Potential Applications of Granular Computing in Knowledge Discovery and Data Mining

Y.Y. Yao
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: yyao@cs.uregina.ca

and

Ning Zhong
Department of Computer Science and Systems Engineering
Faculty of Engineering, Yamaguchi University
Tokiwa-Dai, 2557, Ube 755, Japan
E-mail: zhong@ai.csse.yamaguchi-u.ac.jp

ABSTRACT

In this paper, we argue that granular computing may have many potential applications in knowledge discovery and data mining. Three related basic operations of granular computing are examined: granulation of the universe, characterization of granules, and relationships between granules. Their connections to the tasks of knowledge discovery and data mining are analyzed.

Keywords: Concept formation, data analysis, data mining, granular computing, knowledge discovery.

1. INTRODUCTION

Basic ingredients of granular computing are subsets, classes, and clusters of a universe [15, 20]. There are at least three fundamental issues in granular computing: granulation of the universe, description of granules, and relationships between granules. These issues have been considered either explicitly or implicitly in many fields, such as data and cluster analysis, concept formation, and knowledge discovery and data mining. Granulation of a universe involves the decomposition of the universe into parts, or the grouping of individual elements into classes, based on available information and knowledge. Elements in each granule may be interpreted as instances of a concept. They are drawn together by indistinguishability, similarity, proximity or functionality [19, 20]. One may easily establish some connections between tasks of granular computing and those of concept formation, knowledge discovery, and data mining. One of the tasks of concept formation may be viewed as the representation, characterization, description, and interpretation of granules representing certain concepts. An important function of knowledge discovery and data mining is to establish relationships between granules, such as association and causality. The main objective of this paper is to study these issues in more detail using a simple framework, based on the theory of rough sets [8]. In particular, potential applications of granular computing in knowledge discovery and data mining are discussed.

2. BASIC FRAMEWORK

We assume that information about objects in a finite universe are given by an information table [8, 18], in which objects are described by their values on a finite set of attributes. Formally, an information table is a quadruple:

$$S = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}),$$

where

U is a finite nonempty set of objects,
At is a finite nonempty set of attributes,
V_a is a nonempty set of values for $a \in At$,
$I_a : U \rightarrow V_a$ is an information function.

Each information function I_a is a total function that maps an object of U to exactly one value in V_a. The

Object	Height	Hair	Eyes	Class
o_1	short	blond	blue	+
o_2	short	blond	brown	-
o_3	tall	red	blue	+
o_4	tall	dark	blue	-
o_5	tall	dark	blue	-
o_6	tall	blond	blue	+
o_7	tall	dark	brown	-
o_8	short	blond	brown	-

Table 1: An information table

rows of the table correspond to objects of the universe, the columns (from the second to the last) correspond to a set of attributes, and each cell is the value of an object with respect to an attribute. An information table represents all available information and knowledge. Objects are only perceived, observed, or measured by using a finite number of properties. Similar representation schemes can be found in many fields, such as decision theory, pattern recognition, machine learning, data analysis, data mining, and cluster analysis [8].

Example 1 Table 1 is an example of information table, taken from an example from Quinlan [10]. Each object is described by four attributes. The column labeled by Class denotes an expert's classification of the objects. □

With an information table, we can define a certain language for describing objects or a group of objects of the universe. For example, an object can be represented as a conjunction of attribute-value pairs. A subset of objects can be similarly described. We adopt the decision logic language (DL-language) studied by Pawlak [8]. In the DL-language, an atomic formula is given by (a, v), where $a \in At$ and $v \in V_a$. If ϕ and ψ are formulas in the DL-language, then so are $\neg\phi$, $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \rightarrow \psi$, and $\phi \equiv \psi$. The semantics of the DL-language can be defined in Tarski's style through the notions of a model and satisfiability. The model is an information table S, which provides interpretation for symbols and formulas of the DL-language. The satisfiability of a formula ϕ by an object x, written $x \models_S \phi$ or in short $x \models \phi$ if S is understood, is given by the following conditions:

(1) $x \models (a, v)$ iff $I_a(x) = v$,
(2) $x \models \neg\phi$ iff not $x \models \phi$,
(3) $x \models \phi \wedge \psi$ iff $x \models \phi$ and $x \models \psi$,
(4) $x \models \phi \vee \psi$ iff $x \models \phi$ or $x \models \psi$,
(5) $x \models \phi \rightarrow$ iff $x \models \neg\phi \vee \psi$,
(6) $x \models \phi \equiv \psi$ iff $x \models \phi \rightarrow \psi$ and $\psi \rightarrow \phi$.

