Granular Computing on Binary Relations In Data Mining and Neighborhood Systems

T. Y. Lin

Department of Mathematics and Computer Science San Jose State University
San Jose, California 95192-0103
And
Department of Electric Engineering and Computer Science University of California Berkeley, California
E-mail: tylin@cs.sjsu.edu or tylinCcs.berkeley.edu

Article Outline

Glossary
1. Introduction
   1.1 Partitions, Coverings, Topology and Neighborhoods
2. Some Applications to Databases
   2.1 Classification
   2.2 Approximate Retrieval
   2.3 Mining High Level Rules
   2.4 Mining Soft Rules
   2.5 Conclusions in Applications
3. Mathematical Structures of Granulation
   3.1 Binary Relations and Binary Neighborhood Systems
   3.2 Fuzzy Binary Relations and Fuzzy Binary Neighborhood systems
   3.3 Crisp/Fuzzy Neighborhood System
   3.4 Fundamentals of Neighborhood Systems
4. Granular Structures-A Conclusion
Glossary

**Neighborhood system**: A point \( x \) is the collection of all neighborhoods for the point \( x \)

**Information granulation**: It is a collection of granules, with a granule being clump of objects (points) which are drawn towards an object. In other words each objects is associated a family of clumps.

**Partition**: It is a collection of pair-wise disjoint subsets whose union \( U \)

**Chunking**: A method of splitting content into short

**Clustering**: Technique for data analysis by partitioning a data set into subsets whose elements share common traits

**Homomorphism**: It is a map from one algebraic structure to another of the same type that preserves all the relevant structure.

**Binary Neighborhood System**: To each object \( p \in V \), we associate a crisp subset \( B_p \subset U \).

In notation, \( B:V \to 2^U : p \to B_p \). The map \( B \) or the collection \( \{B_p\} \) is referred to as a binary neighborhood system.

1 Introduction

Since his 1994 keynote speech at the Third International Workshop on Rough Sets and Soft Computing, Lotfi Zadeh has been continuously pointing out the importance of information granulation [34, 35]. Implicitly the idea is earlier; it is an essential ingredient in fuzzy logic. Granulation is a very natural concept and appears almost everywhere in different names, such as chunking, clustering, data compression, divide and conquer, information hiding, interval computations, and rough set theory, just to name a few. In general, however, the computing theory on information granulation has not been fully explored in its own right. Having asserted that, we should add that in mathematics, the notion of partitions is well explored. For examples, "homomorphisms", "triangulation", and "spectral sequences" are advanced notions using partitions. In this paper, we will focus on granulations that are more general than crisp partitions'. According to Lotfi Zadeh [35],

- "information granulation involves partitioning a class of objects(points) into granules, with a granule being a clump of objects (points) which are drawn together by indistinguishability, similarity or functionality."

We will literally take Zadeh's informal words as a formal definition of clumps. Implicitly the phrase "drawn together" implies certain level of symmetry in the clumps. Namely, if \( P \) is drawn towards \( Q \), then \( Q \) is also drawn towards \( P \). Such symmetry, we believe, is imposed by imprecise-ness of natural language. To avoid that we will use the phrase "drawn toward the object \( P \)," so that it is clear the reverse may or may not be true. Further, we also observe that there are no constraints on how many clumps an object may be associated. So the notion of granulation can be rephrase as

- information granulation is a collection of granules, with a granule being clump of objects (points) which are drawn towards an object. In other word each objects is associated a family of clumps.

Does such a seemingly unstructured collection of clumps have some mathematical meaning? Interestingly, in crisp worlds, it does. It is the notion of neighborhood system (NS); see Section 3 for formal definitions. We would like to comme that if all clumps are non-empty, then NS is a covering (Cov). If there is at me one clump per object, then NS is defined by a binary relation, and is call a binary neighborhood system (BNS). If we assume the binary relation is equivalence relation, the neighborhood system is a rough set system (RS). If,
Assume a NS satisfies certain axioms, then NS defines a topological space; such NS will be called topological neighborhood system (TNS).

