.124	0.817 0.073	0.767 0.09	0.9 0.043	0.867 0.056	1, 0

Intuitionistic fuzzy proximity relation R_6 corresponding to attribute 'ERD' is given in Table 9.

Table 9. Intuitionistic fuzzy proximity relation for attribute SU

R_6	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
C_1	1, 0	0.631 0.115	0.831 0.06	0.821 0.063	0.953 0.018	0.726 0.143	0.692 0.167	0.499 0.342	0.812 0.09	0.407 0.462
C_2	0.631 0.115	1, 0	0.8 0.056	0.81 0.053	0.679 0.097	0.357 0.242	0.323 0.262	0.13 0.395	0.443 0.197	0.039 0.476
C_3	0.831 0.06	0.8 0.056	1, 0	0.99 0.003	0.879 0.042	0.557 0.196	0.523 0.218	0.33 0.372	0.643 0.147	0.239 0.47
C_4	0.821 0.063	0.81 0.053	0.99 0.003	1, 0	0.869 0.045	0.547 0.199	0.513 0.221	0.32 0.373	0.633 0.15	0.229 0.47
C_5	0.953 0.018	0.679 0.097	0.879 0.042	0.869 0.045	1, 0	0.679 0.16	0.644 0.183	0.452 0.352	0.764 0.108	0.36 0.465
C_6	0.726 0.143	0.357 0.242	0.557 0.196	0.547 0.199	0.679 0.16	1, 0	0.966 0.026	0.773 0.248	0.914 0.056	0.681 0.434
C_7	0.692 0.167	0.323 0.262	0.523 0.218	0.513 0.221	0.644 0.183	0.966 0.026	1, 0	0.807 0.227	0.88 0.081	0.716 0.427
C_8	0.499 0.342	0.13 0.395	0.33 0.372	0.32 0.373	0.452 0.352	0.773 0.248	0.807 0.227	1, 0	0.687 0.287	0.909 0.327
C_9	0.812 0.09	0.443 0.197	0.643 0.147	0.633 0.15	0.764 0.108	0.914 0.056	0.88 0.081	0.687 0.287	1, 0	0.596 0.446
C_{10}	0.407 0.462	0.039 0.476	0.239 0.47	0.229 0.47	0.36 0.465	0.681 0.434	0.716 0.427	0.909 0.327	0.596 0.446	1, 0

On considering the degree of dependency values $\alpha \geq 0.92$, $\beta < 0.08$ for membership and non-membership functions, the intuitionistic fuzzy approximation space $(U, R(\alpha, \beta))$ have the following equivalence classes. This is termed as the granular computing view of the information system.

$$U / R_1^{\alpha,\beta} = \{\{C_1, C_2, C_3\}, \{C_4, C_5\}, \{C_6, C_7, C_8\}, \{C_9, C_{10}\}\}$$

$$U / R_2^{\alpha,\beta} = \{\{C_1, C_2, C_3\}, \{C_4, C_5, C_6\}, \{C_7\}, \{C_8, C_9, C_{10}\}\}$$

$$U / R_3^{\alpha,\beta} = \{\{C_1, C_2, C_4\}, \{C_3, C_5\}, \{C_6\}, \{C_7, C_8, C_9, C_{10}\}\}$$

$$U / R_4^{\alpha,\beta} = \{\{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}\}\}$$

$$U / R_5^{\alpha,\beta} = \{\{C_1, C_2, C_3, C_5\}, \{C_4, C_6, C_7, C_8, C_9\}, \{C_{10}\}\}$$

$$U / R_6^{\alpha,\beta} = \{\{C_1, C_5\}, \{C_2\}, \{C_3, C_4\}, \{C_6, C_7\}, \{C_8\}, \{C_9, C_{10}\}\}$$

From the above analysis, it is clear that the attribute ERD classify the universe into six categories. Let it be low, average, good, very good, excellent, and outstanding and hence can be ordered. Similarly, the attributes EM, ED, and EA classify the universe into four categories. Let it be low, average, high and very high and hence can be ordered. The attribute SU classify the universe into three categories namely good, very good, and excellent. Since the equivalence class $U / R_4^{\alpha,\beta}$ contains only one group, the universe is (α,β) -indiscernible according to the attribute EMi and hence do not require any ordering while establishing association rules between the attribute values. Therefore, the ordered information table of the small universe Table 3 is given in Table 10.