If ϕ is a formula, the set $m_S(\phi)$ defined by:

$$m_S(\phi) = \{x \in U \mid x \models \phi\}, \tag{1}$$

is called the meaning of the formula ϕ in S. If S is understood, we simply write $m(\phi)$. Obviously, the following properties hold [8]:

(a) $m(a, v) = \{x \in U \mid I_a(x) = v\}$,
(b) $m(\neg\phi) = -m(\phi)$,
(c) $m(\phi \wedge \psi) = m(\phi) \cap m(\psi)$,
(d) $m(\phi \vee \psi) = m(\phi) \cup m(\psi)$,
(e) $m(\phi \rightarrow \psi) = -m(\phi) \cup m(\psi)$,
(f) $m(\phi \equiv \psi) = (m(\phi) \cap m(\psi)) \cup (-m(\phi) \cap -m(\psi))$.

The meaning of a formula ϕ is therefore the set of all objects having the property expressed by the formula ϕ. In other words, ϕ can be viewed as the description of the set of objects $m(\phi)$. Thus, a connection between formulas of the DL-language and subsets of U is established.

A formula ϕ is said to be true in an information table S, written $\models_S \phi$, if and only if $m(\phi) = U$, namely, ϕ is satisfied by all objects in the universe. Two formulas ϕ and ψ are equivalent in S if and only if $m(\phi) = m(\psi)$. By definition, the following properties hold [8]:

(i) $\models_S \phi$ iff $m(\phi) = U$,
(ii) $\models_S \neg\phi$ iff $m(\phi) = \emptyset$,
(iii) $\models_S \phi \rightarrow \psi$ iff $m(\phi) \subseteq m(\psi)$,
(iv) $\models_S \phi \equiv \psi$ iff $m(\phi) = m(\psi)$.

Thus, we can study the relationships between concepts described by formulas of the DL-language based on the relationships between their corresponding sets of objects.

Example 2 Consider Table 1. The following expressions are some of the formulas of the DL-language:

(Height, tall),
(Height, short),
(Hair, dark),
(Height, tall) $\vee$ (Height, short),
(Height, tall) $\wedge$ (Hair, dark),
(Height, tall) $\vee$ (Hair, dark),
(Hair, dark) $\rightarrow$ (Height, tall),
(Hair, dark) $\equiv$ (Height, tall).

The subsets of objects satisfying these formulas are given by:

$$
\begin{aligned}
&m(\text{Height, tall}) = \{o_3, o_4, o_5, o_6, o_7\},\\
&m(\text{Height, short}) = \{o_1, o_2, o_8\},\\
&m((\text{Height, tall}) \vee (\text{Height, short})) = U,\\
&m(\text{Hair, dark}) = \{o_4, o_5, o_7\},\\
&m((\text{Height, tall}) \wedge (\text{Hair, dark})) = \{o_4, o_5, o_7\},\\
&m((\text{Height, tall}) \vee (\text{Hair, dark})) =\\
&\qquad \{o_3, o_4, o_5, o_6, o_7\},\\
&m((\text{Hair, dark}) \rightarrow (\text{Height, tall})) = U,\\
&m((\text{Hair, dark}) \equiv (\text{Height, tall})) =\\
&\qquad \{o_1, o_2, o_4, o_5, o_7, o_8\}.
\end{aligned}
$$

Among these formulas, two are true in the information table, namely:

$$
\begin{aligned}
&\models_S (\text{Height, tall}) \vee (\text{Height, short}),\\
&\models_S (\text{Hair, dark}) \rightarrow (\text{Height, tall}).
\end{aligned}
$$

The first represents the fact that in the information table an object's Height is either tall or short. The second represents the fact that if an object's Hair is dark, then its Height is tall. It is interesting to note that the second formula can be used to analyze the relationships between concepts. Conceptually, an important task of knowledge discovery and data mining may be formulated as searching for such formulas with respect to an information table. □