All of these notions can be easily fuzzified, so we have fuzzy neighborhood systems (FNS), fuzzy coverings (FCov) and fuzzy rough sets (FRS). However, we are interested in qualitative fuzzy notion only, then a FNS can be convert into crisp NS by using the notion of cr-cuts. So, in the qualitative fuzzy work the theory of NS may be adequate. For convenience of readers, let us collect some acronyms; relevant notions are defined in Section 3.

**Acronyms**

1. NS: general neighborhood system(s)
2. BNS: binary neighborhood system(s)
3. TNS: neighborhood system(s) of topological spaces
4. RS: rough set theory or systems
5. Cov: covering(s)
6. FNS: fuzzy neighborhood system(s)
7. FBNS: fuzzy binary neighborhood system(s)
8. FTNS: fuzzy neighborhood system(s) of fuzzy topological spaces
9. FRS: fuzzy rough set theory or systems
10. FCov: fuzzy covering(s)
11. BR: binary relation(s)
12. ER: equivalence relation(s)
13. FBR: fuzzy binary relation(s)
14. FER: fuzzy equivalence relation(s)

**Acknowledgement**

This research is partially supported by Electric Power Research Institute, Palo Alto, California, San Jose State University, NASA Grant NCC2-275, ONR Grant N00014-96-1-0556, LLNL Grant 442427-26449, ARO Grant DAAH04-961-03z, and BISC Program of UC Berkeley.

1.1 Partitions, Coverings, Topology and Neighborhoods

An equivalence relation is an algebraic concept, and its geometric equivalence a partition. A partition is a collection of pair-wise disjoint subsets whose union U. Algebraically, a natural generalization of an equivalence relation is a binary relation. On the other hand, a "common" geometric generalization of a partition is a covering. A covering is a collection of non-empty subsets \( \{X_j\} \) (may not be disjoint) such that \( U = \bigcup X_j \). Each \( X_j \) is called a cover. Unfortunately, a covering is not the geometric equivalence of a binary relation. The equivalent one is the more elaborate notion, called the binary (basic) neighborhood system: Let \( R \) be a binary relation. Let \( N_p = \{u \mid u \in U \text{ and } p R u\} \) be the set that consists of all elements that are related to \( p \) by \( R \). The map \( p \rightarrow N_p \) is a binary neighborhood system. The collection \( \{N_p\} \) forms a partial covering. If \( R \) is serial it is a covering. Intuitively, each \( N_p \) is more than a cover, it is a "cover with a center at \( p \)." The notion of a "center at \( p \)" plays an essential role in providing the mapping from an object to "the cover with a center at this object." In notation,

\[ p \rightarrow N_p \subseteq U. \]

Such a map is the key for extending rough set methodology (information table processing) to general binary relations. We will explore such representations in the next paper [22]. Note that coverings can be regarded as a special form of NS. In such a general case, the analogous map is multivalued [23].

Mathematically \( \{N_p\} \) is an abstraction of c-neighborhoods in analysis. The notion of abstract neighborhoods is the primitive (the building block) of the theory of (pre-)topological spaces [26]. So NS, including BNS, Cov and RS, are generalized (pre-)topological spaces. They are geometric versions of generalized rough set models [3, 27, 9, 28, 25, 29, 30, 31].
Cattaneo gives a very comprehensive axiomatic approach, in which the underlying structure is more general than a lattice. We believe the only non-lattice type known examples are NS [14]. Intuitively the notion of neighborhoods is a common one, it has been used in many different contexts. In database applications, neighborhoods are used to formulate approximate retrieval, query relaxation and data mining since mid-late 80's [1, 10, 12, 11, 4].