Table 10. Ordered information table

Companies	EM (a_1)	ED (a_2)	EA (a_3)	SU (a_5)	ERD (a_6)
C_1	Very high	Very high	Very high	Excellent	Outstanding
C_2	Very high	Very high	Very high	Excellent	Excellent
C_3	Very high	Very high	High	Excellent	Very good
C_4	High	High	Very high	Very good	Very good
C_5	High	High	High	Excellent	Outstanding
C_6	Average	High	Average	Very good	Good
C_7	Average	Average	Low	Very good	Good
C_8	Average	Low	Low	Very good	Average
C_9	Low	Low	Low	Very good	Poor
C_{10}	Low	Low	Low	Good	Poor

$\prec_{EM}$: Very high $\prec$ High $\prec$ Moderate $\prec$ Low

$\prec_{ED}$: Very high $\prec$ High $\prec$ Moderate $\prec$ Low

$\prec_{EA}$: Very high $\prec$ High $\prec$ Moderate $\prec$ Low

$\prec_{SU}$: Excellent $\prec$ Very good $\prec$ Good

$\prec_{ERD}$: Outstanding $\prec$ Excellent $\prec$ Very good $\prec$ Good $\prec$ Poor

The granular computing view of the ordered information system is given in the following Table 11. To check whether a set of attribute values is an association rule or not, we perform the AND operation among the bit representation of the attribute values of this set. If number of 1's in the result of AND operation is greater than or equal to minimum support, then it is an association rule, otherwise it is not an association rule. For example, let us assume the minimum support is 2. To check whether {EM-very high, SU- excellent} is an association rule, we perform AND operation between bit representation of the attribute values EM-very high and SU- excellent as shown in the Table 12. Since the number of 1's in the result of

AND operation is greater than or equal to the minimum support 2, thus {EM-very high, SU- excellent} is an association rule.

Table 11. Granular computing view of the ordered information table

Granules based on	Attribute Values	Granules as List	Granules as Bits
EM	Very high	$\{C_1, C_2, C_3\}$	1110000000
	High	$\{C_4, C_5\}$	0001100000
	Average	$\{C_6, C_7, C_8\}$	0000011100
	Low	$\{C_9, C_{10}\}$	0000000011
ED	Very high	$\{C_1, C_2, C_3\}$	1110000000
	High	$\{C_4, C_5, C_6\}$	0001110000
	Average	$\{C_7\}$	0000001000
	Low	$\{C_8, C_9, C_{10}\}$	0000000111
EA	Very high	$\{C_1, C_2, C_4\}$	1101000000
	High	$\{C_3, C_5\}$	0010100000
	Average	$\{C_6\}$	0000010000
	Low	$\{\{C_7, C_8, C_9, C_{10}\}$	0000001111
SU	Excellent	$\{C_1, C_2, C_3, C_5\}$	1110100000
	Very good	$\{C_4, C_6, C_7, C_8, C_9\}$	0001011110
	Good	$\{C_{10}\}$	0000000001
ERD	Outstanding	$\{C_1, C_5\}$	1000100000
	Excellent	$\{C_2\}$	0100000000
	Very good	$\{C_3, C_4\}$	0011000000
	Good	$\{C_6, C_7\}$	0000011000
	Average	$\{C_8\}$	0000000100
	Poor	$\{C_9, C_{10}\}$	0000000011

Table 12. AND Operation between two attribute values

Attribute values	Granules
EM-Very high	1 1 1 0 0 0 0 0 0 0
SU-Excellent	1 1 1 0 1 0 0 0 0 0
AND	1 1 1 0 0 0 0 0 0 0

Let us consider the attribute values EM-very high, ED-high, and SU-good. The AND operation among bit representation of attribute values EM-very high, ED-high, and SU-good is shown in the Table 13. Since number of 1's in the AND operation is less than the minimum support 2, {EM-very high, ED-high, SU-good} is not an association rule. It indicates that there are no two objects satisfying this association rule.

Table 13. AND Operation between two attribute values

Attribute values	Granules
EM-Very high	1 1 1 0 0 0 0 0 0 0
ED-High	0 0 0 1 1 1 0 0 0 0
SU-Excellent	0 0 0 0 0 0 0 0 0 1
AND	0 0 0 0 0 0 0 0 0 0

From the analysis it is clear that {EM-very high, SU- excellent} is an association rule with bit representation 1110000000. It means that, the attribute values EM-very high, and SU- excellent is an association rule of length 2 since it consists of two attribute values of two different attributes. The objects of the universe that is supporting to this association rule are $C_1, C_2,$ and C_3. In order to find the association between the attribute and attribute values, we find all possible association rules of length two, and then all possible association rules of length three and so on.

Conclusion

Association rule granulation is a model for searching knowledge about the universe. We can not discern some objects of the universe from others due to the lack of sufficient information about the universe. The discovery of the knowledge about the universe requires classification of the objects of the universe based on the almost indiscernibility relation rather than the indiscernibility relation. To this end, the proposed model uses rough set on intuitionistic fuzzy approximation spaces with ordering rules to find the associations between any combinations of attribute values that are almost indiscernible. For the degree of dependency value $\alpha = 1$ and $\beta = 0$, it reduces to Pawlak's [Paw98] granulation of knowledge and hence is a better model. Also, association rule granulation using granular computing is more efficient than conventional method while studying large databases.