3. GRANULATION

When objects are represented through a finite set of attributes, some objects may have the same description and cannot be distinguished. Based on the attribute values, we can cluster objects in the universe. For example, one may group objects based equality, equivalence, or similarity of attribute values [18]. More generally, a fuzzy similarity relation on attribute values may be used. This process in fact involves the granulation of attribute values, which is then used to granulate the universe. A granulation of the attribute values can be supplied by experts, representing additional knowledge about the information table. For example, attribute values can be granulated based on some concept hierarchies [3]. Granulated views of the universe represent our knowledge about the universe such that each granule represents a certain concept. Granulation of the universe can be flat or hierarchical. In this section, we present a simple granulation method.

Consider an attribute $a \in At$. It may happen that two objects x and y have the same value on a, namely, $I_a(x) = I_a(y)$. In this case, one cannot differentiate x from y based solely on their values on attribute a. Thus, they may be put into the same granule. For a value $v \in V_a$, one obtains a granule with respect to an atomic formula of the DL-language [9]:

$$m(a, v) = \{x \mid I_a(x) = v\}. \tag{2}$$

It consists of all objects whose value on attribute a equals to v, and may be interpreted as the granule defined by an equality constraint in the sense discussed by Zadeh [20]. The family of granules,

$$\pi_{\{a\}} = \{m(a, v) \neq \emptyset \mid v \in V_a\}, \tag{3}$$

form a partition of the universe [8]. The corresponding equivalence relation $E_{\{a\}}$ on U is given by:

$$xE_{\{a\}}y \Longleftrightarrow I_a(x) = I_a(y). \tag{4}$$

A granule is indeed an equivalence class of the relation $E_{\{a\}}$. The equivalence class containing $x \in U$, written $[x]_{E_{\{a\}}}$, is:

$$
\begin{aligned}
[x]_{E_{\{a\}}} &= m(a, I_a(x))\\
&= \{y \in U \mid I_a(y) = I_a(x)\}.
\end{aligned} \tag{5}
$$

It consists of all objects whose value on attribute a is the same as that of the object x. The partition $\pi_{\{a\}}$ of the universe is also referred to as a quotient set of U and is denoted by $U/E_{\{a\}}$. It represents a granulated view of the universe.

The sets in $\pi_{\{a\}}$ are called elementary granules, as they are the smallest granules derivable based on values of attribute a. From the elementary granules, larger granules may be built by taking union of families of elementary granules. That is, it is possible to build a hierarchy of granules. If the empty set $\emptyset$ is added, one obtains a sub-Boolean algebra of the Boolean algebra formed by the power set of U.

The elementary granules constructed by using values of V_a may be too large. In order to resolve this problem, additional attributes are used. For a pair of attributes $a, b \in At$ and two values $v \in V_a, w \in V_b$, one can obtain the following granule with respect to the formula $(a, v) \wedge (b, w)$ of the DL-language [9]:

$$
\begin{aligned}
m((a, v) \wedge (b, w)) &= \{x \mid I_a(x) = v \wedge\\
&\qquad I_b(x) = w\}.
\end{aligned} \tag{6}
$$

The granule is defined by two equality constraints. The family of granules:

$$\pi_{\{a,b\}} = \{m((a, v) \wedge (b, w)) \neq \emptyset \mid v \in V_a, w \in V_b\}, \tag{7}$$

is a partition of the universe. The corresponding equivalence relation is given by $E_{\{a,b\}} = E_{\{a\}} \cap E_{\{b\}}$, namely,

$$xE_{\{a,b\}}y \Longleftrightarrow I_a(x) = I_a(y) \wedge I_b(x) = I_b(y). \quad (8)$$

Granules in the partition $\pi_{\{a,b\}}$ may be smaller than granules in partitions $\pi_{\{a\}}$ and $\pi_{\{b\}}$.