It has been explicitly and implicitly used in genetic algorithm texts, papers in modal logic, rough sets and many others [2, 6, 13, 31]. More recent studies have been summarized in [21] and further study have been conducted by others [33]. Pawlak, the creator of rough sets, was aware of these general results, and commented about them in early 90's [24].

2 Some Applications to Databases

To illustrate the usefulness and strength of binary neighborhood systems (binary relations). We discuss some motivational applications. For formal theory see next section.

2.1 Classification

This example illustrates the needs of using a binary relation of two universes. The fundamental idea is essentially in Viveros' thesis [32]. Her goal was to build a system that can automatically classify research articles in computer science.

The relevant idea is as follows: Let V be a list of the titles of research articles, and U be a set of keywords. The goal is to classify V by keywords. To each article p E V, we associate a subset of key words Bp, called elementary neighborhood. In her thesis, such an association is learned from training data. In the terminology of this paper, the association defines a binary neighborhood system (BNS) for V on U. Note that these elementary neighborhoods may have non-empty intersections, so the BNS may not be a partition on U. However, it does induce a partition on V [22]; this is the desired classification for the research articles. Each elementary neighborhood (a subset of key words) is supposed to characterize a subfield of computer science. We assign the subfield name as the name of the neighborhood. In the context of this paper, the name represents an elementary concept. In pattern recognition the theory of clusters seems similar to this approach. However, there are intrinsic differences. Traditional approaches focus on numerical data, while we are dealing symbolic data [5, 8].

Let us enumerate the objects in V, namely, the titles of research articles:

1. L1= Finding Reduct for Very Large Databases
2. L2= Design Optimization of Rough-Fuzzy Controllers Using a Genetic Algorithms
3. L3= Graded Rough Set Approximations Based on Nested Neighborhood Systems.
4. L4= Belief Function Based on Multivalued Random Variables.
Next let us enumerate the subsets of keywords:

1. $K_1 = \{\text{reduct, rough set, data, relation}\}$
2. $K_2 = \{\text{control, genetic algorithm, rough set, neural network, fuzzy logic}\}$
3. $K_3 = \{\text{rough set, fuzzy set, approximation, neighborhood}\}$
4. $K_4 = \{\text{belief function, random variables, probability, rough set}\}$

Table 1 illustrates the idea of classification. The first column is the list of papers to be classified. The second column is the set of lists of key words or phrases used to classify the articles. The third column is the name of each class or subfields.

### 2.2 Approximate Retrieval

The idea is essentially in [10, 11, 12]. Let us consider the following Restaurant Database. For each attribute there are four binary relations, very-close-neighborhood (VCN), close-neighborhood (CN), far-neighborhood (FN), and very-far-neighborhood (VFN). Note that each binary relation defines a BNS on the set of values of each attribute. Together four binary relations define a NS. These four binary relations on LOCATION-attribute are represented in Table 3.

<table>
<thead>
<tr>
<th>Object Space $V$</th>
<th>Elementary Neighborhood Key Word Space $U$</th>
<th>Elementary Concept Subfield Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>$K_1$</td>
<td>rough set theory</td>
</tr>
<tr>
<td>L2</td>
<td>$K_2$</td>
<td>intelligent control</td>
</tr>
<tr>
<td>L3</td>
<td>$K_3$</td>
<td>theory of neighborhood systems</td>
</tr>
<tr>
<td>L4</td>
<td>$K_4$</td>
<td>theory of evidence</td>
</tr>
<tr>
<td>L5</td>
<td>$K_5$</td>
<td>.....</td>
</tr>
</tbody>
</table>