References

[De99] **S. K. De** – *Some Aspects of Fuzzy sets, Rough sets and Intuitionistic Fuzzy sets*, Ph.D. thesis, IIT, Kharagpur, 1999

[DG98] **I. Duntsch, G. Gediga** - *Uncertainty Measures of Rough Set Prediction*, Artificial Intelligence, Vol. 106, pp. 109 – 137, 1998

[JKJ06] **G. Jeon, D. Kim, J. Jeong -** *Rough Sets Attributes Reduction Based Expert System in Interlaced Video Sequences*, IEEE Transactions Consumer Electronics, Vol. 52 (4), pp. 1348 – 1355, 2006

[JQ07] **R. Jensen, Q. Shen** – *Fuzzy Rough Sets Assisted Attribute Selection*, IEEE Transactions on Fuzzy Systems, Vol. 15(1), pp. 73 – 89, 2007

[LLT05] **E. Louie, T. Y. Lin, Xiaohua Tony Hu** – *Analysis of Using Granules to Find Association Rules*, IEEE International Conference on Granular Computing, Beijing, China, pp. 556-560, 2005

[Paw82] **Z. Pawlak** – *Rough Sets*, International Journal of Computer and Information Sciences, Vol. 11(5), pp. 341-356, 1982

[Paw91] **Z. Pawlak** - *Rough Sets, Theoretical aspect of Reasoning about Data*, Kluwer Academic Publishers, 1991

[Paw98] **Z. Pawlak** – *Granularity of Knowledge, Indiscernibility and Rough Sets*, IEEE World Congress on Computational Intelligence, Vol. 1, pp. 106 – 110, 1998

[PS07a] **Z. Pawlak, A. Skowron** - Rudiments of rough sets, Information Science, An International Journal, Vol. 177(1), pp.3–27, 2007

[PS07b] **Z. Pawlak, A. Skowron** - Rough sets some extensions, Information Science, An International Journal, Vol. 177(1), pp.28–40, 2007

[PS07c] **Z. Pawlak, A. Skowron** - *Rough sets and Boolean reasoning*, Information Science, An International Journal, Vol. 177(1), pp.41–73, 2007

[Reb06] **M. Rebolledo** - *Rough intervals-Enhancing intervals for Qualitative Modeling of Technical Systems*, Artificial Intelligence, Vol. 170, pp. 667-685, 2006

[SC02] **Q. Shen, A. Chouchoulas** - *A Rough Fuzzy Approach for Generating Classification Rules*, Pattern Recognition, Vol. 35 (11), pp. 2425 – 2438, 2002

[SS96] **A. Skowron, J. Stepaniuk** – *Tolerance Approximation Spaces*, Fundamenta Informaticae, Vol. 27, pp. 245 – 253, 1996

[Tri06] **B. K. Tripathy** – *Rough sets on Intuitionistic Fuzzy Approximation Spaces*, Notes on Intuitionistic Fuzzy Sets, Vol. 12, No. 1, pp 45-54, 2006

[Yao00] **Y. Y. Yao** – *Information tables with neighborhood semantics*, in Data Mining and Knowledge Discovery: Theory, Tools, and Technology, Vol. 2, B. V. Dasarathy (Ed.), Society for Optical Engineering, Bellingham, Washington, pp. 108–116, 2000

[Yao98] **Y. Y. Yao** – *Generalized Rough Set Models*, In: Rough Sets in Knowledge Discovery, L. Polkowski, and A. Skowron (Eds.), Physica-Verlag, Heidelberg, pp. 286-318, 1998

[YL96] **Y. Y. Yao, T. Y. Lin** – *Generalization of rough sets using modal logic*, International Journal of Intelligent Automation and Soft Computing, Vol. 2, pp.103–129, 1996

[YY01] **Y. Y. Yao, Ying Sai** – *Mining Ordering Rules using Rough Set Theory*, Bulletin of International Rough Set Society, Vol. 5, pp. 99–106, 2001

[Zad94] **L. Zadeh** – *Fuzzy Graphs, Rough Sets and Information Granularity*, Third International Workshop on Rough Sets and Soft Computing, San Jose, November 10 – 12, 1994

[Zad96a] **L. Zadeh** – The Key Rules of Information Granulation and Fuzzy Logic in Human Reasoning, Concept Formulation and Computing with Words, Fifth IEEE International Conference of Fuzzy Systems, New Orleans, September 8 – 11, 1996

[Zad96b] **L. Zadeh** – Information Granulation, Fuzzy Logic and Rough *Sets*, Fourth International Workshop on Rough Sets, and Machine Discovery, Tokyo, November 6 – 8, 1996

[ZPZ06] **A. Zeng, D. Pan, Q. L. Zheng** – *Knowledge Acquisition Based on Rough Set Theory and Principal Component Analysis*, IEEE Intelligent Systems, pp. 78 – 85, 2006