The argument for constructing granules can be easily extended to a subset of attributes $A = \{a_1, \ldots, a_m\} \subseteq At$. The equivalence relation is given by $E_A = \bigcap_{i=1}^{m} E_{\{a_i\}}$, each equivalence class (granule) is defined by the equality constraints $\bigwedge_{i=1}^{m} I_{a_i}(x) = v_i$, where $v_i \in V_{a_i}$. The algebra $(\{E_A\}_{A \subseteq At}, \cap)$ is a lower semilattice with the zero element E_{At} [7]. For two subsets of attributes $A, B \subseteq At$, if $E_A \subset E_B$, we say that the partition π_A is *finer* than π_B, or π_B is *coarser* than π_A. We will also say that π_A is a specialization, or refinement, of π_B, or π_B is a generalization, or coarsing, of π_A [8]. The order relation of the semilattice represents the generalization-specialization relationships between partitions, i.e., families of elementary granules. The empty set $\emptyset$ produces the coarsest equivalence relation, i.e., $E_\emptyset = U \times U$, where $\times$ denotes the Cartesian product of sets. The entire set of attributes produces the finest relation E_{At}. In the formulation of granules, the addition of an attribute leads to a specialization, and hence smaller elementary granules. Conversely, the deletion of an attribute leads to a generalization, and hence larger elementary granules.

Example 3 In Table 1, if the attribute $A = \{\text{Hair}\}$ is chosen, we can partition the universe into equivalence classes:

$$\{o_1, o_2, o_6, o_8\}, \quad \{o_3\}, \quad \{o_4, o_5, o_7\},$$

indicating the colour of Hair being blond, red and dark, respectively. These granules correspond to formulas (Hair, blond), (Hair, red), and (Hair, dark). Similarly, the use of attribute Height produces the partition:

$$\{o_1, o_2, o_8\}, \quad \{o_3, o_4, o_5, o_6, o_7\}.$$

When the pair of attributes, Height and Hair, is used, we have the following formulas of the DL-language:

$$\begin{aligned}
&(\text{Height, short}) \wedge (\text{Hair, blond}),\\
&(\text{Height, tall}) \wedge (\text{Hair, blond}),\\
&(\text{Height, short}) \wedge (\text{Hair, red}),\\
&(\text{Height, tall}) \wedge (\text{Hair, red}),\\
&(\text{Height, short}) \wedge (\text{Hair, dark}),\\
&(\text{Height, tall}) \wedge (\text{Hair, dark}).
\end{aligned}$$

They produce the partition of the universe:

$$\{o_1, o_2, o_8\}, \quad \{o_3\}, \quad \{o_4, o_5, o_7\}, \quad \{o_6\}.$$

This partition is finer than the ones produced by using either Height or Hair. □

There are restrictions on the granulation structures defined by using the trivial equality relation $=$ on attribute values. We seek granulation structures characterized by partitions of the universe. Given a fixed information table, a subset of attributes defines a partition. The converse is not necessarily true. For an arbitrary partition, one may not be able to find a subset of the attributes producing the same partition. Nevertheless, one may easily generalize the discussion by considering other types of binary relations on the attribute values, in order to obtain additional granulation structures [14, 16, 18].

4. CONCEPT FORMATION, KNOWLEDGE DISCOVERY, AND DATA MINING

In the granulation process introduced in the last section, we start from a formula ϕ of the DL-language and find the corresponding subset of objects satisfying the formula. The granule obtained has a very clear meaning in the sense that the formula ϕ may be considered as a description of the granule $m(\phi)$. In this way, one assigns a name to each granule so that elements of the granule are instances of the named category or concept [5]. The hierarchical granulation of universe represents the hierarchical organization of concepts. Larger granules represents more general concepts. Such a granulation process is relatively an easy task. In many situations, we are often faced with the more difficult reverse problems. Given a granule (i.e., subset) of the universe representing certain concept, it is necessary to find a proper description of the granule using the DL-language. This is, in fact, one of the typical problems for machine learning and knowledge discovery. In general, given a granulation of the university consisting of a family of granules, one is required to find descriptions of these granules, and their relationships.

From the viewpoint of granular computing, concept formation, knowledge discovery, and data mining can be regarded as characterizing individual granules and finding relationships between these granules. Several types of relationships can be identified, such as one-way and two-way implications, and their strength can be quantified [17]. These relationships are normally represented as if-then type rules.

With granulated views of the universe, we can formulate different levels of rules, depending on the various granulations of the universe.