Table 1. A Binary Relation on Two Universes

<table>
<thead>
<tr>
<th>RESTAURANT</th>
<th>TYPE</th>
<th>LOCATION1</th>
<th>PRICE</th>
<th>ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wendy</td>
<td>American</td>
<td>West wood</td>
<td>inexpensive</td>
<td>keep</td>
</tr>
<tr>
<td>Le Chef</td>
<td>French</td>
<td>West LA</td>
<td>Moderate</td>
<td>raise</td>
</tr>
<tr>
<td>Great Wall</td>
<td>Chinese</td>
<td>St Monica</td>
<td>moderate</td>
<td>May-raise</td>
</tr>
<tr>
<td>Kiku</td>
<td>Japanese</td>
<td>Hollywood</td>
<td>moderate</td>
<td>raise</td>
</tr>
<tr>
<td>South Sea</td>
<td>Chinese</td>
<td>Los Angeles</td>
<td>expensive</td>
<td>may-raise</td>
</tr>
</tbody>
</table>

Table 2. A Restaurant Database
Equivalently, these four binary relations can also be expressed in terms of BNS: see Section 3.1. They are listed below.

1. The binary neighborhood system of LOCATION-attribute:

\[
\text{VCN}_\text{Westwood} = \{ \text{Westwood, Santa Monica, West LA} \}
\]

\[
\text{CN}_\text{Westwood} = \{ \text{Westwood, Santa Monica, West LA, Hollywood, Los Angeles} \}
\]

<table>
<thead>
<tr>
<th>LOCATION</th>
<th>LOCATION</th>
<th>Binary Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westwood</td>
<td>St Monica</td>
<td>Very close</td>
</tr>
<tr>
<td>Westwood</td>
<td>West LA</td>
<td>very close</td>
</tr>
<tr>
<td>Westwood</td>
<td>Hollywood</td>
<td>close</td>
</tr>
<tr>
<td>Westwood</td>
<td>Los Angeles</td>
<td>close</td>
</tr>
<tr>
<td>Westwood</td>
<td>San Francisco</td>
<td>far</td>
</tr>
<tr>
<td>Westwood</td>
<td>New York City</td>
<td>very far</td>
</tr>
<tr>
<td>.....</td>
<td>.....</td>
<td>.....</td>
</tr>
<tr>
<td>.....</td>
<td>.....</td>
<td>.....</td>
</tr>
</tbody>
</table>

\[
\text{FN}_\text{Westwood} = \{ \text{Westwood, Santa Monica, West LA, Hollywood, Los Angeles, San Francisco} \}
\]

\[
\text{VFN}_\text{Westwood} = \{ \text{Westwood, Santa Monica, West LA, Hollywood, Los Angeles, San Francisco, New York City} \}
\]

2. The binary neighborhood systems of TYPE-attribute:

\[
\text{VCN}_{\text{Chinese}} = \{ \text{Chinese, Vietnamese} \}
\]

\[
\text{CN}_{\text{Chinese}} = \{ \text{Chinese, Vietnamese, Japanese} \}
\]

\[
\text{FN}_{\text{Chinese}} = \{ \text{Chinese, Vietnamese, Japanese, American} \}
\]

\[
\text{VFN}_{\text{Chinese}} = \{ \text{Chinese, Vietnamese, Japanese, American, French} \}
\]

3. The binary neighborhood systems of PRICE-attribute:

\[
\text{VCN}_{\text{moderate}} = \{ \text{moderate, expensive} \}
\]

\[
\text{CN}_{\text{moderate}} = \{ \text{moderate, expensive, inexpensive} \}
\]

\[
\text{FN}_{\text{moderate}} = \{ \text{moderate, expensive, inexpensive} \}
\]

\[
\text{VFN}_{\text{moderate}} = \{ \text{moderate, expensive, inexpensive, French} \}
\]
Now, let us consider the following query:

Ques 1: Select a Restaurant which is Japanese, moderate in price and located in Westwood.

The traditional database will return a null answer. Then the user may vary his condition slightly and issue another query until either the user is too tired to proceed or he finds his answers. For our proposed system, it will supply an approximate answer for the query Q1, namely, the restaurant

"Great Wall"

because the three neighborhood relations indicated that

1. Chinese and Japanese are very close (in TYPE)
2. St Monica and Westwood are very close (in LOCATION)
3. Moderate (in PRICE)

If the user is willing to extend his neighborhood of tolerance to a bigger neighborhood, say close-neighborhood, then two additional restaurants

"Kiku" and "South Sea"

will be included in the approximate answers.