Let $A \subseteq U$ be a subset of the universe representing a certain concept ϕ_A, and G a family of granules whose descriptions are known. An essential part of concept formation involves the description of A in terms of granules in G. For a granule $g \in G$ with description ϕ_g, i.e., $m(\phi_g) = g$, we may have some of following situations:

$$\begin{array}{ll} \text{(I)} & g \cap A = \emptyset, \\ \text{(II)} & g \cap A \neq \emptyset, \\ \text{(II1)} & g \subseteq A, \\ \text{(II2)} & g \supseteq A, \\ \text{(II3)} & g = A. \end{array}$$

Case (I) shows that g and A are not related. However, we have:

$$g \subseteq -A. \tag{9}$$

By property (iii), we have:

$$\models_S \phi_g \rightarrow \neg\phi_A. \tag{10}$$

Hence, we can establish an if-then type rule:

$$\text{IF } \phi_g \text{ THEN not } \phi_A. \tag{11}$$

This rule enables us to decide if an instance of ϕ_g is not an instance of A. It gives the properties that make an element of U not to be an instance of A. Cases (II1)-(II3) are special subcases of (II). For these cases, respectively, we have:

$$\begin{array}{l} \models_S \phi_g \rightarrow \phi_A, \\ \models_S \phi_A \rightarrow \phi_g, \\ \models_S \phi_g \equiv \phi_A. \end{array} \tag{12}$$

By properties (iii) and (iv), we can form the following set of rules:

$$\begin{array}{r} \text{IF } \phi_g \text{ THEN } \phi_A, \\ \text{OIF } \phi_g \text{ THEN } \phi_A, \\ \text{IIF } \phi_g \text{ THEN } \phi_A, \end{array} \tag{13}$$

where OIF stands for "only if" and IIF stands for "if and only if". We expressed the rules slightly different from the conventional way, in order to see the difference between them. The first rule enables us to decide if an element of the universe is an instance of A. It shows the properties that make an element of U to be an instance of A. The second rule, which is normally expression as:

$$\text{IF } \phi_A \text{ THEN } \phi_g, \tag{14}$$

tells us the properties that an instance of A must have. The third rule is the combination of the first two rules. It summarizes the properties that instances of A, and only instances of A, must have. The first two rules may be interpreted as one-way implication, and the third rule as two-way implication. In knowledge discovery and data mining, one may be interested in different rules depending on the situation. Typically, the first rule is referred to as a *decision* rule, while the second rule as a *characteristic* rule.

Example 4 Suppose we are interested in the concept (Class, +) in Table 1, which corresponds to the subset $\{o_1, o_3, o_6\}$. If one uses only granules produced based on attributes Height, Hair, and Eyes, as examples, one can obtain the following of rules:

(1) IF (Hair, red) THEN (Class, +),
(2) IF (Hair, blond) $\wedge$ ((Eyes, blue) THEN (Class, +),
(3) IF (Hair, dark) THEN $\neg$(Class, +),
(4) OIF (Eyes, blue) THEN (Class, +),
(5) OIF (Hair, red) $\vee$ (Hair, blond) THEN (Class, +),
(6) IIF (Hair, red) $\vee$ (Hair, blond) $\wedge$ (Eyes, blue) THEN (Class, +).

It is interesting to note that these rules are not independent with each other. For example, (6) is related to (1) and (2). □

5. UNCERTAIN RULES

The rules discussed in the last section are certain rules, which reflect the logical relationships between concepts or granules. In some situations, even though a strict logical connection does not exist, there may still exist some connection between two granules. This corresponds to the case where $g \cap A \neq \emptyset$ and neither $g \subseteq A$ nor $g \supseteq A$ is true. In order to characterize such associations between two concepts ϕ and ψ, one may generalize logical rules to association rules of the following form:

$$\text{IF } \phi \text{ THEN } \psi \quad \text{with } \alpha_1, \ldots, \alpha_m, \tag{15}$$

where $\alpha_1, \ldots, \alpha_m$ denote the degree or strength of relationships [22]. Although keywords such as IF and THEN are used, one should not interpret the rules as expressing logical implications. Instead, these keywords are used to simply link concepts together [20]. For clarity, we also simply write $\phi \rightarrow \psi$. The values

	ψ	$\neg\psi$	Totals
ϕ	$\lvert m(\phi) \cap m(\psi)\rvert$	$\lvert m(\phi) \cap m(\neg\psi)\rvert$	$\lvert m(\phi)\rvert$
$\neg\phi$	$\lvert m(\neg\phi) \cap m(\psi)\rvert$	$\lvert m(\neg\phi) \cap m(\neg\psi)\rvert$	$\lvert m(\neg\phi)\rvert$
Totals	$\lvert m(\psi)\rvert$	$\lvert m(\neg\psi)\rvert$	$\lvert U\rvert$

	ψ	$\neg\psi$	Totals
ϕ	a	b	$a+b$
$\neg\phi$	c	d	$c+d$
Totals	$a+c$	$b+d$	$a+b+c+d=n$

Table 2: Contingency table for rule $\phi \rightarrow \psi$

$\alpha_1, \ldots, \alpha_m$ quantifies different types of uncertainty and properties associated with the rule. Examples of quantitative measures include confidence, uncertainty, applicability, quality, accuracy, and interestingness of rules. A recent systematic study on uncertain rules was given by Yao and Zhong [17].

Using the cardinalities of sets, we obtain the contingency Table 2, representing the quantitative information about the rule $\phi \rightarrow \psi$, where $|\cdot|$ denotes the cardinality of a set. The values in the four cells are not independent. They are linked by the constraint $a+b+c+d=n$. The 2×2 contingency table has been used by many authors for representing information of rules [1, 4, 11, 12, 21].

From the contingency table, we can define some basic quantities. The *generality* of concept ϕ is defined by:

$$G(\phi) = \frac{|m(\phi)|}{|U|} = \frac{a+b}{n}, \tag{16}$$

which indicates the relative size of the concept ϕ. A concept is more general if it covers more instances of the universe. If $G(\phi)=\alpha$, then $(100\alpha)\%$ of objects in U satisfy ϕ. The quantity may be viewed as the probability of a randomly selected element satisfying ϕ. Obviously, we have $0 \leq G(\phi) \leq 1$.

The *absolute support* of ψ provided by ϕ is the quantity:

$$\begin{aligned} AS(\psi|\phi) &= \frac{|m(\psi) \cap m(\phi)|}{|m(\phi)|} \\ &= \frac{a}{a+b}. \end{aligned} \tag{17}$$

The quantity, $0 \leq AS(\psi|\phi) \leq 1$, shows the degree to which ϕ implies ψ. If $AS(\psi|\phi)=\alpha$, then $(100\alpha)\%$ of objects satisfying ϕ also satisfy ψ. It may be viewed as the conditional probability of a randomly selected element satisfying ψ given that the element satisfies ϕ. In set-theoretic terms, it is the degree to which $m(\phi)$ is included in $m(\psi)$. Clearly, $AS(\psi|\phi)=1$, if and only if $m(\phi) \subseteq m(\psi)$. The *change of support* of ψ provided by ϕ is defined by:

$$\begin{aligned} CS(\psi|\phi) &= AS(\psi|\phi) - G(\psi) \\ &= \frac{an-(a+b)(a+c)}{(a+b)n}. \end{aligned} \tag{18}$$

Unlike the absolute support, the change of support varies from -1 to 1. One may consider $G(\psi)$ to be the prior probability of ψ and $AS(\psi|\phi)$ the posterior probability of ψ after knowing ϕ. The difference of posterior and prior probabilities represents the change of our confidence regarding whether ϕ actually causes ψ. For a positive value, one may say that ϕ causes ψ; for a negative value, one may say that ϕ does not cause ψ. The *mutual support* of ψ and ϕ is defined by:

$$\begin{aligned} MS(\phi,\psi) &= \frac{|m(\phi) \cap m(\psi)|}{|m(\phi) \cup m(\psi)|} \\ &= \frac{a}{a+b+c}. \end{aligned} \tag{19}$$

One may interpret the mutual support, $0 \leq MS(\phi,\psi) \leq 1$, as a measure of the strength of the double implication $\phi \leftrightarrow \psi$. It measures the degree to which ϕ causes, and only causes, ψ.