My Example:
Let us consider an apartment database

<table>
<thead>
<tr>
<th>APARTMENT</th>
<th>TYPE</th>
<th>LOCATION</th>
<th>PRICE</th>
<th>ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scotch Hollow</td>
<td>individual</td>
<td>San Mateo</td>
<td>inexpensive</td>
<td>keep</td>
</tr>
<tr>
<td>Crest View</td>
<td>condominium</td>
<td>Menlo park</td>
<td>Moderate</td>
<td>raise</td>
</tr>
<tr>
<td>Madison</td>
<td>condominium</td>
<td>Foster city</td>
<td>moderate</td>
<td>May-raise</td>
</tr>
<tr>
<td>Beach Cove</td>
<td>duplex</td>
<td>Hillsdale</td>
<td>moderate</td>
<td>raise</td>
</tr>
<tr>
<td>Sand cove</td>
<td>individual</td>
<td>San Jose</td>
<td>expensive</td>
<td>may-raise</td>
</tr>
</tbody>
</table>

Let us consider the query: select an apartment in San Mateo and price is moderate

If it is a relational database the result will be null, but for the proposed system we will get the result as “Beach Cove” because the following three neighborhood relations indicate that

1. Individual and duplex are close (in type)
2. San Mateo and Hillsdale are very close (in location)
3. Moderate (in price)

If the user is willing to extend his neighborhood of tolerance to a bigger neighborhood, say close-neighborhood, then other apartments result i.e. “Madison” as San Mateo and Foster city are average (in location)
2.3 Mining High Level Rules

This is an example of attribute oriented generalization. The following table is a high level table derived from the restaurant database by replacing all attributes values by its elementary concepts. For example, West wood and West LA are replaced by western area, and St Monica and Hollywood by northern area. The high level table is a table of elementary concepts. Processing of such a table is a computing of elementary concepts (words). It results in high level rules of the restaurant database [18, 17].

<table>
<thead>
<tr>
<th>RESTAURANT</th>
<th>TYPE</th>
<th>LOCATION1</th>
<th>PRICE</th>
<th>ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wendy</td>
<td>western</td>
<td>western area</td>
<td>inexpensive</td>
<td>probably raise</td>
</tr>
<tr>
<td>Le Chef</td>
<td>western</td>
<td>western area</td>
<td>Moderate</td>
<td>definitely raise</td>
</tr>
<tr>
<td>Great Wall</td>
<td>Chinese-type</td>
<td>northern area</td>
<td>moderate</td>
<td>probably raise</td>
</tr>
<tr>
<td>Kiku</td>
<td>Japanese-type</td>
<td>northern area</td>
<td>moderate</td>
<td>definitely raise</td>
</tr>
<tr>
<td>South Sea</td>
<td>Chinese-type</td>
<td>central area</td>
<td>expensive</td>
<td>probably raise</td>
</tr>
</tbody>
</table>

Table 4. The Generalization of Restaurant Database (Italic terms are generalized items)

Proposition. Decision rules in a high level table (e.g., Table 4) are the high level rules of the original information table (e.g., Table 2).

2.4 Mining Soft Rules

If we want to make decision based on elementary concepts, not data, of decision attributes, then we replace only the data of decision attributes by their elementary concepts. Then all rules in this new table are soft rules in the original table. Table processing of the new table (called concept decision table) results in soft rules of the Restaurant Database [17].