The degree of *independence* of ϕ and ψ is measured by:

$$\begin{aligned} IND(\phi,\psi) &= \frac{G(\phi \wedge \psi)}{G(\phi)G(\psi)} \\ &= \frac{an}{(a+b)(a+c)}. \end{aligned} \tag{20}$$

It is the ratio of the joint probability of $\phi \wedge \psi$ and the probability obtained if ϕ and ψ are assumed to be independent. One may rewrite the measure of independence as [2]:

$$IND(\phi, \psi) = \frac{AS(\psi|\phi)}{G(\psi)}. \quad (21)$$

It shows the degree of the deviation of the probability of ψ in the subpopulation constrained by ϕ from the probability of ψ in the entire data set [6, 13]. With this expression, the relationship to the change of support becomes clear. Instead of using the ratio, the latter is defined by the difference of $AS(\psi|\phi)$ and $G(\psi)$. When ϕ and ψ are probabilistic independent, we have $CS(\psi|\phi) = 0$ and $IND(\phi, \psi) = 1$. Moreover, $CS(\psi|\phi) \geq 0$ if and only if $IND(\phi, \psi) \geq 1$, and $CS(\psi|\phi) \leq 0$ if and only if $IND(\phi, \psi) \leq 1$. This provides further support for use of CS as a measure of confidence that ϕ causes ψ.

All measures introduced so far have a probabilistic interpretation. They can be roughly divided into three classes:

generality:	G,
one-way association:	AS, CS,
two-way association:	MS, IND.

Each type of association measures can be further divided into absolute support and change of support. The measure of absolute one-way support is AS, and the measure of absolute two-way support is MS. The measures of change of support are CS for one-way, and IND for two-way. It is interesting to note that all measures of change of support are related to the *deviation* of joint probability of $\phi \wedge \psi$ from the probability obtained if ϕ and ψ are assumed to be independent. In other words, a stronger association is presented if the joint probability is further away from the probability under independence. The association can be either positive or negative.

Example 5 Suppose we are interested in association between two concepts $\phi =$ (Hair, blond) and $\psi =$ (Class, +) in the information table 1. With respect to the proposed measures, we have:

$G(\phi) = 1/2,$	$G(\psi) = 3/8,$
$AS(\psi \mid \phi) = 1/2,$	$AS(\phi \mid \psi) = 2/3,$
$CS(\psi \mid \phi) = 1/8,$	$CS(\phi \mid \psi) = 1/6,$
$MS(\phi, \psi) = 2/5,$	$IND(\phi, \psi) = 4/3.$

From the values of these measures, one can conclude that there exists a positive association between the two concepts. Consider now another concept $\phi' =$ (Height, tall). In this case, we have:

$G(\phi') = 5/8,$	$G(\psi) = 3/8,$
$AS(\psi \mid \phi') = 2/5,$	$AS(\phi' \mid \psi) = 2/3,$
$CS(\psi \mid \phi') = 1/40,$	$CS(\phi' \mid \psi) = 1/24,$
$MS(\phi', \psi) = 1/3,$	$IND(\phi', \psi) = 16/15.$

The obtained values indicate that the association of ϕ' and ψ is not as strong as that of ϕ and ψ. □

6. CONCLUSION

In this paper, we examine some basic issues of concept formation, knowledge discovery, and data mining from the view point of granular computing. The emphasis is on a simple framework for interpreting many fundamental issues of the former. Our preliminary studies show that granular computing may have many potential applications in knowledge discovery and data mining. We also emphasize the processes of granulation and concept formation. They have not received sufficient attention in knowledge discovery and data mining, where the main concern is to find rules for representing relationships between concepts. With the framework of granular computing, we are able to consider these inter-related issues.

7. ACKNOWLEDGMENTS

The authors would like to thank Cory Butz for the constructive comments on the manuscript.

References

[1] Gaines, B.R. The trade-off between knowledge and data in knowledge acquisition, in: Piatetsky-Shapiro, G. and Frawley, W.J. (Eds.), *Knowledge Discovery in Databases*, AAAI/MIT Press, 491-505, 1991.

[2] Gray, B. and Orlowska, M.E. CCAIIA: clustering categorical attributes into interesting association rules, in: Wu, X., Kotagiri, R,. and Bork, K.B. (Eds.), *Research and Development in Knowledge Discovery and Data Mining*, Springer-Verlag, Heidelberg, Germany, 132-143, 1998.

[3] Han, J., Cai, Y., and Cercone, N. Data-driven discovery of quantitative rules in data bases, *IEEE Transactions on Knowledge and Data Enginering*, **5**, 29-40, 1993.

[4] Ho, K.M. and Scott, P.D. Zeta: a global method for discretization of continuous variables, *Proceedings of KDD-97*, 191-194, 1997.

[5] Jardine, N. and Sibson, R., 1971, *Mathematical Taxonomy*, Wiley, New York.

[6] Liu, H., Lu, H., and Yao, J. Identifying relevant databases for multidatabase mining, in: Wu, X., Kotagiri, R,. and Bork, K.B. (Eds.), *Research and Development in Knowledge Discovery and Data Mining*, Springer-Verlag, Heidelberg, Germany, 211-221, 1998.

[7] Orlowska, E. Logic of indiscernibility relations, *Lectures Notes in Computer Science*, vol. 208, Springer-Verlag, Berlin, 177-186, 1985.

[8] Pawlak, Z. *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.

[9] Polkowski, L. and Skowron, A. Towards adaptive calculus of granules, *Proceedings of 1998 IEEE International Conference on Fuzzy Systems*, 111-116, 1998.

[10] Quinlan, J.R., Learning efficient classification procedures and their application to chess endgames, in: Michalski, J.S., Carbonell, J.G., and Mirchell, T.M. (Eds.), *Machine Learning: An Artificial Intelligence Approach*, Vol. 1, Morgan Kaufmann, Palo Alto, CA, 463-482, 1983.

[11] Silverstein, C., Brin, S., and Motwani, R. Beyond market baskets: generalizing association rules to dependence rules, *Data Mining and Knowledge Discovery*, **2**, 39-68, 1998.

[12] Tsumoto, S. and Tanaka, H. Automated discovery of functional components of proteins from amino-acid sequences based on rough sets and change of representation, *Proceedings of KDD-95*, 318-324, 1995.

[13] Yao, J. and Liu, H. Searching multiple databases for interesting complexes, in: Lu, H., Motoda, H., and Liu, H. (Eds.), *KDD: Techniques and Applications*, World Scientific, Singapore, 1997.

[14] Yao, Y.Y. Generalized rough set models, in: *Rough Sets in Knowledge Discovery*, Polkowski, L. and Skowron, A. (Eds.), Physica-Verlag, Heidelberg, pp. 286-318, 1998.

[15] Yao, Y.Y. Granular computing using neighborhood systems, in: *Advances in Soft Computing: Engineering Design and Manufacturing*, Roy, R., Furuhashi, T., and Chawdhry, P.K. (Eds), Springer-Verlag, London, pp. 539-553, 1999.

[16] Yao, Y.Y., Wong, S.K.M., and Lin, T.Y. A review of rough set models, in: Lin, T.Y. and Cercone, N. (Eds.), *Rough Sets and Data Mining: Analysis for Imprecise Data*, Kluwer Academic Publishers, Boston, 47-75, 1997.

[17] Yao, Y.Y. and Zhong, N., An analysis of quantitative measures associated with rules, *Third Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 1999.

[18] Yao, Y.Y. and Zhong, N., Granular computing using information tables, manuscript, 1999.

[19] Zadeh, L.A. Fuzzy sets and information granularity, in: Gupta, N., Ragade, R. and Yager, R. (Eds.), *Advances in Fuzzy Set Theory and Applications*, North-Holland, Amsterdam, 3-18, 1979.

[20] Zadeh, L.A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, **19**, 111-127, 1997.

[21] Zembowicz, R. and Żytkow, J.M. From contingency tables to various forms of knowledge in database, in: Fayyad, U.M, Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (Eds.), *Advances in Knowledge Discovery and Data Mining*, AAAI Press / MIT Press, California, 39-81, 1996.

[22] Zhong, N., Dong, J., Fujitsu, S., and Ohsuga, S. Soft techniques for rule discovery in data, *Transactions of Information Processing Society of Japan*, **39**, 2581-2592, 1998.