<table>
<thead>
<tr>
<th>RESTAURANT</th>
<th>TYPE</th>
<th>LOCATION1</th>
<th>PRICE</th>
<th>ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wendy</td>
<td>American</td>
<td>West wood</td>
<td>inexpensive</td>
<td>probably raise</td>
</tr>
<tr>
<td>Le Chef</td>
<td>French</td>
<td>West LA</td>
<td>Moderate</td>
<td>definitely raise</td>
</tr>
<tr>
<td>Great Wall</td>
<td>Chinese</td>
<td>St Monica</td>
<td>moderate</td>
<td>probably raise</td>
</tr>
<tr>
<td>Kiku</td>
<td>Japanese</td>
<td>Hollywood</td>
<td>moderate</td>
<td>definitely raise</td>
</tr>
<tr>
<td>South Sea</td>
<td>Chinese</td>
<td>Los Angeles</td>
<td>expensive</td>
<td>probably raise</td>
</tr>
</tbody>
</table>

Table 5. A Concept Decision Table of Restaurant Database

Proposition Decision rules in a decision concept table (e.g., Table 5) are soft rules in the original information table (e.g., Table 2).
Soft rules are rules which conclude to nearby data. In other words, the conclusions are very close to correct answers.

2.5 Conclusions in Applications

We have shown that by utilizing the notion of neighborhood systems (granulation), we can retrieve
1. data
2. soft data (approximation)
3. rules
4. soft rules
5. high level rules

So the notion of neighborhood systems seem a useful conceptual framework for data mining and retrieval. These initial applications warranty the need for further studies.

3 Mathematical Structures of Granulation

Based on the informal analysis in the introduction, we will propose several form of granulation.

3.1 Binary Relations and Binary Neighborhood Systems

A binary neighborhood system (BNS) is: To each object p E V, we associate a crisp subset Bp C U. In notation,

\[ B : V \rightarrow 2^U : p \rightarrow B_p \]

The map B or the collection \{Bp\} is referred to as a binary neighborhood system for V on U. We have called Bp a basic neighborhood and B or \{Bp\} a basic neighborhood system respectively [21]. However, we will use the term binary neighborhood to emphasize its relationship with binary relation.

A binary relation (BR) is: Let R C V x U be a binary relation. For each object p E V, we associate a binary subset Np C U, where

\[ N_p = \{ u | R(p, u) \} \]

that consists of all elements u that are related to p by R.

Proposition. Given a binary neighborhood system \{Bp\} for V on U, there is a binary relation R C V x U such that

\[ N_p = B_p \]

and vise versa. So from now on B or R is BNS as well as BR. Normally we are interested in the case V = U. Two universe formulation is needed for some applications; see Section 2.1.

If the binary relation R is an equivalence relation E, then the binary set N7, is the equivalence class [p]E. In rough set theory an equivalence class is called an elementary set. So the binary neighborhood N7, may also be called an elementary neighborhood of p. A subset X is a definable set if it is a union of equivalence classes. So a subset X is called a definable neighborhood, if X is a union of elementary neighborhoods. If the definable neighborhood X contains the elementary neighborhood B2, of p, it is a definable neighborhood of p.
3.2 Fuzzy Binary Relations and Fuzzy Binary Neighborhood Systems

So far implicitly all discussions are in the crisp world. Lotfi Zadeh notes that human information processing is often fuzzy. So it is desirable these notions are fuzzified [36]. In [15], we have discussed various fuzzy sets. In this paper, a fuzzy set is uniquely defined by its membership function. So a fuzzy set is a w-sofset, if we use the language of the cited paper. By fuzzify previous discussions, we should have

A fuzzy binary neighborhood (FBNS) is: To each object \( p \in V \), we associate a fuzzy subset (a clump), denoted by \( FB_p \). In other words, we have a map

\[
NB : V \rightarrow \mathcal{F}(U) : p \rightarrow FB_p,
\]

where \( \mathcal{F}(U) \) means all fuzzy sets on \( U \). \( FB_p \) is called an fuzzy elementary neighborhood or fuzzy binary neighborhood, \( FB \) a fuzzy binary neighborhood system.

If we take \( \alpha \)-cuts of every \( FB_p \), then the new neighborhood system is a (genera) neighborhood system defined in Section 3.3. So if the actual numerical \( \alpha \)-values are unimportant, crisp neighborhood system can handle the fuzzy applications.

A fuzzy binary relation (FBR) is: Let \( I \) be the unit interval \([0, 1]\). Let \( FR \) be a fuzzy binary relation whose membership function, denoted by the same notation again, is

\[
FR : V \times U \rightarrow I : (p, u) \rightarrow r.
\]

To each \( p \in V \), we associate a fuzzy set (a clump) \( FN_p \) whose membership function is,

\[
N_p : U \rightarrow I \text{ is defined by } FN_p(u) = FR ((p, u)).
\]

As in crisp case, we have

**Proposition.** Given a fuzzy binary neighborhood system \( \{FB_p\} \) for \( V \) on \( U \), there is a fuzzy binary relation \( FR \) such that

\[
FN_p = FB_p
\]

and vise versa.

So as in the crisp case, from now on we will use algebraic and geometric terms interchangeably. \( FB \) is a fuzzy binary neighborhood system as well as a fuzzy binary relation \( FR \).

3.3 Crisp/Fuzzy Neighborhood Systems

In many real world applications, we often need to consider different granules from different contexts simultaneously. So in stead of one granule at a time, we consider a family of granules.

A crisp/fuzzy neighborhood system (NS/FNS) is: To each object \( p \in V \), we associate a (empty, finite or infinite) family of clumps (crisp/fuzzy subsets). The mathematical system defined by these families of clumps is called crisp/fuzzy neighborhood system or simply neighborhood system, and these clumps associated to \( p \) are called fundamental neighborhoods of \( p \). Note that fundamental neighborhoods are called elementary neighborhoods or elementary sets, if the crisp/fuzzy neighborhood system is a binary neighborhood systems, or a rough set system respectively.

In single level granulation, the algebraic (i.e., binary relations) and topological (binary neighborhood systems) notions correspond in one-to-one fashion, and are interchangeable. Intrinsically, however, they are different; binary relations are global in nature, while binary neighborhood systems are local. In multilevel cases, such a correspondence is no longer reversible, it is a many to
one relationship. It is clear that a selection of one clump at each point (a local
selection) gives rise to a binary relation (a global notion). Different selections
correspond to different binary relations. It is also clear that these different se-
lections are not intrinsically different. So for multilevel granulation, we adopt
topological formulation only. This subject will be take up in next paper [23]

3.4 Fundamentals of Neighborhood Systems

Let us summarize some fundamental notions on neighborhood systems [21]

Definitions and Properties

1. A subset X is a definable neighborhood if it is a union of fundamental
neighborhoods; the subset is a definable neighborhood of p, if the union
contains a fundamental neighborhood of p.

2. A neighborhood system of an object p, denoted by NS(p), is the maximal
family of definable neighborhoods of p. If NS(p) is an empty family, we
simply say that p has no neighborhood.

3. A neighborhood system of U, denoted by NS(U) is the collection of NS(p)
for all p in U. For simplicity a set U together with NS(U) is called a neigh-
borhood system space (NS-space) or simply neighborhood system.

4. A subset X of U is open if for every object p in X, there is a neighborhood
Np C X. A subset X is closed if its complement is open.

5. NS(p) and NS(U) are open if every neighborhood is open. NS(U) is topo-
logical, if NS(U) are open and U is the usual topological space [26]. In such
a case both NS(U) and the collection of open sets are called topology -see
next proposition.

6. An object p is a limit point of a set E, if every neighborhood of p contains
a point of E other than p. The set of all limit points of E is call derived set.
E together with its derived set is a closed set.

7. NS(U) is discrete, if NS(U) is the power set.

8. NS(U) is indiscrete, if NS(U) is a singleton {U}.

Specific Neighborhood systems

Let us recall some definitions on binary (or basic) neighborhood systems [21]

1. B is serial, if Vp, Bp is non-empty,
2. B is reflexive, if Vp, p E Bp;
3. B is symmetric, if Vp,Vq, q E Bp = p E Bq;
4. B is transitive, if Vp, Vq, Vr, q E Bp and r E Bq = r E Bp;
5. B is Euclidean, if q E Bp, and r E Bp, = r c Bq;
6. B is rough set (clopen), if it is reflexive, symmetric, and transitive.

Examples

Let us give few simples examples.

1. Let U be the Euclidean plane, we assign each point p a family of neighbor-
hoods by open solid disks: Vp and Vc, where d is the usual Euclidean dis-
tance. The family of such Np(c) is the neighborhood systems of the usual topology of Euclidean plane.

2. Next, we assign each point p in the Euclidean plane U a unique neighbor-
hood, namely, a open solid disk of radius 2:

Such an assignment gives U a neighborhood system, but not a topological
space.

Definition

Let X be a subset of U.

I[X] = {p : 3 Np C X} = the interior of X,
i.e., $I[X]$ is the largest open set contained in $X$,

$$C[X] = \{ p : 'N1, X n Np 01 = \text{the closure of } X \}$$

i.e., $C[X]$ is the smallest closed set contains $X$. $I[X]$ and $C[X]$ are precisely the lower and upper approximation, if the neighborhood system is a rough set system.

We will collect few simple properties of topology and neighborhood systems, and point out their differences.

**Proposition**

1. A topological space is a neighborhood system space (NS-space), but not the converse.
2. Intersections and finite unions of closed sets are closed in NS-spaces.
3. In topological spaces, unions and finite intersections of open sets are open.
   In NS-spaces, unions is open, but intersections may not be open.
4. In a topological space NS($U$) determines and is determined by the collection of all open sets. This property may or may not be true for neighborhood systems.
5. A neighborhood system is a subbase (see below) of a topology. An open set in NS-space is also open in this topology, the converse may not be true.

A family $B$ of subsets is a base for a topology, if every open set is a union of members of $B$. A family $S$ of subsets is a subbase, if the finite intersections of members of $S$ is a base. Note that the union and intersection of empty family is the whole space and empty set respectively. We can take a neighborhood system as a subbase of some topology. Note that two distinct neighborhood systems may give the same topology.

4 Granular Structures-A Conclusion

Let us summarize our study by offering the following formal definitions. A granular structure consists of 4-tuple

$$(V, U, B, C)$$

where $V$ is called the object space, $U$ is the data space ($V$ and $U$ could be the same set), $B$ is a crisp/fuzzy neighborhood system, and $C$ is the concept space which consists of all the names of the fundamental neighborhoods of $B$. If $B$ is a binary neighborhood system (binary relation), then the 4-tuple $(V, U, B, C)$ is called a binary granular structure.

From the point of view of category theory, we have defined the objects. So we should define a proper notion of morphisms and form the category of granular
structures. This is rather straightforward, so we will not go into the details. In fact, we may, more generally, propose to impose some sort of "computing structure," whatever it means, onto the two universes U and V, and form a category. We will report the study if we find interesting applications in real world problems.

In this paper, we take Zadeh's informal definition of granulation almost literally and form a mathematical system. The system turns out to be essentially a known mathematical structure, called neighborhood system (we use two universes instead of one). We report some of its applications to databases, and propose, based on these applications, a very general notion of neighborhood systems which is slightly beyond the classical theory of (pre-)topological spaces. Some real world applications systems are under way [16, 19, 20]. In the next two papers [22, 23], we will examine the granular structures and their applications from the point of views of representations by words.

References

7. W. L. Hwang and T. Y. Lin, Noisy Data Processing in Information Table, Preprint, 1997


34. Lotfi Zadeh, Fuzzy Graph, Rough sets and Information Granularity. In: Proceedings of the Third International Workshop on Rough Sets and Soft Computing, San Jose, Nov. 10-12, 1, 1994
